

データ同化における大規模グラフィカルモデルの推定について

上野 玄太^{1,2}

(受付 2012年8月17日; 改訂 10月22日; 採択 11月7日)

要 旨

多変量の変化をモデル化する際に、共分散行列が導入されることが多い。本稿では、大規模な時系列解析であるデータ同化における共分散行列の構築方法を解説する。データ同化では、初期状態、モデル化された物理過程、観測データの重みを規定する際に共分散行列が用いられる。共分散行列は、標本共分散行列に何らかの構造を課して変換することで作成される。共分散構造のモデル化は、標本共分散行列の正則化と物理的拘束条件の付加からなる。正則化は、特異な標本共分散行列を非特異なものへと変換し、遠地点間の偽の相関を除去し、行列指定に必要なパラメータ数を減少させるために必要となる。従来の研究では、正則化は物理空間、周波数空間、ウェーブレット空間でなされてきた。本稿では、ガウシアングラフィカルモデルを用いた正則化法を紹介する。これは、逆行列空間での正則化と考えられる。各変数に対して、近傍とする変数を選定し、非近傍の変数とは条件付き独立という形でモデル化する。条件付き独立性は、共分散行列の逆行列にゼロ要素を指定することで表現される。非ゼロ要素はニュートン法を用いて、数値的に最尤推定する。適切な近傍は、AIC や BIC など情報量規準に照らして選択する。さらに、共分散行列が巨大次元である場合の実装法についても紹介する。単純な 3×3 行列を用いた簡単な例、および人工衛星観測による海面高度データへの適用例を示す。

キーワード：ガウシアングラフィカルモデル、共分散選択、ニュートン法、並列計算。

1. はじめに

本稿は、共分散行列の推定方法を扱う。時系列解析や空間統計学などに見られるように、観測されていない変数の推定・予測を行う際には、観測されている変数と観測されていない変数の関係を事前に与えておく必要がある。さらに、統計モデル内部では、観測されていない変数同士の関係を与えておく必要も生じる。これらの関係づけの方法で、基本的でかつ広く用いられるのが、変数同士の振る舞いの相関を与えておくこと、やや広く言えば、変数の変動に関する共分散行列を与えておくことである。筆者はこれまでに時系列解析、特に大規模物理シミュレーションモデルと人工衛星観測を主としたグローバルな観測データをもとにした状態推定の問題を扱ってきた。シミュレーションモデルはそれ自体で(データはなくとも)状態の時間発展の予測が可能であるため、そこに敢えてデータの情報を入れ込んだ状態推定を行うということで、この推定作業はデータ同化と呼ばれている。

¹ 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

² 科学技術振興機構 さきがけ

第1段階: 統計量の計算 (標本共分散)

- モデル出力/アンサンブル
- 観測データ (生データ, スケール変換, トレンド除去)
- モデルと観測の差

第2段階: 共分散構造のモデル化

- 正則化
 1. 物理空間での正則化
 - ・ 対角 (独立)/ブロック対角
 - ・ 相関モデル
 - ・ デジタルフィルタ
 - ・ コンパクトな台/moderation 関数
 2. 周波数空間
 - ・ 球面調和関数展開 (等方性)
 - ・ 対角/ブロック対角
 - ・ コンパクトな台
 3. ウェーブレット空間
 - ・ 対角 (局所平均)
 - ・ 対角 + 非対角の一部
 4. 逆行列空間 (本研究)
 - ・ 対角 + 非対角の一部
 - (条件付き独立)
- 力学的拘束条件

図 1. データ同化における共分散行列の指定法.

データ同化では, 初期状態, システムノイズ, 観測ノイズの性質を規定するために共分散行列が導入される. それぞれの共分散行列の逆行列は, 初期推定値, モデル化された物理過程, 観測値の重みとして働く. それらの共分散行列を適切に指定することは, よい推定値を得る際に本質的であり, 誤って指定することでデータに当てはまりが強すぎたり弱すぎたりし, データ同化そのものが失敗に終わってしまう (Fukumori, 2001). 本稿は, データ同化における共分散行列の構築の方法を扱う.

図 1 は, データ同化における共分散行列の指定法を示す. 指定プロセスは, (1) 統計量 (標本共分散) の計算, (2) 共分散構造のモデル化の 2 つの段階からなる. この見方にしたがって, これまでの共分散指定の技術を振り返ってみたい. 本稿では, 後者すなわち共分散構造のモデル化に寄与する技術を 3 節以降で紹介する.

1.1 統計量の計算 (標本共分散)

共分散行列の指定のための第 1 の段階は, モデル出力, 観測, もしくはそれらから求められる量を用いて標本共分散行列を計算することである. モデル出力の標本共分散行列は, 初期状態の不確実性を表現する共分散行列として使われることが多く, 変分型データ同化の文献では背景誤差共分散行列と呼ばれる. NMC 法 (Parrish and Derber, 1992) は, 異なるリードタイムの 2 つのモデル推定値から標本共分散行列を計算し, 背景誤差共分散行列としている. CQ 法 (Polavarapu et al., 2005; Jackson et al., 2008) はモデル実行結果は 1 つしか使わないが, ある決められた時間間隔でのモデル出力値間の差の標本共分散行列を計算する. モデルアンサンブルも標本共分散行列の計算に用いられ, 各メンバーは観測値, パラメータ, 初期・境界条件な

どに摂動を加えることで生成する(背景誤差共分散については Houtekamer et al., 1996; Fisher, 2003; Buehner, 2005; Borovikov et al., 2005; Alves and Robert, 2005; システムノイズ(モデル誤差)については Trémolet, 2007)。

観測ノイズ共分散行列に関しては、標本共分散行列はデータそのもの、データの定数倍(Oke et al., 2002)、トレンドを除いたデータ(Ueno et al., 2007)から計算される。共分散一致法(Fu et al., 1993)やその拡張版(Menemenlis and Chechelnitsky, 2000)では、モデル出力と観測の差が用いられる。この方法によれば、線形システムモデル・線形観測モデルに対するシステムノイズと観測ノイズの共分散行列を与える。

しかし、標本共分散行列は、多くの応用例ではそのままでは使用に耐えない。これは、おもに以下の3つの理由による。第1の理由は数学的なものである。アンサンブルメンバー数や全タイムステップ数は状態変数の総数や観測地点の総数と比べて小さいことが通常であるため、このように計算された標本共分散行列は特異である。利用に際して、特異な共分散行列には逆行列が存在しないため、目的関数に含まれる背景項や観測項を評価することができないなどの不便が生じる。

第2の理由は、標本共分散は遠く離れた格子点間での偽の相関を含むことである。その相関は物理的感覚に照らすと不合理と考えられるものの、サンプル数が限られているためにどうしても発生してしまう。第3の理由は応用上のもので、メモリ容量に関係する。当然のことながら、共分散行列の要素は変数の数の2乗のオーダーであり、例えば、状態ベクトルが 10^7 に上る最近の大气シミュレーションモデルに対しては、背景誤差共分散行列の要素数は 10^{14} に上る。必要とされるメモリのサイズが大きいと、計算資源の上限により、データ同化計算の実行が不可能になりうる。

1.2 共分散構造のモデル化

標本共分散行列に関するこれら3つの難点により、共分散設定の第2の段階が始動する。すなわち、共分散構造のモデル化である。この段階では、共分散に対して正則化の条件や物理的拘束条件を課す。

1.2.1 正則化

ここでいう正則化とは、標本共分散行列を変換して、非特異であり、偽の相関に汚染されておらず、変数の数の2乗よりも少ない数で指定されるような行列を得ることをいう。同時に、その行列はもとの標本共分散の性質をある程度維持していることが期待されている。以下で、正則化の手続きをまとめることとする。正則化がどの空間で行われたかで区別して、(1)物理空間(格子空間)、(2)周波数空間(フーリエ空間)、(3)ウェーブレット空間と分けて紹介する。

標本共分散行列を正則化するもっとも単純な方法は、その行列を物理空間内で対角としてしまうことである。すなわち、標本共分散行列の非対角成分は強制的にゼロとするのである。対角とすれば行列は明らかに非特異で、その表現に必要なパラメータ数は変数の数と等しい。このやりかたは多くのデータ同化実験で広く採用されている。しかし、対角の仮定というのは、離れた2地点間の偽の相関を消すだけでなく、近くのモデル格子点や観測地点間の現実的な相関構造も無視してしまうことになる。

共分散行列の正則化の第2の方法は、パラメータ化された解析関数により行列の要素が表現されると仮定することである。そのような関数は、正值定符号共分散を与えることが知られているものから選び、例えば空間統計学で知られる Matérn 族などの共分散関数族から選ぶ(例えば Diggle and Ribeiro, Jr., 2007)。データ同化では、ガウス関数や自己回帰関数が相関関数としてよく知られている(例えば Daley, 1991, p. 117)。Frehlich (2006)は、構造関数を用いて、空間変化のある観測ノイズ行列を構築した。

デジタルフィルタにより、ガウス分布を近似する共分散を構築できる。反復的なラプラスフィルタ (Derber and Rosati, 1989; Egbert et al., 1994; Weaver and Courtier, 2001) やガウス再帰フィルタ (Lorenc, 1992; Lorenc, 1997; Purser et al., 2003a; Purser et al., 2003b) が開発された。デジタルフィルタを用いた相関は、複雑な横の境界を持つ海洋モデルに適している。一方、横の境界がない大気モデルでは、後述するような球面調和展開が使われることが多い。最近の研究では、デジタルフィルタは、非常に小さいアンサンブルの標本共分散行列に存在する高周波短距離振動の減衰に用いられている (Keppenne et al., 2008; Raynaud et al., 2008)。

コンパクトな台を持つ相関関数は、あるカットオフ距離以遠で本質的にゼロとなる相関を表すが、これは偽の長距離相関を消すのに使われる (Gaspari and Cohn, 1999; Gaspari et al., 2006; Gneiting, 1999; Gneiting, 2002)。コンパクトな台を持つ相関関数は、アンサンブルカルマンフィルタ (EnKF, Evensen, 2003) の実装でしばしば使われ、相関関数とアンサンブル共分散とのシュア積 (アダマール積) をとることで、小さいアンサンブルサイズによる偽のアンサンブル相関を減少させる (Houtekamer and Mitchell, 2001; Hamill et al., 2001)。流れに依存した緩和関数 (moderation function) も提案されている (Bishop and Hodyss, 2007)。

第2の正則化の空間は周波数空間である。Hollingsworth and Lönnberg (1986) および Lönnberg and Hollingsworth (1986) は、相関関数を円柱調和関数 (フーリエ=ベッセル展開) でモデル化し (例えば Daley, 1991 の Appendix G) これは球面上の接平面上の相関関数の局所解析に用いられた。球面上そのものにおいては、球面調和関数 (ルジャンドル展開) を用いて、等方相関関数がモデル化された。周波数空間での対角行列を用いて、Parrish and Derber (1992) は一様等方な相関を表現して、観測ノイズに用いた。Courtier et al. (1998) はブロック対角行列を用いて背景誤差共分散をモデル化し、鉛直相関スケールを水平スケールに依存させた。コンパクトな台をもつ相関関数は、周波数空間でも使われており、Buehner and Charron (2007) は、サンプリング誤差を減少させるために、周波数の局所化が重要であることを指摘した。

第3の正則化空間、ウェーブレット空間は、ウェーブレットを用いた物理空間と周波数空間の正則化であり、両空間で局所化がなされる。ウェーブレット空間において、標本共分散は対角な共分散行列 (Fisher and Andersson, 2001; Deckmyn and Berre, 2005) もしくは対角要素と一部の非対角要素をもつ共分散行列 (Nychka et al., 2002; Rhodin and Anlauf, 2007) により正則化される。コンパクトな台をもつ相関関数と同様に、ウェーブレットは標本共分散の偽の誤差を落とすことができる (Pannekoucke et al., 2007)。

1.2.2 力学的拘束条件

共分散正則化に加えて、力学的な拘束条件を課すことも共分散構造のモデル化の方法の一つである。準地衡風理論に基づいて、Desroziers (1997) は地衡風空間で等方相関を仮定し、それを物理空間に座標変換し、非等方な流れ依存相関をモデル化した。つり合い演算子を導入して、変数の非つり合い部分に関する水平に一様・等方な共分散を変換し、力学的関係に従って変数のつり合い部分の効果も含めた変数の共分散を得ている (Gauthier et al., 1998; Derber and Bouttier, 1999; Cullen, 2003)。Riishøjgaard (1998) は、背景状態推定値間の差に依存した要素をもつ行列として場の項を導入し、流れの依存を考慮に入れている。

1.3 逆行列空間での正則化

本稿では、最近 Ueno and Tsuchiya (2009) により提案された新たな共分散行列の正則化の方法を解説する。正則化を行う空間は第4の空間、逆行列空間である。標本共分散行列に基づいて、逆行列をモデル化し、有効な共分散行列を得る。後述するように、逆行列は変数の条件付き独立性の概念に関係している。この概念は、先述した物理、周波数、ウェーブレット空間で

