

# 転移学習による空間予測の高精度化：犯罪データへの応用

村上大輔\*・梶田真実\*\*

## An Improvement of Spatial Prediction via Transfer Learning: Application to Crime Data

Daisuke Murakami\*, Mami Kajita\*\*

Although spatial prediction is widely performed for urban and environmental monitoring, the accuracy is often insufficient if only small samples are available in the target area. The objective of this study is to overcome the limitation through a transfer learning using larger samples obtained outside the target area. Our proposal is to pre-train spatially dependent features, which are hard to transfer, and apply them as additional features in the subsequent transfer learning. The proposed method estimates the spatial/local features, which are specific to the target area, using a spatial regression while global features that are common across the target and other areas using a gradient boosting. The performance of the proposed method is examined through a crime prediction. The result suggests that the proposed method successfully improves predictive accuracy and stably compared to alternatives.

**Keywords:** 転移学習 (transfer learning), 空間回帰 (spatial regression), 犯罪予測 (crime prediction), 勾配ブースティング (gradient boosting)

### 1. はじめに

今日、多種多様な空間データが防犯、防災、COVID-19 対応などの実問題に役立てられている。空間データは観測点や標本数が限られている場合があり、必要に応じて空間統計モデルによる空間予測（補間）が行われてきた。例えば Cattle ら（2002）は限られたデータから土壌汚染状況を面的に予測し、堤ら（2016）は公示地価データを用いて地価を面的に予測した。

一般に、空間予測の精度は標本数が少なるにつれて悪化する。これに対し転移学習（Weiss et al., 2016 参照）は、対象地域外（以下、他地域と呼称）のより大きな標本を活用することで、対象地域における予測精度を改善しようという手法である。転移学習の有用性は、データが限られている場合に精度を高める手法として空間予測の文脈でも確認されてきた（例えば Zhang et al. 2019）。

一方、十分な標本を確保する以外に、空間相関（近所と類似した傾向を持つという空間データの一般的性質）を考慮することもまた空間予測の精度を高

めるうえで重要である。位置座標を特徴量として組み込めば、ニューラルネットワークやブースティング（GB: Gradient boosting）などの転移学習によく応用される汎用の学習器でも、データの空間パターンが学習できる。しかしながら、対象地域の標本数が少ない場合、空間相関を明示的に考慮しないそれらの手法で、データ空間パターンを捉えることは必ずしも容易ではない。例えば図 1 は、空間統計モデルと GBDT (GB decision tree; 広く用いられる学習手法 (特徴量: 緯度, 経度)) を用いて、空間相関を持つ疑似データによる空間予測を実施した結果である。標本数  $N$  は 100, 1,000, 4,000 とした。空間統計モデルは  $N=100$  の場合でも空間パターンが精度よく予測できている。対照的に、GBDT は視覚的に不自然な予測結果となっている。残念ながら、対象地域と他地域におけるデータの空間分布は互いにほぼ独立と考えられる。従って位置座標を特徴量として組み込んで転移学習したとしても精度改善はあまり見込めない。対象地域における空間相関パターンが精度良くとらえられるような、新たな転移学習手法が必要で

\* 正会員 株式会社 Singular Perturbations (Singular Perturbations Co. Ltd.)

〒102-0074 東京都千代田区九段南 1-5-6 りそな九段ビル 5F KS フロア E-mail : [dmuraka@ism.ac.jp](mailto:dmuraka@ism.ac.jp)

\*\* 正会員 株式会社 Singular Perturbations (Singular Perturbations Co. Ltd.)

