

NTT Data

ユーザーコンファレンス 学生研究奨励賞

# 地理的可変パラメータモデル のスケールに関する考察

泊 将史

神戸大学 大学院工学研究科  
市民工学専攻 博士前期過程(M2)

# 目次

- 分析の背景
- 分析の目的
- 実データを用いた実験
- 分析結果の考察と今後の課題

# 分析の背景

- 実証分析では、回帰モデルにおいて、回帰係数パラメータを**地理的に可変**としたい場合が少なくない。
  - 例) 対象領域が大きく、回帰係数が対象領域全体で「均一」と想定するのが難しい場合.

$$y_i = \beta_0 + \sum_{h=1}^{k-1} x_{h,i} \beta_h + \varepsilon_i \quad , \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$



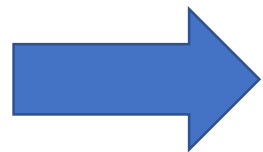
回帰係数を標本*i*ごとに与える

$$y_i = \beta_{0,i} + \sum_{h=1}^{k-1} x_{h,i} \beta_{h,i} + \varepsilon_i \quad , \varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

# 分析の背景(2)

- 地理的な可変パラメータモデルは、様々な分野で提案されてきている。

- 統計学: Gelfand et al. (2003); Franco-Villoria et al. (2018)
- 経済学: McMillen (1996)
- 機械学習: Bussas et al. (2017)
- 地理学: Brunson et al. (1996); Murakami et al. (2017)



中でも、最も実証研究で用いられているのが Brunson et al. (1996) の **GWRモデル** である

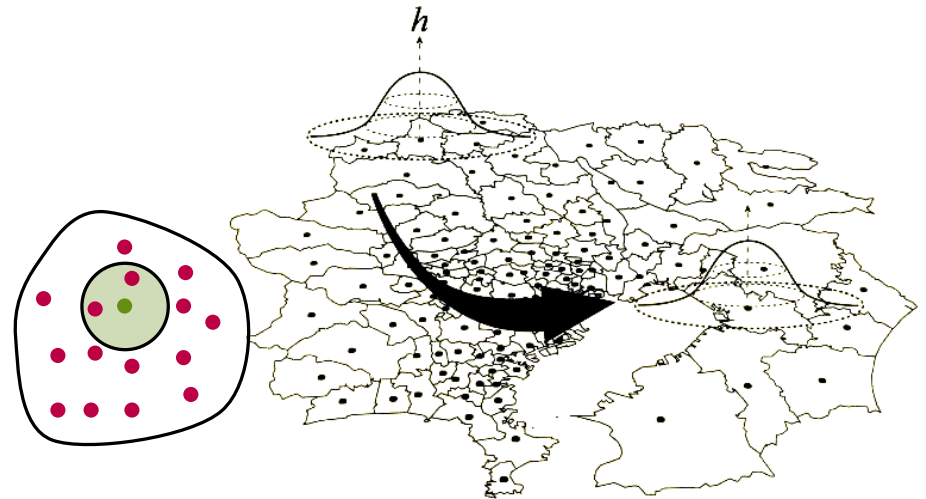
# GWRモデルとは?

- 方法: 近くの観測値に着目する地点*i*からの距離に応じた重みをつける.

$$W_i y = W_i X \beta_i + \varepsilon$$

$$W_i = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{in} \end{pmatrix}$$

where:  $w_{ij} = \exp\left(-\frac{d_{ij}^2}{\theta^2}\right)$   
 $\theta$ : カーネルバンド幅



杉浦 (2003) より

修正AICやCV(クロスバリデーション)を基にした  
繰り返し計算で求める



全ての観測点において回帰係数の  
推定値が得られる.

