

空間統計モデルを用いた つくばエクスプレス沿線の地価の分析

筑波大学大学院
システム情報工学研究科1年
都市交通研究室
瀬谷 創

つくばエクスプレス(TX)

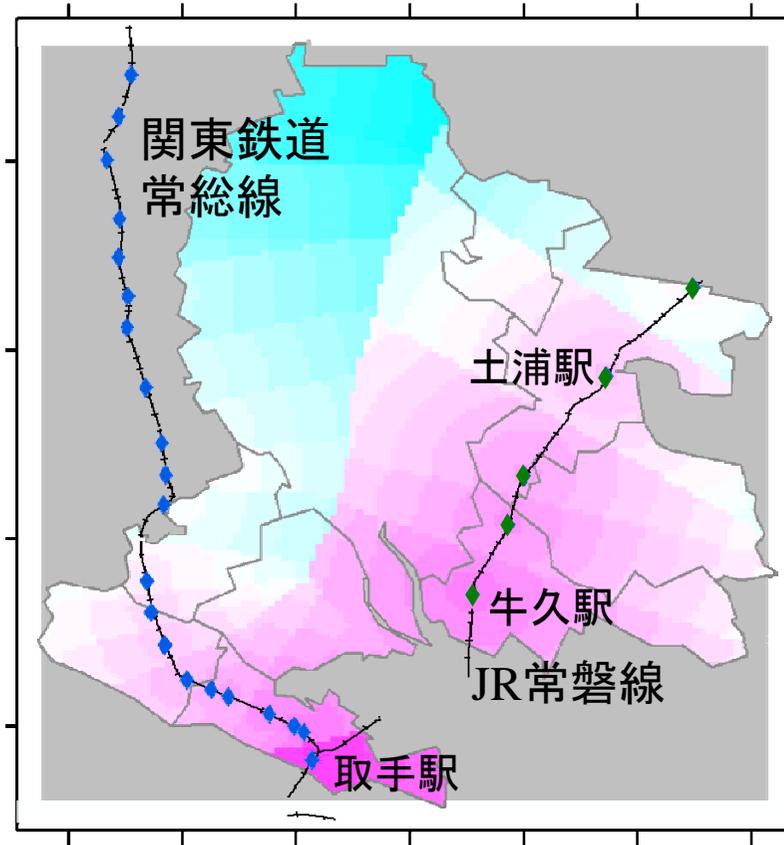
- **2005年8月24日つくばエクスプレス開業**
 - つくばと秋葉原を最短45分で結ぶ
 - 首都圏の都市鉄道では我が国で最後といわれる大規模プロジェクト
- **TXが沿線地域に与えた影響の一例**
 - 交通体系の変化
 - 時間費用の低下
 - ライフスタイルの変化
 - 観光客の増加
 - 地価の上昇



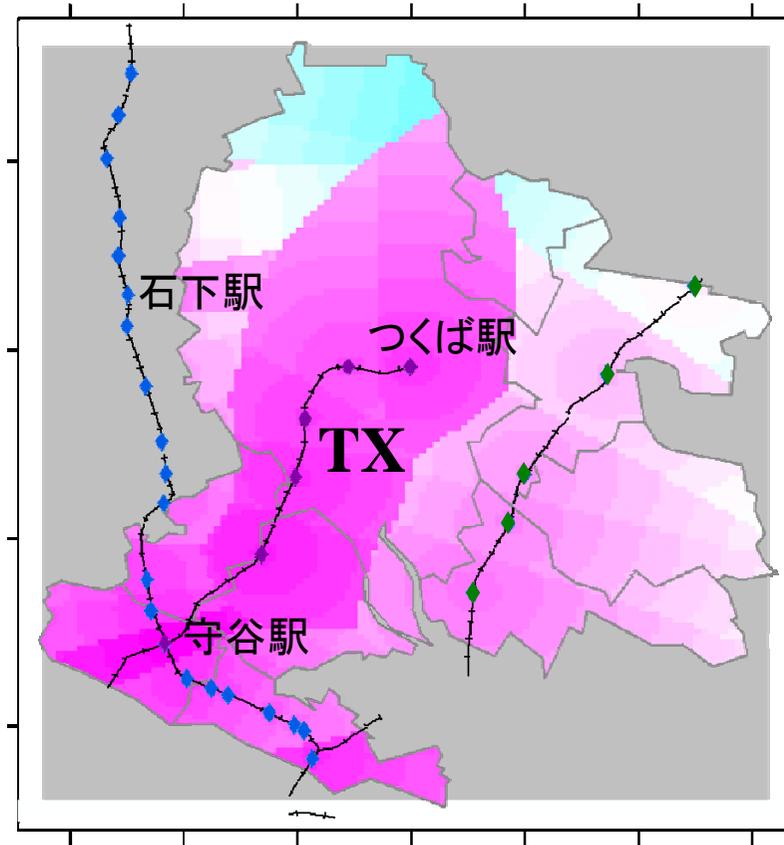
TXによる時間短縮効果 (1)

- 秋葉原駅までの所要時間 (S-plus levelplot関数を使用)

【TX開業前】



【TX開業後】



※ 計算の前提:

- ① 最寄駅までの所要時間: 全国平日混雑時旅行速度35.3km(平成17年度道路交通センサス速報値より)を、各最寄駅までの直線距離に乗じることによって算出
- ② 最寄駅から秋葉原駅までの所要時間: gooの路線検索を用いて、平日午前9:00に秋葉原駅到着という条件で検索

(分)