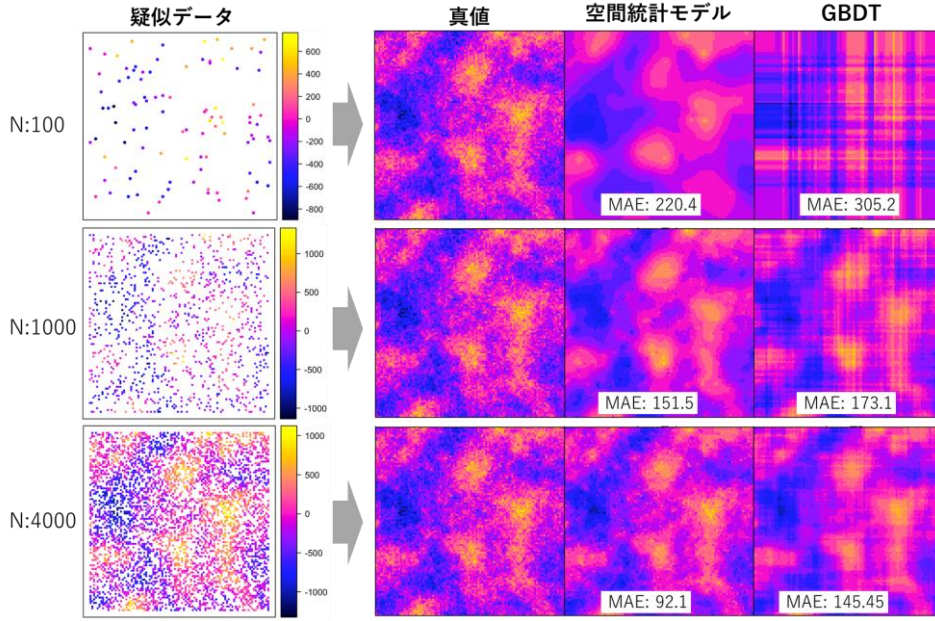


図1. 疑似データを用いた空間統計モデルとGBDTによる空間予測の結果比較 (Nは標本数)

ある.

そこで本研究では, 空間統計モデルで対象地域内の小標本から空間相関パターンを学習し, GB で他地域との共通傾向を学習するような, 新たな転移学習手法を開発し, その予測精度を犯罪データへの応用を通して検証する.

## 2. 提案手法

### 2.1. モデル

本研究では, 小標本しか得られない対象地域と, より大きな標本が得られる  $G$  個の他地域を想定し, 後者からの転移学習により, 対象地域における空間予測精度を改善することを目標とする. なお, 被説明変数と説明変数は地域毎に基準化 (平均 0, 分散 1) されているものとする.

対象地域の地点  $i$  における被説明変数  $y_i$  と,  $g$  番目の他地域内の地点  $I(g)$  における被説明変数  $Y_{I(g)}$  が下式に従うと仮定する:

$$y_i = f_{\theta}(\mathbf{x}_i, z_i) + u_i, \quad (1)$$

$$Y_{I(g)} = f_{\theta}(\mathbf{X}_{I(g)}, Z_{I(g)}) + U_{I(g)}, \quad (2)$$

$\mathbf{x}_i \in \{x_{i,1}, \dots, x_{i,K}\}$  と  $\mathbf{X}_{I(g)} \in \{X_{I(g),1}, \dots, X_{I(g),K}\}$  はそれぞれ対象地域と他地域  $g$  における説明変数 (既知) を表す.  $z_i$  と  $Z_{I(g)}$  は各地の空間相関パターンなどを

捉える特徴量 (未知),  $u_i$  と  $U_i$  はノイズ項である.

空間相関パターンは対象地域と他地域がオーバーラップしていない限りはほぼ独立に決まると考えられる. そこで  $z_i$  と  $Z_{I(g)}$  は以下のような空間統計モデルで事前に学習/推定することとする:

$$y_i = z_i + e_i, \quad e_i \sim N(0, \sigma^2 w_i^{-1}), \quad (3)$$

$$Y_{I(g)} = Z_{I(g)} + E_{I(g)}, \quad E_{I(g)} \sim N(0, \sigma^2 w_{I(g)}^{-1}), \quad (4)$$

