

誤差要素モデルを援用した Kriging による空間内挿

東京大学大学院 工学系研究科 正会員 堤 盛人
日本政策投資銀行 正会員 井出 裕史
東京大学大学院 工学系研究科 正会員 清水 英範

1. はじめに

地理的・空間的に連続して分布するデータ（空間データ）を扱う分析においては、結果の表示や分析自体のために、何らかの内挿を必要とする場合が少なくない。空間データを合理的に内挿する手法の一つとして、鉱物学や地質学など自然科学と関わりの深い空間統計学の分野において、Kriging と呼ばれる手法が確立されつつある（Cressie(1993), Bailey and Gatrell (1995)）。Kriging は、ある条件下で最良線形不偏推定量を与えるという意味で統計学的に優れた手法である。Kriging の適用に際しては、データの空間的な相関を表す covariogram と呼ばれる共分散関数が必要となる。一方、社会経済分析と関わりの深い空間計量経済学と呼ばれる分野においては、空間データを用いた回帰分析において誤差項に何らかの相関が生じる場合に、誤差要素モデル(Error Component Model : ECM)と呼ばれるモデルが用いられる（Anselin (1988)）。ECM は、誤差項における未知の空間的な系列相関に対し、パラメータを用いたモデル化によって対処しようとするものである。

本稿では、Kriging における covariogram の同定に際しての、ECM の援用可能性について検討を行う。

2. Kriging による空間データの内挿

Kriging では、ある任意の地点 o における値 $\hat{\varepsilon}_0$ は周囲の観測値 ε_i の線形和として、次式のように書けるものと仮定する。

$$\hat{\varepsilon}_0 = \sum_{i=1}^n \lambda_i \varepsilon_i \quad (1)$$

ここで、任意の 2 地点 i, j における値 $\varepsilon_i, \varepsilon_j$ の共分散は、その 2 地点間の距離 d_{ij} のみによって決まるという定常性を仮定する。

$$\text{cov}(\varepsilon_i, \varepsilon_j) = c(d_{ij}) \quad (2)$$

式(2)の関数は covariogram と呼ばれる。期待二乗誤差最小化から式(1)における重み λ_i が次のように導かれる。

$$\min E[\varepsilon_0 - \hat{\varepsilon}_0]^2 \Rightarrow \lambda = C^{-1}c \quad \text{ただし、} c, C : \text{covariogram} \quad (3)$$

本稿で扱う線形回帰モデル $y = Xb + e$ においては、トレンドのパラメータ b は式(2)を分散共分散行列とした一般化最小二乗法等により求められ、ある地点 o における推定値 $\hat{y}_0$ が次式により求められる。

$$\hat{y}_0 = x_0 \beta_{GLS} + \lambda^t \varepsilon \quad \text{ただし、} \beta_{GLS} = (X^t C^{-1} X)^{-1} X^t C^{-1} Y, \lambda = C^{-1}c \quad (4)$$

3. 誤差要素モデル

社会経済分析を対象とする空間計量経済学において、回帰モデルの誤差項に系列相関が生じている場合の対処法としては、誤差項同士の影響関係をモデル化するという方法が用いられる。誤差項に移動平均モデル(Moving Average)を仮定した場合には、誤差項の分散共分散行列は次式のように与えられる。

$$\hat{a}_i = \hat{n} \sum_{j=1}^n w_{ij} u_j + u_i \Rightarrow \text{Var}[\hat{a}] = \left[(I + \hat{n}W)(I + \hat{n}W)^t \right] \rho_u^2 \quad (5)$$

ただし、 u_i : 正規分布 $N(0, \sigma_u)$ に従う確率変数、 σ_u : u の分散、 ρ : パラメータ

このとき、 w_{ij} は地点間の結びつきの強さを示し、 W は結合行列と呼ばれる。 w_{ij} は、地点間の距離 d_{ij} の関数

キーワード：空間内挿、空間統計学、空間計量経済学、Kriging、ECM

連絡先：〒113-8656 東京都文京区本郷 7-3-1 TEL:03-5841-6128 FAX:03-5841-7453

$w_{ij} = f(d_{ij})$ と表されることが多い。

4 . Kriging における誤差要素モデル援用の可能性

1.で述べたように、Kriging においては covariogram を何らかの方法で決める必要があり、データのプロットから目視によって決めるなどの方法が用いられることが多い。本章では、covariogram の同定に、3.で述べた ECM の考え方を援用することを試みる。

式(5)に示す ECM における誤差項の共分散行列の各非対角要素は、次式のように表わされる。

$$\text{cov}(\hat{a}_i, \hat{a}_j) = \left[\tilde{n}(w_{ij} + w_{ji}) + \tilde{n}^2 \sum_{k=1}^n w_{ik} w_{kj} \right] \sigma_u^2 \quad (i \neq j)$$
 (6)

