

空間データの形態が分析結果に及ぼす 影響の分析：空間フィルタリングを用いて

筑波大学大学院
システム情報工学研究科
村上大輔

時空間データの多様化

■ 時空間データの多様化

- Point data (例: 公示地価)
- Areal data (例: 市町村別人口)
- Line data (例: 道路交通量)
- Surface data (例: DEM)



多様な時空間データの分析

- 形態の異なる複数のデータを同時に取り扱う必要が生じる
- データの形態の差異が分析に及ぼす影響は・・・？

計量地理学

- 形態の影響は議論すべき問題(Openshaw and Taylor, 1981)
- 古くから議論されてきた

(空間)統計学

- 形態の影響は議論すべき問題ではない(Tobler, 1979)
- 殆ど注目されてこなかった(Gelfand, 2010)

代表例：集計問題

- 分析結果が集計単位の取り方に応じて変化する問題

人口分布

5	10	15
0	0	0
0	0	0



集計された人口

5	10	15
---	----	----

犯罪数分布

5	0	0
0	10	0
0	0	15



集計された犯罪数

5	10	15
---	----	----

相関係数：-0.3

相関係数：1.0

計量地理学と集計問題

■ 代表的な対処法

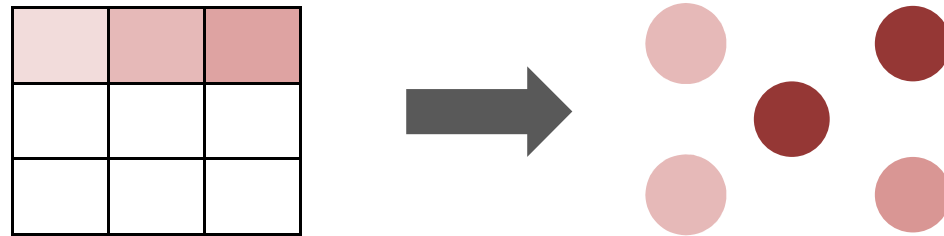
(Swift et al., 2008)

- GWRの適用
- 非集計データを使用する
- ミクロな共分散構造を仮定
- シミュレーション等により判断
- 感度分析を行う

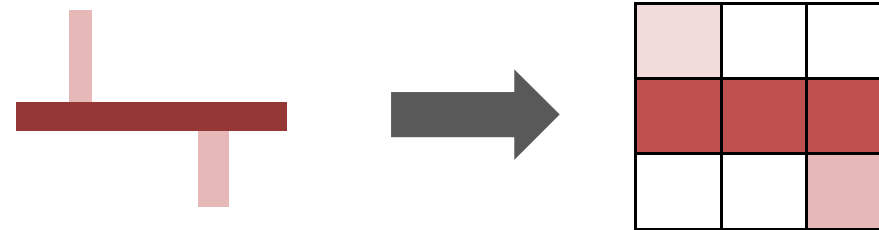
- それぞれ短所があり、**計量地理学では必ずしも理論的に十分な対処法は存在しないとされることが多い** (e.g., McMaster and Sheppard, 2004; Swift et al., 2008)

空間統計学と形態変換

- 形態変換の問題は空間統計学ではCOSP (Change of Support Problem) と呼ばれる



計量地理学の手法との
本質的な違い

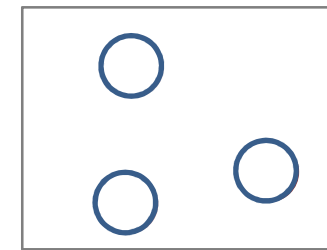
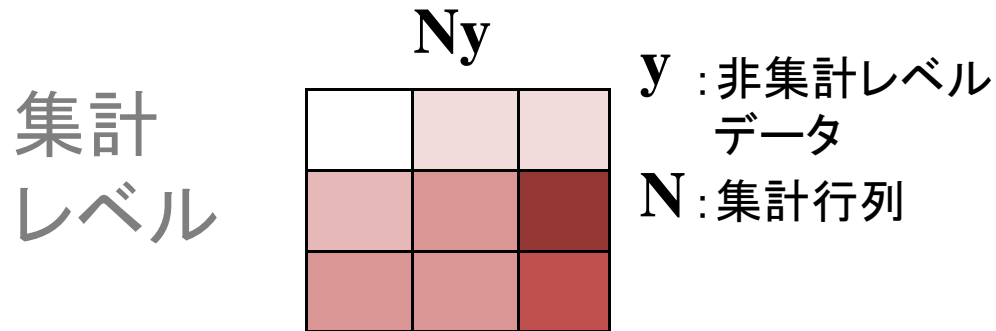


