スパースモデリングに基づく空間相関構造が変化する空間領域境界の抽出

石山 里穂子・井上 亮

Identifying spatial borders where spatial correlation structures change based on sparse modeling

Rihoko ISHIYAMA, Ryo INOUE

Abstract: The structure of spatial correlation is usually considered to be constant all over the target spatial domain when analyzing spatial data, however, the assumption might be questionable in some cases. This paper proposes a new approach to identify the spatial borders where the spatial correlation structures change. The approach first partitions the target spatial domain into subregions, and assumes the second order stationarity in each subregion. Then it estimates the parameters of spatial covariance functions of each subregion by regularizing the difference of parameters between the adjacent subregions based on the generalized fused lasso. The method tends to estimate the same values for the parameters of adjacent subregions if the spatial correlation structures are similar, and identifies the pairs of adjacent subregions whose estimated parameters are different.

Keywords: 空間統計（spatial statistics），定常領域の抽出（detection of spatial region with stationarity），非定常共分散関数（nonstationary covariance function），スパースモデリング (sparse modeling)，罰則項付き尤度（penalized likelihood）

1. はじめに

近年，情報通信技術や測位技術といった衛星リモートセンシング技術の発展によって，空間的位置と関連付けられた多種多様なデータを入手し，活用できる環境が整ってきている．空間データを用いた分析では，空間現象を表現するモデルを構築する際に，分析対象の空間領域全体に本質的定常性や二次定常性を仮定し，空間相関構造が場所によらず一定であるとすることが多い．

しかし，自然現象・社会経済現象に関わらず，ある境界で空間相関構造が異なる現象は存在する．例えば，尾根線を境に異なる降雨量，国境で異なる人口分布などは，境界の両側ではそれぞれ定常な分布をしていたとしても，それらの空間相関構造は異なる場合があると考えられる．

その変化要因を明示的に説明するモデルの構築が困難な場合や，構造が変化している空間境界が事前に分からない場合は，場所によって空間相関構造が異なることを表現できるモデルを構築した上で，そのモデルの推定を通して空間相関構造の変化を捉える分析を行う必要がある．

空間相関構造の主要なモデルの一つは，空間確率場に二次定常性を仮定した上で，二地点の確率変数の共分散を相対的な位置の関数とした共分散関数で表す手法である．また，等方性も仮定して，二地点の確率変数の共分散をその間の距離の関数として表現することが多い．しかし，場所に

石山里穂子
〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06
東北大学博士前期課程
大学院情報科学研究科 人間社会情報科学専攻
E-mail: rihoko.ishiyama.r5@dc.tohoku.ac.jp

依存して空間相関が異なることを表現するためには，場所によって共分散関数が異なるモデルを構築しなければならない．

これまで，非定常な共分散構造のモデル化を対象とした研究は数多くなされてきた．その一つは，分析対象の空間領域全体を小さな小領域に分割し，各小領域内に定常性を仮定した上で，各小地域間の距離に基づいた重み関数やカーネル関数によって定常な共分散関数を重み付けた平均を用いて，場所に依存する非定常な共分散関数を表現する方法である（例えば，Fuentes, 2001, Fuentes, 2002）．一方，Sampson and Guttorp (1992) は，非定常領域を定常領域に投影する deformation と呼ばれるアプローチを提案している．また，Nychka and Saltzman (1998) のように，共分散関数を EOF 解析を通して基底関数で分解する方法も考えられている．さらに，Higdon (1998) は，空間上で可変なカーネルを用いた畳み込みカーネルによって，より柔軟に共分散関数を表現する方法を提案している．この考えを受けて，Paciorek and Schervish (2006) は，地点毎に固有のパラメータを与えた形で，より一般的な非定常な共分散関数を表した．非定常な共分散構造のモデル化に関する詳細なレビューについては，Sampson (2010) を参照されたい．

一般に共分散関数の推定は，計算コストが膨大となりやすい．そこで，近年，パラメータ推定にスパースモデリングを用いる研究も行われている（例えば，Chang et al., 2010; Parker et al., 2016）．

Parker et al. (2016) は，スパースモデリング手法の 1 つである Generalized fused lasso (Tibshirani and Taylor, 2011) を活用した分析手法を提案している．空間領域を小領域に分割した上で，各小領域内では二次定常性を仮定し，それぞれに固有の共分散関数のパラメータを設定した上で，隣接する小領域の共分散関数のパラメータ間の差に関する L_1 正則化項を加えた罰則付き対数尤度を最大化するアプローチを提案している．このモデルによる隣接地域内で共通のパラメータ推定を通して，空間相関構造が共通の定常な空間領域の特定を試みている．しかし，異なる領域に含まれる二地点間の共分散については，各地点が含まれる領域の共分散関数のパラメータの平均とする単純なモデルで表現されており，地域間の共分散構造に関して十分な検討がなされていない．