120

100

80

60

40

3

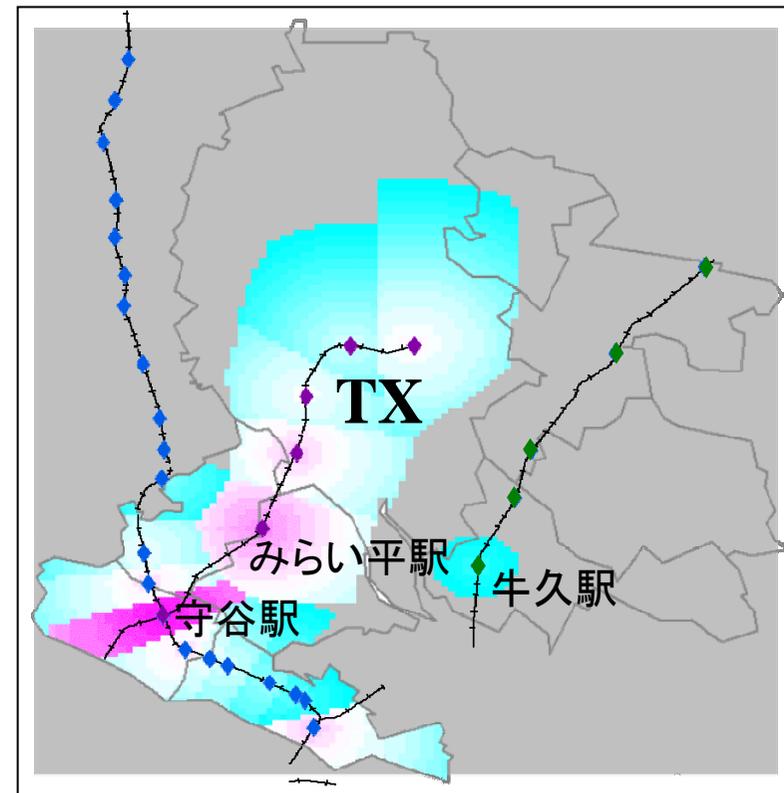
TXによる時間短縮効果 (2)

- 60分以内に秋葉原駅に到達できる範囲

【TX開業前】



【TX開業後】



(分)

55

50

45

40

35

4

※ 算出の前提は前頁と同様である

研究の目的

- 空間統計モデルを用いたつくばエクスプレス沿線地域の公示地価変化の視覚化
- 空間統計モデルを用いたつくばエクスプレス沿線地域の地価モデルの作成

空間統計モデル

- 通常の変帰モデル(最小二乗法)

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \tau^2 I)$$

- 空間統計モデル

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \Sigma)$$

$$\text{and } \Sigma = \tau^2 I + \sigma^2 H(\phi)$$

$$\text{where } (H(\phi))_{ij} = \underline{\rho(\phi; d_{ij})} \quad d_{ij} = \|s_i - s_j\|$$

距離の関数

弱定常性

- 弱定常性(weak stationary)

$$\begin{cases} E[\varepsilon(s_i)] = \mu \text{ const.} \\ \text{Cov}[\varepsilon(s_i), \varepsilon(s_j)] = \underbrace{C(d_{ij})}_{\text{covariogram}} \end{cases}$$

誤差項における
空間的な自己相関
を考慮



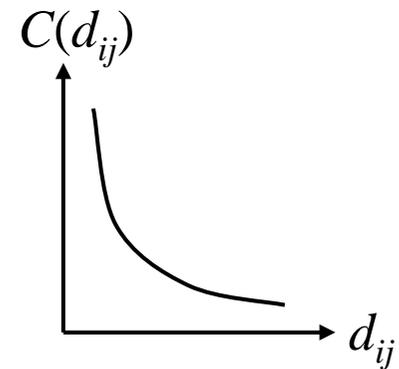
地点 i, j 間の共分散が、 i, j 間の距離 d_{ij} のみに依存

例) 指数型(exponential)の場合

$$(H(\phi))_{ij} = \rho(\phi; d_{ij}) = \exp(-d_{ij} / \phi)$$

if $d_{ij} > 0$, otherwise, $\rho(\phi; d_{ij}) = 0$

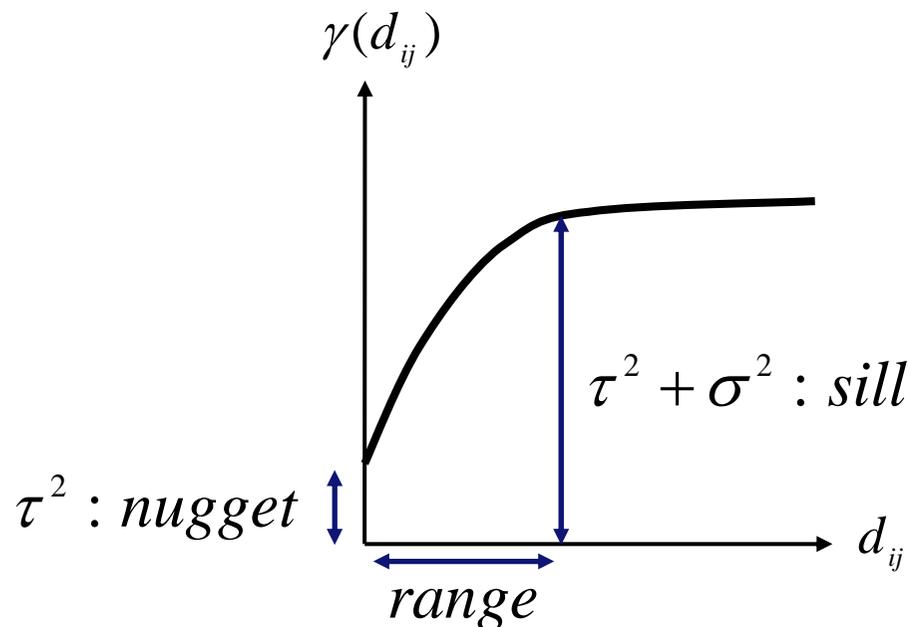
where $d_{ij} = \|s_i - s_j\|$



本質的定常性

- 本質的定常性 (Intrinsically stationary)

$$\begin{cases} E[\varepsilon(s_i) - \varepsilon(s_j)] = 0 \\ \text{Var}[\varepsilon(s_i) - \varepsilon(s_j)] = \underbrace{2\gamma(d_{ij})}_{\text{variogram}} \quad \text{where } d_{ij} = \|s_i - s_j\| \end{cases}$$