- 予測誤差分散の最小化に基づいて非集計レベルのモデルを構築し、これを用いて対処
- 統計学的にかなり有効な集計問題への対処法も存在
- COSPは非常にマイナーであり地理学ではほぼ引用されない

空間統計学と形態変換: アプローチ

変換前データ(既知)

変換後データ(未知)



直接適用

サンプリング
で復元

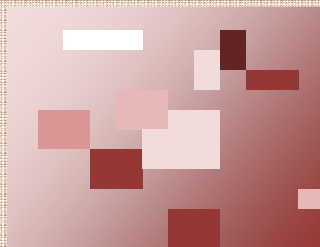
非集計データ(未知)

予測値

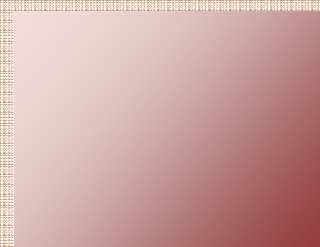
予測不能な要因

$$y = \hat{y} + u$$

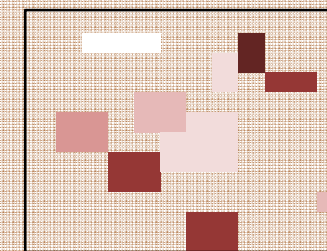
非集計
レベル



=



+



空間統計学と形態変換：主な研究グループ

■ 統計学者：Gelfandら

- ひたすら柔軟なベイズモデルを構築
- 集計データを補助データとして気象データの空間構造を探ることに関心

■ 空間疫学者：Gotway, Youngら

- Geostatistical simulationを用いた研究が多数
- 非集計データを元に集計単位毎の疾病データの空間構造を探ることに関心

■ 地理学者：Kyriakidisら

- 集計データから非集計レベルの空間構造を特定することに関心 (Area-to-point kriging)
- 潜在構造を考慮して回帰係数を与える空間統計モデルを構築 (Geostatistical linear model)