$w_i$  と  $w_{I(g)}$  は各標本に対する重み (既知) である.  $z_i$  と  $Z_{I(g)}$  は本研究では以下のように定式化することとした:

$$z_i = \sum_{k=1}^K x_{i,k} b_k + \sum_{l=1}^L s_{i,l} \gamma_l, \quad \gamma_l \sim N(0, v(\tau^2, \alpha)), \quad (5)$$

$$Z_{I(g)} = \sum_{k=1}^K X_{I(g),k} B_{k(g)} + \sum_{l=1}^L S_{I(g),l} \gamma_{I(g)}, \quad (6)$$

$$\gamma_{(g)l} \sim N(0, v(\tau_g^2, \alpha_g))$$

$b_k$  と  $B_{k(1)}, \dots, B_{k(G)}$  は地域毎の回帰係数である.  $\sum_{l=1}^L s_{i,l} \gamma_l$  は対象地域の空間相関パターンを捉える項であり, 空間基底 (空間パターンを持つ既知の説明変数; 本研究では近接行列の固有ベクトルで与える)  $\{s_{i,1}, \dots, s_{i,L}\}$  を係数  $\{\gamma_1, \dots, \gamma_L\}$  で重みづけることで与える.  $\sum_{l=1}^L S_{I(g),l} \gamma_{I(g)}$  も同様に定義する. 各地の

空間相関の強さは分散パラメータ  $\tau^2, \tau_1^2, \dots, \tau_G^2$  で、空間スケールはパラメータ  $\alpha, \alpha_1, \dots, \alpha_G$  で地域毎に推定する。(5), (6)式の右辺 1, 2 項は、各地のトレンドと空間相関パターンを捉える項である。

## 2.2. 学習手順

次の手順で提案モデルを推定/学習する：

- (i) 空間統計モデル (3,4 式に 5,6 式を代入したものを) を地域毎に個別に推定することで特徴量  $\hat{z}_i, \hat{Z}_{I(1)}, \dots, \hat{Z}_{I(G)}$  を得る。
- (ii) (i)で得られた空間特徴量を(1), (2)式に代入したうえで、両式用いたメタ学習により各地共通の関数  $f_{\theta}(\cdot)$  を推定する。以降では GBDT を用いて学習を行う。

以上の学習のイメージを図 2 に示す。同図のように、通常の転移学習では  $y_i$  と  $Y_{I(g)}$  に共通に作用する特徴量  $\mathbf{x}_i, \mathbf{X}_{I(g)}$  (青領域) を用いて学習を行う場合が多い。例えば「人口が多い地域で犯罪が多い」という共通傾向があれば、同性質を考慮した転移学習ができる。しかし、各地域の空間相関パターンは通常は互いに独立なため (例：東京都の東部で犯罪が多いとしても、大阪府でも東部で犯罪が多いとは限らない) 転移困難である。そこで本研究では、手順(i)で各地の空間相関パターン (図中の緑領域) を事前

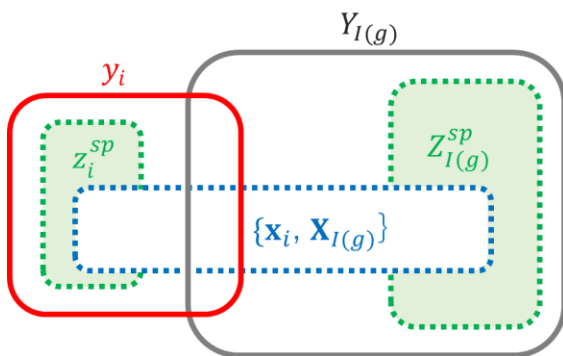


図 2. 問題設定のイメージ。  $z_i^{sp} = \sum_{l=1}^L s_{i,l} \gamma_l$ ,  $Z_{I(g)}^{sp} = \sum_{l=1}^L S_{I(g),l} \gamma_l$  である。赤領域は  $y_i$ , 黒領域は  $Y_{I(g)}$  の総変動を表し、それら内部は総変動の内訳を表す。また 2 変数の領域の重複が大きいことは、変数間の従属関係が強いことを意味する。緑領域は通常の学習では転移困難な各地の空間相関パターンを表す。

学習して特徴量  $\hat{z}_i, \hat{Z}_{I(1)}, \dots, \hat{Z}_{I(G)}$  に埋め込んだ上で、手順(ii)のメタ学習により、それらの影響を評価することとした。それにより、各地の空間相関パターン (緑領域) と共通傾向 (青領域) を考慮したモデルが学習でき、より高精度な転移予測が可能となると考えられる。

