

データ同化におけるハイパーパラメータ推定

予測発見戦略研究センター データ同化グループ

准教授 上野 玄太

1 データ同化とハイパーパラメータ

データ同化は時系列解析において用いられる状態空間モデルの大規模版である。そのため、状態空間モデルの適用において問題となる最適なハイパーパラメータの推定方法などは、データ同化でも等しく議論の対象となり、その大規模性を考慮した推定方法の確立が必要となる。ここでハイパーパラメータとは、システムノイズ、観測ノイズ、初期状態の確率分布を表現するパラメータを指すこととする。

データ同化が広く用いられる気象・海洋分野においては、ハイパーパラメータは (1) イノベーションの最小化、(2) ガウス近似を仮定した最尤法、(3) ベイズ推定、(4) 共分散一致法を用いて推定されてきた。これらは互いに異なる基準ではあるものの、いずれもイノベーションを用いて観測データとモデルの両方を考慮したものであることは特徴的である。

2 アンサンブルベースの最尤法

本稿で提案するのは、(2) を一般化した場合、つまり状態が非ガウス分布をとる場合においても適用できる最尤法である。気象・海洋分野で用いられる運動量方程式に移流項が含まれるように、システムモデルが非線型である場合には、非ガウス型の状態の確率分布を追わねばならない。時系列解析の分野では、過去の研究において、非ガウス型分布を表現するために、粒子フィルタなどアンサンブルベースのフィルタリング手法が提案され、尤度をアンサンブル近似することで最尤法を実施し、ハイパーパラメータの推定も可能であることが示されてきた。

しかし、この尤度のアンサンブル近似法は気象・海洋分野でのデータ同化には適用されてこなかった。難点の一つは、尤度の計算で用いられる観測ノイズの共分散行列を適切に与えることが難しかったことである。人工衛星やレーダーによる観測データは空間的に密に観測点が配置されるため、地理的に近い観測点間の相関を含む共分散行列を設定することが望まれる。ところが、そのような相関構造を有した上で正定性をも満足する行列を作成することは容易ではない。そこで、ガウシアングラフィカルモデルを用いて適切な共分散行列を作成する方法を提案し、数万次元変数でも運用が可能なアルゴリズムを構築した。この共分散行列を用いることで、空間的な相関を考慮した上での尤度の評価が可能となり、非線型過程を含むデータ同化のシステムモデルに対しても、最尤法を用いてハイパーパラメータの推定が可能となった。

応用例として、大気海洋結合モデルに対して、海面高度データを同化する数値実験を取り上げる。ハイパーパラメータとして選んだのは、水温躍層深度アノマリーの強度 (σ_p)、東西・南北方向の相関距離 (L_x, L_y) (以上がシステムノイズ)、観測ノイズの強度 (α) である。図1は、以上のパラメータに対するプロファイル尤度を示す。システムノイズのパラメータの中では、強度が最も尤度に敏感であることがわかる。南北方向の相関距離は2つのピークを示すが、これは特性長の異なる2つの波 (赤道ケルビン波・ロスビー波) が混在している領域であることを示唆している。図2には、システムノイズのパラメータに関して最大化し、観測ノイズの強度のみについてのプロファイル尤度を示す。 $\alpha=20$ に1つのピークがある。以上から、 $(\sigma_p, L_x, L_y, \alpha)=(2m, 20^\circ, 5^\circ, 20)$ の組み合わせが尤度の最大値を与えることがわかる。

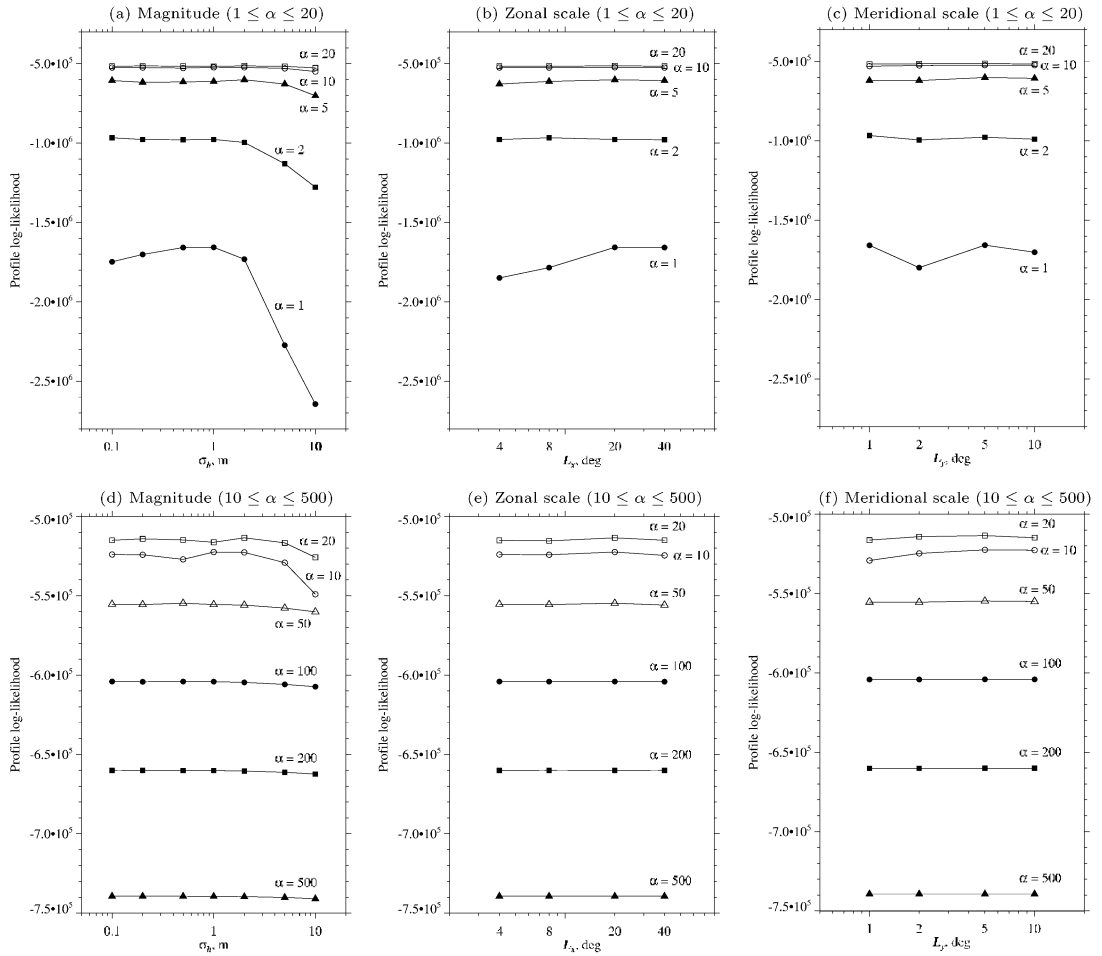


図 1：システムノイズのパラメータ (σ_b, L_x, L_y)、観測ノイズのパラメータ (α) に対するプロファイル尤度。

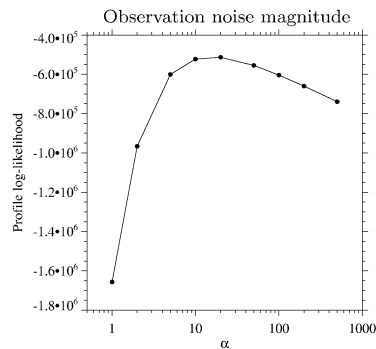


図 2：観測ノイズの係数パラメータ α に対するプロファイル尤度。