

# 関数データ解析への空間統計モデルの援用

瀬谷創<sup>1</sup>・堤盛人<sup>2</sup>

<sup>1</sup>正会員 独立行政法人 国立環境研究所 地球環境研究センター（〒305-8506 茨城県つくば市小野川16-2）  
E-mail: seya.hajime@nies.go.jp

<sup>2</sup>正会員 筑波大学准教授 大学院システム情報工学研究科（〒305-8573 茨城県つくば市天王台1-1-1）  
E-mail: tsutsumi@sk.tsukuba.ac.jp

本研究では、「ある社会資本整備事業によって、実際に何処で地価が上昇したか」という問いに対して、視覚化という観点から答えるための一つの方法論を提示する。具体的には、関数データ解析（functional data analysis (FDA)）の手法に基づき、4次B-スプライン関数を基底関数として用いて公示地価の標準地における対象期間（サービス提供前後）の地価変動を時間軸上の連続関数として表現し、それをGiraldo et al. (2011) 等によって開発された関数データに対するオーディナリ・クリギング手法を用いて内挿することで、各町丁目中心座標における地価変動関数を得る。最後に、それをk-mean法でクラスタリングすることで、地価の変動特性を視覚化する。さらに、Giraldo et al. (2011) の手法に関数データに対する回帰分析手法（function-on-scalar regression）を組み合わせることで、土地属性（用途地域）による地価変動関数の差異を明示的にモデルに取り入れる方法論についても検討する。

**Key Words :** *functional data analysis, k-mean method, function-on-scalar regression, kriging, Tsukuba Express line*

## 1. はじめに

社会資本整備事業によって、どの程度の空間的な範囲で地価が上昇するかを予測することは難しい。帰着ベースで便益額を捉える資産価値法の代表的アプローチの一つであるヘドニック・アプローチにおいてはこの点が特に重要であり、通常は何らかのアクセシビリティ指標を定義し、その変化を用いて便益の帰着額と帰着範囲を求めることが多い（例えば、肥田野ほか, 1992; 肥田野1997; 白柳ほか, 2000; 堤・瀬谷, 2010）。しかしながら、この方法の是非に関する実証的検討は、筆者らレビューした範囲ではほとんど行われていない。本研究では、社会資本整備のサービス開始後に実際に地価が上昇した地域を把握するための実用的な方法について検討することを目的としている。これにより、キャピタリゼーション仮説におけるsmall条件（例えば、金本, 1992; 白柳ほか, 2000）の成立性についても、実証的な観点から示唆を得ることができると考えられる。

具体的には、まず関数データ解析（functional data analysis (FDA)）の手法に基づき、4次B-スプライン関数を基底関数として用いて公示地価の標準地における対象期間（サービス提供前後）の地価変動を時間軸上の連続関数として表現し、それをGiraldo et al. (2011) 等によって開発された関数データのオーディ

ナリ・クリギング手法を用いて内挿することで、各町丁目中心座標における地価変動関数を得る。最後に、それをk-mean法でクラスタリングすることで、地価の変動特性を視覚化する。さらに、本研究ではGiraldo et al. (2011) の手法に関数データに対する回帰分析手法（function-on-scalar regression）を組み合わせることで、土地特性（用途地域）による地価変動関数の差異を明示的にモデルに取り入れる方法論についても検討する。

本研究では、上記の手法について、つくばエクスプレス（TX）事業を例に具体的に示す。以下、第2章では、事業前後の地価変動を捉えるための手法をまとめる。次に、第3章では、まずTX事業に関する既往研究を概観する。TX事業は後述するいくつかの観点から特徴的な事業であることから、その関連研究を網羅的に整理することは重要であるといえる。しかしながらこのような試みは現時点でなされておらず、本研究の目的とは必ずしも一致しないものの、敢えて第3章で簡単な整理を試みたい。その後、本研究で用いるデータと、FDAの手法について説明する。続いて第4章で、実証分析を行う。最後に第5章で結論と今後の課題を述べる。

