

# 時空間可変パラメータモデルの高速なモデル選択法の検討

村上 大輔 統計基盤数理研究系 准教授

## 1. 概要

時空間上で滑らかに変化する回帰係数Spatiotemporally varying coefficient (STVC)を推定する回帰モデルはSTVCモデルと呼ばれる。同モデルでは、通常、回帰係数毎に単一の時空間過程が仮定されるが、現実には複数のプロセスが回帰係数に影響を与えている可能性がある。たとえば、地域毎・時期毎の犯罪頻度を被説明変数とする場合、回帰係数は時間帯、曜日、月といった時間軸毎に時空間パターンを持つ可能性がある(例: 日中は市街地で夜間は郊外で犯罪が増える、夏は全域的に犯罪が増える)。また地理的特性に依存した時間不変の空間パターンや、マクロ経済状況等に応じて変化する移動不変の時空間パターンもあろう。各パターンを適切に検出することは、オーバーフィッティングを抑制しながらSTVCモデルの精度を高めていくうえで重要である。しかし、特に時空間上の相互作用(時間と空間の両方に依存するパターン)を考慮する場合、モデルが複雑になり計算コストが増大する。

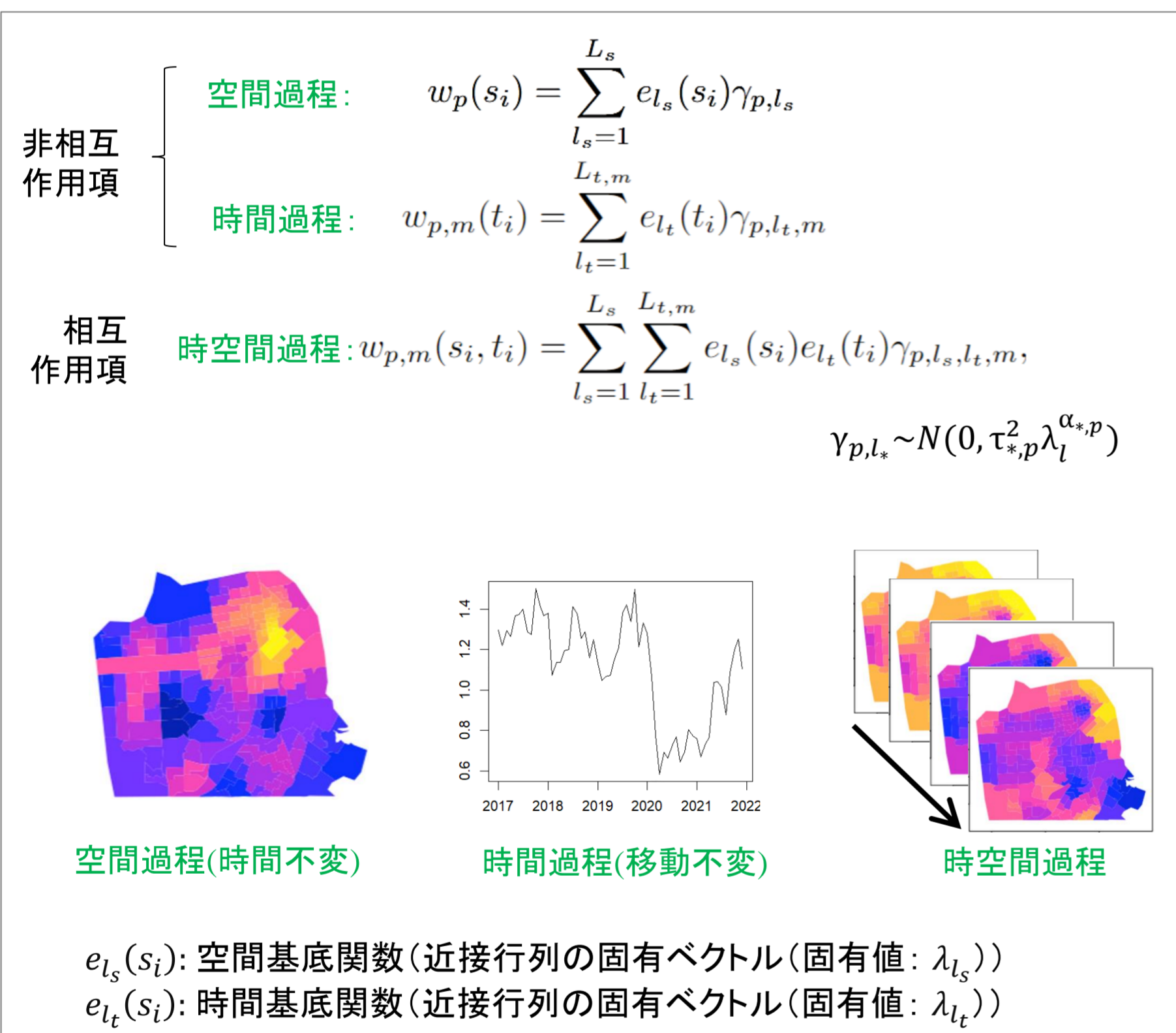
そこで本研究では、各回帰係数の背後にある時空間パターンを計算効率良く選択することで、STVCモデルを高速化する方法を検討する。計算コスト削減のために、時空間上の相互作用項と非相互作用項を区別した上で、reluctant interaction modelingをベースの応用により高速にモデル選択を行う。また、同手法の有用性を犯罪データへの応用を通して検証する。

