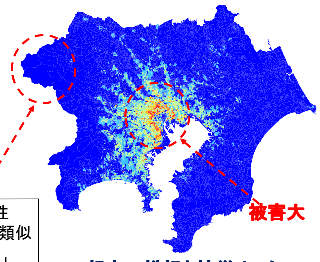
# ビックデータを用いた都市レジリエンス分析のための新たな空間統計手法の開発 :Eigenvector spatial filteringの連続空間への拡張

山形与志樹　国立環境研究所　主席研究員（サービス科学研究センター　客員教授）
村上大輔　　国立環境研究所　特別研究員




都市の様相と被災リスク

## 【研究の背景と目的】

- 近年震災や水害の多発するわが国において、都市の**レジリエンス**（=致命傷の回避; 被害の最小化; 早急な回復）の分析は重要
- レジリエンスの分析に際しては、**都市の様相の詳細な把握**が必要となる

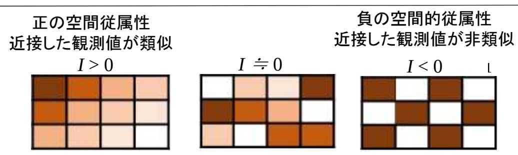
都市レジリエンス分析の第一歩として、ビッグデータを活用して都市の様相を詳細に把握するための空間統計手法を開発する

## 【Eigenvector spatial filtering: ESF】

### モランＩ統計量

- 右式で与えられる空間的従属性の検定等計量: $I[\mathbf{z}]=\dfrac{N}{\mathbf{1}'\mathbf{K}\mathbf{1}}\dfrac{\mathbf{z}'\mathbf{MKMz}}{\mathbf{z}'\mathbf{Mz}}$

$\mathbf{z}$ :変数ベクトル　$\mathbf{K}$: 近接行列　$\mathbf{M}$: 射影行列　$\mathbf{1}$ :を要素に持つベクトル　$N$: サンプル数




### $I$の近接行列MKMの固有ベクトル$\mathbf{E}=[\mathbf{e}_1 \dots \mathbf{e}_n, \dots \mathbf{e}_N]$

- $I[\mathbf{e}_n]$は対応する固有値$\lambda_n$に比例しており、$\mathbf{e}_n$は$I$の基底（空間従属性を説明する成分）と考えることができる

$$I[\mathbf{e}_n]\propto\frac{\mathbf{e}_n'\mathbf{MKMe}_n}{\mathbf{e}_n'\mathbf{Me}_n}=\frac{\mathbf{e}_n'\mathbf{E\Lambda E}'\mathbf{e}_n}{\mathbf{e}_n'\mathbf{Me}_n}=\lambda_n$$

### ESF＝ モランI統計量 + 回帰モデル

- ESFとはMKMの固有ベクトル$\mathbf{E}=[\mathbf{e}_1 \dots \mathbf{e}_n, \dots \mathbf{e}_N]$を説明変数として用いることでデータの空間成分をモデル化するアプローチであり、その基本式は次の通りである。

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{E}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}\qquad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$$

$\mathbf{\Lambda}$:固有値を要素 に持つ対格行列
$\mathbf{y}$ :データ
$\mathbf{X}$: 説明変数
$\mathbf{0}$ :ゼロベクトル
$\mathbf{I}$ :単位行列
$\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2$: パラメータ

## 【ESFの拡張】

- ESFは集計データ等の離散空間上のデータをモデル化する手法である。一方で、ビッグデータは連続空間上に分布したデータの集合である場合が多い（Twitterデータ、人の流れデータ etc.）。そこで本研究は、**ESFを連続空間上のモデルに拡張する**。

### 連続空間上のモランI統計量:$I^+$

- 地域$D$を無限個の地点の集合とみなすと(infill asymptotics)、$D$上のモランＩ統計量は次のように表現できる

$$I^+[z_n]=\frac{1}{\sigma_z^2}\frac{\int_{n'\neq n}\int_n k(s_n,s_{n'})\tilde{z}_n\tilde{z}_{n'}ds_n ds_{n'}}{\int_{n'\neq n}\int_n k(s_n,s_{n'})ds_n ds_{n'}}$$

$s_n$: 地点　$k(s_n, s_{n'})$: 地点間の近接性　$\tilde{z}_n$ : センタリングされた変数$z_n$

### 提案モデル

- 観測地点が地域$D$上に均一に分布するという仮定の下で$D$上の空間成分$\mathbf{E}\boldsymbol{\gamma}$($I^+$の基底)が近似できる。本研究ではこれを説明変数に適用する。

観測地点のモデル
非観測地点$s_0$のモデル
それ以外の全地点のモデル

$$\begin{pmatrix}\mathbf{y}\\ y(s_0)\\ -\end{pmatrix}=\begin{pmatrix}\mathbf{X}\\ \mathbf{x}(s_0)\\ -\end{pmatrix}\boldsymbol{\beta}+\begin{pmatrix}\mathbf{E}\\ \mathbf{e}(s_0)\\ -\end{pmatrix}\boldsymbol{\gamma}+\begin{pmatrix}\boldsymbol{\varepsilon}\\ \varepsilon(s_0)\\ -\end{pmatrix}$$