計量地理学と空間統計学の各方法の比較

■ 計量地理学

- 実用的ではあるものの精緻さでは劣る

■ 空間統計学

- 精緻ではあるものの実用性は極めて低い



- 形態変換は誰しものが直面しうる問題であるため
実用性は重要
- 精緻な手法を適用することが望ましい

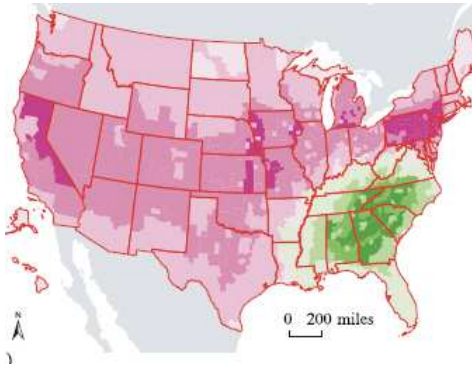
実用的な空間モデル：空間フィルタリング

- 空間的な成分を説明変数で考慮するモデル

$$y_s = \sum_k x_{s,k} \beta_k + \sum_l E_{s,l} \gamma_l + \varepsilon_s$$

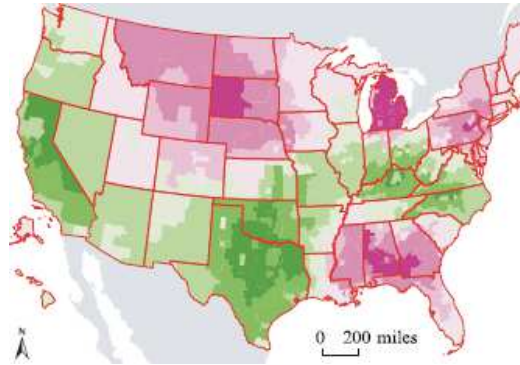
大域的空間成分

$E_{s,1}$



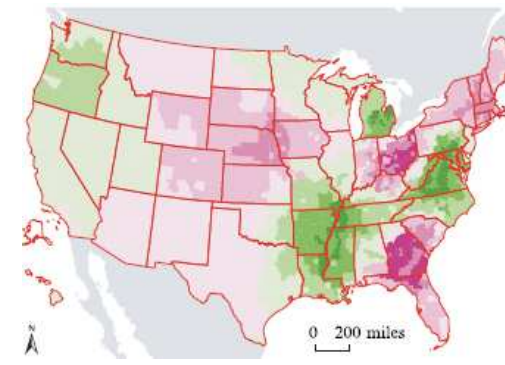
近接行列の固有ベクトル

$E_{s,6}$



局所的空間成分

$E_{s,9}$

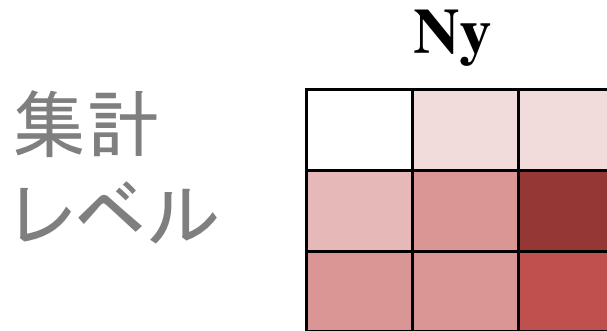


- 本来、集計データのための空間統計モデルの一種
- 非集計データのための手法に独自に拡張

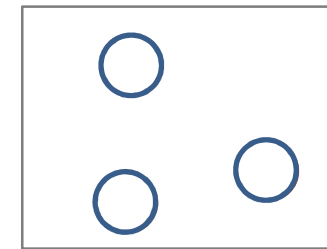
本研究のアプローチ

変換前データ(既知)

変換後データ(未知)



y : 非集計レベル
データ
 N : 集計行列



直接適用

サンプリング
で復元

非集計データ(未知)

$$y = \hat{y} + u$$

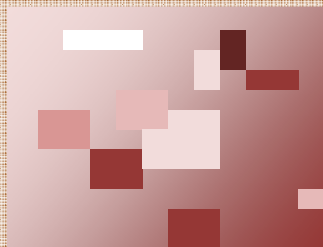
予測値

$$\hat{y} = X\hat{\beta} + E\hat{y}$$

予測されなかった要因

u

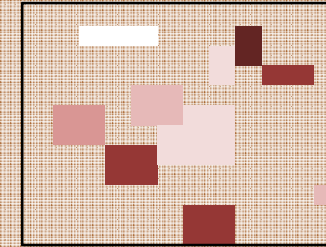
非集計
レベル



=



+



※以前別バージョンの方法を発表したが何かと制約が多かったためこちらに修正した¹¹

提案する方法の特性

	計量 地理学	空間 統計学	空間 フィルタリン グ
予測誤差分散の最小化 (統計学的な望ましさ)	×	○	○
空間的相関の考慮	???	○	○
パラメータ推定法	さまざま	MCMCなど	一般化最小 二乗法
実装の容易さ	◎	×	○