# GWRモデルにおけるパラメータ推定

- バンド幅が決まれば，重みつき最小二乗法(**WLS**)で，観測点*i*ごとの回帰係数推定値を計算できる

$$\hat{\beta}_i = (X'W_iX)^{-1}X'W_iy$$

- GWRモデルは1000件を超える引用があり<sup>1)</sup>，標準的な地理的可変パラメータモデルとして，様々な実証研究で用いられている。
- しかし，問題点も指摘されている。
  - 説明変数間の「ローカルな多重共線性」の存在  
例) 地点*i*の近傍だけを見ると，説明変数間の相関が強い。
  - GWRはバンド幅が**全説明変数で共通**のため，ローカルな多重共線性の存在により，回帰係数の推定値が全説明変数で非常に似通って，解釈が困難になる場合がある。

# MGWRモデル

- Fotheringham et al. (2017) は、バンド幅を説明変数ごとに変えるべきと指摘。  
⇒ **Multiscale GWR (以下, MGWR)を提案**
- 一般化加法モデル: Generalized additive model (GAM) におけるパラメータキャリブレーション法を応用して、バンド幅を説明変数ごとに推計.
- GWRの適用可能性を大きく広げる試みであるが、実データを用いた検証は皆無である。

# 分析の目的

- GWRモデルとMGWRモデルを実データ（日本のマンション分譲価格データ）を用いて比較し、バンド幅を説明変数毎に推計することによって、どのようなインプリケーションの差がもたらされるかについて考察する。