## 2. 事業前後の地価変動を捉えるための手法

資産価値法（例えば、肥田野ほか、1986; 堤, 1997）の観点からすれば、事業便益は、small-open 仮説の下、キャピタリゼーション仮説に基づき地価に帰着するため、社会資本整備事業の影響を地価から見るアプローチが一つの有効な方法と成り得る。事業前後の地価動向（変動）を効果的に把握するためには、地価の空間分布を視覚化し、その時系列的な動向を観察するアプローチが有効である（Tsutsumi and Seya, 2008; 井上ほか, 2009）。しかしながらこの手法で地価変動を把握するには、毎期の地価分布図をアニメーションとして連続的に表示する必要があり（例えば、時空間krigingによる地価変動分析およびその視覚化、井上亮氏、<http://www.plan.civil.tohoku.ac.jp/inoue>）、直感的で分かりやすい反面、作成に手間がかかり、また表示に何らかの電子媒体を要するという側面もある。したがって、地価の空間分布の時系列的な変化を、「一枚の図」で表現するというアプローチが、直感的な分かりやすさと作成コストの観点から有用であるといえる。本研究は、FDAの手法を用いて、この目的を達しようとするものである。

### 3. 研究の方法とFDAの導入

#### 3.1. TX事業の概要と既往研究レビュー

TXは、「東京圏北東部地域の交通体系の整備」、  
「JR常磐線等既設鉄道の混雑緩和」、  
「首都圏における住宅供給の促進」、  
「沿線地域における産業基盤の整備と業務核都市の形成」を目標に、東京都秋葉原と茨城県つくば市の間58.3kmを最速45分で結ぶ高速都市鉄道として2005年8月24日に運転を開始した。TX沿線地域の多くはそれまで長い間鉄道空白域であったため、その開業に伴う影響は広域かつ大規模なものとして期待されている。乗車人員は、開業以来順調に増加し、09年には採算ベースといわれる27万人に達している（松本, 2010）。TX事業の沿革について首都圏都市鉄道株式会社のホームページおよびTX協議会のホームページを参考に表1に概略を記す。TX事業は、都心への通勤鉄道としては首都圏最後の大型鉄道プロジェクトであり（鉄道のための総事業費が約8,300億円）、また、大都市地域における宅地開発及び鉄道整備の一体的推進に関する特別措置法（通称、宅鉄法）に基づき、大規模鉄道敷設と、宅地開発を一体的に行うといった特徴的な事業であることから（鈴木・中井, 2002; 高津ほか, 2006）、その影響について幅広い観点から考察することが重要であるといえる。

TX事業に関しては、現在までに様々な研究が行われている。まず、都市計画の観点からは、TX沿線の一体型土地区画整理事業について整理を行っている高津ほか（2006）、現在までの沿線開発の動向についてまとめている松本（2010）、沿線開発とまちづ

表1 TX事業の沿革

年	内容
85年	運輸政策審議会が「東京圏における高速鉄道を中心とする交通網の整備に関する基本計画について」を答申。常磐新線整備は都市交通対策上喫緊の課題と位置づけられた。
90年	「大都市地域における宅地開発及び鉄道整備の一体的推進に関する特別措置法」成立。
91年	首都圏新都市鉄道株式会社設立。
93年	秋葉原～浅草間工事施工認可。
94年	守谷～みらい平で工事施工認可。
同年	秋葉原において新線の起工式。
98年	流山おおたかの森～守谷で工事施工認可。
99年	みらい平～つくばで工事施工認可。
同年	全線都市計画決定。
00年	全線で工事施工認可。
01年	常磐新線の名称が「つくばエクスプレス」に決定。
同年	全線で工事着手。
03年	つくばエクスプレス全線20駅の駅名を決定。
同年	茨城県内の土木工事が完了。
04年	全線で土木工事が完了。
05年	ダイヤ発表。
同年	つくばエクスプレス開業。