## 2. STVCモデル

- 地点 $s_i$ 、時点 $t_i$ で観測された被説明変数 $y(s_i, t_i)$ を下式でモデル化
- 安定・高速に回帰係数を推定するために、計 $P(1+2M)$ 個の確率過程の中から説明力のあるものを選択し、計算効率良くモデル化したい

$$y(s_i, t_i) = \sum_{p=1}^P x_p(s_i, t_i) \beta_p(s_i, t_i) + \epsilon(s_i, t_i), \quad \epsilon(s_i, t_i) \sim N(0, \sigma^2)$$

$$\beta_p(s_i, t_i) = b_p + w_p(s_i) + \sum_{m=1}^M w_{p,m}(t_i) + \sum_{m=1}^M w_{p,m}(s_i, t_i)$$



## 3. 提案手法

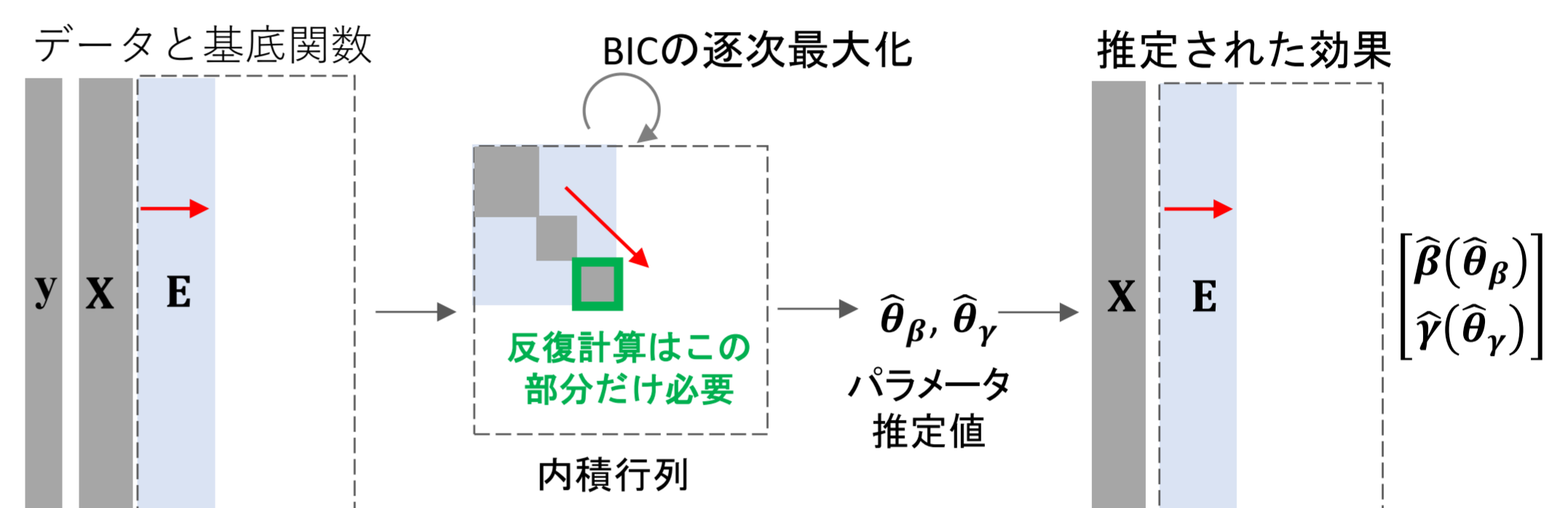
- Reluctant interaction modelingのアイデアを元に「非相互作用項を最初にモデル化し、その後相互作用項を逐次導入」する方法を検討
  - ✓ 非相互作用項: 空間過程( $P$ 個)、時間過程( $PM$ 個)
  - ✓ 相互作用項: 時空間過程( $PM$ 個)