- Variogramのパラメータ

nugget : 微小スケールの変動
測定誤差

range : 相関がある
距離 d_{ij} の最大値

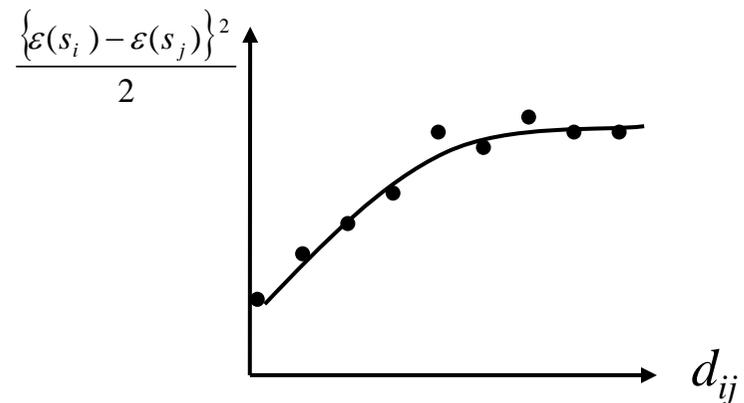
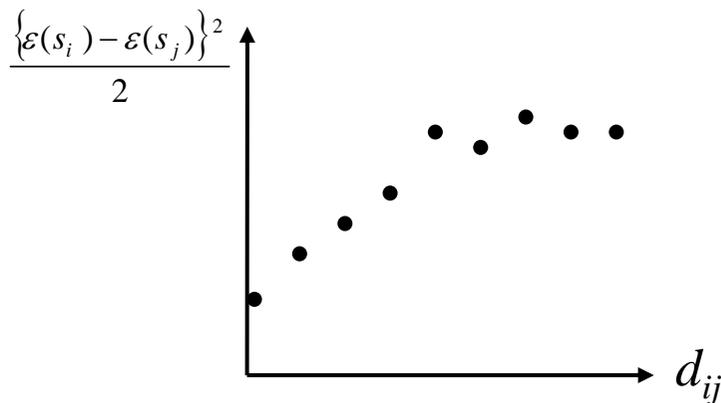
sill : ペア i, j 間の分散 $\gamma(d_{ij})$
の最大値

Covariogram と Variogram

- 確率場が弱定常で、かつエルゴード性を満たす
($C(d_{ij}) \rightarrow 0$ as $d_{ij} \rightarrow \infty$)とき、

$$C(d_{ij}) = C(0) - \gamma(d_{ij})$$

- Variogramのパラメータ推定
 - OLSの残差から経験variogramを作成し、非線形最小二乗法等により理論モデルにあてはめる(パラメータ推定)
 - 分散共分散行列 $\hat{\Sigma}$ 既知のもとでGLSで $\hat{\beta}$ を推定



Ordinary Kriging

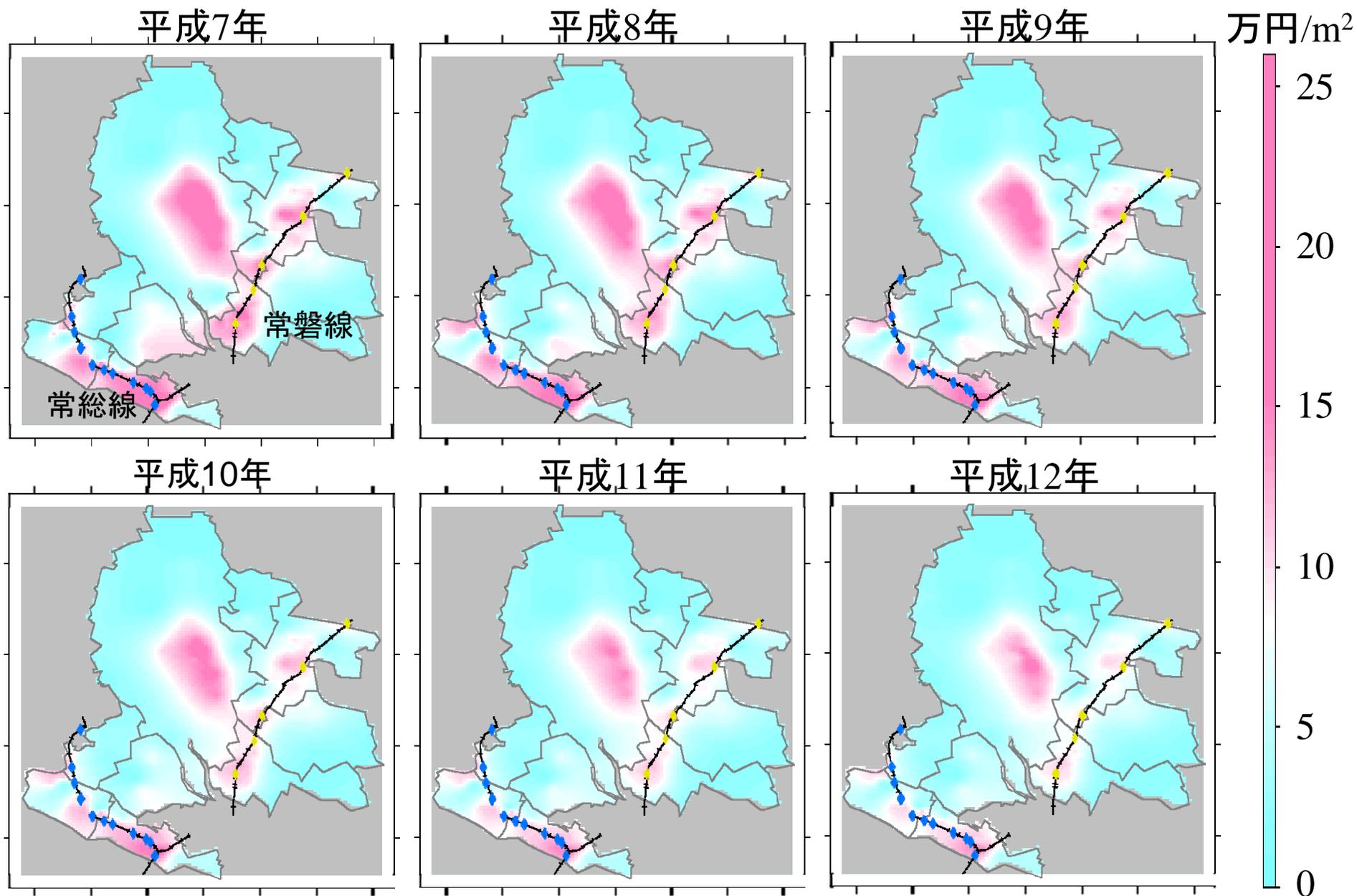
- **Kriging**

- 値が得られていない地点の値の内挿(予測)を統計学的に厳密な方法で行うための手法
 - 最良線形不偏推定量(BLUP)