# 実データを用いた実験

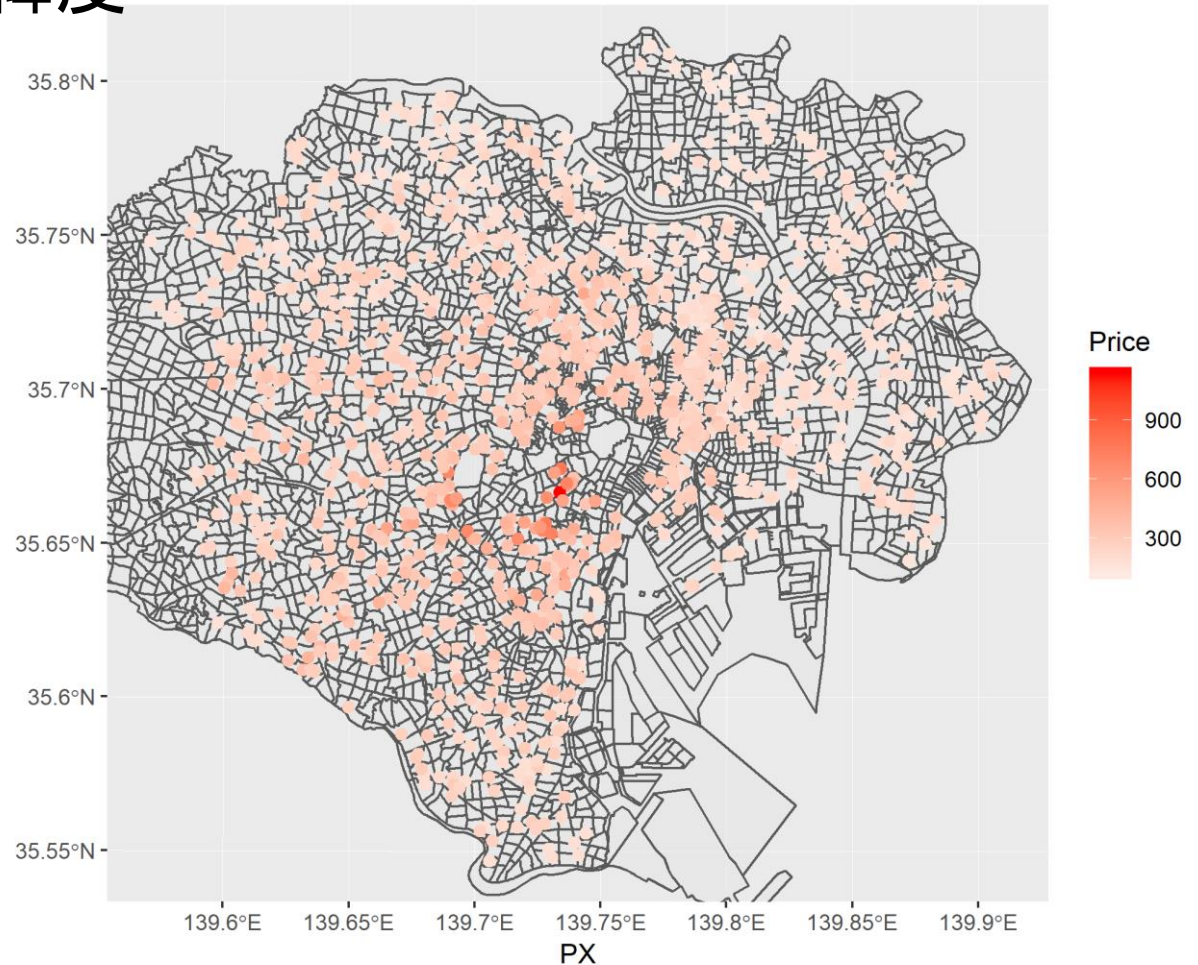
緯度

## 使用データ

- 東京23区の新築住宅分譲価格(登録価格, 2003~2009)
- Seya et al. (2013)で用いられたものの一部を利用
- サンプルサイズ1199

## 使用ソフトウェア

- Visual R Platform  
主要な使用パッケージ  
GWmodel; psych; sf; ggplot2



経度

# 分析の方法

- 東京23区の新築住宅分譲価格を被説明変数に、次のページにあげる変数を説明変数とした  
GWRモデル(バンド幅固定)  
MGWRモデル(バンド幅が説明変数毎に異なる)の  
回帰係数パラメータを推計
- まず、回帰係数パラメータ間の相関係数を計算し、MGWRによって改善が見られるかどうか調べる
  - カーネル関数としては,  $\text{exponential:wgt} = \exp\left(-\frac{vdist}{bw}\right)$
  - $vdist$ : 観測地点間距離,  $bw$ : バンド幅
- いくつかの変数の回帰係数の $t$ 値をRで可視化し、比較・考察する