学習されたモデルによる空間予測は、予測地点における説明変数  $\mathbf{x}_0 \in \{x_{0,1}, \dots, x_{0,k}\}$  を準備したうえで空間統計モデルの予測値  $\hat{z}_0$  を評価し、それを学習済みの予測モデルに代入することで求められる。つまり  $\hat{y}_0 = f_{\theta}(\mathbf{x}_0, \hat{z}_0)$  である。

## 2.3. カウントデータへの応用

被説明変数がガウス分布に従うことを仮定したが (3-4 式)、提案手法は犯罪件数のようなカウントデータ測にも応用可能である。その手順は次のとおりである：

- (a) 被説明変数  $y_i$  を  $i$  番目のカウントデータ  $\tilde{y}_i$  を関数  $\log(\tilde{y}_i + 0.5) - \frac{1+0.5q}{\tilde{y}_i+0.5}$  で変換して基準化したもので与え、重みを  $w_i = \tilde{y}_i + 0.5$  で与える。  $q$  はカウントデータ内のゼロ値に割合である。
- (b) 手順(i)-(ii)による転移学習を実施する。

手順(a)は、(3)-(4)式を(7)-(8)式で置き換えることで得られる疑似ポアソン空間回帰の近似モデルとなることが知られている (Murakami and Matsui, 2021 参照)：

$$y_i \sim \text{quasiPoisson}(z_i, \sigma^2), \quad (7)$$

$$Y_{I(g)} \sim \text{quasiPoisson}(Z_{I(g)}, \sigma^2), \quad (8)$$

ただし  $\sigma^2$  は過分散を調整するパラメータである。

## 3. 犯罪データへの応用

### 3.1. 問題設定

提案手法を米国サンフランシスコ (SF) の犯罪件数 (地区別) の 1 ヶ月前予測に応用する。ただし SF では 1 ヶ月前の犯罪件数データしか得られておらず、シカゴ、デンバー、ボストンでは過去 12 ヶ月の犯罪

件数データが入手可能であると仮定する。対象地域である SF の標本数は 196 (196 地区×1), 他地域の標本数はシカゴ: 9,552 (796×12), デンバー: 1,728 (144×12), ボストン: 2,160 (180×12) である。3都市の犯罪データを考慮することで, SFにおける一か月先予測の精度が改善するかどうかを検証する。同検証は罪種 (Larceny, Burglary, Robbery) 毎にそれぞれ実施する。各罪種について, 予測対象月を 2018 年 1 月から 12 月まで変えながら, 予測を計 12 回行うことで, その精度を Root mean squared error (RMSE) で比較した。また, シカゴを対象とした同様の転移予測も実施して傾向を確認する。

予測に用いる手法は, 提案手法, 対象地域の小標本のみを用いる通常のポアソン回帰 (PM: Poisson regression model) とポアソン空間回帰 (SPPM: Spatial Poisson regression model), ならびに (ポアソン) リッジ回帰に基づく転移学習法である GLMtrans (Tien and Feng, 2022) と GBDT (1-2 式において  $z_i = Z_{I(1)} = \dots Z_{I(G)} = 0$  としたメタモデルを学習) である。各モデルの説明変数は人口, 所得 (中央値), 貧困率, 学校教育を受けていない方の割合, 大学卒業者の割合とした。

### 3.2. 結果

図 3 は, 月別予測の RMSE まとめたものである。

参考までに, ここでは対象地域の小標本のみを用いた場合の GBDT ( $GBDT_{loc}$ ) も比較対象とした。SPPM の RMSE は全ケースで PM よりも小さく, 空間相関の考慮による精度改善が確認された。また, GLMtrans と GBDT の RMSE は PM より小さく, 転移学習による精度改善も確認された。 $GBDT_{loc}$  の精度は GLMtrans と GBDT と同程度となった。以上 5 手法に比べ, 提案手法 (Proposed) の RMSE は, SF・シカゴの両都市で RMSE が最小となるケースが大半であり, 転移学習においても空間相関を考慮することが重要であるとの示唆を得た。