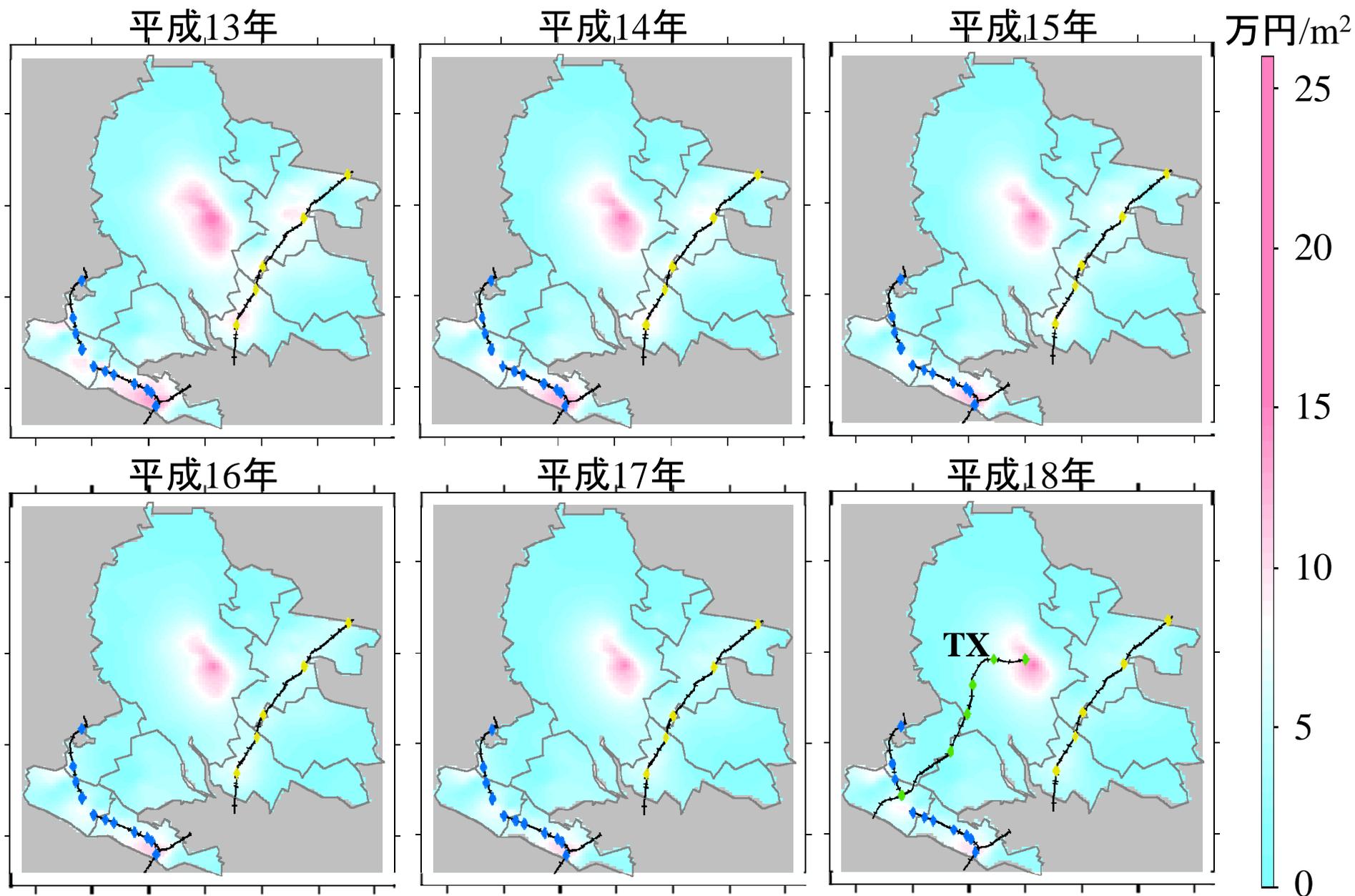
- **Ordinary Kriging**

- トренд項 $X\beta$ がないとする方法
 - 任意地点における説明変数の値が不明な場合に、値の空間的な分布を表現する簡便な手段として有効
- 分析には、S-plusのSpatial Statを利用した。結果の表示も、グラフィック用の様々な関数を用いれば、簡単に分かりやすい図が得られる