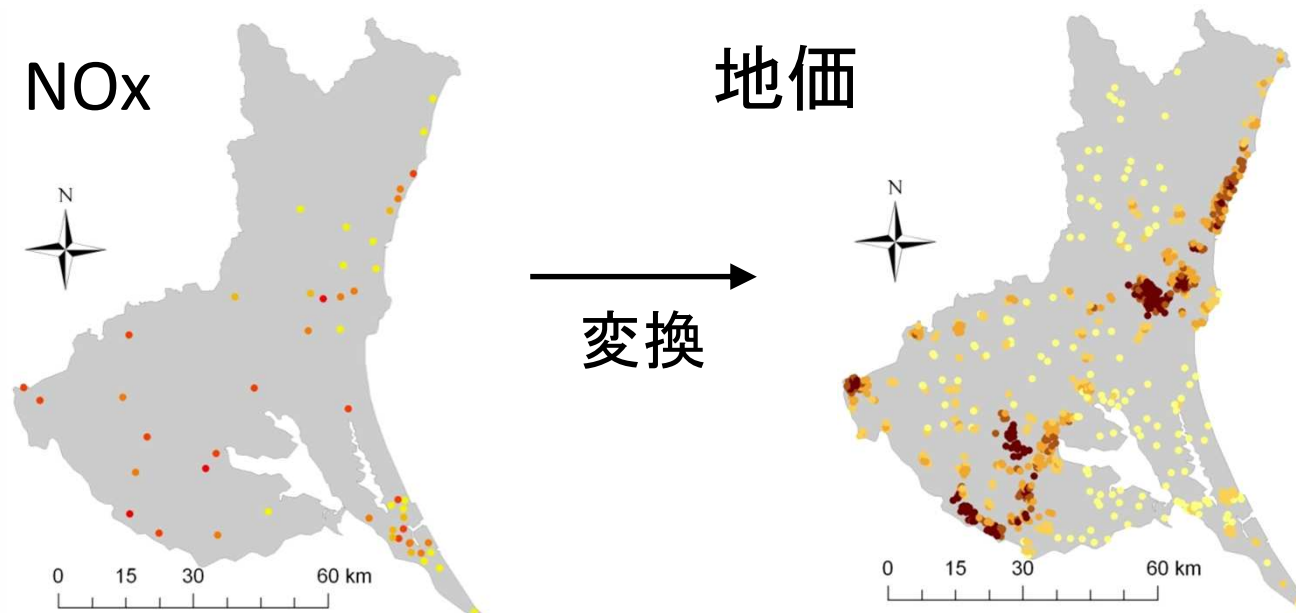
提案手法のヘドニック分析への適用

- NO_x が住宅地価(地価調査+都道府県地価調査:時点調整済み)に与える影響を、茨城県を対象に検証
- 被説明変数
 - 住宅地価
- 説明変数
 - NO_x
 - 東京駅までの鉄道ネットワーク距離
 - 最寄り駅までの距離
 - 所属する1kmメッシュにおける各土地利用面積
- NO_x データと住宅地価データは観測点が異なる

モデル

- 地価観測地点についてのモデルを構築

	地価関数	説明変数のNO _x (\mathbf{x})の与え方
Non-Spatial	線形回帰モデル	予測値: $\hat{\mathbf{x}}$
Spatial	空間フィルタリング	予測値: $\hat{\mathbf{x}}$
COSP-Spatial	空間フィルタリング	予測値 + 予測不能な攪乱: $\hat{\mathbf{x}} + \mathbf{u}$



パラメータ推定結果

COSP-Spatial

Spatial

Non-Spatial

	推定値	t値		推定値	t値		推定値	t値	
定数項	10.30	411.99 ***		10.29	418.09 ***		4.59	41.64 ***	
東京までの距離(km)	0.00	0.00		0.00	0.00		0.00	4.03 ***	
最寄駅まで距離(km)	-0.10	21.80 ***		-0.10	21.81 ***		-0.03	5.24 ***	
田(km ²)	-0.35	6.88 ***		-0.35	6.84 ***		0.00	0.00	
その他農地(km ²)	-0.48	7.14 ***		-0.47	7.05 ***		-0.58	5.36 ***	
森林(km ²)	-0.21	2.89 ***		-0.21	2.82 ***		-0.51	4.50 ***	
荒地(km ²)	0.00	0.00		0.00	0.00		1.99	3.45 ***	
道路(km ²)	0.00	0.00		0.00	0.00		-5.37	3.29 ***	
鉄道(km ²)	5.54	2.52 ***		5.67	2.58 ***		0.00	0.00	
その他用地(km ²)	0.00	0.00		0.00	0.00		0.00	0.00	
河川・湖沼(km ²)	-0.32	3.74 ***		-0.32	3.77 ***		-0.51	3.41 ***	
海水域(km ²)	0.00	0.00		0.00	0.00		-4.49	8.70 ***	
NO2	-0.27	0.31		0.00	0.00		33.69	9.38 ***	
補正R2	0.88			0.88			0.54		

考察

- COSP-SpatialとSpatial
 - 有意となった最寄駅までの距離(－), 田(－), その他の農地(－), 森林(－), 鉄道(+), 河川・湖沼(－)の各符号は直観と整合
 - Simul-SpatialとSpatialの結果はほぼ同じ
- Non-Spatial
 - 東京駅までの距離(+), 荒地(+), NOx(+)の符号が直観と整合しない
 - 精度が低い
- (空間従属性に起因した)データの空間パターンを考慮することがパラメータを適正に評価するうえで重要と再確認
- 本ケースではCOSP-SpatialとSpatialに差異はみられなかった

まとめ

- 空間フィルタリングを用いて形態の異なるデータを適正に取り扱う方法を議論した
- 従来の空間統計学の方法と同様に予測誤差分散最小化に基づき、かつより実用的な方法
- 集計問題等にも適用できるので今後適用したい