のモデル化では確率の概念が含まれていなかったことと対照的である。

2 節では、共分散行列の基本的な統計量である標本共分散行列を振り返り、ガウス分布の共分散行列の最尤推定量であることを見る。3 節では、標本共分散行列のランク不足を克服するため、逆行列の要素をパラメタ化する。これは、共分散選択モデル (Dempster, 1972) の考え方を用いるものであり、最近ではガウシアングラフィカルモデル (Lauritzen, 1996) として知られる。4 節では、 3×3 行列を用いた簡単な例を示す。大規模問題に特に有用な実装技術を 5 節で述べる。6 節では、海面高度観測データから得られた標本共分散行列に対する応用例を示す。これは観測ノイズの共分散行列に使うことができる。7 節では、逆行列空間での正則化共分散の性質を述べ、8 節に結論を示す。

2. 標本共分散行列

標本共分散行列の基本概念を見てみよう。いま、 n 次元ベクトルデータ $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_e}$ が得られたとき、標本平均と標本共分散はそれぞれ

$$(2.1) \quad \bar{\mathbf{x}} = \frac{1}{N_e} \sum_{i=1}^{N_e} \mathbf{x}_i,$$

$$(2.2) \quad \mathbf{S} = \frac{1}{N_e} \sum_{i=1}^{N_e} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})',$$

と定義される。

標本平均と標本共分散はガウス分布の最尤推定量として知られている。平均 $\boldsymbol{\mu}$ と共分散 $\boldsymbol{\Sigma}$ をもつガウス分布

$$(2.3) \quad p(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right],$$

を考える。データ $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N_e}$ が与えられたとき、対数尤度は

$$(2.4) \quad \ell(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{N_e} \log p(\mathbf{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$(2.5) \quad = \sum_{i=1}^{N_e} \left[-\frac{n}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right]$$

$$(2.6) \quad = -\frac{N_e}{2} [n \log 2\pi - \log |\boldsymbol{\Sigma}^{-1}| + \text{tr} \mathbf{S} \boldsymbol{\Sigma}^{-1} + (\bar{\mathbf{x}} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})],$$

となる。 $\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の最尤推定量は、 $N_e \geq n + 1$ (例えば、Magnus and Neudecker, 1999) のとき、それぞれ標本平均と標本共分散

$$(2.7) \quad \tilde{\boldsymbol{\mu}} = \bar{\mathbf{x}},$$

$$(2.8) \quad \tilde{\boldsymbol{\Sigma}} = \mathbf{S}$$

である。不等式 $N_e \geq n + 1$ は、(2.6) で与えられる対数尤度のヘッセ行列が負値定符号となるための必要条件であり、 $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ が厳密に極大であることを保証するものである。もし反対に $N_e \leq n$ の場合、 $\bar{\mathbf{x}}$ と \mathbf{S} は停留点条件を満たすものの (勾配ベクトルがゼロになるものの)、ヘッセ行列は特異になり、その結果 $(\bar{\mathbf{x}}, \mathbf{S})$ が最大点か最小点か、もしくは鞍点となるかはわからない。

3. 逆行列空間での標本平均の正則化

3.1 ガウシアングラフィカルモデル

前節で述べたように、標本共分散 S ((2.2)式)は $N_e \leq n$ のときには必ずしも共分散 Σ の最尤推定量ではない。それに加え、 S は $n \times n$ 行列だがランクは高々 $N_e - 1$ であるため特異になりうる。この望ましからぬ性質が現れるのは、共分散行列の自由パラメータの数が、標本の数と比べて大きいためである。

こういった難点を回避するため、 Σ に構造を仮定して自由度を減らし、 Σ を少ない標本から推定できるようにする。具体的には、 Σ の通常の物理(格子)空間ではなく、逆行列空間に疎な構造を仮定する。すなわち、 Σ^{-1} の大半の要素にゼロを割り当てる。このような共分散行列の逆行列が疎であるようなモデルは共分散選択モデル、もしくはガウシアングラフィカルモデル (Dempster, 1972; Lauritzen, 1996) と呼ばれる。3変数の単純な例を図 2(a) に示す。この例では、 σ^{13} と σ^{31} がゼロと置かれる。ここで、 σ^{ij} は Σ^{-1} の (i, j) -要素を示す。

Σ^{-1} のゼロ要素は、対応する変数の組が、残りの変数が与えられたときに独立であることに同値である (例えば, Lauritzen, 1996 の Proposition 5.2)。例として、図 2(a) に示される共分散行列の逆行列は、変数 2 が与えられたときに変数 1 と変数 3 が条件付き独立であることと等価である。この条件付き独立性は、図 2(b) に示される無向グラフで示され、頂点 1, 2 と頂点 2, 3 間には辺があるが、頂点 1, 3 間には辺がない。

Σ^{-1} の残りの非ゼロ要素は最尤法で推定し $\hat{\Sigma}^{-1}$ を得、その結果 $\hat{\Sigma}$ を得る。最尤推定値 $\hat{\Sigma}^{-1}$ は次のような驚くべき性質がある。それは、物理空間の共分散 $\hat{\Sigma} \equiv (\hat{\Sigma}^{-1})^{-1}$ が正値定符号 (よって非特異) であり、 Σ^{-1} に非ゼロ要素を仮定した位置では、もとの標本共分散行列 S と同一の値をとることである (例えば, Lauritzen, 1996 の Theorem 5.3)。3次元の例で、 $\sigma^{13} = \sigma^{31} = 0$ (図 2(a)) のとき、 $\hat{\Sigma}$ は、(1,1), (1,2), (2,1), (2,2), (2,3), (3,2), (3,3) 要素で S と同一の値を持つ。この位置は、 Σ^{-1} に非ゼロ要素を仮定した位置と一致する。

3.2 逆行列の非ゼロ要素の指定

ここでは、ガウシアングラフィカルモデル Σ^{-1} を、その非ゼロ要素を指定することで構築する。変数間の局所的な相互作用を吟味する前に、何れともあれ、 Σ^{-1} のすべての対角要素は正としてよいであろう。これは、 Σ の正定値性からの結果である。ここで止めれば、 Σ^{-1} は対角となるため、その逆行列である共分散行列 Σ も対角となる。この状況は、どの変数の組も独立で、物理空間で対角行列を仮定した場合と等しい。この条件をグラフで書くと、頂点のみがあって辺がないものとなり、図 3(a) のようになる。

続いて、非対角要素を考える。 Σ^{-1} の非対角要素は、変数の空間配置に基づいて導入する。例えば 2次元格子空間を考えると、東西・南北の 2 方向の最近格子点 4 点との局所相互作用を仮定するのは自然だろう。この関係をグラフで書くと、中央の頂点から 4 つの隣接頂点へ 4 本の辺を持つものとして表現される (図 3(b))。これを 4 近傍モデルと呼ぶことにする。このグ

(a) Inverse covariance matrix (b) Undirected graph

$$\Sigma^{-1} = \begin{pmatrix} \sigma^{11} & \sigma^{12} & 0 \\ \sigma^{12} & \sigma^{22} & \sigma^{23} \\ 0 & \sigma^{23} & \sigma^{33} \end{pmatrix} \quad \begin{array}{ccc} \circ & \text{---} & \circ & \text{---} & \circ \\ 1 & & 2 & & 3 \end{array}$$

図 2. 3 変数 (変数 1, 2, 3 とする) に対して、変数 2 が与えられたとき、変数 1, 3 が独立であることの (a) 共分散行列の逆行列による表現、および (b) 無向グラフによる表現。

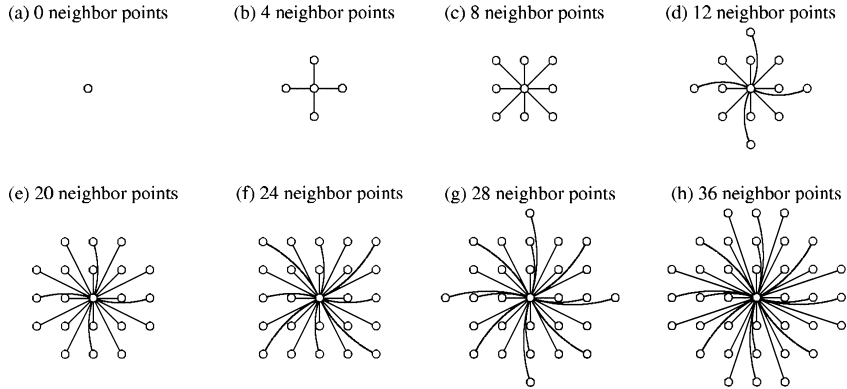


図 3. 変数間の独立性のグラフ表現. 中央の頂点は他の変数と直接の関係がある(辺により示す). 図(a)は中央の変数は他の変数と独立である. それ以外の図では, 中央の変数は近傍の変数と相互作用をする. (b) 4 近傍, (c) 8 近傍, (d) 12 近傍, (e) 20 近傍, (f) 24 近傍, (g) 28 近傍, (h) 36 近傍の各モデル.

ラフが示す配置は, 4つの隣接変数の外側にある変数との条件付き独立性も示している. 共分散行列の逆行列 Σ^{-1} は, 4つの隣接変数に関して非ゼロ要素を持ち, より遠くの変数に関してはゼロ要素を持つ.

斜めの位置にある格子点の変数との相互作用を含める場合は, 斜めの頂点も近傍とみなす(すなわち, 中央の頂点と辺で結ばれる) (図 3(c)). これは 8 近傍モデルとなり, 中央の変数はそれを取り囲む 8 変数とのみ相互作用が許され, それら 8 変数より外側の変数とは条件付き独立である. 図 3(d)–(h) のように, 12 近傍モデル, 20 近傍モデル, ..., 36 近傍モデルと, 同様に頂点間の辺を加えていくことで, より複雑な構造を持つモデルも構成できる.

どのようなグラフ配位にしても, いったん Σ^{-1} の非ゼロ要素を仮定すれば, 逆行列は定数行列の線形結合で書くことができ, それらの係数の関数

$$(3.1) \quad \Sigma^{-1}(\beta) = \sum_{k=1}^m \mathbf{A}_k \beta_k$$

として表現できる. ここで, \mathbf{A}_k ($k=1, \dots, m$) は相互作用する変数の組の非ゼロ要素を示す定数行列, β_k ($k=1, \dots, m$) はこれから推定するスカラー係数, $\beta \equiv (\beta_1 \dots \beta_m)'$ は係数の集合からなるベクトルである. 具体的には, 一つの対角要素が 1 をとる行列, もしくは相互作用をする組に関する 2 つの非対角要素が 1 をとる行列から \mathbf{A}_k ($k=1, \dots, m$) のそれぞれを選ぶことができる. また, 一般性を失うことなく, 対角要素の値は 1 ではなく 2 とすることもできる. このように選ぶと, アルゴリズムの構築に便利である(5 節).

3.3 パラメータの最尤推定

(3.1) 式に示すように, $\Sigma^{-1} = \Sigma^{-1}(\beta)$ の条件のもとで, 最尤法により μ と β を推定する. 対数尤度関数は

$$(3.2) \quad \ell(\mu, \beta) = -\frac{N_e}{2} \left[n \log 2\pi - \log |\Sigma^{-1}(\beta)| + \text{tr} \mathbf{S} \Sigma^{-1}(\beta) + (\bar{\mathbf{x}} - \mu)' \Sigma^{-1}(\beta) (\bar{\mathbf{x}} - \mu) \right].$$

μ の停留点は, ここでも

$$(3.3) \quad \hat{\mu} = \bar{\mathbf{x}},$$

であり、 Σ が非特異であれば最尤推定量となる。 β に関しては、停留点を解析的に得ることはできない。そこで、最適な β を、

$$(3.4) \quad \ell(\bar{x}, \beta) = -\frac{N_e}{2} [n \log 2\pi - \log |\Sigma^{-1}(\beta)| + \text{tr} \mathbf{S} \Sigma^{-1}(\beta)]$$

を最大化するものとして数値的に探索する。

(3.4)式で与えられる対数尤度を最大化することは、以下で示す目的関数を最小化することと同値である：

$$(3.5) \quad f(\beta) = \text{tr} \mathbf{S} \Sigma^{-1}(\beta) - \log |\Sigma^{-1}(\beta)|.$$

ここでは、 $f(\beta)$ を最小化する解 $\hat{\beta}$ をニュートン法を用いて見つける。 $\Sigma^{-1}(\beta)$ が式(3.1)のように表現されているとき、目的関数(3.5)の勾配ベクトル、ヘッセ行列が解析的に得られることを利用する。ニュートン方向 $\Delta\beta$ は

$$(3.6) \quad \Delta\beta = -[\nabla^2 f(\beta)]^{-1} \nabla f(\beta)$$

により与えられる。ここで、勾配ベクトル $\nabla f(\beta) = (\partial f / \partial \beta_i)$ およびヘッセ行列 $\nabla^2 f(\beta) = (\partial^2 f / \partial \beta_i \partial \beta_j)$ は次のように得られる：

$$(3.7) \quad \frac{\partial f}{\partial \beta_i} = \text{tr} \mathbf{S} \mathbf{A}_i - \text{tr} \Sigma(\beta) \mathbf{A}_i,$$

$$(3.8) \quad \frac{\partial^2 f}{\partial \beta_i \partial \beta_j} = \text{tr} \Sigma(\beta) \mathbf{A}_i \Sigma(\beta) \mathbf{A}_j.$$