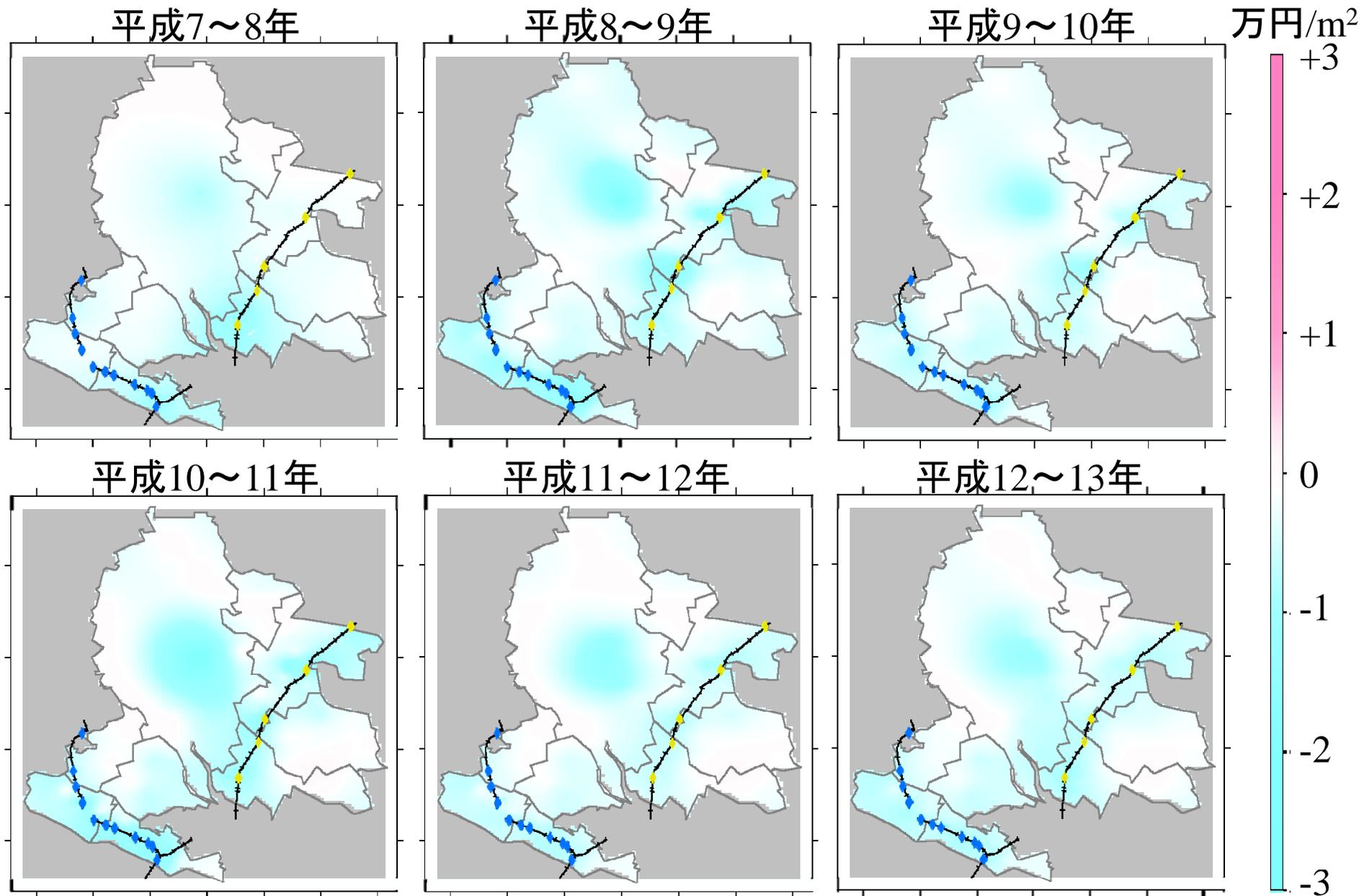
Krigingによる住宅地公示地価の視覚化



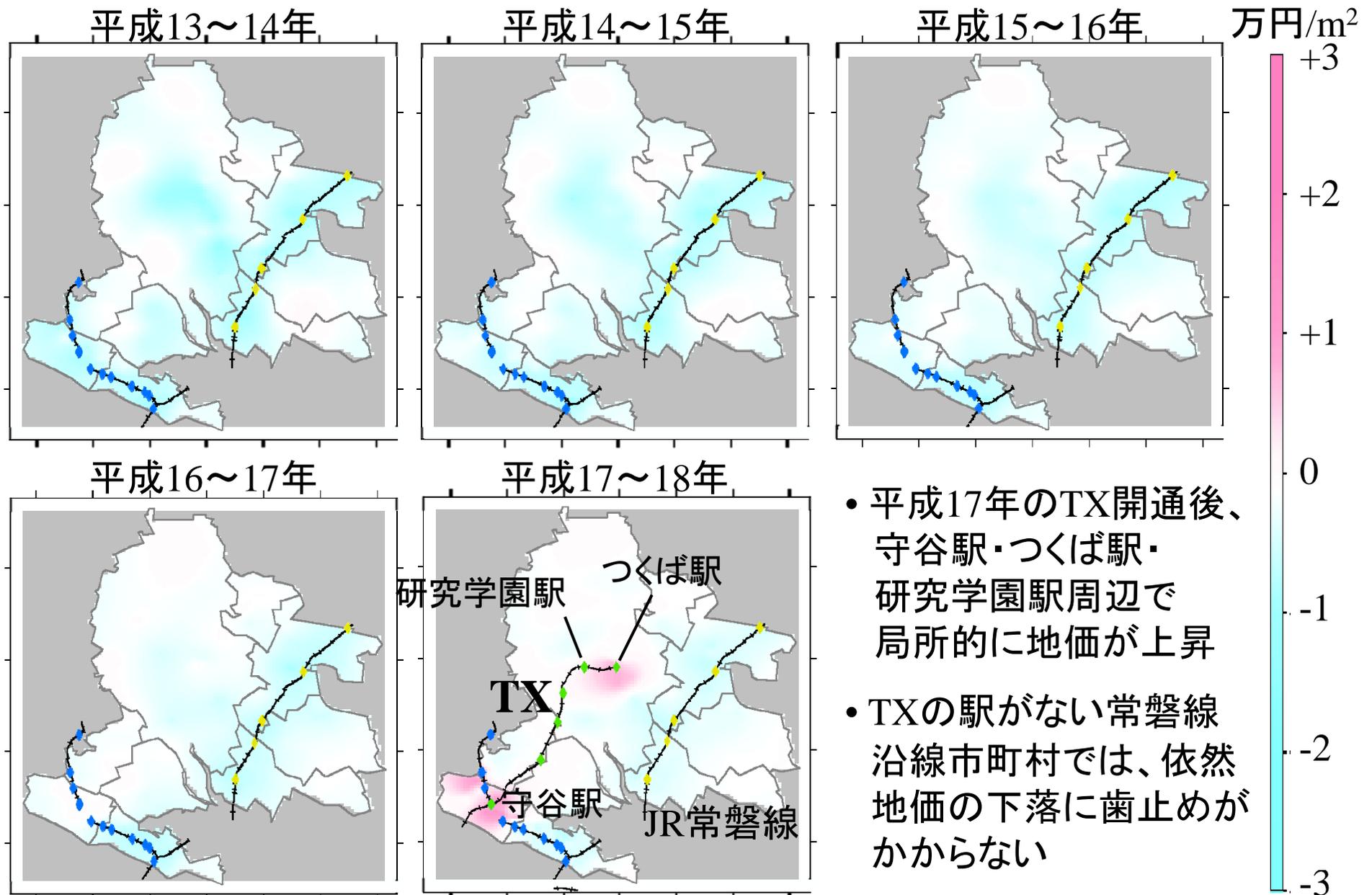
Krigingによる住宅地公示地価の視覚化



Krigingによる地価上昇率の視覚化



Krigingによる地価上昇率の視覚化



Universal Kriging

- トрендがある場合に、ある地点0のBLUP (最良線形不偏推定量)を得るための手法
- 期待二乗誤差最小化 $\min : E[\{y(s_0) - y(\hat{s}_0)\}^2]$



$$\hat{y}_0 = x'_0 \hat{\beta} + \hat{\gamma}' \hat{\Sigma}^{-1} (\mathbf{y} - X \hat{\beta})$$

$$\text{Var}(\hat{y}_0) = \hat{\tau}^2 + \hat{\sigma}^2 - \hat{\gamma} \hat{\Sigma}^{-1} \hat{\gamma}$$