図 4~6 は SF における 2018 年 6 月の一月先予測の結果を地図化したものである。小標本のみを考慮する PM と SPPM は予測値の分散が真値 (True) よりも大きく不安定な傾向がある。一方, 他地域も考慮する GLMtrans と GBDT は, 対象地域における空間相関パターンを明示的に考慮しないため, 真値よりも分散が小さくなる傾向がある。それに対し, 提案手法は真値と最も近い予測分布になっている。

最後に, シカゴを対象としたシミュレーション実験を追加で行った。ここでは, シカゴで利用可能な標本 (過去一か月) は無作為抽出された 50/796 個のみと仮定し, 同標本を用いた一期先予測を 30 回繰り返すことで, 予測値の RMSE 誤差を評価した。また同様の操作を標本数が 400/796 個の場合について

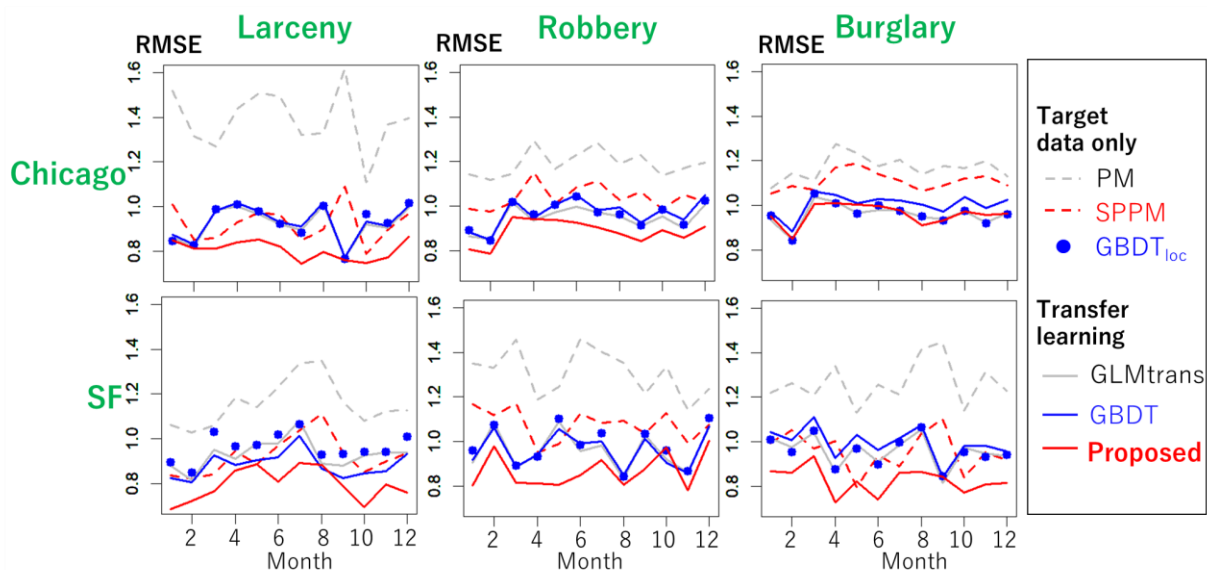


図 3. 都市別・罪種別・月別の一期先予測の RMSE 誤差の比較

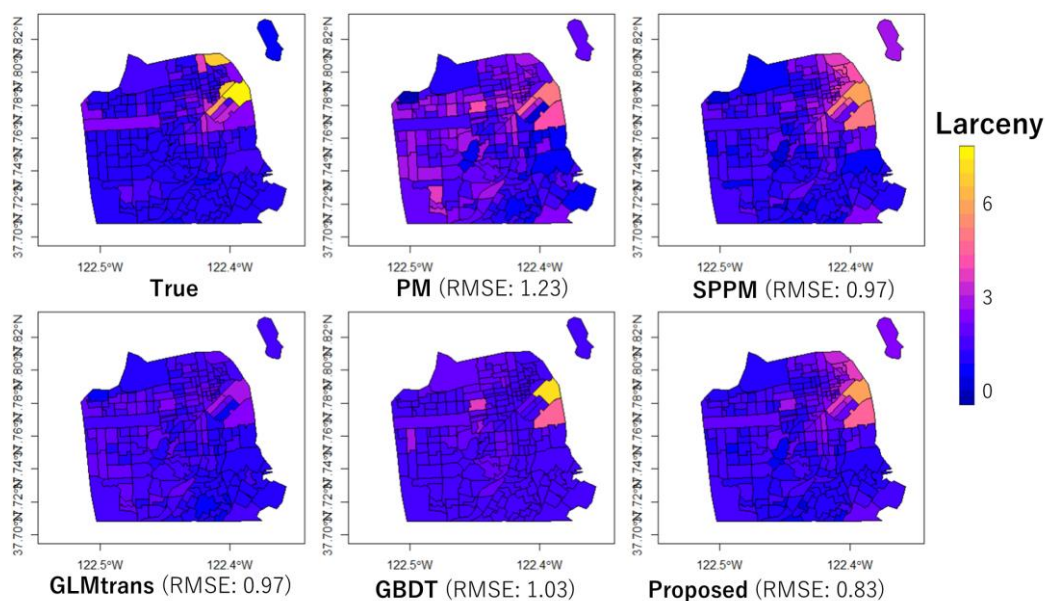


図 4. Larceny の予測結果