これら微分係数の導出は Ueno and Tsuchiya (2009) を参照されたい。

最尤推定は凸最適化問題として定式化され(例えば、Boyd and Vandenberghe, 2004 の Chapter 4)、ニュートン法により尤度関数を(大域的に)最大化することで有効かつ厳密に計算が可能である。近年の凸最適化理論によると、この最適化問題は多項式時間で解ける(Vandenberghe et al., 1998; Boyd and Vandenberghe, 2004; Tsuchiya and Xia, 2007)。

3.4 近傍の数の選択

図3に示すような異なる近傍を持つモデルを用いて正規化共分散行列を得ることができるが、データ同化の実施には、ひとつの共分散行列があれば十分である。そこで、異なるモデルから推定される共分散行列のなかから好ましいものを選ぶ。一つの方法は、情報量規準を用いることである(例えば、Konishi and Kitagawa, 2007)。係数の数 $m = \dim \beta$ を AIC (Akaike, 1974)

$$(3.9) \quad \text{AIC} = -2\ell(\bar{x}, \hat{\beta}) + 2m,$$

もしくは BIC (Akaike, 1977; Schwarz, 1978)

$$(3.10) \quad \text{BIC} = -2\ell(\bar{x}, \hat{\beta}) + m \log N_e,$$

を用いて選ぶ。ここで、

$$(3.11) \quad \hat{\beta} = \underset{\beta}{\text{argmax}} \ell(\bar{x}, \beta)$$

とおいた。異なる近傍を持つ最尤推定モデルの中で、AIC もしくは BIC を最小化するモデルを見つけることで、最もよいモデルを選択することができる。

4. 簡単な例

特異である標本共分散行列の数値例を用いて、正則化の手続きを紹介する。3×3 標本共分散行列

$$(4.1) \quad \mathbf{S} = \begin{pmatrix} 0.884 & 0.780 & 0.028 \\ 0.780 & 0.876 & 0.458 \\ 0.028 & 0.458 & 1.005 \end{pmatrix},$$

を考える。これは、3つの標本

$$(4.2) \quad \mathbf{x}_1 = (0.573 \quad 0.223 \quad -1.366)',$$

$$(4.3) \quad \mathbf{x}_2 = (0.190 \quad 0.930 \quad 1.042)',$$

$$(4.4) \quad \mathbf{x}_3 = (-1.585 \quad -1.312 \quad -0.578)',$$

とそれらの標本平均

$$(4.5) \quad \bar{\mathbf{x}} = (-0.274 \quad -0.053 \quad -0.301)'$$

から求めたものである。S のランクは 2 であるから (標本の大きさマイナス 1), S はランク不足であり、実際

$$(4.6) \quad |\mathbf{S}| = 0.000$$

である。S を正則な行列 Σ に変換したい。

図 2 に示すガウシアングラフィカルモデルを考える。変数 1 と変数 3 は、変数 2 が与えられたときに独立であり、正則化された逆行列 Σ^{-1} は (1,3)-要素および (3,1)-要素にゼロを持つ：

$$(4.7) \quad \Sigma^{-1}(\boldsymbol{\beta}) = \begin{pmatrix} \beta_1 & \beta_4 & \\ \beta_4 & \beta_2 & \beta_5 \\ & \beta_5 & \beta_3 \end{pmatrix},$$

ここで、 $\boldsymbol{\beta} = (\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5)'$ であり、空白はゼロ要素を示す。式 (3.1) に示したように、 $\Sigma^{-1}(\boldsymbol{\beta})$ は 5 つの行列 $\mathbf{A}_1, \dots, \mathbf{A}_5$:

$$(4.8) \quad \Sigma^{-1}(\boldsymbol{\beta}) = \sum_{k=1}^5 \mathbf{A}_k \beta_k,$$

の線形結合により表現できる。ここで、

$$(4.9) \quad \begin{aligned} \mathbf{A}_1 &= \begin{pmatrix} 1 & & \\ & & \\ & & \end{pmatrix}, & \mathbf{A}_2 &= \begin{pmatrix} & & \\ & 1 & \\ & & \end{pmatrix}, \\ \mathbf{A}_3 &= \begin{pmatrix} & & \\ & & \\ & & 1 \end{pmatrix}, & \mathbf{A}_4 &= \begin{pmatrix} & & \\ & 1 & \\ 1 & & \end{pmatrix}, \\ \mathbf{A}_5 &= \begin{pmatrix} & & \\ & & \\ & & 1 \\ 1 & & \end{pmatrix} \end{aligned}$$

とおいた。

対数尤度 (3.4) は,

$$(4.10) \quad \ell(\bar{\mathbf{x}}, \boldsymbol{\beta}) = -\frac{N_e}{2} [3 \log 2\pi - \log |\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta})| + \text{tr} \boldsymbol{S} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta})]$$

となる.

式 (4.7) の形の逆行列を考え, 対数尤度 (4.10) を最大化することで $\beta_{1:5}$ を推定すると, 逆行列

$$(4.11) \quad \hat{\boldsymbol{\Sigma}}^{-1} = \begin{pmatrix} 5.293 & -4.715 & 0.000 \\ -4.715 & 5.700 & -0.684 \\ 0.000 & -0.684 & 1.307 \end{pmatrix}$$

を得る. $\hat{\boldsymbol{\Sigma}}^{-1}$ は確かに正則である ($|\boldsymbol{\Sigma}^{-1}| = 7.900$) ので, 逆行列を計算することができて, 正則化された共分散行列

$$(4.12) \quad \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.884 & 0.780 & 0.408 \\ 0.780 & 0.876 & 0.458 \\ 0.408 & 0.458 & 1.005 \end{pmatrix},$$

を得る. 当然のことながら, この行列も正則で, 行列式は有限値 $|\hat{\boldsymbol{\Sigma}}| = 0.127$ をとる.

表 1 は, いくつかのグラフィカルモデルを用いた時の正則化行列をまとめたものである. モデル (a) は, 3 変数が独立であることを仮定し, 逆行列は対角になる. 3 つのパラメータ $\beta_{1:3}$ で逆行列を表現できる.

表 1. 標本共分散行列 \boldsymbol{S} (4.1) から構成された正則化行列のまとめ. 各列は, 仮定したグラフィカルモデル, 推定した逆行列 ($\hat{\boldsymbol{\Sigma}}^{-1}$), 推定した共分散行列 ($\hat{\boldsymbol{\Sigma}}$), 共分散行列の行列式 ($|\hat{\boldsymbol{\Sigma}}|$), 最大尤度値, パラメータ数 (m), AIC, BIC を示す. AIC も BIC もモデル (b) で最小になる.

Model	Inverse: $\hat{\boldsymbol{\Sigma}}^{-1}$	Covariance: $\hat{\boldsymbol{\Sigma}}$	$ \hat{\boldsymbol{\Sigma}} $	$\ell(\bar{\mathbf{x}}, \hat{\boldsymbol{\beta}})$	m	AIC	BIC
(a)	$\begin{pmatrix} 1.131 & & \\ & 1.142 & \\ & & 0.995 \end{pmatrix}$	$\begin{pmatrix} 0.884 & & \\ & 0.876 & \\ & & 1.005 \end{pmatrix}$	0.778	-12.394	3	30.787	28.083
(b)	$\begin{pmatrix} 5.293 & -4.715 & \\ -4.715 & 5.342 & \\ & & 0.995 \end{pmatrix}$	$\begin{pmatrix} 0.884 & 0.780 & \\ 0.780 & 0.876 & \\ & & 1.005 \end{pmatrix}$	0.167	-10.079	4	28.158	24.553
(c)	$\begin{pmatrix} 1.132 & & -0.032 \\ & 1.142 & \\ -0.032 & & 0.996 \end{pmatrix}$	$\begin{pmatrix} 0.884 & & 0.028 \\ & 0.876 & \\ 0.028 & & 1.005 \end{pmatrix}$	0.778	-12.392	4	32.785	29.179
(d)	$\begin{pmatrix} 1.131 & & \\ & 1.500 & -0.684 \\ & -0.684 & 1.307 \end{pmatrix}$	$\begin{pmatrix} 0.884 & & \\ & 0.876 & 0.458 \\ & 0.458 & 1.005 \end{pmatrix}$	0.593	-11.985	4	31.969	28.364
(e)	$\begin{pmatrix} 5.295 & -4.715 & -0.032 \\ -4.715 & 5.342 & \\ -0.032 & & 0.996 \end{pmatrix}$	$\begin{pmatrix} 0.883 & 0.779 & 0.028 \\ 0.779 & 0.875 & 0.025 \\ 0.028 & 0.025 & 1.005 \end{pmatrix}$	0.166	-10.078	5	30.156	25.649
(f)	$\begin{pmatrix} 5.293 & -4.715 & \\ -4.715 & 5.700 & -0.684 \\ & -0.684 & 1.307 \end{pmatrix}$	$\begin{pmatrix} 0.884 & 0.780 & 0.408 \\ 0.780 & 0.876 & 0.458 \\ 0.408 & 0.458 & 1.005 \end{pmatrix}$	0.127	-9.670	5	29.340	24.833
(g)	$\begin{pmatrix} 1.132 & & -0.032 \\ & 1.500 & -0.684 \\ -0.032 & -0.684 & 1.308 \end{pmatrix}$	$\begin{pmatrix} 0.884 & 0.013 & 0.028 \\ 0.013 & 0.876 & 0.458 \\ 0.028 & 0.458 & 1.005 \end{pmatrix}$	0.592	-11.983	5	33.966	29.460

モデル (b)–(d) は 1 本の辺を持つ. 変数 1 と変数 2 が独立なのがモデル (b) で, モデル (c) 及び (d) も同様な仮定から作られている. 対応する逆行列はブロック対角で, 4 つのパラメータ $\beta_{1:4}$ を持つ. それぞれの共分散行列, すなわち逆行列の逆行列も, ブロック対角になっている. これらの 1 本の辺を持つ 3 つのモデルの中では, モデル (b) が最大の尤度値を与えるので, モデル (c) および (d) よりもよいと考えられる. 加えて, AIC もしくは BIC の値を見ると, モデル (b) はモデル (a) よりも好ましいことが分かる. モデル (c) とモデル (d) は, AIC および BIC 値がモデル (a) より大きいため, これらのモデルに加えた頂点は不要であることを意味している.

モデル (e)–(g) は 2 本の辺を持つ. モデル (e) では, 変数 2 と変数 3 が, 変数 1 が与えられたもとで条件付き独立, モデル (f), (g) でも同様な条件を仮定している. モデル (f) は本節の前半で用いたモデルである. これら 2 辺モデルの中では, モデル (f) が最大の尤度値を与える. しかし, AIC, BIC 値はモデル (b) より大きい. これより, モデル (f) は辺が 1 本のモデル (b) よりよくないことがわかる. モデル (g) は 2 本の辺を持つにも関わらず, 辺がないモデル (a) より悪いとみなされる. 以上のモデルで AIC, BIC 値を比べると, モデル (b) が, 正則条件を満たす, 最も好ましい共分散行列として選ばれる.

5. 実装方法

ベクトル x は状態ベクトルか観測ベクトルに対応するので, 次元は大きい, すなわち n は大きい. 近傍はないと仮定したとき, 非ゼロ要素の数 m は n と等しくなるが, 通常は近傍を仮定して m は n より大きいことより, n が大きい場合には m も大きくなる. 本節では, n が大きいときに特に役立つ実装技術を述べる.

定数行列 A_k ($k=1, \dots, m$) は, 次のような行列の集合から選ぶものとする: (1) 対角要素の一つに 2 を持つ (4 節のように 1 ではなく), または (2) 相互作用をする変数の 1 組を形成する対称位置にある 2 つの非対角要素に 1 を持つ. 図 2 に示すモデルを考えると, 正則化された逆行列 Σ^{-1} は次の 5 つの行列の線形結合で表現できる:

$$(5.1) \quad \begin{aligned} A_1 &= \begin{pmatrix} 2 & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}, & A_2 &= \begin{pmatrix} & & & & \\ & 2 & & & \\ & & & & \\ & & & & \\ & & & & \end{pmatrix}, \\ A_3 &= \begin{pmatrix} & & & & \\ & & & & \\ & & & & \\ & & & 2 & \\ & & & & \end{pmatrix}, & A_4 &= \begin{pmatrix} & & & & \\ & & & & \\ & & & 1 & \\ & & 1 & & \\ & & & & \end{pmatrix}, \\ A_5 &= \begin{pmatrix} & & & & \\ & & & & \\ & & & 1 & \\ & & 1 & & \\ & & & & \end{pmatrix}. \end{aligned}$$

A_k が非ゼロの値を $(p(k), q(k)), (q(k), p(k))$ 要素に持つとすると, A_k は次のように表される:

$$(5.2) \quad A_k = e_{p(k)} e_{q(k)}' + e_{q(k)} e_{p(k)}',$$

ここで, e_i は n 次元単位ベクトルで, 第 i 要素に 1 をもち, 他の要素は 0 を持つものとする. (5.2) 式を用いると, 目的関数 (3.5), 勾配ベクトル (3.7), ヘッセ行列 (3.8) を簡単にすることができる:

$$(5.3) \quad f(\beta) = 2 \sum_{k=1}^m \beta_k (S)_{p(k)q(k)} - \log |\Sigma^{-1}|,$$

$$(5.4) \quad \frac{\partial f}{\partial \beta_i} = 2(\mathbf{S})_{p(i)q(i)} - 2(\boldsymbol{\Sigma})_{p(i)q(i)},$$

$$(5.5) \quad \frac{\partial^2 f}{\partial \beta_i \partial \beta_j} = 2(\boldsymbol{\Sigma})_{p(i)p(j)} (\boldsymbol{\Sigma})_{q(i)q(j)} + 2(\boldsymbol{\Sigma})_{p(i)q(j)} (\boldsymbol{\Sigma})_{q(i)p(j)}.$$