これから、一般に、ECM によってモデル化された誤差項の共分散は、Kriging で前提とされる定常性を満足しないことが分かる。何故なら、任意の 2 地点 i, j 間における共分散は、それぞれの地点と別の地点 k との相関によって影響されるため、これらの地点がある特殊な配置に分布している場合を除いて、2 地点 i, j 間の距離 d_{ij} のみの関数とはならないからである。そこで、図 1 に示すように、地点が、実際の観測地点を含む無限に広がる格子点状に配置されており、観測されていない地点はそのうちのデータが欠損している地点であると仮定する。このとき、式(6)によって計算される共分散は定常となるため、Kriging の前提と整合する。この ECM の拡張による Kriging 内挿手法を、ここでは便宜的に ECM-Kriging と呼ぶこととする。

次に、通常の Kriging (詳しくは GLM) と ECM-Kriging を、実証データへの適用を通して比較する。データは、東京大学生産技術研究所・虫明研究室提供による、タイにおける過去 30 年の年平均降雨量を用いた。観測データの範囲はおよそ 500km × 500km で、478 の観測データを半分に分け、一方をパラメータ推定に、他方を validation に用いた。モデルの推定にはそれぞれ最尤法を用いている。各々のパラメータの推定結果、モデルによる内挿結果は、表 1 と図 2 に示すとおりであり、ECM-Kriging の適用可能性が示唆される。

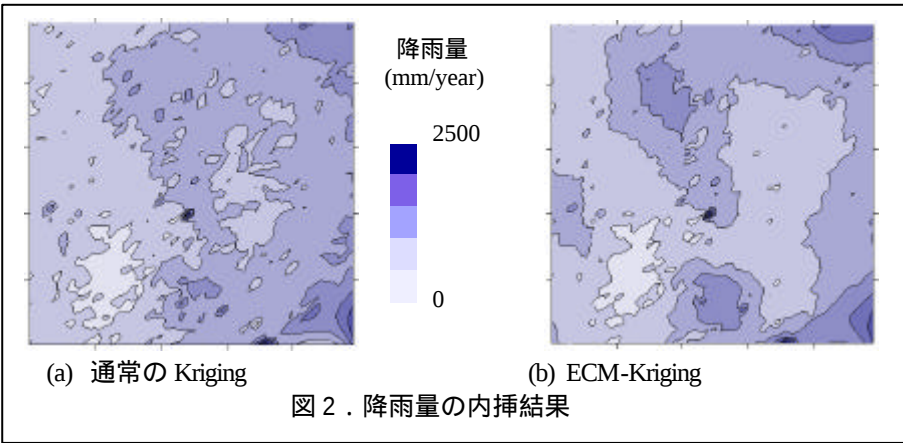
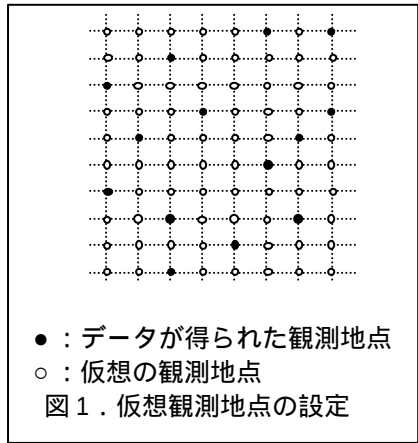
表 1 . パラメータ推定結果

回帰モデル： 降雨量(mm/year) = 経度 + 傾斜面 + 海岸からの距離 + log(標高)

()内は t 値	共分散パラメータ		trend 関数パラメータ (mm/year/)			
	lag	残差の分散	緯度(Km)	傾斜面(度)	海岸への距離(Km)	標高(log)(m)
(a) 通常の Kriging	—	—	1.31 (12.86)	0.51 (1.87)	-0.61 (-2.26)	141.97 (3.86)
(b) ECM-Kriging	0.10	46000	1.39 (9.08)	0.21 (0.84)	-0.61 (-1.75)	141.72 (3.51)

(a)における covariogram として、spherical 型： $c(d_{ij}) = c_0 + c_1 \left\{ (3/2)(d_{ij}/R) - (1/2)(d_{ij}/R)^3 \right\}$ (c_0, c_1, R : パラメータ)を用いた。

validation data に対して推定値と実測値との残差二乗和を求めると、(a) 1.69×10^7 、(b) 1.49×10^7 であった。これは、本ケースでは通常の Kriging の方が局所的な変動に対して過敏に影響を受けたためであると考えられる。



参考文献 Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, Kluwer Academic.
Bailey, T. C. and Gatrell, A. C. (1995): *Interactive Spatial Data Analysis*, Longman.
Cressie, N. (1993): *Statistics for Spatial Data*, Revised Edition, John Wiley & Sons.