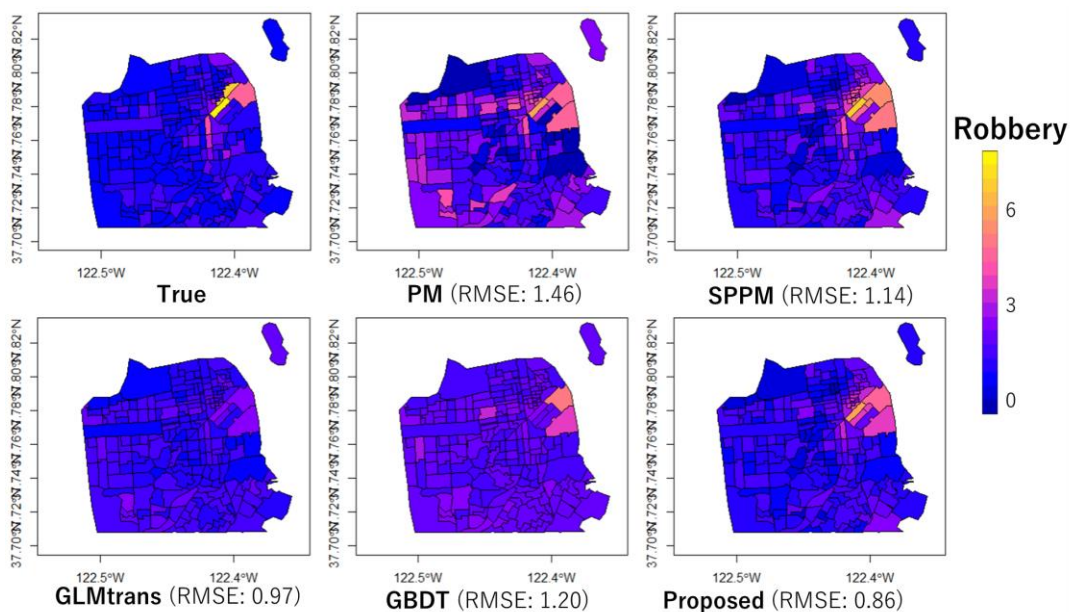


図 5. Robbery の予測結果

も同様に実施した。図 7 は RMSE の箱ひげ図である、この図から PM や SPPM は予測精度が低く不安定であることや、GLMtrans や GBDT がより安定的であることなどが確認された。また提案手法は PM や SPPM よりも安定しており、GLMtrans や GBDT よりも精度がよくことなどを確認した。一方で、GLMtrans や GBDT に比べると RMSE のばらつきが大きく、安定性の改善が今後の課題であることも明らかとなった。

#### 参考文献

- 堤盛人・村上大輔・嶋田章 (2014) 我が国の三大都市圏を対象とした住宅地価分布図の作成。「GIS-理論と応用」, 22 (2), 69-79.
- Cattle, J. A., McBratney, A. B., & Minasny, B. (2002). Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *Journal of Environmental Quality*, 31(5), 1576-1588.