導出は Ueno and Tsuchiya (2009) を参照されたい。これら 3 本の式より、式(3.5), (3.7), (3.8) に現れる行列の乗算が不要であることが分かる。

5.1 アルゴリズム

ニュートン法により、目的関数(5.3)を最小化する。

ステップ 1: 初期値 $\beta^{(0)} \in \mathbb{R}^m$ と収束判定基準値 $\varepsilon > 0$ を選ぶ。反復カウンターを $r = 0$ とする。

ステップ 2: ニュートン方向 $\Delta\beta^{(r)}$ を、線形方程式

$$(5.6) \quad [\nabla^2 f(\beta^{(r)})] \Delta\beta^{(r)} = -\nabla f(\beta^{(r)})$$

の解として求める。

ステップ 3: $\delta^2(\beta^{(r)}) \leq \varepsilon$ ならば終了。ここで、

$$(5.7) \quad \delta^2(\beta^{(r)}) = -\nabla f(\beta^{(r)})' \Delta\beta^{(r)}$$

とおいた。

ステップ 4: $\beta^{(r)}$ から $\beta^{(r+1)}$ へ更新する:

$$(5.8) \quad \beta^{(r+1)} = \beta^{(r)} + 2^{-t} \Delta\beta^{(r)}$$

ここで、

$$(5.9) \quad f(\beta^{(r)} + 2^{-t} \Delta\beta^{(r)}) < f(\beta^{(r)}).$$

を満たす最小の $t \in \{0, 1, \dots\}$ を用いる。

ステップ 5: $r = r + 1$ とし、ステップ 2 に進む。

5.2 注意

変数間で標本分散が大きく異なるときは、数値的安定性のため、標本共分散行列 \mathbf{S} の代わりに標本相関行列 \mathbf{R} を正則化する方がよい。ここで、 $(\mathbf{R})_{ij}$ は $(\mathbf{S})_{ij} / \sqrt{(\mathbf{S})_{ii}(\mathbf{S})_{jj}}$ と定義する。

ステップ 1 で、初期値 $\beta^{(0)}$ は $\boldsymbol{\Sigma}^{-1}(\beta^{(0)})$ が正定値行列となるように選ぶ必要がある。これは、(5.9) 式の右辺 $f(\beta^{(0)})$ が定義されるための必要条件である。正定値性を得る簡便な方法は、対角行列を仮定することである。一つの例は、

$$(5.10) \quad \sum_{k=1}^m \mathbf{A}_k \beta_k^{(0)} = \text{diag } \mathbf{S}^{-I},$$

となるように $\beta^{(0)} = (\beta_1^{(0)} \dots \beta_m^{(0)})'$ を選ぶことである。ここで、 \mathbf{S}^{-I} は標本共分散行列 \mathbf{S} の一般逆行列である。もう一つの方法は、計算負荷が少なく済む方法で、 $\beta^{(0)}$ を

$$(5.11) \quad \sum_{k=1}^m \mathbf{A}_k \beta_k^{(0)} = \frac{n}{\text{tr } \mathbf{S}} \mathbf{I}_n,$$

となるように選ぶ。ここで、 \mathbf{I}_n は $n \times n$ の単位行列である。

ステップ 2 で、勾配ベクトル $\nabla f(\beta^{(r)})$ とヘッセ行列 $\nabla^2 f(\beta^{(r)})$ を式(5.4)および(5.5)を用いて計算する。これら 2 つの式には $\boldsymbol{\Sigma}$, すなわち $\boldsymbol{\Sigma}^{-1}(\beta^{(r)})$ の逆行列が必要である。 $\boldsymbol{\Sigma}^{-1}(\beta^{(r)})$ は疎であることを仮定しているので、 n が大きい場合でもこの計算は実行可能である。

(5.6)式は m 変数の 1 次連立方程式である．通常 m は n より大きく，ヘッセ行列 (5.5) は密行列であるため，(5.6) を解く際にもっとも計算負荷がかかる． m が大きいとき，ヘッセ行列は複数メモリに分散させる必要がある．ヘッセ行列は正定値対称であるため，並列線形計算ソルバー，例えば ScaLAPACK に含まれる PDPOSV を用いることができる．PDPOSV は，並列処理によって，コレスキー分解により分散保持された正定値行列を分解するルーチンである．

ステップ 3 では， $\delta(\beta^{(r)})$ はニュートン減少量を表すが，これはもともと

$$(5.12) \quad \delta(\beta^{(r)}) = \sqrt{\nabla f(\beta^{(r)})' \nabla^2 f(\beta^{(r)})^{-1} \nabla f(\beta^{(r)})}$$

として定義されている (Boyd and Vandenberghe, 2004, Section 9.5.1)．収束判定基準に $\delta^2(\beta^{(r)})$ を用いているが，その理由は $\delta^2(\beta^{(r)})$ が $f(\beta^{(r)}) - \min_{\beta} f(\beta)$ の推定値であること， $-\delta^2(\beta^{(r)})$ がニュートン方向での $\beta^{(r)}$ での $f(\beta)$ の方向微分と解釈できることによる (Boyd and Vandenberghe, 2004, Section 9.5.1)．以下の応用では，収束判定基準値を $\varepsilon = 10^{-10}$ とする (6 節)．

ステップ 4 では，ニュートン方向 $\Delta\beta^{(r)}$ を 2^{-t} 倍して用いる．この 2^{-t} はスケールパラメータと呼ばれ，移動距離を徐々に短くしながら，ニュートン方向に沿った最小点をおおざっぱに探索する．スケールパラメータはニュートン法の原型には存在しないが，2 次式で近似し切れない目的関数の形状の可能性を考慮して，目的関数値が確実に減少するように導入しているものである．反復の最初の方では (r が小さいとき) 大きい t が選ばれ，反復が終了に近づくにつれて t が小さくなる (例えば $t=0$) 傾向がある．

式 (5.9) の計算には $\log|\Sigma^{-1}|$ が必要である．ここでも， Σ^{-1} が疎行列であるため，この項は膨大な計算負荷を必要とせずに計算できる．ここで注意しておきたいのは， $\beta^{(r)}$ の更新時に (5.9) が満たされている限り， $\Sigma(\beta^{(r)})$ の正定性は保証されていることである．もし，少なくとも一つの固有値がゼロか負になった場合，(5.3) 式の第 2 項は定義できず，その結果 (5.9) の条件は満たされない．

6. 適用例

ここでは適用例として，人工衛星 TOPEX/POSEIDON (T/P) による海面高度計データから作られた標本共分散行列に対して，この正則化法を適用する．標本共分散行列は海面高度 (SSH, sea-surface height の略) からトレンド成分を差し引いて得た残差の系列から計算したものである (Ueno et al., 2007)．全時間ステップ数は $N_e = 364$ (1992 年 9 月 23 日から 2002 年 8 月 11 日にわたる，T/P サイクルに対応している) である．

以下の 2 つの分節ではそれぞれ，太平洋赤道域および全球での標本共分散行列を正則化する．それらの適用例で，変数の次元は $n = 503$ および $n = 8585$ である．いずれの場合においても， $n \times n$ 行列はランク不足 (ランクは $N_e - 1$ 以下であるため) であり，したがって特異である．初めの適用例では，正則化のすべての手続きを示し，得られた行列を吟味する．第 2 の適用例では，大規模共分散行列へも正則化プロセスが適用可能であることを示すことを目的とする．

6.1 適用例 1: 太平洋赤道域

6.1.1 標本共分散行列

太平洋赤道域を東西方向 $L_x = 8^\circ$ 間隔，南北方向 $L_y = 2^\circ$ 間隔のデータから作成した標本共分散行列を正則化する．データ点の総数は $n = 503$ となる．図 4 (a) は，トレンド成分を差し引いたデータの標本共分散行列 S を示している．対角要素が卓越しているが，非対角要素も意味のある値を持っているように見える．図 4 (a) で破線で囲んだ領域 $[278, 364] \times [278, 364]$ 内の要素を拡大して，図 4 (b) に示す．非対角要素にも意味がありそうなパターンがあることがわかる．例えば，(293, 323) 要素および (322, 353) 要素周りの正負の組，(290, 350) 要素周りの正の

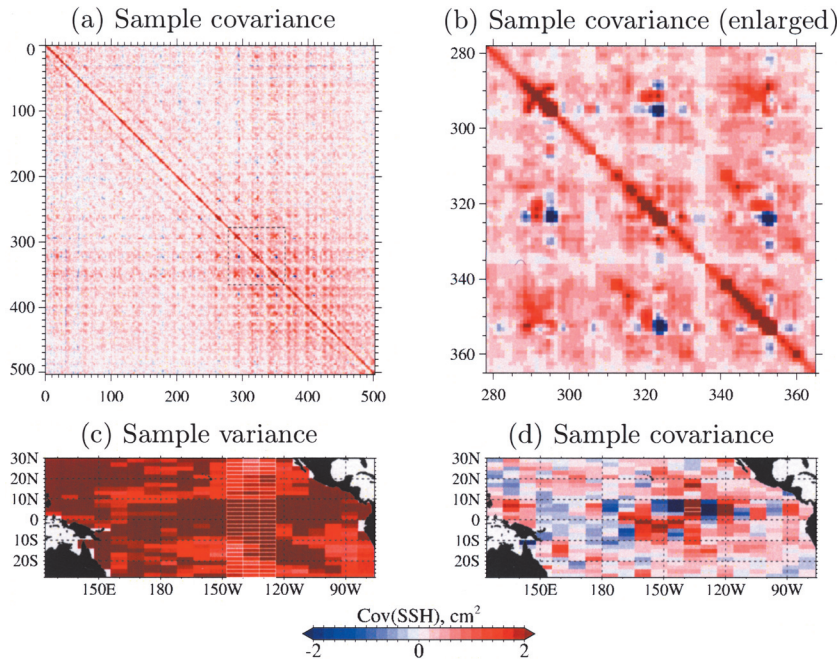


図4. 海面高度 (SSH) 観測データからトレンド成分を除去して計算した標本共分散行列 (a) 標本共分散行列そのもの, (b) [278, 364] に含まれる要素, (c) 分散 (共分散行列の対角要素), (d) 323 列目の要素 ((136°W, 5°N) に対応, 白で縁取り). パネル (b) で拡大した領域は, パネル (a) の破線の正方形およびパネル (c) の白く縁取りしたメッシュに対応している.

要素などである. 図 4(c) には対角要素, すなわち分散を示す. 148°E–124°W, 28°S–30°N 内の白で縁取ったメッシュは図 4(b) の領域に対応している. 図 4(d) は 323 列目 (136°W, 5°N に対応, 白で縁取りしている) に並ぶ分散・共分散要素を示している. 南北方向には正の共分散が現れるが, 東西方向に並ぶ隣接観測点では強い負の共分散が現れる. 図 4(b) の (293, 323) 周りの正負のペアは (144°W, 1°S) 周りの正の共分散と (144°W, 7°N) 周りの負の共分散に対応している.

6.1.2 逆行列空間での正則化

図 4 に示す標本共分散行列の正則化を行う. 図 3 に示す 8 つのグラフィカルモデルを適用し, それらのパラメータを最尤法とニュートン法を用いて推定する. 図 5 に示すのは, 推定された共分散行列の逆行列 Σ^{-1} と, それに対応した共分散行列 $\hat{\Sigma}$ である. 逆行列の要素であらかじめ指定した非ゼロ要素は縁取りしてある. 近傍を仮定しないとき (図 5(a)), 逆行列は対角となり, 対応する共分散行列は当然のことながら対角になる.

近傍を仮定すると, 共分散行列は全要素に有限値を持つようになる. 4 近傍を仮定すると (図 5(b)), 逆行列の大半の要素はゼロであるが, 対応する共分散行列は標本共分散行列 (図 4(b)) と似たパターンを持ち, 例えば (325, 350) 要素周りの負の要素も見られる. 共分散要素のパターンは, 8 近傍の場合も同様である (図 5(c)). 12 近傍の場合 (図 5(d)), 対応する共分散の値の大きさは小さくなり, 大きい値は逆行列の非ゼロ要素の周りのみに存在する. 20-, 24-, 28-, 36-近傍モデルを仮定すると (図 5(e), (f), (g), (h)), 共分散パターンの色は非ゼ

表 2. ガウシアングラフィカルモデルを用いて推定した共分散行列のまとめ. 異なる近傍数を仮定したときの近傍数, $\log|\hat{\Sigma}|$, 最大尤度値 ($\ell(\hat{\alpha}, \hat{\beta})$), パラメータ数 ($m = \dim(\beta)$), AIC, BIC 値. AIC の極小は 12 近傍モデルで得られ, 最小は 28 近傍モデルで得られる. BIC については, 極小は 4 近傍で得られ, 最小は 12 近傍モデルで得られる.