くりについて考察している大村（2007）等がある。また、五十嵐（2007）は、TX開業に至るまでの政治過程を詳細に分析している。

TX事業に関する既往研究のほとんどは、TX事業の影響に関するものである。TX開業が住民の交通行動や意識に与えた影響に関する研究としては、大規模アンケートの集計・分析を行っている岡本ほか（2008）、Kawada et al.（2010）、パーソンプロブ調査による眞浦ほか（2005）がある。谷口ほか（2009）は、TX開業と同時に筑波大学で運行が開始された新学内交通システムが、必ずしも有効に活用されていないことを念頭に、バス停近くの居住地選択を促すことが可能か否かの実験を行っている。谷口ほか（2009）の事例は、TX事業の波及効果（金銭的外部効果）の一つといえるであろう。社会資本整備が土地利用に大きな影響を与えることは広く認識されており（上田ほか, 2009）、TX開業前後の土地利用変化に関しても、多くの研究が蓄積されている（王尾ほか, 2005; 小川, 2007; 淡野ほか, 2009; 小荒井・中埜,

2010) . また, 藤谷 (2006), 日下ほか (2009) は, 土地利用変化がヒートアイランドに与える影響について考察している. 以上, TX事業に関する既往研究を簡単にまとめた. しかしながら言うまでもなく, ここで取り上げた論文は, 膨大なTX事業に関する研究のほんの一部であろうことを断っておきたい. 今後, 事後評価の観点 (例えば, 金子・伊東 (2007)) から, 体系的にTX事業の評価を行う取り組みが必要となろう.

筆者らは現在まで, 地価に注目しながら, TX事業の影響把握に関する研究を行ってきた (瀬谷・堤, 2007; Tsutsumi and Seya, 2008; 堤・瀬谷, 2010) . 瀬谷・堤 (2007) では, 地価のヘドニック・アプローチに地理的加重回帰モデルを適用し, パラメータ推定値の時系列的な変化について考察した. Tsutsumi and Seya (2008) では, 開業以前から, TXの駅周辺で地価が上昇する, いわゆる「アナウンスメント効果」の存在を確認している. 堤・瀬谷 (2010) では, 空間・ヘドニックアプローチを用いて, TX事業の便益評価を試みている. 他にも, 資産価値法に基づくTX事業の便益計測を行っている例には, ヘドニック・アプローチを用いたPior et al. (1998), Pior and Shimizu (2001), 土地利用・交通モデルを用いたAhsan et al. (2001) がある. しかしながらこれらの研究は, 開業前から開業直後の影響に着目した研究であり, 開業後の影響についてはほとんど考察がなされていない. 開業から5年がたち, 順調に発展している地域と, 予定通りの開発が進んでいない地域の差が生まれてきており, これらの差は, 住宅・オフィスの需供調整により, 最終的には地価に反映され得るため, 地価の時系列的な推移・変動パターンの地域差に関する時空間分析が重要であるといえる.

### 3. 2. 研究の方法

以下, 本節では, 研究の方法について述べる. 図1は, TX沿線の, 茨城県と千葉県にかかる部分における住宅地公示地価の標準地のうち, 開業3年前にあたる2002年から2010年まで連続で観測されている地点を示したものである. これらの標準地における地価変動を示したのが, 図2 (茨城県), 図3 (千葉県) である. ただし, 比較のため, 2002年の地価 (円/m<sup>2</sup>) に対する比率として表示し, 以下の分析においてもこのデータを用いることとする. 無論, 地価の相対的な地域差と, その時系列的な推移 (地価変動) について考察するためのデータの変換方法はこれに留まるものではなく, この与え方に対する議論も重要である (Tsutsumi and Seya, 2008) . なお, 物価は対象期間において安定的であるため, 本研究では物価変動の影響については考慮していない.