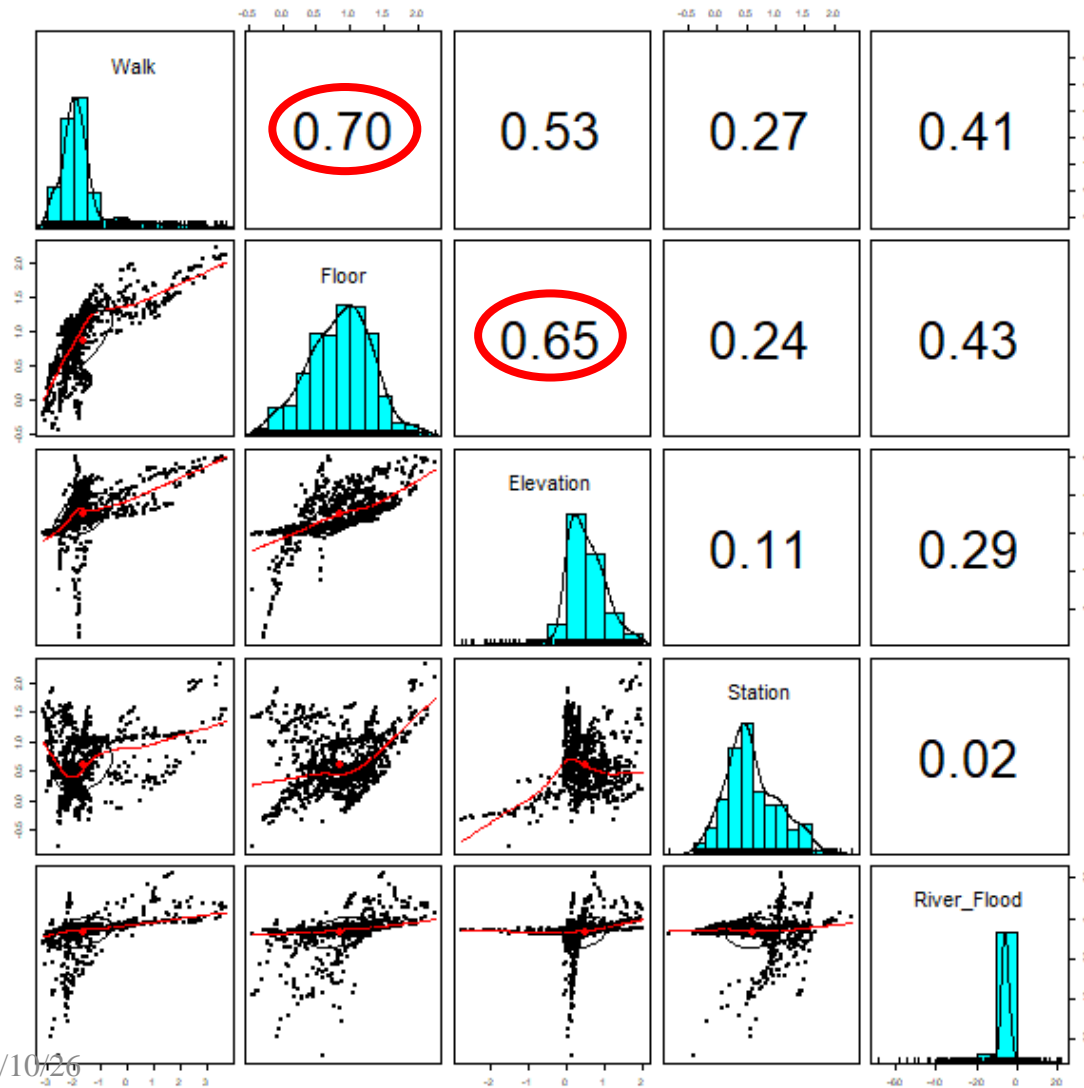
# モデルの被説明変数と説明変数

変数	単位	説明
Price	10,000 円/ 坪	Averaged housing price per tsubo (1 tsubo = 3.306m <sup>2</sup> )
Intercept		
Area	m <sup>2</sup>	敷地面積/1000
Walk	minute	最寄り駅への徒歩時間
Station	#	最寄り駅の年間乗客数(人/1000)2008年データ
Urban_Flood	m	都市水害リスク (0, 0.5, 1, 2, 3)
River_Flood	m	河川水害リスク (0, 0.5, 1, 2, 3)
Elevation	m	標高
Floor	#	階数
Num_Elevator	#	エレベーターの数
Developer	(Yes:1, No:0)	デベロッパーによる開発
Denentoshi	(Yes:1, No:0)	田園都市線が最寄であるかどうか
Toyoko	(Yes:1, No:0)	東横線が最寄であるかどうか
R1K1R	(Yes:1, No:0)	ワンルームとセパレートタイプのキッチン
R1DK	(Yes:1, No:0)	ワンルームとダイニングキッチン
R1LDK	(Yes:1, No:0)	ワンルーム, リビングとダイニングキッチン
R2LDK	(Yes:1, No:0)	ツールーム, リビングとダイニングキッチン