以上を踏まえ，本研究は，二次定常性および等方性を仮定した共分散関数を用いた空間相関構造の表現を前提とした上で，地域内，並びに，地域間の空間相関構造が共通の領域を，スパースモデリングに基づいて抽出する分析手法を開発する．具体的には，事前に分割した小領域毎に設置した共分散関数のパラメータと，その隣接地域の共分散関数のパラメータとの間の差に対する罰則項を新たに加え，パラメータが共通の値となる範囲を探索して，空間相関構造が共通の領域，変化する境界の検出を目指す．

2. 既往の非定常な空間相関構造の表現

地点 $\mathbf{s}_i$ における観測値 y_i を要素とする観測値ベクトル $\mathbf{y} = (y_1,\ldots,y_n)^{\mathrm{T}}$ が，平均ベクトル $\boldsymbol{\mu}(\boldsymbol{\beta})$，共分散行列 $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ の多変量正規分布に従うとする．このとき，対数尤度は式(1)で表される．ただし，$\boldsymbol{\beta},\boldsymbol{\theta}$ はモデルのパラメータである．

$$\ell(\boldsymbol{\beta},\boldsymbol{\theta};\mathbf{y}) = -\frac{1}{2}\Big[n\log(2\pi) + \log|\Sigma(\boldsymbol{\theta})| + [\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})]^{\mathrm{T}}\Sigma(\boldsymbol{\theta})^{-1}[\mathbf{y}-\boldsymbol{\mu}(\boldsymbol{\beta})]\Big] \tag{1}$$

2.1 定常な共分散関数

二次定常性及び等方性を仮定した場合，$\phi > 0$, $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)^{\mathrm{T}}$ として，二地点 $\mathbf{s}_i$, $\mathbf{s}_j$ の共分散を表す共分散関数は式(2)で表わされる．

$$C(\mathbf{s}_i,\mathbf{s}_j;\boldsymbol{\theta}) = \tau^2 I(i=j) + \sigma^2 R^s\left(d_{ij}\right) \tag{2}$$

ただし，d_{ij} は

$$d_{ij} = \frac{|\mathbf{s}_i - \mathbf{s}_j|}{\phi} \tag{3}$$

を表し，$|\ \ |$ は二地点間の距離を表す．なお，相

関関数 $R^s(d_{ij})$は，通常指数型で $R^s(d_{ij}) = \exp(-d_{ij})$とすることが多い．また，式(2)中の τ^2, σ^2, ϕ は，それぞれナゲット，パーシャルシル，レンジと呼ばれ，これら三つのパラメータによって共分散関数の形状が表される．空間領域全体で同じ空間相関構造が成立している場合には，同じパラメータで地点間の共分散を表現することができる．

2.2 Parker et al. (2016)

Paciorek and Schervish (2006)による地点毎に三つのパラメータを設定した共分散関数を基に，Parker et al. (2016) はパラメータ数を削減するため，対象領域を小地域 R_r $(r = 1,\ldots,m)$ に分割し，地域毎に三つのパラメータを与えた共分散関数を用いている．したがって，$\mathbf{s}_i \in R_a$，$\mathbf{s}_j \in R_b$ とすると，等方性を仮定した共分散関数は式(4)で表わされる．

$$C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}) = \tau_a^2 I(i=j) + \sigma_a \sigma_b R^{NS}\left(\mathbf{s}_i, \mathbf{s}_j\right) \quad (4)$$

ただし，

$$R^{NS}\left(\mathbf{s}_i, \mathbf{s}_j\right) = \phi_a \phi_b \left\{\frac{1}{2}\left(\phi_a + \phi_b\right)\right\}^{-\frac{1}{2}} R^S\left(D_{ij}\right) \quad (5)$$

$$D_{ij} = \frac{\left|\mathbf{s}_i - \mathbf{s}_j\right|}{\frac{1}{2}\left(\phi_a + \phi_b\right)} \quad (6)$$

したがって，パラメータ $\boldsymbol{\theta}$ は $3 \times m$ の行列であり，$\theta_{1r} = \log(\tau_r^2)$，$\theta_{2r} = \log(\sigma_r^2)$，$\theta_{3r} = \log(\phi_r)$ とする．このとき，地域毎に推定する三つのパラメータについて，隣接地域のパラメータ内に L_1 正則化項を設けた罰則項付き対数尤度（式(7)）を最大化して，隣接地域で共通のパラメータを推定する．

$$\ell_P(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}, \boldsymbol{\lambda}) = \ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) - \sum_{i=1}^{3} \lambda_i \sum_{m \sim n} \left|\theta_{im} - \theta_{in}\right| \quad (7)$$

ただし，$m \sim n$ は小地域 R_m と小地域 R_n が隣接していることを表し，λ_i は罰則の強さを制御するハイパーパラメータである．

式(4)～(6)のように，このモデルは各地域内の共分散関数については固有のパラメータを設定している一方で，地域をまたいだ地域間の共分散関数については，二地域のパラメータの平均を用いた表現となっており，固有のパラメータを設定していない．しかし，例えば隣接する 2 地域で，地域内の空間相関構造は同じパラメータで表現できるけれども，地域間については相関が小さい現象があり得る．そのため，式(5)による地域間の共分散構造のモデル化は必ずしも適切とは言えない．