Number of neighbors	$\log \hat{\Sigma} $	$\ell(\hat{\alpha}, \hat{\beta})$	m	AIC	BIC
0	418.4	-295336.7	503	591679.3	593574.8
4	54.1	-237047.2	1457	477008.4	482498.9
8	43.6	-235373.2	2375	475496.5	484446.2
12	9.2	-229860.5	3278	466277.0	478629.5
20	-1.7	-228122.3	5028	466300.5	485247.7
24	-6.7	-227324.6	5863	466375.2	488468.9
28	-15.0	-225985.6	6716	465403.2	490711.2
36	-22.8	-224741.8	8374	466231.5	497787.4

ロ要素の数の増加に応じて薄くなる.

6.1.3 好ましい近傍数

図 5 に示す共分散行列のうち, AIC もしくは BIC を用いて好ましい共分散行列を選択する. 図 6 は, 近傍数の関数として AIC および BIC の変化を示し, 実際の値は表 2 にまとめてある. AIC, BIC もいずれも, 最小近傍数 4 を仮定しただけで値が大きく減少する (およそ 10^5). ここからわかるのは, 単なる対角行列 (0 近傍モデル) と比べると, 非対角行列はずっと好ましいということである. 8 近傍モデルでは, AIC は減少を続ける一方で, BIC は増加し, 斜め方向の近傍は有効に働かないと考えられる. ここでグラフィカルモデルの適用を止めて, 4 近傍モデルを好ましいモデルとして選択することも可能であろう.

12 近傍モデルを適用すると, AIC も BIC も極小値をとる. BIC については, 12 近傍モデルの値は 4 近傍モデルの値よりも小さく, さらにより多くの近傍を持つモデルよりも小さいことから, 12 近傍モデルは最小 BIC 値を持つとわかる. AIC は, 28 近傍モデルが最小値を示す (表 2). AIC および BIC の吟味から, 次のうち一つを選ぶことができよう.

- (1) 4 近傍モデルを選択, 理由は BIC が最初の極小をとるため.
- (2) 12 近傍モデルを選択, 理由は AIC が最初の極小をとる, または BIC が最小をとるため.
- (3) 28 近傍を選択, 理由は AIC が最小をとるため.

ここでは, 8 つの試行モデルのうちで好ましいモデルとして, 12 近傍モデル (図 5(d)) を選択して以下で見てみることにする. 図 7 は, 選択した共分散行列を示す. 左上から右下にかけてのバンド構造が強調される一方で, バンド間の要素は小さい値になっている. これは, 遠地点間の相関が小さくなったことを示している (図 7(d)). 図 7(c) に示す推定分散は, 標本統計量のそれと同一の値である (図 4(c)).

6.2 適用例 2: 全球海洋

6.2.1 標本共分散と正則化された共分散

全球海洋の高度データからトレンドを除いた量に関する標本共分散に対して, グラフィカルモデルを適用する. 観測地点は格子点上に分布しており, 空間分解能は, 東西方向に $L_x = 2^\circ$, 南北方向に $L_y = 2^\circ$ である. データ地点の総数は $n = 8,585$ である. 簡単のため, 経度 0° をまたぐ変数間の隣接関係は無視している. 提案アルゴリズム (5 節) が次元の大きい行列に対しても適用可能であることを示したいので, この仮定は妥当であろう.

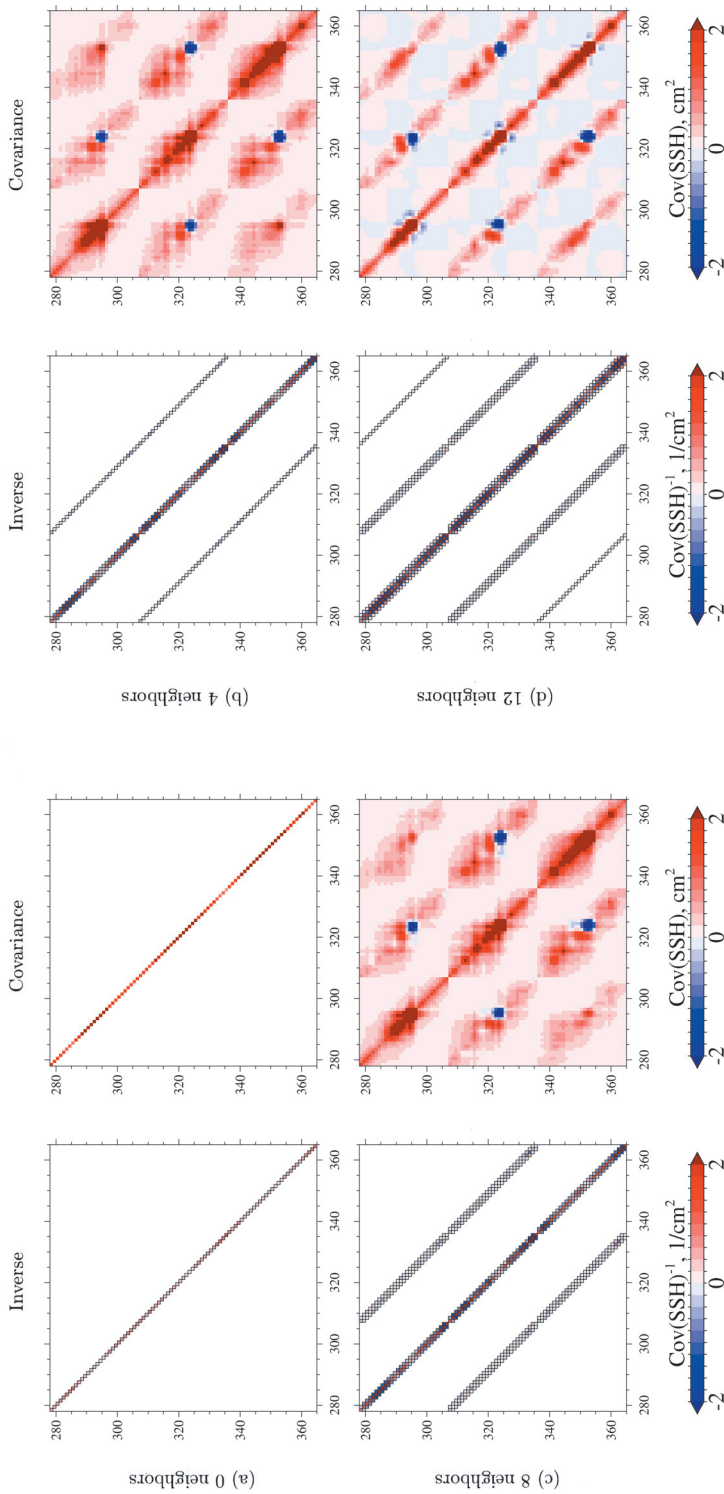


図 5. 推定された逆行列と共分散行列. (a) 0 近傍, (b) 4 近傍, (c) 8 近傍, (d) 12 近傍, (e) 20 近傍, (f) 24 近傍, (g) 28 近傍, (h) 36 近傍モデルをそれぞれ仮定. [278, 364] 以内の要素のみ示している. 逆行列の要素であらかじめ指定した非ゼロ要素は緑取りしてある.

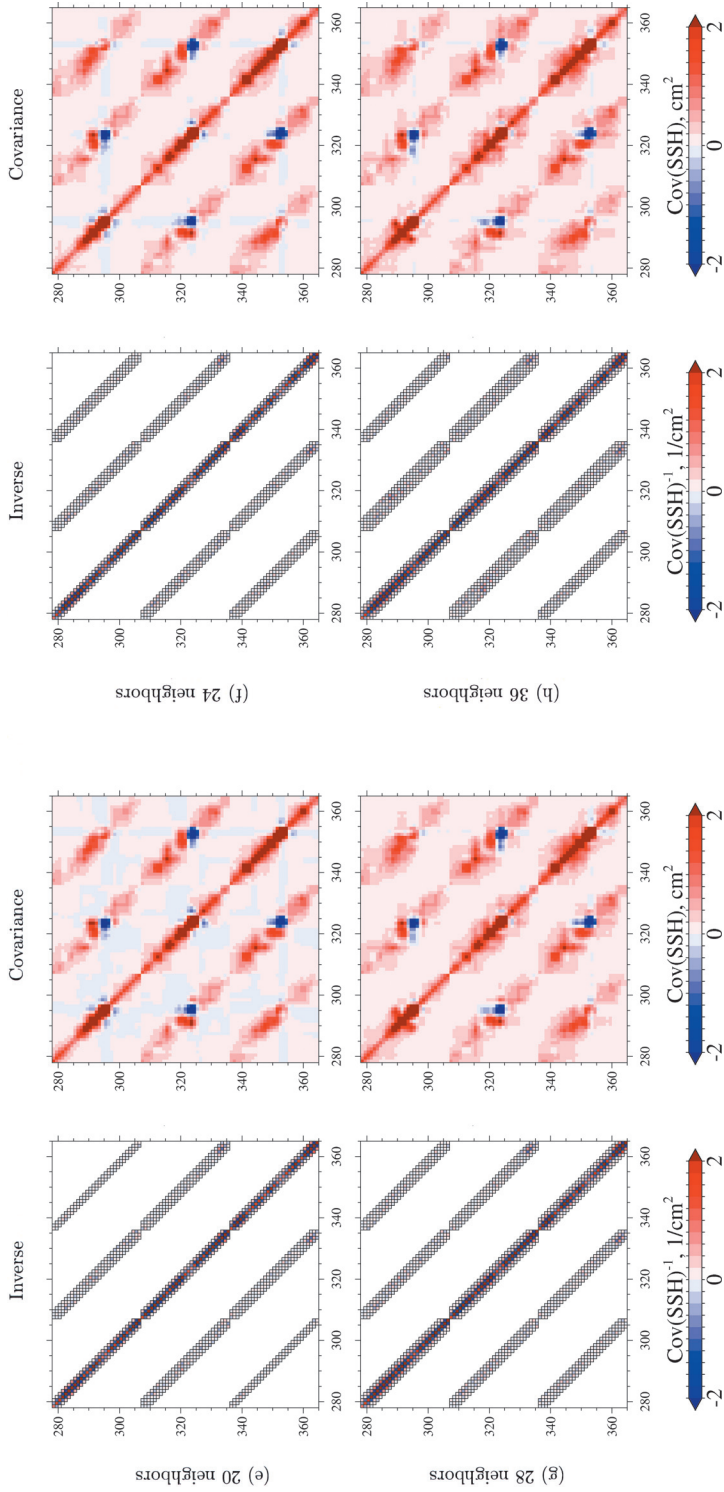


図 5. (つづき)

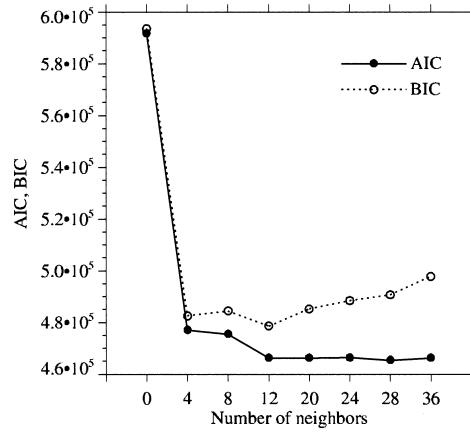


図 6. 太平洋赤道域での適用例に関して、近傍数の関数として AIC と BIC の値を示したもの。

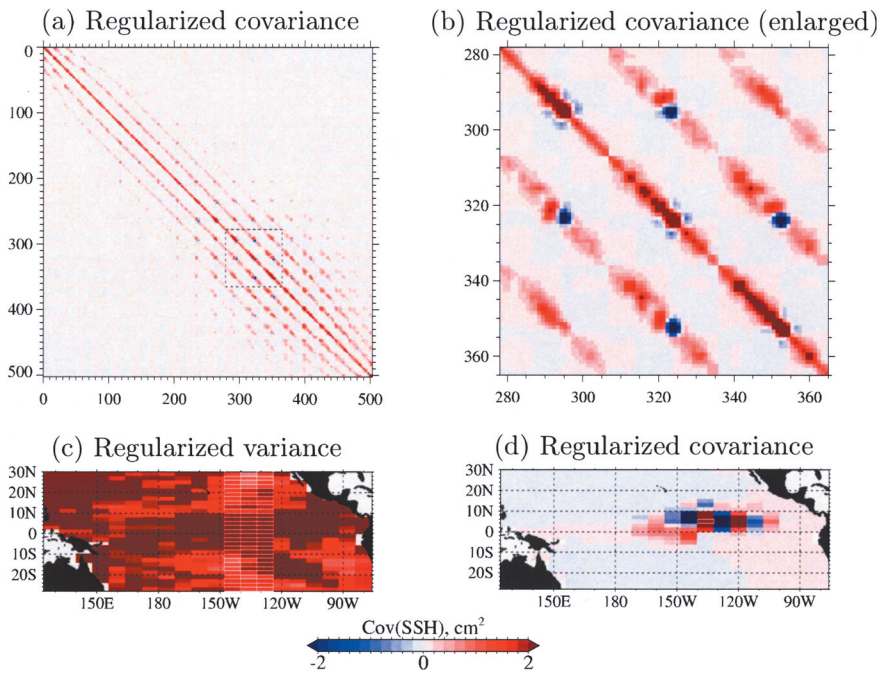


図 7. 標本共分散行列(図 4)および 12 近傍モデル(図 3d)を用いて得た正則化共分散行列. 形式は図 4 と同じ.