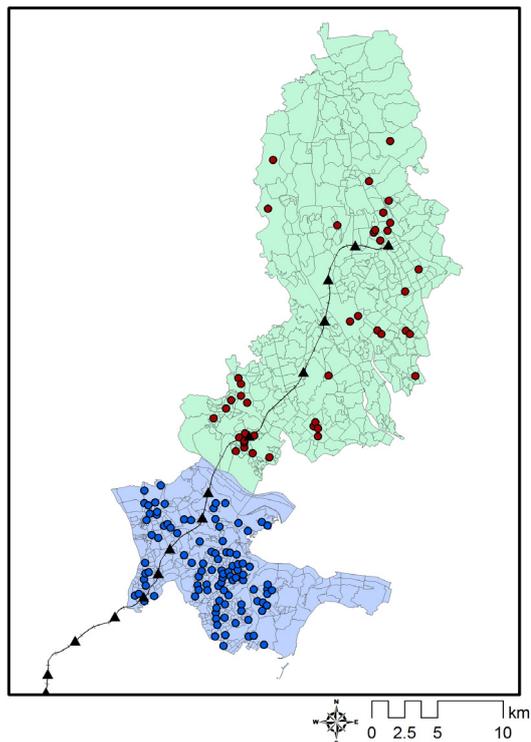


図1 2002-2010年まで観測されている公示地点

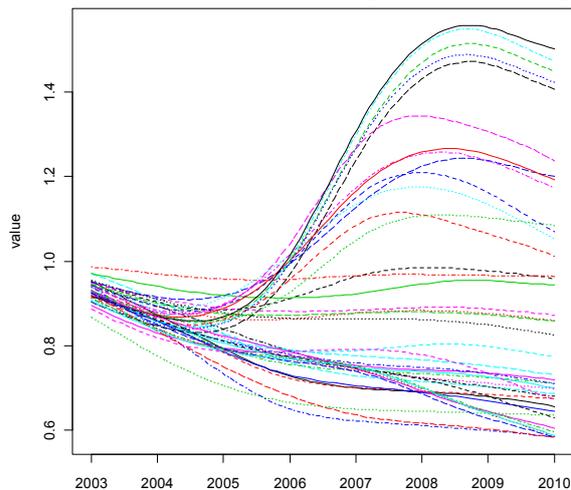


図2 2003-2010年までの地価変動 (茨城)

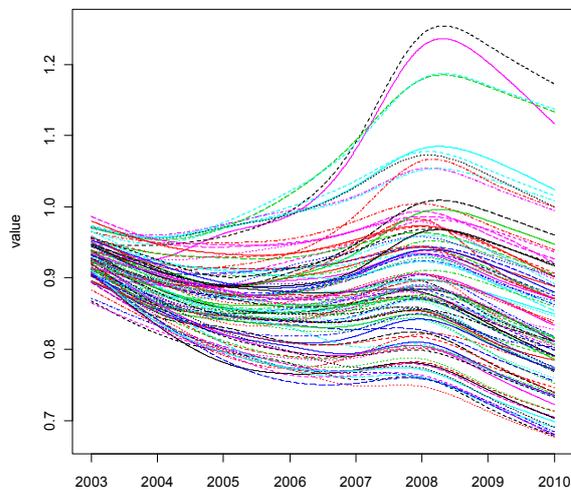


図3 2003-2010年までの地価変動 (千葉)

図2より、茨城県内においては、特に開業直後の2005年から2008年にかけて地価が上昇する地点が確認される一方で、多くの地点では一貫して減少していることが分かる。他方千葉県では、ほとんどの地域で2005年から2008年にかけて地価が上昇し、その後2010年にかけては下落している。

図1と、図2、3を結びつけることで、標準地における地価変動を捉えることができる。しかしながらキャピタリゼーションの空間範囲の把握という観点からは、標準地だけでなく、non-sampleの地点（地域）における地価変動にも興味を持たれよう。そこで本研究では、図2、3のデータの変動を関数として捉え、FDAの方法を用いて関数をクリギング内挿する方法、具体的にはDelicado et al. (2010), Giraldo et al. (2011) の手法を用いて、各町丁目における地価変動を推計する。続いて、それをk-mean法を用いてクラスタリングすることで、地価の変動構造の視覚化/マッピングを試みる。

以下、3.3において、FDAについて概説した後、3.4でDelicado et al. (2010), Giraldo et al. (2011) によって考案された関数データに対するクリギングについて説明する。