3. 提案アプローチ

前節で指摘した Parker et al. (2016) の問題点を踏まえ，本研究では地域内及び地域間のそれぞれに対して二次定常性および等方性を仮定した上で，各地域内・地域間に固有の共分散関数のパラメータを与えたモデルを設定し，地域毎に設けた共分散関数のパラメータとその隣接地域間との共分散関数のパラメータ間に罰則項を加え，共通の値となる範囲を探索する分析アプローチを提案する．

地域内及び地域間で，固有の共分散関数のパラメータを与えた共分散関数は，$\mathbf{s}_i \in R_a$，$\mathbf{s}_j \in R_b$，相関関数を指数型とし，地域毎に設定するパラメータを $\boldsymbol{\theta}_E$，地域間毎に設定するパラメータを $\boldsymbol{\theta}_F$ で区別した上で，式(8)で表される．

$$C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_E, \boldsymbol{\theta}_F) = \tau_a^2 I(i=j) + \sigma_{(a,b)}^2 \exp\left(-\frac{\left|\mathbf{s}_i - \mathbf{s}_j\right|}{\phi_{(a,b)}}\right) \quad (8)$$

このモデルは，$a = b$ ならば式(8)は，小地域 R_a 内で二次定常性を仮定したモデルに等しい．したがって，パラメータ $\boldsymbol{\theta}_E$ は $3 \times m$ の行列，パラメータ $\boldsymbol{\theta}_F$ はナゲットを除いた $2 \times {}_mC_2$ の行列で表される．これより，地域毎に設けたパラメータについて，その隣接地域内にパラメータの差を取った L_1 正則化項と，地域間毎に設けたパラメータについて，その隣接地域間において，地域毎に設定したパラメータとの差を取った L_1 正則化項を加えた罰則項付き対数尤度（式(9)）の最大化を考える．

$$
\begin{aligned}
&\ell'_{P}(\boldsymbol{\beta},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2;\mathbf{y},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) \\
&=\ell(\boldsymbol{\beta},\boldsymbol{\theta}_1,\boldsymbol{\theta}_2;\mathbf{y}) \qquad (9)\\
&-\sum_{i=1}^{3}\lambda_{1i}\sum_{m\sim n}\left|\theta_{Eim}-\theta_{Eirn}\right|-\sum_{i=1}^{2}\lambda_{2i}\sum_{m\sim n}\left|\theta_{Fi(m,n)}-\theta_{Eim}\right|
\end{aligned}
$$

ただし，$\theta_{E1r}=\log(\tau_r^2)$，$\theta_{E2r}=\log(\sigma_r^2)$，$\theta_{E3r}=\log(\phi_r)$ 及び $\theta_{F1(a,b)}=\log(\sigma^2_{(a,b)})$，$\theta_{F2(a,b)}=\log(\phi_{(a,b)})$ である．

4. おわりに

本研究では，地域内及び地域間で，固有の共分散関数のパラメータを与えたモデルを考えた上で，地域毎のパラメータの隣接地域内に設定した L_1 正則化項に加えて，地域間毎に設けたパラメータについて，その隣接地域間において，地域毎に設定したパラメータとの差を取った L_1 正則化項も加えた罰則項付き対数尤度を最大化することで，空間相関構造が共通な範囲を探索する分析アプローチを提案した．今後については，シミュレーションデータと実データを用いながら，結果の解釈や，アプローチの再考を検討していきたい．さらに，平均構造に関して一次の非定常性を考慮したモデルの構築を試みる．

謝辞

本研究は, JSPS 科研費 19K21982 の助成を受けた．

参考文献

Fuentes, M., 2001. A high frequency kriging approach for non-stationary environmental process, *Environmetrics*, **12**, 469-483.

Fuentes, M., 2002. Spectral methods for nonstationary spatial processes, *Biometrika*, **89**(1), 197–210.

Sampson, P. D. and Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure, *Journal of the American Statistical Association*, **87**(417), 108–119.

Nychka, D. and Saltzman, N., 1998. Design of air-quality monitoring networks, in *Case studies in environmental statistics*, Springer, 51–76.

Higdon, D., 1998. A process-convolution approach to modelling temperatures in the North Atlantic Ocean, *Environmental and Ecological Statistics*, **5**, 173–190.

Paciorek, C. J. and Schervish, M. J., 2006. Spatial modelling using a new class of nonstationary covariance functions, *Environmetrics*, **17**, 483–506.

Sampson, P. D., 2010. Constructions for nonstationary spatial processes, in *Handbook of Spatial Statistics*, eds. Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P., CRC Press, chap. 9.

Chang, Y.-M., Hsu, N.-J., and Huang, H.-C., 2010. Semiparametric estimation and selection for nonstationary spatial covariance functions, *Journal of Computational and Graphical Statistics*, **19**(1), 117–139.

Parker, R. J., Reich, B., J., and Eidsvik, J., 2016. A fused lasso approach to nonstationary spatial covariace estimation, *Journal of Agricultural, Biological, and Environmental Statistics*, **21**(3), 569–587.

Tibshirani, R. and Taylor, J., 2011. The solution path of the generalized lasso. *Annals of Statistics*, **39**(3), 1335–1371.