

確率過程に基づく各種内挿法の実務上の汎化性の考察

Consideration of generalization performance of interpolation methods based on stochastic process model

新潟大学大学院 学生会員 ○居城 英孝

新潟大学 正会員 大竹 雄

国立研究開発法人土木研究所 正会員 小出 央人

1. 目的

本研究は、情報の不確実性の定量化のための基礎的な研究と位置づけて、観測地点間の空間内挿法の基本的な考え方を整理し、簡易な試算に基づきモデルの得失について議論する。具体的には、伝統的な空間分布推定手法であるSimple Kriging¹⁾と近年、時系列解析等で注目される状態空間モデル²⁾を取り上げ、仮想データを用いて、2つのモデルによる内挿誤差を比較することで、実務上の汎化性について考察を行う。

2. 各種内挿補間法の基本的な考え方

2.1. 手法1: Simple Kriging (ベイズモデル)

Kriging は、様々な手法が提案されているが、本研究では、Simple Kriging を用いる。Simple Kriging を採用した理由は、ベイズ推定として解釈することができるためである。手法2では、階層ベイズモデルを適用することから、解析結果を比較しやすい手法を選定した。なお、Simple Kriging では、確率場のパラメータである、平均、分散、自己相関距離を事前に設定する必要がある。通常、観測された、限られたデータからの標本値を採用する場合が多い。

2.2. 手法2: ローカルレベルモデル (階層ベイズモデル)

状態空間モデルは階層ベイズモデルと位置づけられ、通常我々には見にできない「状態」空間を定義したモデルである。このモデルは、与えられた観測ベクトル $\mathbf{y} = \{y_1, \dots, y_T\}$ に基づいて状態ベクトル $\mathbf{x} = \{\mu_1, \dots, \mu_T\}$ を推定する時系列モデルである。今回用いるローカルレベルモデルは状態空間モデルの中でも単純な構造なモデルであり、以下の式で定義される。

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (1)$$

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (2)$$

(1),(2)式はそれぞれ状態方程式と観測方程式であり、 η_t と ε_t はそれぞれ状態誤差、観測誤差である。互いの誤差は独立で、それぞれ平均0、分散 $\sigma_\eta^2, \sigma_\varepsilon^2$ の正規分布に従い、初期の状態 μ_0 も正規分布に従う場合、状態ベクトル $\mathbf{x}$ は全て正規分布にしたがうことが知られている。このような過程のもとでは、カルマンフィルタによる逐次計算で解析解が与えられる。紙面の都合上、アルゴリズムについての詳細は省略する。なお、ローカルレベルモデルでは、事前に設定する確率場パラメータはなく、階層ベイズ法により、データから、確率場パラメータが決定される。

3. 簡易な試算による各種内挿法の比較方法

3.1. 解析に用いるデータ

本研究では、確率過程にもとづいて「正解データ」を2種類生成し、その一部のデータを間引いた、「間引き調整データ」から、間引いたデータの推定精度を議論することで、各モデルの得失の議論や汎化性能について考察を行う。なお、間引き調整データを生成した諸元は以下の通りである。
間引き調整データ 1: トレンド一定($y=0$), 分散0.70, 自己相関距離 $\theta=10$, 定常確率過程から生成したある1つの時系列データ
間引き調整データ 2: トレンド関数正弦波, 残差

キーワード 状態空間モデル

連絡先 〒950-2181 新潟県新潟市西区五十嵐2の町8050番地 新潟大学大学院

表 1 各モデルにおけるパラメータ推定結果と正解データのパラメータ値

	トレンド	分散	自己相関距離
間引き調整データ 1	トレンド関数一定	分散 0.70	10
間引き調整データ 2	トレンド関数正弦波	残差の分散 0.70	10

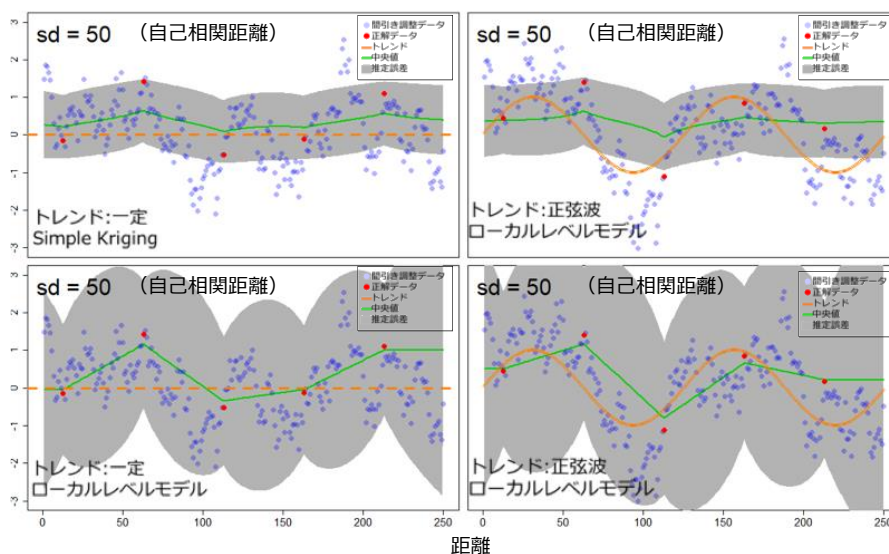


図2 各モデルにおける間引き調整データの内挿結果

の分散 0.70, 自己相関距離 10, の定常確率過程から生成した 1 つの時系列データである。

図 2 にトレンド関数及び間引き調整データを示している。図中の青色のプロットが正解データで、赤色のプロットが間引いた後のデータである。

3.2. 内挿精度の算出方法

3.1. で作成した間引き調整データを用いて, 各モデルの内挿精度を検証する。ここでは, 内挿精度の表現方法を統一し, $bias$ (正解データと推定値の比の平均値) とその変動係数 COV_{bias} の 2 つの指標で内挿精度の特性を表現する。着目は, 期待値のデータの再現性だけでなく, 推定分散が実データを包含しているか, という観点である。

4. 簡易な試算による解析結果

表 1 に各モデルに用いるパラメータの値を表にまとめている。Simple Kriging のパラメータは間引き調整データより直接計算した標本平均, 標本分散, 自己相関距離を, ローカルレベルモデルのパラメータは最尤推定法で求めたシステムノイズと観測ノイズの分散を示している。なお比較のために, Simple Kriging の観測誤差はローカル

レベルモデルの観測誤差の平均値を採用している。

図 2 には, それぞれの間引き調整データより求めた推定結果を示す。図 2 を見ると, Simple Kriging では定常性を仮定しているためうまく推定できていない箇所があるのに対し, ローカルレベルモデルは, 正解データを包含するように推定分散が計算されることが分かる。紙面の都合上, 詳細な考察は記載しないが, 講演時には, 3.1 に示した推定誤差の結果を含めて, 2 つのモデルの得失について発表する予定である。

5. 結論と今後の研究の展望

今後は, 推定される観測誤差とシステム誤差の特徴を分析し, Simple Kriging と階層ベイズ法の関係性を考察するとともに, 階層ベイズ法の推定結果を伝統的な統計学的の正規標本論から考察を加えることにより, 実用化を目指す。

参考文献

- 1) 間瀬茂: 地球統計学とクリギング法 -R geoR によるデータ解析-, 2010
- 2) 野村俊一: カルマンフィルタ -R を使った時系列予測と状態空間モデル-, 2016
- 3) 馬場真哉: 時系列分析と状態空間モデルの基礎 R と Stan で学ぶ理論と実装, 2018