# モデルの被説明変数と説明変数(つづき)

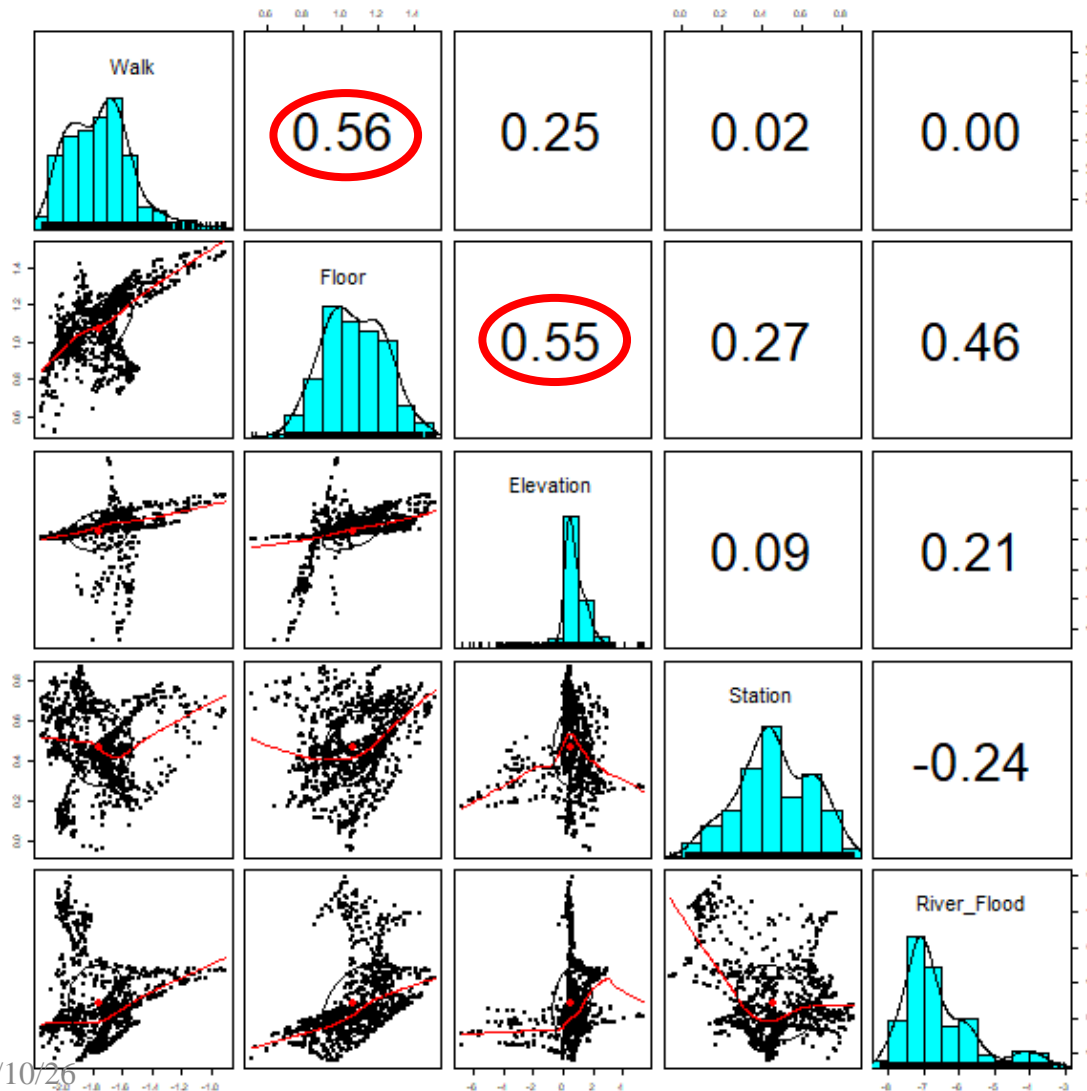
変数	単位	説明
R1	(Yes:1, No:0)	第一種住居地域
Com	(Yes:1, No:0)	商業地域
MR1	(Yes:1, No:0)	第一種中高層住居専用地域
RR1	(Yes:1, No:0)	第一種低層住居専用地域
R2	(Yes:1, No:0)	第二種住居地域
MR2	(Yes:1, No:0)	第二種中高層住居専用地域
RR2	(Yes:1, No:0)	第二種低層住居専用地域
Y103	(Yes:1, No:0)	2003年登録
Y104	(Yes:1, No:0)	2004年登録
Y105	(Yes:1, No:0)	2005年登録
Y106	(Yes:1, No:0)	2006年登録
Y107	(Yes:1, No:0)	2007年登録
Y108	(Yes:1, No:0)	2008年登録
Y109	(Yes:1, No:0)	2009年登録