→ Kriging分散 $\text{Var}(\hat{y}_0)$ は、 $\hat{\Sigma}^{-1}$ を既知としており、パラメータに関する不確実性を含まない

Bayesian Kriging

- パラメータを推定する代わりに、パラメータに事前分布をおき、データを用いてパラメータを更新する(事後分布)
- パラメータを確率変数として、その不確実性をKriging内挿に組み込む
- **Step**
 1. 事前分布の選定
 2. データを用いて事前分布を更新(事後分布)
 3. x_0, X, y が与えられた下での y_0 の条件付分布(予測分布)を導出し、これをもとに内挿
→ Bayesian Kriging

Bayesian Kriging

- **モデル**

$$\mathbf{y} | \boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \tau^2 I + \sigma^2 H(\boldsymbol{\phi}))$$

$$\text{where } \boldsymbol{\theta} = (\boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\phi})'$$

- **事前分布**

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\tau^2)p(\sigma^2)p(\boldsymbol{\phi})$$

Banerjee et al.(2004)を参考に、

$\boldsymbol{\beta}$ は正規分布、 $\tau^2, \sigma^2, \boldsymbol{\phi}$ は逆ガンマ分布に従うとする

- **周辺事後分布の例($\boldsymbol{\beta}$)**

$$p(\boldsymbol{\beta} | \mathbf{y}) = \iiint p(\boldsymbol{\beta}, \tau^2, \sigma^2, \boldsymbol{\phi} | \mathbf{y}) d\tau^2 d\sigma^2 d\boldsymbol{\phi}$$

- **予測分布**

$$\begin{aligned} p(y_0 | x_0, X, \mathbf{y}) &= \int p(y_0, \boldsymbol{\theta} | \mathbf{y}, X, x_0) d\boldsymbol{\theta} \\ &= \int p(y_0 | \boldsymbol{\theta}, \mathbf{y}, x_0) p(\boldsymbol{\theta} | \mathbf{y}, X) d\boldsymbol{\theta} \end{aligned}$$

MCMC (マルコフ連鎖モンテカルロ)

- 予測分布の推計値を得るためには、通常MCMC(マルコフ連鎖モンテカルロ)法が用いられる
- ベイズ統計学においては事後分布が多次元であることが多く、あるパラメータに関する周辺事後密度を数値積分により求めたり、それに基づく推論を行うことは難しい。このため、事後分布からの確率標本 $\theta^{(t)}$ (添え字 t は反復ステップ回数) をMCMC法によりサンプリングし、得られた確率標本を用いることにより、事後分布に関する要約や、事後分布に推論を行う(伊庭他(2005))
- 事後分布からのサンプル $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(T)}$ を用いて、bayes. Kriging による予測値は次のように求められる

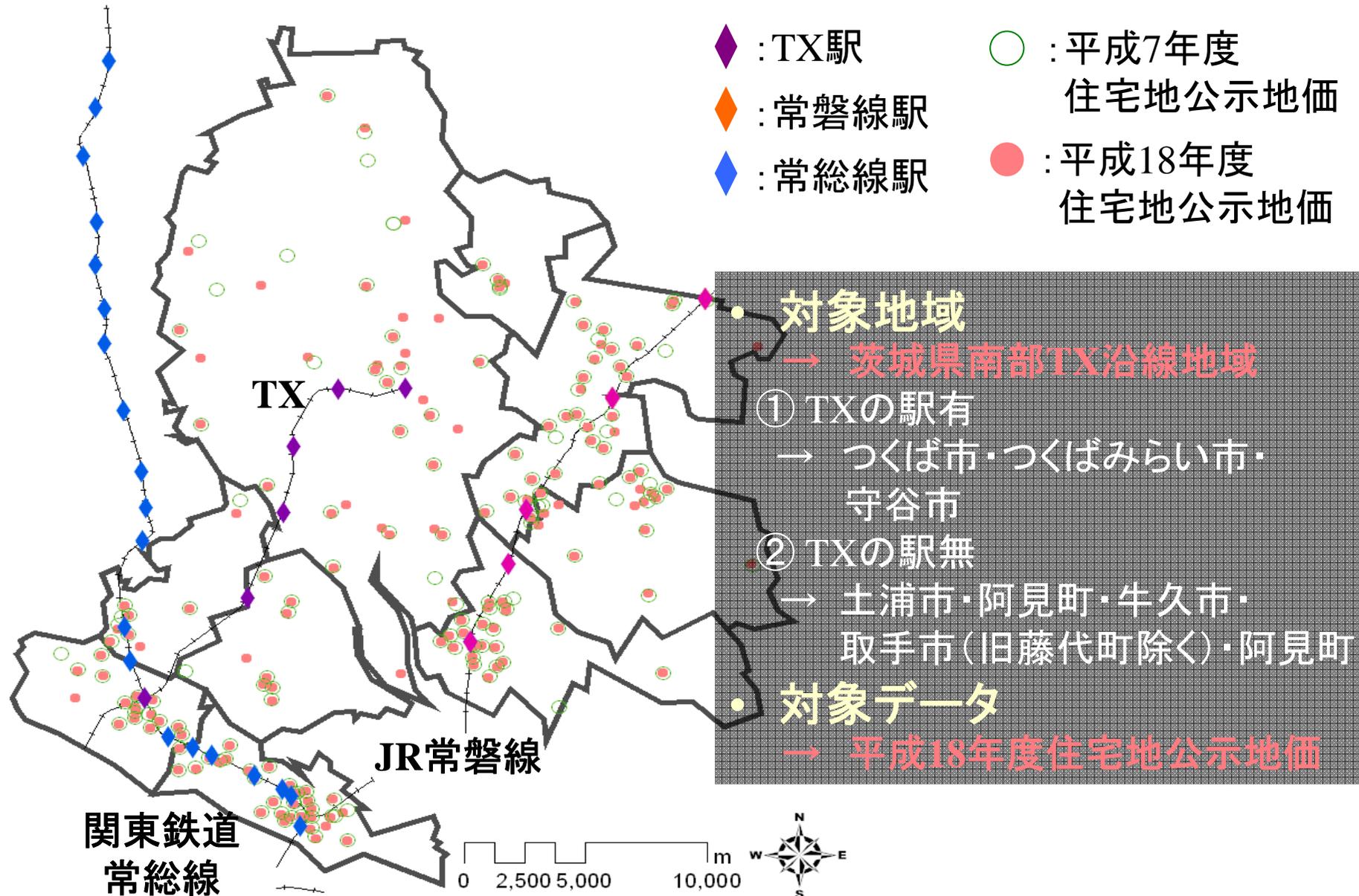
(※今回 $T = 11,000$ とし、初期値に依存する最初の1000回は捨てた)

$$\hat{p}(y_0 | x_0, X, \mathbf{y}) = \frac{1}{T} \sum_{t=1}^T p(y_0 | \mathbf{y}, \theta^{(t)}, x_0)$$