標準的な仮定の下では $\mathbf{e}(s_0)=(\mathbf{k}(s_0)-\mathbf{1}'\mathbf{K}/N)(\mathbf{I}-\mathbf{11}'/N)\mathbf{E\Lambda}^{-1}$

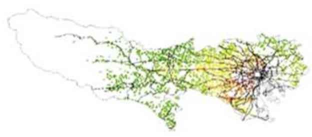
### 従来の空間統計モデルとの関係

- 従来モデルが変量効果として空間成分を考慮するのに対し、提案モデルは固定効果としてこれを考慮する。ただし、両モデルは同じ形式で記述できる。また、$\boldsymbol{\beta}$を既知とした場合、非観測地点のデータ$y(s_0)$の予測量は、両モデルで同一となる。

従来モデル: $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{u}$　$\mathbf{u}\sim N(\mathbf{0},\tau^2\mathbf{K}+\sigma^2\mathbf{I})$を展開することで導出

**提案モデル:** $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{E}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}\qquad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$
空間成分（固定効果）

**従来モデル:** $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{E}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}\qquad \boldsymbol{\gamma}\sim N(\mathbf{0},\tau^2\mathbf{\Lambda})\qquad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$
空間成分（変量効果）



地価公示データ(2000)

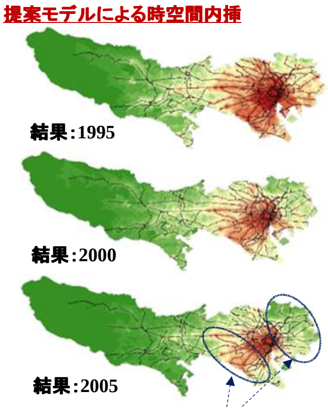
## 【実証分析】

- 提案モデルを、以下のモデルとともに公示地価データ（1995-2006: 標本数24,120）の分析に適用する。

Basic model（BM）: $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\qquad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{I})$

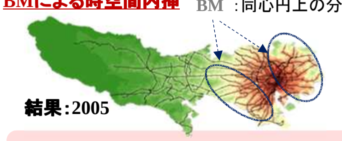
**時空間kriging（従来の空間統計モデル）:** $\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\qquad \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\mathbf{K}_{s,t})$

### 提案モデルによる時空間内挿




**提案**:西高東低状の分布
BM :同心円上の分布

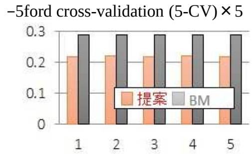
### BMによる時空間内挿




### RMSEによる予測誤差の比較

- **提案** vs BM

–5ford cross-validation (5-CV)×5




- **提案**vs Kriging

–Krigingの計算負荷に対処するために、標本を5分割し各1/5標本に1回ずつ5-CV

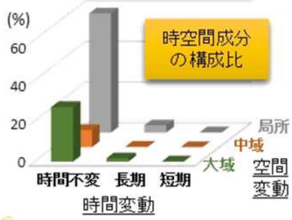

### 計算時間の比較

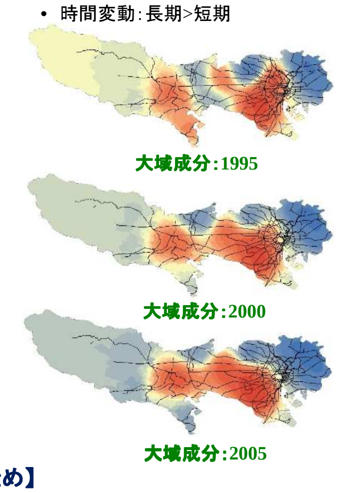
| | | 全標本 | 1/5標本 |
|---|---|---|---|
| 標本数 | | 24,120 | 3,859 |
| 内挿地点数 | | 66,060 | 965 |
| 計算時間（秒） | 提案 | 36.36 | 1.69 |
| | Kriging | 計算不可 | 32.53 |

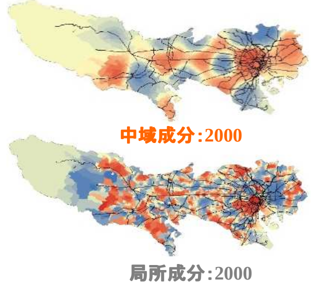
以上より、提案はKrigingと同等の精度を有するとともに、大規模データを高速に処理する手法であることを明らかとした。

### 提案モデルによる時空間成分(Eγ)の抽出

- 空間変動:局所>大域>中域
- 時間変動:長期>短期




## 【まとめ】

- 空間統計モデルの一種であるESFを、大規模データモデリングの手法に拡張
- 拡張手法の有用性を時空間内挿と時空間成分抽出の2つの分析より確認
- 今後、提案手法を都市レジリエンスの分析に適用予定