# 推定値の相関プロット(GWR)



FloorとWalk  
FloorとElevation間で  
やや大きな相関がみられる

# 推定値の相関プロット(MGWR)

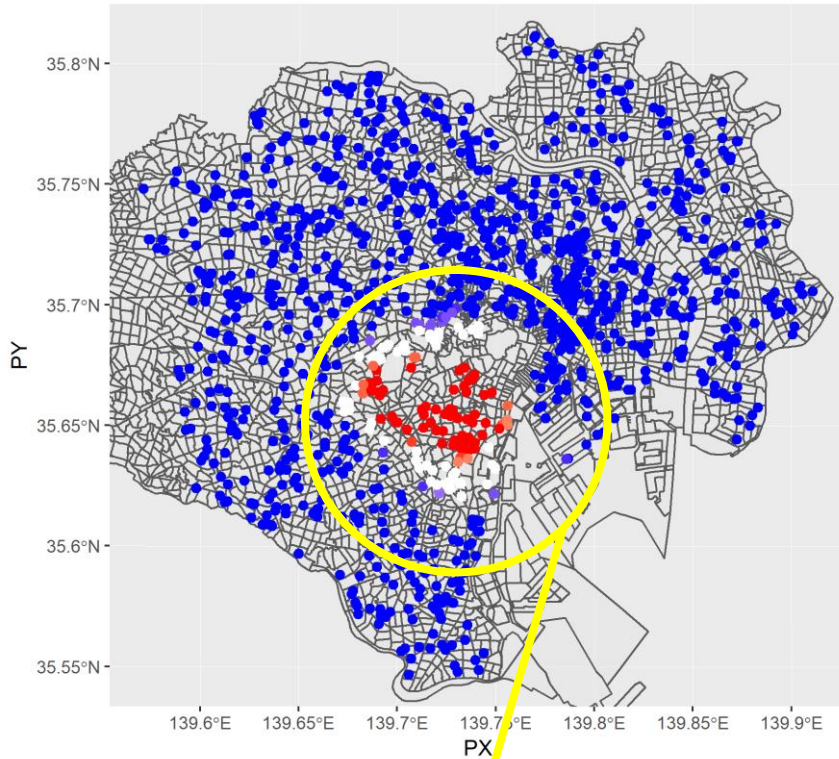


0.6を超えるようなやや  
大きな相関は改善  
されている



可変回帰係数の解釈が  
しやすくなっている可能性