### 計算手順

- 手順1: 非相互作用項のみのSTVCモデルを推定
- 手順2: 相互作用項を1つ追加したSTVCモデルを、全相互作用項について推定
- 手順3: 手順2でBayesian Information Criterion (BIC)を最小にした相互作用項を1つモデルに追加。ここで選択された相互作用項のハイパーパラメータ( $\tau_{s,p}^2, \alpha_{s,p}$ )は、以降で固定
- 手順4: 手順2-3をBICが改善しなくなるまで反復

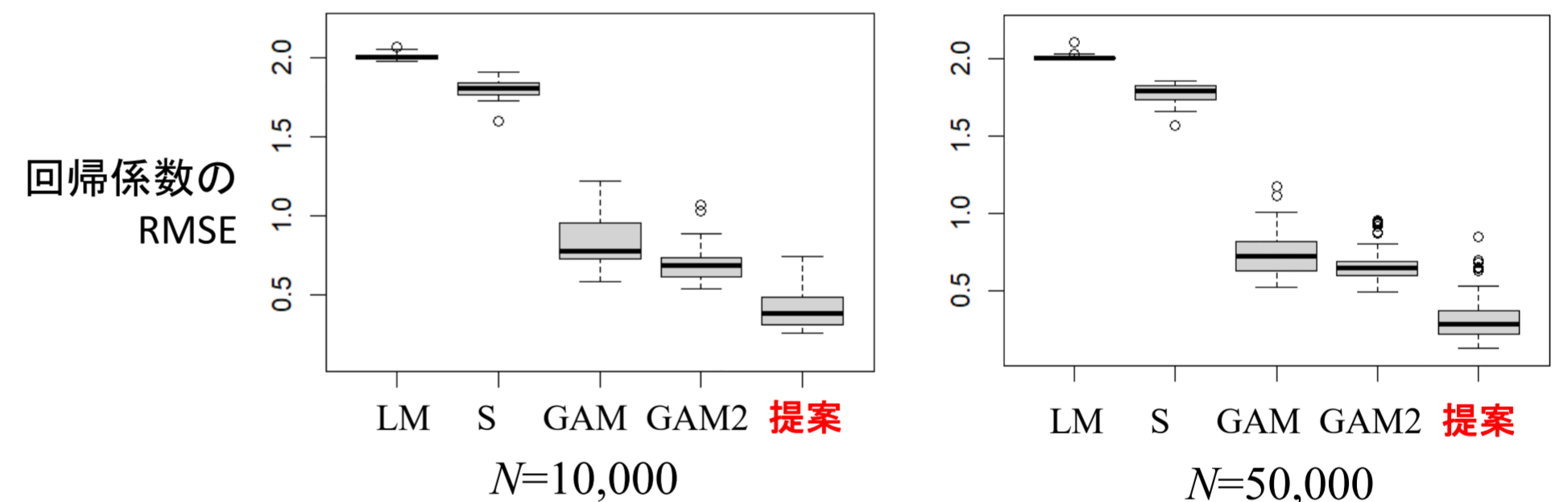
### 計算コスト削減

- 各基底関数を用いずに、それらの内積行列を用いて一連の計算が可能
  - サンプルサイズに依存しない計算コストでモデルが推定・選択可
- 手順2では、内積行列のうち、追加する相互作用項についての部分のみを繰り返し評価すればハイパーパラメータが推定できる
  - 選択された相互作用項の増加に伴う計算コスト増加が小さい



## 4. シミュレーション実験

- STVCの推定精度: 既存手法を上回ることを確認
- 計算時間: 例えば $N=50,000$ の場合の計算時間が789秒であり、7,023秒を要した従来法(一般化加法モデル)よりも高速であることを確認



## 5. 犯罪分析への応用

- サンフランシスコを対象に、Larcenyの要因分析に提案したSTVCモデルを適用
- 被説明変数: Larceny件数(面積あたり対数値、時間別・月別)
- 説明変数: 前期のLarceny件数、人口密度、所得、人種多様性

### 結果(一部)

- 回帰係数(前期の件数): 近接反復傾向がコロナ禍で弱まったこと、北西の中心部で反復傾向が強いことをなど確認
- 回帰係数(人口密度): 一部郊外で人口集中地区における犯罪リスク増が顕著と確認

→ 提案したモデル選択法で、解釈しやすい推定結果が得られることを確認