### 3.3. 関数データ解析(FDA)

FDAは、従来の多次元データ解析法を拡張し、関数を明示的にデータとして扱う手法として1990年代にRamseyらに提案された手法である（水田, 2005）。荒木・小西(2004)の定義を借りてより具体的に述べれば、各対象に対して離散点で経時的に観測・測定された一組のデータを滑らかな関数として捉え、この関数の集合からなるデータ、すなわち関数データから有効に情報を抽出する手法である。経時的データを関数化する際には、フーリエ級数やB-スプライン関数を基底関数として用い、関数の滑らかさの調整はクロス・バリデーション(cross-validation(CV))によって調整することが多い。現在までに、回帰分析、主成分分析等、標準的な統計学の様々な手法がFDAに拡張されており、書籍もいくつか出版されている(Ramsay and Silverman, 2002, 2005; Ferraty and Vieu, 2006; Ramsay et al., 2009; Ferraty and Romain, 2011)。

### 3.4. 関数データに関するクリギング

Delicado et al. (2010), Giraldo et al. (2011) は、関数データに対するオーディナリ・クリギング手法（以下、関数OK）を提案している。関数OKでは、距離の近い関数は似たような形状を示すという考え方にに基づき、関数データ間の空間的自己相関関係をバリオグラムを用いて構造化する。以下では、Delicado et al. (2010), Giraldo et al. (2011) を参照し

ながら、関数OKを導入する。

まず、関数ランダム過程（functional random process） $\{\chi(s) : s \in D \subseteq \mathfrak{R}^d\}$  における実現値が、 $T = [a, b] \subseteq \mathfrak{R}$  上で定義され、次に属する関数であると仮定しよう。

$$L_2(T) = \left\{ f : T \rightarrow \mathfrak{R}, \text{ s.t. } \int_T f(t)^2 dt < \infty \right\}, \quad (1)$$

ここで、 $L_2(T)$  は、内積が  $\langle f, g \rangle = \int_T f(t)g(t)dt$  で定義されるヒルベルト空間である（例えば、篠崎ほか, 1994）。また、ある固定された  $t_0 \in T$  に対して、 $\{\chi_{t_0}(s) : s \in D \subseteq \mathfrak{R}^d\}$  は  $\mathfrak{R}$  において値を持つ、スカラー値ランダム過程（scalar-valued random process）とする。我々の最終的な目的は、標準地で観測された地価変動  $\chi(s_i), i=1, \dots, I$  を用いて、任意の地点における関数、 $\chi(s_0) : [2003, 2010] \rightarrow \mathfrak{R}$  を予測することである。ここで、関数ランダム過程が次の意味で弱定常性を満たすと仮定する。

$$\begin{aligned} E(\chi_{t_0}(s)) &= m_{t_0}, V(\chi_{t_0}(s)) = \sigma_{t_0}^2, \forall t_0 \in T, s \in D. \\ \text{Cov}(\chi_{\tilde{t}_0}(s+h), \chi_{t_0}(s)) &= C_{\tilde{t}_0, t_0}(h), \forall t_0, \tilde{t}_0 \in T, h. \\ (C_{\tilde{t}_0, t_0}(h) : T \times T \rightarrow \mathfrak{R} &\text{は二乗可積分とする}) \end{aligned}$$

これらの仮定により、次式に示すバリオグラムが定義される。

$$\gamma_{\tilde{t}_0, t_0}(h) = \frac{1}{2} \text{Var}(\chi_{\tilde{t}_0}(s+h) - \chi_{t_0}(s)). \quad (2)$$

$\chi(s_0)$  の最良線形不偏予測量（best linear unbiased predictor (BLUP)）は、

$$\hat{\chi}(s_0) = \sum_{i=1}^I \lambda_i \chi(s_i), \lambda_1, \dots, \lambda_I \in \mathfrak{R}, \quad (3)$$