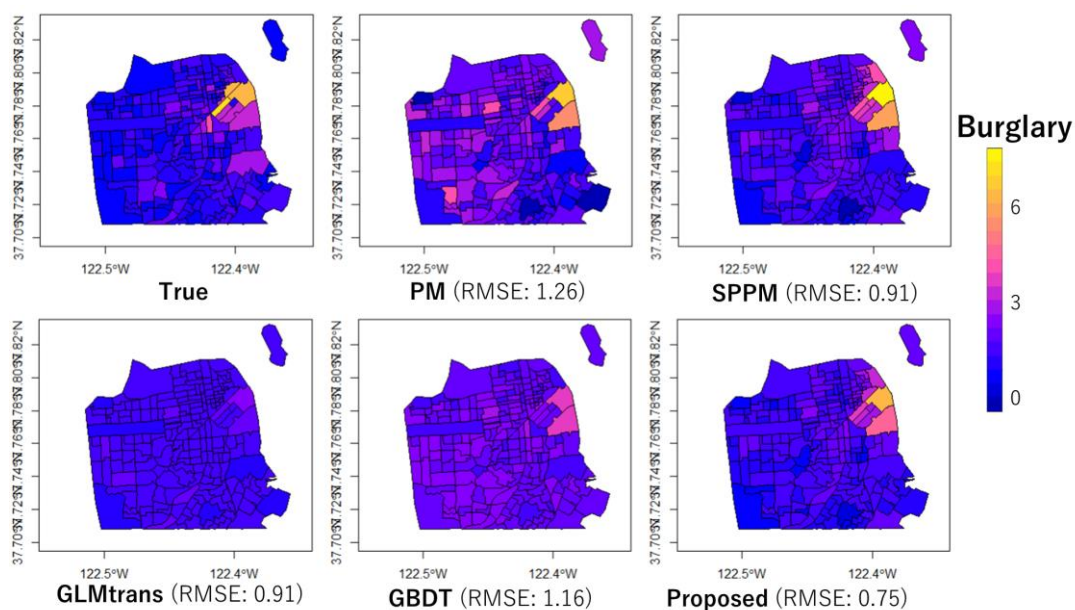


図 6. Burglary の予測結果

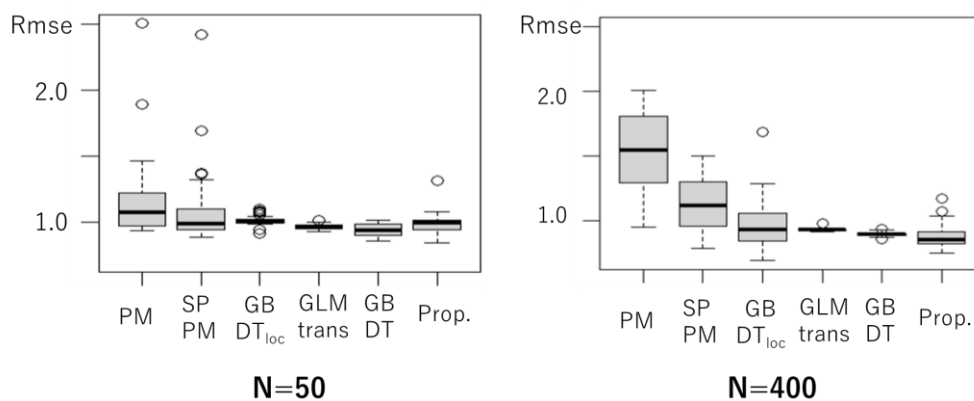


図 7. RMSE の箱ひげ図

Murakami, D., & Matsui, T. (2022). Improved log-Gaussian approximation for over-dispersed Poisson regression: Application to spatial analysis of COVID-19. *PloS one*, 17(1), e0260836.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1), 1-40.

Zhang, J., Li, W., Ogunbona, P., & Xu, D. (2019). Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective. *ACM Computing Surveys (CSUR)*, 52(1), 1-38.