# “Walk” の $t$ 値: GWR vs. MGWR

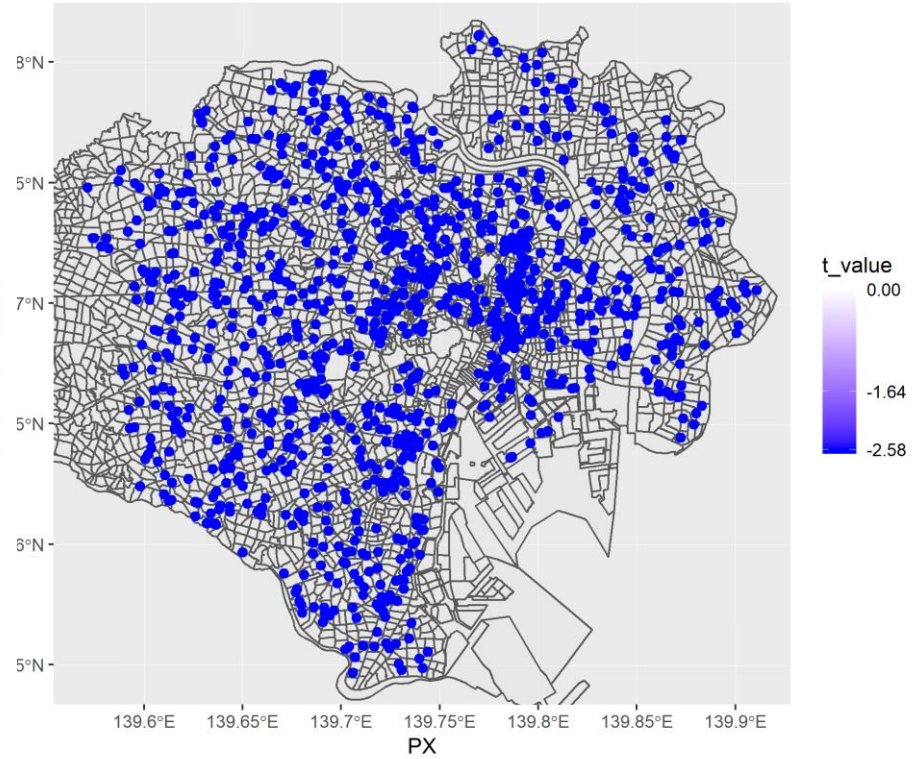


最寄り駅までの徒歩  
時間が長いほど住宅  
価格が高い

GWR

バンド幅: 1.15km

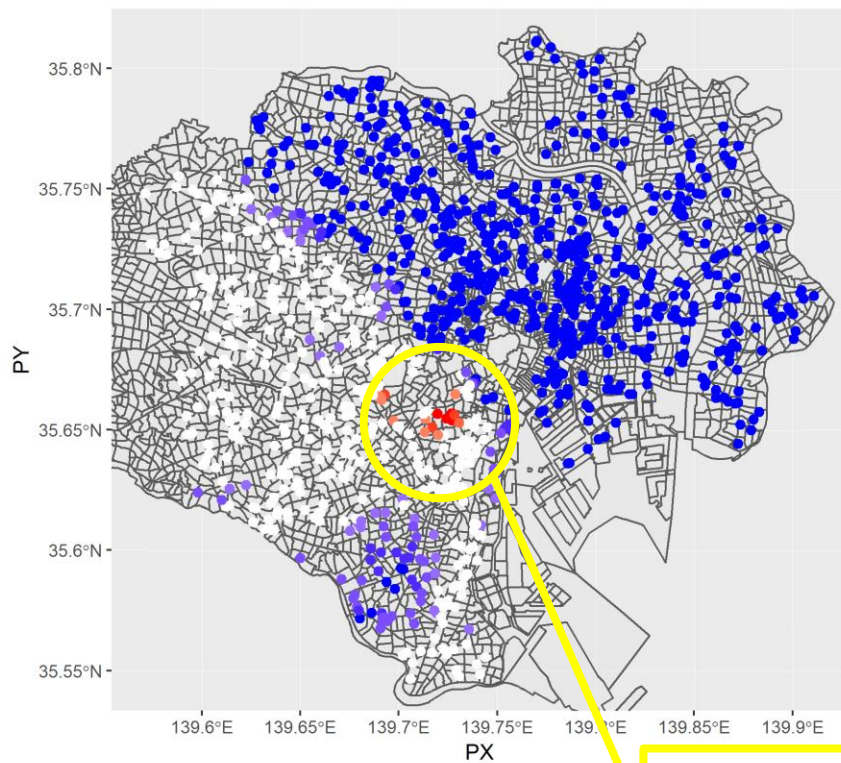
(すべての説明変数共通)