本研究の位置づけ

- 社会経済データに関するBayes. Krigingの実証研究は非常に少ない(例えば、Knight et al.(1998),Valente et al.(2005))
- 国内ではBayes. Krigingに関する研究は皆無
 - **TX沿線地域の住宅地公示地価を用いて、Univ.KrigingとBayes.Krigingの実証比較を行う**
- Univ.Krigingには、空間統計モデルの汎用ソフトウェアとして最も有名な**S+SpatialStat**を用いる
- Bayesian Krigingには、ベイズ分析のフリーソフトウェアである**WinBUGS**を用いる。WinBUGSは、S-Plusの文法に従っており、S-plusユーザーに使いやすい。WinBUGSでは、事後分布の密度関数の計算にMCMC(マルコフ連鎖モンテカルロ)を用いる。通常、WinBUGSのMCMCのoutputに対して、S-plusのライブラリ**coda**を用いて、マルコフ連鎖の収束判定を行う(Gewekeの値等)

対象地域・対象データ



モデルの前提

- 使用するデータ

$$y = X\beta + \varepsilon$$

y : \ln (平成18年度住宅地公示地価)
 x_1 : TX最寄り駅距離 (m), x_2 : TX以外の最寄り駅距離 (m),
 x_3 : 容積率 (%), x_4 : 地積 (m²), x_5 : 水道ダミー (有:1, 無:0),
 x_6 : ガスダミー (有:1, 無:0), x_7 : 下水道ダミー (有:1, 無:0)

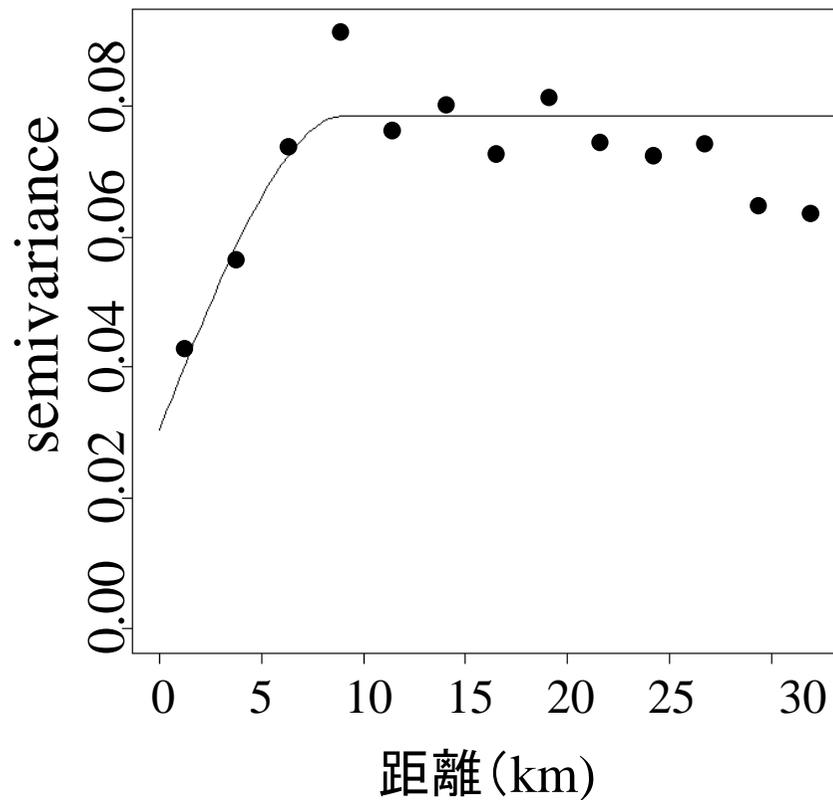
- 評価の方法 → **Validation(RMSEによる比較)**
 - 対象地域全地点数170点の50%を用いて
パラメータを推定し、残りの50%の値を内挿
 - Validationは5回行った

ある試行におけるパラメータ推定結果

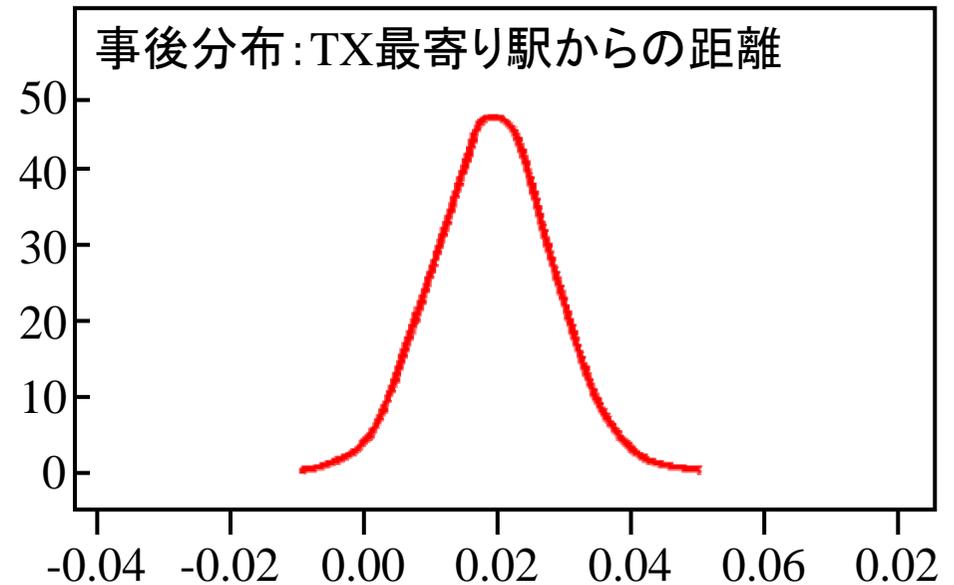
	OLS		Univ. Kriging		Bayes. Kriging		Geweke
	係数	標準誤差	係数	標準誤差	係数	標準誤差	
切片	10.9812	0.3067	11.0860	0.3043	11.0280	0.3759	0.85
TX最寄駅距離	2.51×10^{-2}	9.70×10^{-3}	1.90×10^{-2}	9.87×10^{-3}	1.98×10^{-2}	8.90×10^{-3}	0.33
TX以外最寄駅距離	-8.65×10^{-2}	1.27×10^{-2}	-7.46×10^{-2}	1.36×10^{-2}	-6.52×10^{-2}	1.70×10^{-2}	0.46
容積率	4.35×10^{-4}	6.11×10^{-4}	-2.97×10^{-4}	6.24×10^{-4}	-3.02×10^{-4}	5.84×10^{-4}	0.31
地積	-2.85×10^{-4}	1.39×10^{-4}	-3.16×10^{-4}	1.36×10^{-4}	-3.21×10^{-4}	1.22×10^{-4}	0.45
水道ダミー	-0.4208	0.2359	-0.4345	0.2329	-0.4641	0.2446	0.56
ガスダミー	0.2692	6.80×10^{-2}	0.1712	5.74×10^{-2}	0.1461	6.06×10^{-2}	0.88
下水道ダミー	0.1533	9.27×10^{-2}	0.1785	0.1033	0.1763	0.1092	0.08
残差分散	0.0812		-----		-----		
nugget	-----		0.0128		0.0138		0.20
Partial-sill(sill-nugget)	-----		0.0656		0.0871		0.18
phi	-----		0.4065		0.1597		0.24