図 8 は、標本共分散 S と正則化された共分散 \hat{S} を比較している。 \hat{S} の分散成分は、図 8 (b)–(i) の左側のパネルに示すように、 S の分散成分と同じように見える(図 8(a)). これが示すのは、 β が各グラフィカルモデルにおいて、最適値に収束しているということである。

図 8(a) の右側は、 $(180^\circ, 0^\circ)$ における変数に関する共分散である。共分散成分は、正の相関がある変数は 162°E – 162°W 、 6°N – 6°S に集まっており、負の相関がある変数は赤道道上 105° – 136°W

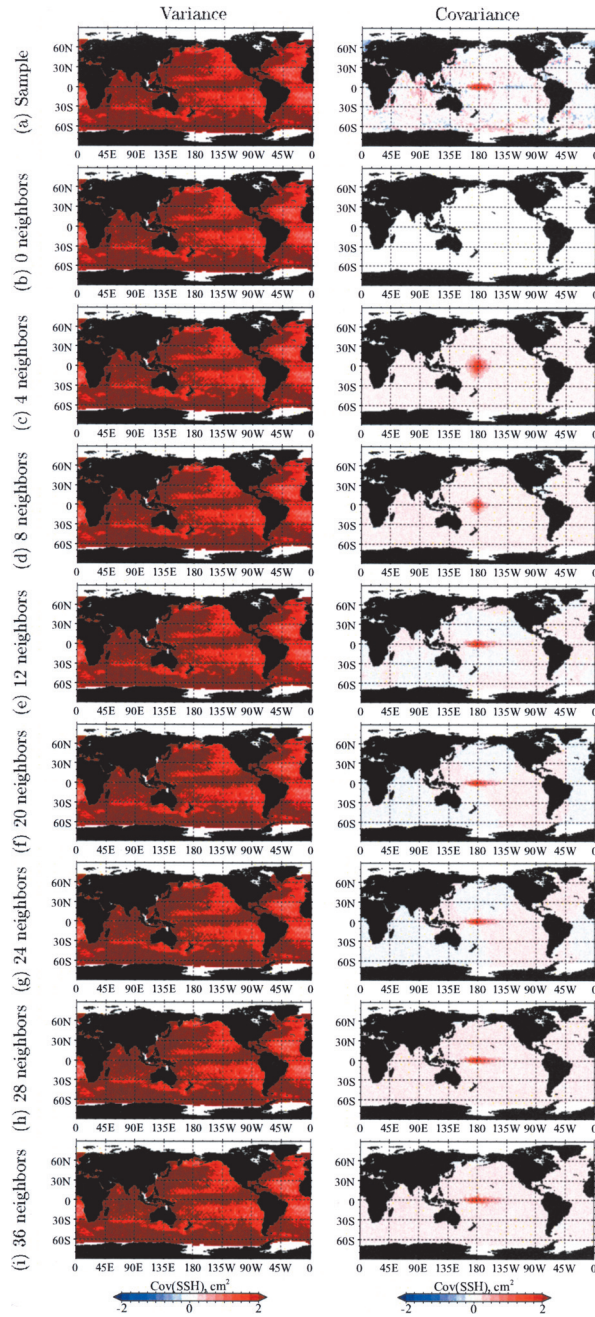


図 8. トレンド成分を除いた海面高度(SSH)観測データの分散および共分散. 全球海洋上 $2^\circ \times 2^\circ$ 格子点上で与えられている. 共分散要素は, $(180^\circ, 0^\circ)$ における変数に対してのものである. パネル (a) は標本共分散行列 S を示す. パネル (b)–(j) は正規化された共分散 $\hat{\Sigma}$ を示し, それぞれ仮定した近傍数は, 0, 4, 8, 12, 20, 24, 28, 36 である.

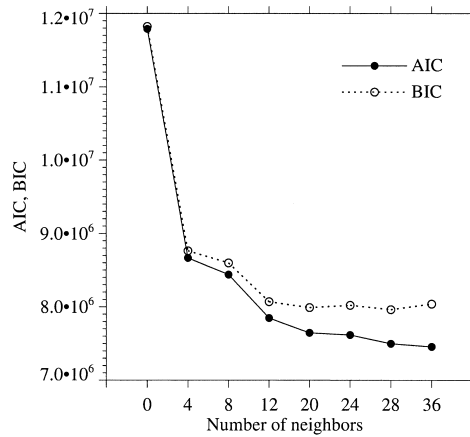


図 9. 全球海洋に適用したときの近傍数と AIC, BIC の関係.

表 3. ガウシアングラフィカルモデルを用いた共分散行列のまとめ: 近傍数, $\log |\hat{\Sigma}|$, 最大尤度値 ($\ell(\hat{\alpha}, \hat{\beta})$), パラメータ数 ($m = \dim \beta$), AIC, BIC 値を示す.

Number of neighbors	$\log \hat{\Sigma} $	$\ell(\hat{\alpha}, \hat{\beta})$	m	AIC	BIC
0	8891.2	-5886018.3	8585	11789206.7	11822424.6
4	-24.7	-4307907.8	25049	8665913.5	8762835.6
8	-758.9	-4177959.6	41234	8438387.3	8597933.8
12	-2519.3	-3866364.8	57189	7847107.6	8068388.8
20	-3266.4	-3734132.7	88726	7645717.4	7989024.6
24	-3434.6	-3704350.7	104190	7617081.5	8020223.5
28	-3856.4	-3629697.3	119682	7498758.7	7961843.9
36	-4150.0	-3577730.6	150381	7456223.2	8038092.0

に現れている. 加えて, 衛星軌道に沿った縞模様の構造が, 太平洋およびインド洋の低緯度から中緯度領域に対して見られる. 小さいサンプルサイズによる偽の共分散構造が, 大西洋の低緯度領域に現れている.

ゼロ近傍で推定された共分散行列 $\hat{\Sigma}$ は対角行列で, $\text{diag } S$ に等しく, 共分散変数は推定されない(図 8 (b)). 4 近傍モデルでは, 半径およそ 20° の円内で相関を持つ変数が推定される(図 8 (c)). 同様の円領域は 8 近傍モデルでも見られるが, 半径は小さくなって 15° 程度である(図 8 (d)). 12 近傍を仮定すると, 相関のある変数の領域は楕円形となり, 東西方向に 160°E - 148°W の範囲に広がる一方, 南北方向には 6°N - 6°S に限定される(図 8 (e)). 20 近傍モデル(図 8 (f))でも東西方向に伸びた楕円領域が推定されるが, 南北方向にはより短くなる. こうした楕円領域は 24-近傍, 36-近傍モデルでも現れるが(図 8 (g)-(i)), これらの領域の見目は変わらない. 近傍の数にかかわらず, 正規化された共分散は中央から東岸の赤道太平洋での負の相関を持つ変数や, 太平洋・インド洋の縞模様構造, 大西洋中緯度での偽の構造は再現していない.

6.2.2 好ましい近傍数

図 9 は, 近傍数を変化させたときの AIC, BIC 値であり, 表 3 は結果をまとめたものである. AIC も BIC も, 4 近傍を仮定すると, 近傍なしと仮定した場合と比べて値が劇的に減少す

る. 8 近傍を仮定すると, AIC と BIC は減少するが, 値は 4 近傍モデルでのものとさほど変わらない. 図 8(c) および (d) に示すように, 4 近傍および 8 近傍モデルは, 相関変数による円状の領域を推定していた. 12 近傍を仮定すると, AIC, BIC は 8 近傍モデルとくらべて大きく減少し, $AIC(8) - AIC(12)$ ($BIC(8) - BIC(12)$) の値は $AIC(4) - AIC(8)$ ($BIC(4) - BIC(8)$) より大きい. さらに, 12 近傍かそれ以上の近傍を仮定すると, AIC, BIC それぞれ近傍の数にあまり依存せず同等の値をとる. これは, 12 近傍から 36 近傍モデルはいずれも相関変数の楕円領域を推定していることと対応しているのであろう. 12 近傍もしくはそれ以上のモデルの中では, AIC 近傍数とともに減少を続け, 36 近傍モデルが最も適切なモデルであるということになる. 一方で BIC によれば, 20 近傍モデルで最初の極小が現れ, 28 近傍モデルで第 2 の極小(かつ最小)が現れる. BIC によれば, 好ましいモデルとして 20 近傍モデルか 28 近傍モデルを選ぶことができる.

7. 考察

本稿では, 標本共分散行列をグラフィカルモデルを用いて逆行列空間で正則化する方法を述べた. ここでは, 以下について述べる. (1) 物理空間でのこのような正則化を行うことは難しいこと, (2) 正則化による偽の相関へのフィルター効果, (3) 特異値分解 (SVD) による行列近似との違い, (4) 逆行列空間モデル化によりもたらされるデータ同化における計算コストの削減, (5) 正則化共分散行列を用いた変数の前処理, (6) より大規模な問題への展望について述べる.

7.1 独立性のモデル化と条件付き独立性のモデル化

逆行列空間での正則化のもっとも特筆すべき点は, 共分散行列はガウス分布を規定するパラメータであることを明確に意識していることである. 変数がガウス分布に従うなら, 2 つの変数間の条件付き独立性は, 共分散行列の逆行列のゼロ要素に対応する (例えば, Lauritzen, 1996 の Proposition 5.2). 条件付き独立性のモデル化は, 独立性と比べて実施が容易である. その理由は, 変数の組の局所的な関係のみ考慮すればよいからである. 例えば, 図 2 に示す 3 変数の関係は, 変数 2 を与えたもとで変数 1 と 3 は条件付き独立であるが, この関係を通常の共分散でモデル化することは難しい. 変数 1, 3 は独立ではない (すなわち, 変数 2 を介して依存関係にある) ため, 共分散行列で明示的に制約を課す (例えば, ある要素をゼロとするなど) ことはできないからである.

さらに, 独立性を直接指定するやり方では, 対角行列かブロック対角行列にする以外では, 数学的な意味で不適切な共分散行列にもなりかねない. 対角でもブロック対角でもない共分散の例として, 変数 1, 3 間の独立性を仮定する. このとき, 推定される共分散は次のような構造を持つ:

$$(7.1) \quad \hat{\Sigma} = \begin{pmatrix} s_{11} & s_{12} & \\ s_{12} & s_{22} & s_{23} \\ & s_{23} & s_{33} \end{pmatrix}.$$

ガウス分布の定義により, $\hat{\Sigma}$ は正定値性を持たねばならない. この条件は, $\hat{\Sigma}$ のコレスキー分解が存在することと等価であるが, これは常に満たされるわけではなく, 次が必要である.

$$(7.2) \quad \frac{s_{12}^2}{s_{11}s_{22}} + \frac{s_{23}^2}{s_{22}s_{33}} < 1.$$

明らかに, この条件は常に成り立つものではなく, S に依存しているため, 式 (7.1) のような共分散行列の要素を直接ゼロと置く (独立性を明示的にモデル化する) やり方はうまくいかないことがあることが分かる.

7.2 偽の相関のフィルタリング

1節で記したように、コンパクトな台を持つ相関関数により、限られた数の標本(アンサンブルメンバー)による偽の相関を消すことができる。われわれの逆行列空間での正則化も同様の効果があり、偽の遠距離相関をフィルタすることができる。

この様子を 3×3 行列で見よう。標本共分散行列

$$(7.3) \quad \mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix}$$

に対して、式(4.7)の形の逆行列を考えると、正則化された共分散行列は解析的に得られ、

$$(7.4) \quad \hat{\Sigma} = \begin{pmatrix} s_{11} & s_{12} & s_{12}s_{23}/s_{22} \\ s_{12} & s_{22} & s_{23} \\ s_{12}s_{23}/s_{22} & s_{23} & s_{33} \end{pmatrix},$$

となる(導出は Ueno and Tsuchiya, 2009 を参照されたい)。正則化された共分散行列(7.4)は、次の相関行列に対応する。

$$(7.5) \quad \hat{P} = \begin{pmatrix} 1 & \frac{s_{12}}{\sqrt{s_{11}s_{22}}} & \frac{s_{12}}{\sqrt{s_{11}s_{22}}} \frac{s_{23}}{\sqrt{s_{22}s_{33}}} \\ \frac{s_{12}}{\sqrt{s_{11}s_{22}}} & 1 & \frac{s_{23}}{\sqrt{s_{22}s_{33}}} \\ \frac{s_{12}}{\sqrt{s_{11}s_{22}}} \frac{s_{23}}{\sqrt{s_{22}s_{33}}} & \frac{s_{23}}{\sqrt{s_{22}s_{33}}} & 1 \end{pmatrix}.$$

相関行列 \hat{P} が示すのは、変数 1,3 間の相関は、変数 1,2 間の相関より強度が小さいということである。なぜならば、

$$(7.6) \quad \left| \frac{s_{12}}{\sqrt{s_{11}s_{22}}} \frac{s_{23}}{\sqrt{s_{22}s_{33}}} \right| = \left| \frac{s_{12}}{\sqrt{s_{11}s_{22}}} \right| \left| \frac{s_{23}}{\sqrt{s_{22}s_{33}}} \right|$$

$$(7.7) \quad \leq \left| \frac{s_{12}}{\sqrt{s_{11}s_{22}}} \right|.$$

これは、格子点間距離に応じて変数間の直接の関係を仮定すると、単調に減少する相関を意味している。さらに、3節で示したように、正則化された相関は、近傍の格子点上で、逆行列にゼロを仮定する要素では標本相関と同じ値をとる、これらの2つの性質は、近傍の格子点間でもとの相関は維持しつつも、離れた格子点での偽の相関を消すというフィルター関数に期待される性質である。この性質は標本相関において常に期待できるものではない。変数 1,3 間の相関 $s_{13}/\sqrt{s_{11}s_{33}}$ は、偽の相関に汚染される可能性があるが、変数 1,2 間の相関 $s_{12}/\sqrt{s_{11}s_{22}}$ より大きくなる可能性があるからである。