で与えられ、 $\lambda_i$  は、 $E(\hat{\chi}(s_0) - \chi(s_0)) = 0$  を満たしながら、次式を最小化するように決定される。

$$E \left[ \int_T (\hat{\chi}(s_0) - \chi_t(s_0))^2 dt \right]. \quad (4)$$

$E(\hat{\chi}(s_0) - \chi(s_0)) = 0$  の制約下での式(4)の最適化問題は次式のように定式化できる。

$$\min_{\lambda_1, \dots, \lambda_I} \int_T \text{Var}(\hat{\chi}_t(s_0) - \chi_t(s_0)) dt, \text{ s.t. } \sum_{i=1}^I \lambda_i = 1. \quad (5)$$

$\sum_{i=1}^I \lambda_i = 1$  は、不偏性を満たすための制約である。

ここで、式(3)のように、重みが時間  $t$  に依存しないとする利点は、トレースバリオグラム (trace-variogram)  $\gamma(h) = \int_T \gamma_t(h) dt$  のみを求めればよいため、計算負荷が少ないという点にある。実際には、重みを  $\lambda_{i,t}$  と、 $t$  に依存する形で与えることも可能であるが、この方法では重みの時間軸上での連続関

数を別途考える必要があり、計算負荷が大きい。また、Giraldo et al. (2010) は、重みを時間に依存する関数として与えたモデルと、トレースバリオグラムを用いたモデル (continuous time-varying kriging (CTVK)) を比較し、結果に大きな差異がないことを実証している。

さて、重みを  $\lambda_i$  と与えたとしても、実際に観測地が与えられているのは、いくつかの離散的な時点についてのみであるため、 $\int_T (\chi_i(\mathbf{s}_i) - \chi_i(\mathbf{s}_j))^2 dt$  を計算するために、 $\chi_i(\mathbf{s})$  を何らかの関数で近似する必要がある。関数クリギングに関する古典的な研究である Goulard and Voltz (1993) は、 $\chi_i(\mathbf{s})$  をパラメトリックな関数  $\chi_i(\mathbf{s}; \hat{\theta}_s)$  で置き換え、この積分を  $\int_T (\chi_i(\mathbf{s}_i; \theta_{s_i}) - \chi_i(\mathbf{s}_j; \theta_{s_j}))^2 dt$  で与える方法を提案している。しかしながら、パラメトリック関数は、データの推移が複雑な場合、当てはまりの点で問題がある。一方、近年セミパラメトリック統計学 (Ruppert et al., 2003, 2009; Wood, 2011) や、FDA手法の発展によって、フーリエ級数やB-スプライン関数を基底関数として用いたノンパラメトリックな関数近似に関する方法論が発展し、実用化されている。Giraldo et al. (2011) は、特に4次B-スプライン基底関数を用いる手法を提示している。

今、4次B-スプライン基底関数を用いて、地点  $\mathbf{s}_i$  におけるデータが次式のように表現できるとする (例えば、井元・小西, 1999; 柳原・大瀧, 2004)。

$$\tilde{\chi}_i(\mathbf{s}_i) = \sum_{k=1}^{L+4} a_k(\mathbf{s}_i) \mathfrak{g}_{k,t} = \mathbf{a}(\mathbf{s}_i)^{Tr} \mathfrak{g}_t, \quad (6)$$

ここで、 $a_k(\mathbf{s}_i)$  は回帰係数、 $\mathfrak{g}_{k,t}$  ( $k=1, \dots, K$ ) は、4次B-スプラインの基底関数を示す。また、 $(\cdot)^{Tr}$  はベクトル、行列の転置を示す。今、観測が行われた時点が、 $t_1, \dots, t_M \in T$  であるとしよう。このとき、回帰係数は、

$$\min_{\mathbf{a} \in \mathfrak{R}^{L+4}} \sum_{m=1}^M (\chi_{t_m}(\mathbf{s}_i) - \tilde{\chi}_{t_m}(\mathbf{s}_i))^2 + \mu \int_T (\tilde{\chi}_i''(\mathbf{s}_i))^2 dt, \quad (7)$$