※ Gewekeは、収束診断のための母集団の差の仮説検定の p 値であり、マルコフ連鎖が収束していることが分かる

アウトプットの一例



ある試行における
Univ. KrigingのVariogram



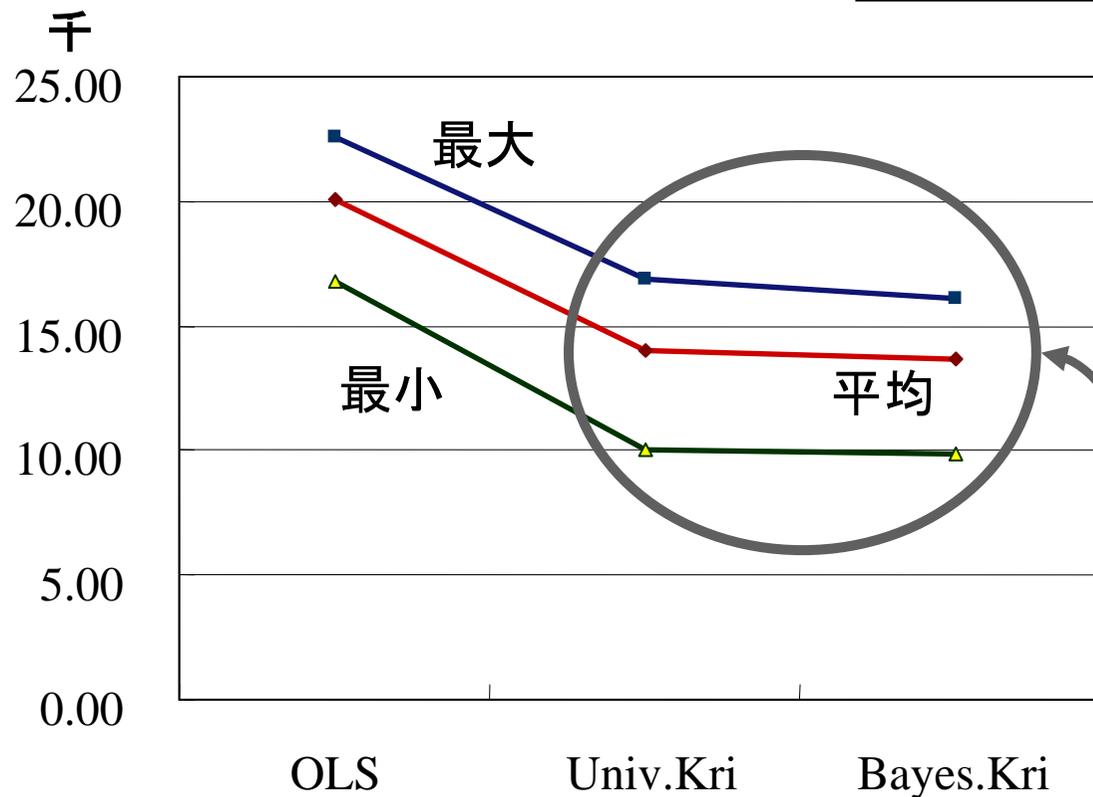
ある試行における
Bayes. Krigingの事後分布の例

Validationの結果

RMSE:

$$\sqrt{\frac{1}{85} \sum_{j=1}^{85} [(y_{0,obs}) - \exp(\hat{y}_0)]^2}$$

	ave.	max.	min.
OLS	20093	22591	16780
Univ.Kri	14000	16835	10011
Bayes.Kri	13669	16080	9846



Bayes.Kriのほうが
精度が若干高い
という結果

結論とまとめ

- S-plusとOrdinary Krigingを用いて、TX沿線地域における住宅地公示地価の視覚化を行った
- 空間統計モデルを用いて、TX沿線地域の地価モデルを作成した
- Univ. Kriging と Bayes.Krigingを実証比較し、Bayes. Krigingによって高い精度の内挿が可能であることを示した

参考文献

- Cressie, N.A.C. (1993), *Statistics for Spatial Data, revised edition.*, John Wiley & Sons
- Banerjee, S., Carlin, B.P., and Gelfand, A.E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC
- Knight, J. R., Sirmans, C. F., Gelfand, A. E., and Ghosh S. K. (1998), Analyzing Real Estate Data Problems Using the Gibbs Sampler, *Real Estate Economics* 26 (3), pp.469-492
- Valente, J. Wu, S.S. Gelfand, A., and Sirmans, C. F. (2005), Apartment Rent Prediction Using Spatial Modeling, *Journal of Real Estate Research*, 27, pp.105-136
- 伊庭幸人他 (2005), 計算統計Ⅱ, 岩波書店
- 井上亮・木越尚之・清水英範 (2005), 時空間クリギングの地価推定への適用可能性の検討, 地理情報システム学会講演論文集, Vol.14, pp.39-42