7.3 SVD 近似行列との比較

グラフィカルモデルの推定は、標本共分散行列の近似の一つの方法であると考えられる。行列の近似では、特異値分解(SVD)が広く使われてきた。SVD を用いて、ある行列を特異ベクトルと特異値に分解し、小さい特異値をゼロとおいて無視し、改めて特異ベクトルと修正後の特異値を掛け合わせて近似行列を作成する。近似行列は、標本共分散行列のみに基づき、格子点間の距離は考慮に入れない。したがって、SVD 近似は遠地点間の偽の相関を積極的に消す操作として働くとは期待できない。実際、図 10 が示すのは SVD 近似した標本共分散行列で、全特異値の和の 10^{-2} 倍より大きい 23 個の特異値を用いた。遠地点の相関は残っており(図 10(d))、一方でもとの標本共分散行列(図 4(c))と比べて分散は小さくなっている(図 10(c))。

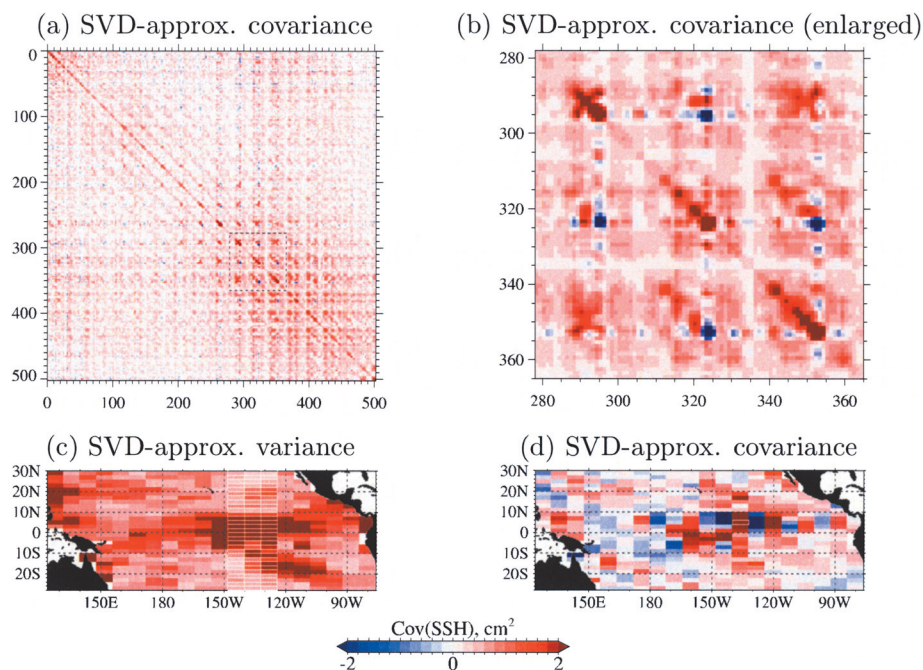


図 10. SVD 近似標本共分散行列. 全特異値の和の 10^{-2} 倍より大きい特異値を用いた. 形式は図 4 と同じ.

SVD 近似で用いた特異値の数の依存性を図 11 に示す. 分散は用いた特異値の数とともに徐々に増加し, 主要 23,155 項を用いた時にもとの標本共分散の 49%, 94%となる(図 11 (h)). 一方で, (136°W, 5°N)での変数に関する共分散要素は少なくとも 2つの主たる特異値で十分表現されているようである(図 11 (b)). 2つの主要項でも, 遠地点の距離の相関も再現されている. これより, SVD 近似は長距離相関のフィルターとしては働かないことが示されている.

加えて, 逆行列も SVD で近似することができる(一般化逆行列と呼ばれるもの). 近似逆行列は, 一般に全要素に非ゼロ値を持つ. これが意味するのは, SVD 近似は, グラフィカルモデルが行っているような条件付き独立の仮定とは一致するものではない.

7.4 パラメータ数の削減

共分散行列の要素の数は, 変数の数の 2 乗であり, データ同化では 10^{14} に上る. 共分散行列の正則化の目的の一つは, 共分散行列を指定するパラメータの数を減らすことである. 1 節でレビューした正則化法と同様に, 提案した逆行列空間での正則化法でも行列指定のためのパラメータの数を大幅に減らすことが可能である. ガウシアングラフィカルモデルの場合, パラメータ数は非ゼロ要素数に対応する. 4 近傍モデルを仮定すると, 必要なパラメータ数は変数の数のおよそ 4 倍で, これは変数の数の 2 乗よりずっと小さいものである.

逆行列でのパラメータ数の減少は, 変分型データ同化の評価関数の計算負荷をも減らす. 背景誤差共分散行列や観測誤差共分散行列は, その逆行列が評価関数で評価されるため, 逆行列でモデル化された多くのゼロ要素は乗算の数を著しく減少させる.

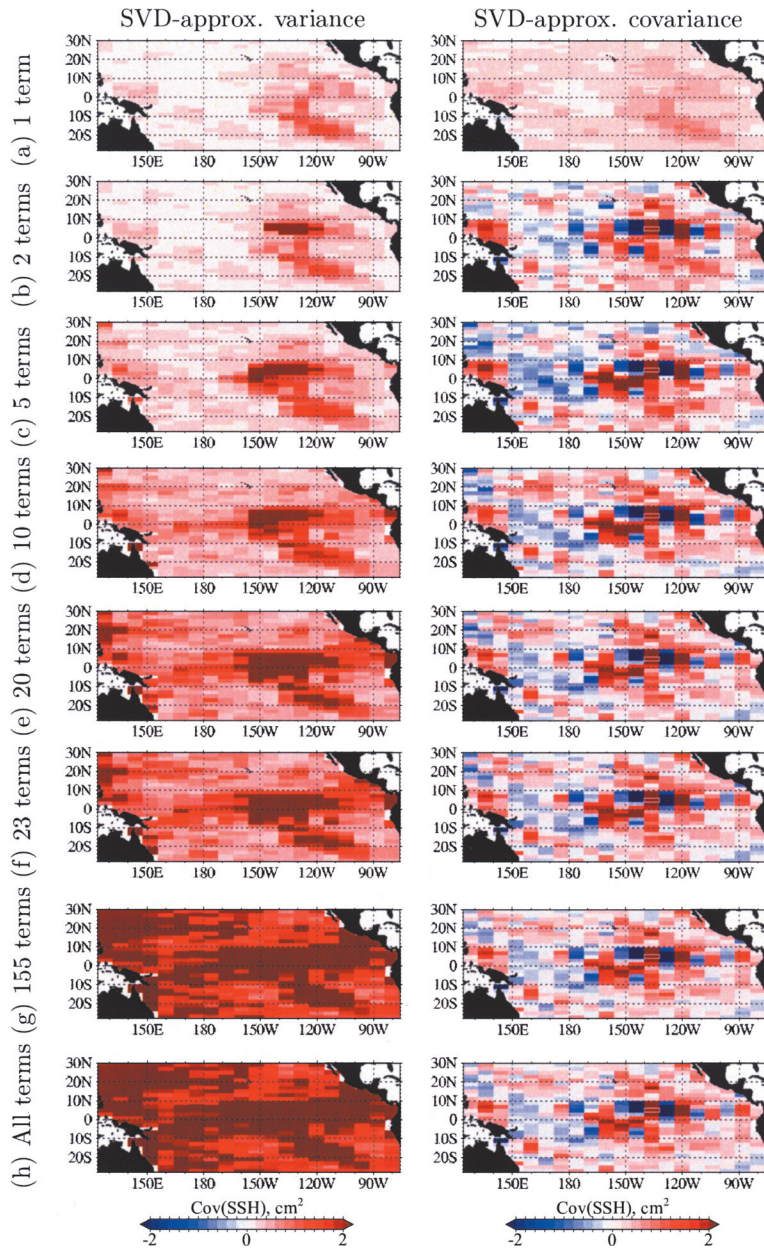


図 11. SVD 近似した標本共分散行列. それぞれ用いた特異値の個数は, 大きい方から (a) 1, (b) 2, (c) 5, (d) 10, (e) 20, (f) 23, (g) 155, (h) すべてである. 左側のパネルは分散 (共分散行列の対角要素), 右側のパネルは $(136^{\circ}\text{W}, 5^{\circ}\text{N})$ の変数 (白色で縁取り) に関する共分散要素. 主要項 23 および 155 (図 (f), (g)) は, 残した特異値が全特異値の和の 10^{-2} および 10^{-3} 倍より大きい場合に対応している. すべての特異値を用いた場合 (図 (h)) は, 標本共分散と同一である.

7.5 逆行列空間での正則化共分散行列を用いた前処理

逆行列空間での正則化共分散は、前処理 (Courtier et al., 1994) にも用いることができる。4 次元変分法において、背景誤差共分散行列を B としたとき、平方根行列 $B^{1/2}$ が前処理行列として選ばれる。 x と x_b をそれぞれ、状態ベクトルとその背景推定値 (事前設定平均) とすると、評価関数の背景項は $(x - x_b)' B^{-1} (x - x_b) / 2$ と表される。新しい変数として $u = B^{-1/2} (x - x_b)$ を定義すると、この項は $u'u / 2$ となる。すなわち、変換後の変数 u を用いることで、背景コストにおける B^{-1} を陽的に評価することが不要となり、評価関数のヘッセ行列の条件数を減らすことができる。これにより、評価関数最小化の反復過程の収束率を上げることができる。一方で観測項では状態ベクトル x は u を用いて評価する必要があり、 $x = x_b + B^{1/2} u$ を用いる必要がある。

B^{-1} の近似として正則化された背景誤差共分散行列の逆行列 $\hat{\Sigma}^{-1}$ を用いても、同様の前処理が可能である。前処理後、共分散行列の逆行列 $\hat{\Sigma}^{-1}$ は評価関数に現れないが、 $\hat{\Sigma}^{1/2} u$ は必要である。後者は、次の 2 ステップで得られる。(1) $\hat{\Sigma}^{-1}$ をコレスキー分解し、 $\hat{\Sigma}^{-1/2}$ を得る。(2) 1 次連立方程式 $u = \hat{\Sigma}^{-1/2} v$ を v について解く。得られた v が $\hat{\Sigma}^{1/2} u$ であり、これは観測項で必要である。 $\hat{\Sigma}$ が疎であるため、コレスキー分解は実施可能である。線形方程式 $u = \hat{\Sigma}^{-1/2} v$ も、 $\hat{\Sigma}^{-1/2}$ が疎で三角であるため大きい計算負荷をかけずに解くことができる。

$\hat{\Sigma}^{1/2}$ による前処理は $\hat{\Sigma}$ の性質を引き継ぐため、もとの共分散行列 B に存在した偽の相関を除去することができる。

7.6 より大規模問題への展望

6.2 節で示した適用例では、 $n = 8,585$ -次元共分散行列が $m = 150,381$ 個のパラメータ (非ゼロ要素) を持っていた。この数は確かに小さいものではないが、 $n = O(10^7)$ 変数を持つ最先端の力学モデルを直接扱うにはまだ不十分である。そのギャップは 10^3 のオーダーであるが、提案法の上限は実装手続きの改良によって 10^6 程度まで上がることが予想される。実際、 $1^\circ \times 1^\circ$ 格子点での全球 SSH 観測データに対して、4 近傍モデルの適用に成功している。このときの総格子点数は $n = 34,300$ で、非ゼロ要素数は $m = 101,310$ である。

大規模問題を扱う代替法は、全領域を複数の領域に分割することである。ブロック対角構造を共分散行列 Σ に仮定すれば、その逆行列 Σ^{-1} もブロック対角となり、計算負荷が大きく減少する。領域別アプローチは、適用例でも見たように (Section 6)、有効な相関領域は限定されていることから正当化される。有効相関領域の大きさは、もとの標本共分散行列 S から推定できる (例えば、Pannekoucke et al., 2008)。

また、複数の分割した領域の一つに注目し、その中に含まれるパラメータのみを推定し、それ以外の領域のパラメータは固定としておくという方法も考えられる。このやり方を、注目領域を変えながら繰り返すことで、全領域のパラメータの推定値を得ることが期待できる。

8. まとめ

ガウシアングラフィカルモデルを用いて、逆行列空間で共分散行列を正則化する方法を紹介した。グラフィカルモデル構築のポイントは、各変数に対して、近傍変数を仮定することである。対象とする変数は近傍変数と直接関係し、近傍変数以外の変数とは条件付き独立である。条件付き独立は、共分散行列の逆行列にゼロ要素を指定することで表現できる。非ゼロ要素は、最尤法とニュートン法を用いて数値的に推定する。ニュートン法で必要となる勾配ベクトルとヘッセ行列は、このガウシアングラフィカルモデルで解析的に導出される。適切な近傍変数の数は、AIC や BIC などの情報量規準により選択できる。また、変数の次元が大きい場合に有用な実装技術を紹介した。提案法を用いて、図 4 に示す標本共分散行列を正則化共分散行列