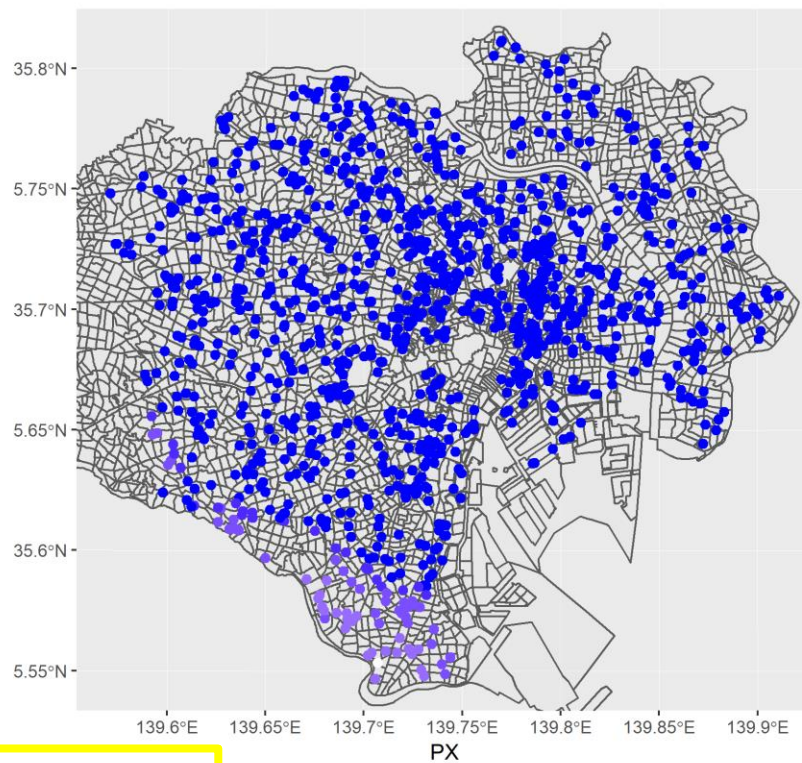
MGWR

バンド幅: 2.46km

# “River Flood”の $t$ 値: GWR vs. MGWR



t\_value  
2.58  
1.64  
0.00  
-1.64  
-2.58



t\_value  
0.00  
-1.64  
-2.58

河川水害リスクが高い  
ほど価格が高い

GWR  
バンド幅: 1.15km  
(すべての説明変数共通)

MGWR  
バンド幅: 2.50km



# 分析結果の考察と今後の課題

- $t$ 値の分布図において、GWRモデルでは、Walk(最寄り駅までの徒歩時間)、River\_Flood(河川水害リスク)において、直感と整合しない係数が得られているが、MGWRモデルでは改善している。
- 実証研究で用いられることが多いGWRモデルであるが、バンド幅が説明変数で共通というのは非常に強い仮定であることが実データ分析から示せた。

# 参考文献

- 杉浦芳夫 (2003) 地理空間分析 (シリーズ・人文地理学), 朝倉書店.
- Brunson, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281-298.
- Bussas, M., Sawade, C., Kühn, N., Scheffer, T., & Landwehr, N. (2017). Varying-coefficient models for geospatial transfer learning. *Machine Learning*, 106(9-10), 1419-1440.
- Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale geographically weighted regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247-1265.
- Franco-Villoria, M., Ventrucci, M., & Rue, H. (2018). Bayesian varying coefficient models using PC priors. *arXiv preprint arXiv:1806.02084*.
- Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462), 387-396.
- McMillen, D. P. (1996). One hundred fifty years of land values in Chicago: A nonparametric approach. *Journal of Urban Economics*, 40(1), 100-124.
- Murakami, D., Yoshida, T., Seya, H., Griffith, D. A., & Yamagata, Y. (2017). A Moran coefficient-based mixed effects approach to investigate spatially varying relationships. *Spatial Statistics*, 19, 68-89.
- Seya, H., Yamagata, Y., & Tsutsumi, M. (2013). Automatic selection of a spatial weight matrix in spatial econometrics: Application to a spatial hedonic approach. *Regional Science and Urban Economics*, 43(3), 429-444.