(図7)に変換した。加えて、 $8,585 \times 8,585$ 標本共分散行列の正則化を実証し、パラメータ数は150,381まで上る。これは、全球海洋海面高度データの $2^\circ \times 2^\circ$ 分解能に当たる。この正則化により、非特異共分散行列が得られ、遠地点間の偽の相関を消すことができる。それに加え、正則化共分散行列を指定するために必要なパラメータの数は、変数の2乗と比べて小さい。

謝 辞

政策研究大学院大学の土谷隆教授には貴重なコメントを頂いた。TOPEX/POSEIDON 海面高度偏差データは、アメリカ航空宇宙局ジェット推進研究所から提供頂いた (<http://podaac.jpl.nasa.gov/>)。本研究は、JST 戦略的創造研究推進事業さきがけの一環として行われたものである。

参 考 文 献

- Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **AC-19**, 716–723.
- Akaike, H. (1977). On entropy maximization principle, *Applications of Statistics, Proceedings of the symposium held at Wright State University, Dayton, Ohio, 14-18 June 1976* (ed. P. R. Krishnaiah), 27–41, North-Holland Publishing Company, Amsterdam.
- Alves, O. and Robert, C. (2005). Tropical Pacific Ocean model error covariances from Monte Carlo simulations, *Quarterly Journal of the Royal Meteorological Society*, **131**, 3643–3658, doi:10.1256/qj.05.113.
- Bishop, C. H. and Hodyss, D. (2007). Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation, *Quarterly Journal of the Royal Meteorological Society*, **133**, 2029–2044, doi:10.1002/qj.169.
- Borovikov, A., Rienecker, M. M., Keppenne, C. L. and Johnson, G. C. (2005). Multivariate error covariance estimates by Monte Carlo simulation for assimilation studies in the Pacific Ocean, *Monthly Weather Review*, **133**(8), 2310–2334, doi:10.1175/MWR2984.1.
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press, Cambridge.
- Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational nwp setting, *Quarterly Journal of the Royal Meteorological Society*, **131**, 1013–1043.
- Buehner, M. and Charron, M. (2007). Spectral and spatial localization of background-error correlations for data assimilation, *Quarterly Journal of the Royal Meteorological Society*, **133**, 615–630.
- Courtier, P., Thépaut, J. N. and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367–1387.
- Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. (1998). The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, *Quarterly Journal of the Royal Meteorological Society*, **124**, 1783–1807.
- Cullen, M. J. P. (2003). Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential-vorticity representation, *Quarterly Journal of the Royal Meteorological Society*, **129**: 2777–2796, doi:10.1256/qj.02.10.
- Daley, R. (1991). *Atmospheric Data Analysis, Cambridge Atmospheric and Space Science Series*,

- Vol. 2, Cambridge University Press, Cambridge.
- Deckmyn, A. and Berre, L. (2005). A wavelet approach to representing background error covariances in a limited-area model, *Monthly Weather Review*, **133**(5), 1279–1294, doi:10.1175/MWR2929.1.
- Dempster, A. P. (1972). Covariance selection, *Biometrics*, **28**(1), 157–175.
- Derber, J. and Bouttier, F. (1999). A reformulation of the background error covariance in the ECMWF global data assimilation system, *Tellus*, **51A**(2), 195–221.
- Derber, J. and Rosati, A. (1989). A global oceanic data assimilation system, *Journal of Physical Oceanography*, **19**, 1333–1347.
- Desroziers, G. (1997). A coordinate change for data assimilation in spherical geometry of frontal structures, *Monthly Weather Review*, **125**, 3030–3038.
- Diggle, P. J. and Ribeiro Jr., P. J. (2007). *Model-based Geostatistics*, Springer Series in Statistics, Springer-Verlag, New York.
- Egbert, G. D., Bennett, A. F. and Foreman, M. G. G. (1994). TOPEX/POSEIDON tides estimated using a global inverse model, *Journal of Geophysical Research*, **99**(C12), 24821–24852.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367.
- Fisher, M. (2003). Background error covariance modelling, *Seminar on Recent Developments in Data Assimilation for Atmosphere and Ocean, 8-12 September 2003*, ECMWF, Shinfield Park, Reading, 45–64.
- Fisher, M. and Andersson, E. (2001). Developments in 4D-Var and Kalman filtering, Technical Memorandum 347, ECMWF, Shinfield Park, Reading.
- Frehlich, R. (2006). Adaptive data assimilation including the effect of spatial variations in observation error, *Quarterly Journal of the Royal Meteorological Society*, **132**, 1225–1257, doi:10.1256/qj.05.146.
- Fu, L. L., Fukumori, I. and Miller, R. N. (1993). Fitting dynamic models to the Geosat sea level observations in the tropical Pacific Ocean. Part II: A linear, wind-driven model, *Journal of Physical Oceanography*, **23**, 2162–2181.
- Fukumori, I. (2001). Data assimilation by models, *Satellite Altimetry and Earth Sciences, International Geophysics Series*, Vol. 69 (eds. L. L. Fu and A. Cazenave), 237–265, Academic Press, San Diego.
- Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.
- Gaspari, G., Cohn, S. E., Guo, J. and Pawson, S. (2006). Construction and application of covariance functions with variable length-fields, *Quarterly Journal of the Royal Meteorological Society*, **132**, 1815–1838, doi:10.1256/qj.05.08.
- Gauthier, P., Buehner, M. and Fillion, L. (1998). Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses, *ECMWF Workshop on Diagnosis of Data Assimilation Systems, 2-4 November 1998*, 131–145, ECMWF, Shinfield Park, Reading.
- Gneiting, T. (1999). Correlation functions for atmospheric data analysis, *Quarterly Journal of the Royal Meteorological Society*, **125**, 2449–2464.
- Gneiting, T. (2002). Compactly supported correlation functions, *Journal of Multivariate Analysis*, **83**, 493–508, doi:10.1006/jmva.2001.2056.
- Hamill, T. M., Whitaker, J. S. and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, **129**, 2776–2790.

- Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus*, **38A**(2), 111–136.
- Houtekamer, P. L. and Mitchell, H. L. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, **129**(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- Houtekamer, P. L., Lefaiivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. (1996). A system simulation approach to ensemble prediction, *Monthly Weather Review*, **124**(6), 1225–1242, doi:10.1175/1520-0493(1996)124<1225:ASSATE>2.0.CO;2.
- Jackson, D. R., Keil, M. and Devenish, B. J. (2008). Use of Canadian Quick covariances in the Met Office data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, **134**, 1567–1582, doi:10.1002/qj.294.
- Keppenne, C. L., Rienecker, M. M., Jacob, J. P. and Kovach, R. (2008). Error covariance modeling in the GMAO ocean ensemble Kalman filter, *Monthly Weather Review*, **136**, 2964–2982, doi:10.1175/2007MWR2243.1.
- Konishi, S. and Kitagawa, G. (2007). *Information Criteria and Statistical Modeling*, Springer Series in Statistics, Springer, New York.
- Lauritzen, S. L. (1996). *Graphical models*, *Oxford Statistical Science Series*, Vol. 17, Oxford University Press, Oxford.
- Lönnberg, P. and Hollingsworth, A. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors, *Tellus*, **38A**(2), 137–161.
- Lorenc, A. C. (1992). Iterative analysis using covariance functions and filters, *Quarterly Journal of the Royal Meteorological Society*, **118**, 569–591.
- Lorenc, A. C. (1997). Development of an operational variational assimilation scheme, *Journal of the Meteorological Society of Japan*, **75**(1B), 339–346.
- Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised ed., Wiley Series in Probability and Statistics, John Wiley & Sons, West Sussex.
- Menemenlis, D. and Chechelnitsky, M. (2000). Error estimates for an ocean general circulation model from altimeter and acoustic tomography data, *Monthly Weather Review*, **128**(3), 763–778, doi:10.1175/1520-0493(2000)128<0763:EEFAOG>2.0.CO;2.
- Nychka, D., Wikle, C. and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions, *Statistical Modelling*, **2**(4), 315–331, doi:10.1191/1471082x02st037oa.
- Oke, P. R., Allen, J. S., Miller, R. N., Egbert, G. D. and Kosro, P. M. (2002). Assimilation of surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical Research*, **107**(C9), doi:10.1029/2000JC000511.
- Pannekoucke, O., Berre, L. and Desroziers, G. (2007). Filtering properties of wavelets for local background-error correlations, *Quarterly Journal of the Royal Meteorological Society*, **133**, 363–379, doi:10.1002/qj.33.
- Pannekoucke, O., Berre, L. and Desroziers, G. (2008). Background-error correlation length-scale estimates and their sampling statistics, *Quarterly Journal of the Royal Meteorological Society*, **134**: 497–508, doi:10.1002/qj.212.
- Parrish, D. F. and Derber, J. C. (1992). The national meteorological center’s spectral statistical-interpolation analysis system, *Monthly Weather Review*, **120**, 1747–1763, doi:10.1175/1520-0493(1992)120<1747:TnMCSS>2.0.CO;2.
- Polavarapu, S., Ren, S., Rochon, Y., Sankey, D., Ek, N., Koshyk, J. and Tarasick, D. (2005). Data assimilation with the Canadian Middle Atmosphere Model, *Atmos.-Ocean*, **43**(1), 77–100.

- Purser, R. J., Wu, W. S., Parrish, D. F. and Roberts, N. M. (2003a). Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances, *Monthly Weather Review*, **131**(8), 1524–1535, doi:10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2.
- Purser, R. J., Wu, W. S., Parrish, D. F. and Roberts, N. M. (2003b). Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances, *Monthly Weather Review*, **131**(8), 1536–1548, doi:10.1175//2543.1.
- Raynaud, L., Berre, L. and Desroziers, G. (2008). Spatial averaging of ensemble-based background-error variances, *Quarterly Journal of the Royal Meteorological Society*, **134**, 1003–1014, doi:10.1002/qj.245.
- Rhodin, A. and Anlauf, H. (2007). Representation of inhomogeneous, non-separable covariances by sparse wavelet-transformed matrices, *ECMWF Workshop on Flow-dependent Aspects of Data Assimilation, 11–13 June 2007*, 169–183, ECMWF, Shinfield Park, Reading.
- Riishøjgaard, L. P. (1998). A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus*, **50A**(1), 42–57.
- Schwarz, G. (1978). Estimating the dimension of a model, *The Annals of Statistics*, **6**(2), 461–464.
- Trémolet, Y. (2007). Model-error estimation in 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, **133**, 1267–1280, doi:10.1002/qj.94.
- Tsuchiya, T. and Xia, Y. (2007). An extension of the standard polynomial-time primal-dual path-following algorithm to the weighted determinant maximization problem with semidefinite constraints, *Pacific Journal of Optimization*, **3**(1), 165–182.
- Ueno, G. and Tsuchiya, T. (2009). Covariance regularization in inverse space, *Quarterly Journal of the Royal Meteorological Society*, **135**, 1133–1156.
- Ueno, G., Higuchi, T., Kagimoto, T. and Hirose, N. (2007). Application of the ensemble Kalman filter and smoother to a coupled atmosphere-ocean model, *SOLA*, **3**(1), 5–8.
- Vandenberghe, L., Boyd, S. and Wu, S. P. (1998). Determinant maximization with linear matrix inequality constraints, *SIAM Journal on Matrix Analysis and Applications*, **19**(2), 499–533.
- Weaver, A. and Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation, *Quarterly Journal of the Royal Meteorological Society*, **127**, 1815–1846.

Estimation of Large-scale Graphical Models for Data Assimilation

Genta Ueno

The Institute of Statistical Mathematics;
Japan Science and Technology Agency, PRESTO

Covariance matrices are often introduced when we construct statistical models that have multiple variables. In the present article, I review covariance matrices used in large-scale time-series analysis, specifically data assimilation. In data assimilation, covariance matrices are introduced in order to prescribe the weights of the initial state, model dynamics, and observation, and suitable specification of the covariances is known to be essential for obtaining sensible state estimates. The covariance matrices are specified by sample covariances and converted according to an assumed covariance structure. Modeling of the covariance structure consists of the regularization of a sample covariance and the constraint of a dynamic relationship. Regularization is required for converting the singular sample covariance into a non-singular sample covariance, removing spurious correlation between variables at distant points, and reducing the number of parameters required to specify the covariances. In previous studies, regularization of sample covariances has been carried out in physical (grid) space, spectral space, and wavelet space. I herein review a method for covariance regularization in inverse space, in which we use the covariance selection model (the Gaussian graphical model). For each variable, we assume neighboring variables, i.e., a targeted variable is directly related to its neighbors and is conditionally independent of non-neighboring variables. Conditional independence is expressed by specifying zero elements in the inverse covariance matrix. The non-zero elements are estimated numerically by the maximum likelihood using Newton's method. Appropriate neighbors can be selected with the AIC or BIC information criteria. I address some techniques for implementation when the covariance matrix has a large dimension. I present an illustrative example using a simple 3×3 matrix and an application to a sample covariance obtained from sea surface height (SSH) observations.