によって求めることができる。ここで、「 $\int$ 」は微分演算子を示す。 $\mu$  は、滑らかさの度合いを決めるパラメータであり、この値が小さすぎる場合や、基底関数の個数が多めに選択されている場合において、データへの過剰適合の問題が起こる (柳原・大瀧, 2004)。これら、滑らかさを定めるパラメータは、以下の関数クロスバリデーションスコア (FCV) の最小化によって行う。

$$FCV(L, \mu) = \sum_{i=1}^I \sum_{m=1}^M (\chi_{t_m}(\mathbf{s}_i) - \tilde{\chi}_{t_m}(\mathbf{s}_i)^{(i)})^2. \quad (8)$$

ただし、 $\tilde{\chi}_{t_m}(\mathbf{s}_i)^{(i)}$  は、 $\mathbf{s}_i$  以外における観測値を用いて予測した、 $t_m$  期の  $\mathbf{s}_i$  における関数の予測値である。Giraldo et al. (2011) 同様、本研究でも、FCVを用いていることとする。しかしながら、 $\mu$  と  $L$  という2つの変数に関する最適値の探索は計算負荷が大きい。以下の実証では、茨城県と千葉県を別々に計算、考察している。したがって今後、効率的な探索方法を開発することが課題である。例えば、荒木・小西 (2004) は、情報量基準に基づく方法を提示しており、Reiss et al. (2010) は、セミパラメトリック統計学の方法に基づき、制限付き最尤法 (restricted maximum likelihood (REML)) によって推定する方法を提示している。これらはそれぞれ、参考になるアプローチであると考えられる。

### 3.5. 土地特性の考慮

Giraldo et al. (2011) の関数OKでは、関数ランダム過程が弱定常性を満たすと仮定された。しかしながら、アクセシビリティや土地利用等の土地特性によって、関数の期待値は異なると考えるのが自然であろう。本研究では、関数データに対する線形回帰分析 (functional linear regression (FLR)) によってトレンド項を明示的に取り入れることを試みる。FLRのアプローチは、大きく分けて、[A] 従属変数がスカラーであり、説明変数が関数の場合、[B] 従属変数が関数であり、説明変数がスカラーの場合、[C] 両者が関数の場合の3つがあるが (例えば、Ramsay and Silverman, 2002)、本研究では対象期間で変化しない用途地域ダミーを説明変数として用いるため、[B] のアプローチである function-on-scalar regression を用いることとする。以下、Reiss et al. (2010) を参照しながらこのアプローチについて説明する。function-on-scalar regressionの基本モデルは、次式により与えられる。

$$\chi_i(\mathbf{s}_i) = \sum_{q=1}^Q z_q(\mathbf{s}_i) \beta_{q,t} + \varepsilon_i(\mathbf{s}_i), \quad (9)$$

ここで、 $\chi_i(\mathbf{s}_i)$  は従属変数の関数、 $z_q(\mathbf{s}_i)$  は説明変数のスカラー、 $\beta_{q,t}$  は、回帰係数関数、 $\varepsilon_i(\mathbf{s}_i)$  は、ランダム誤差関数とする。回帰係数関数は、次のように4次B-スプラインの基底関数を用いて表現できるとする。

$$\beta_{q,t} = \mathbf{b}_q^{Tr} \mathfrak{g}_t. \quad (10)$$

ここで、 $\mathbf{b}_q$  は回帰係数  $b_{q,k}$  から成る  $K \times 1$  ベクトルである。

生データ  $\chi_i(\mathbf{s}_i)$  は、外れ値や異常値を含むことが多い。実際には  $\tilde{\chi}_i(\mathbf{s}_i)$  を用いて分析を行うことが多い。  $\tilde{\chi}_i(\mathbf{s}_i)$  からなる  $I \times 1$  ベクトルを、 $\mathbf{y}_i$  とする

と、式 (9) は、 $y_t = \mathbf{Z}\boldsymbol{\beta}_t + \boldsymbol{\varepsilon}_t$  と書き換えることが可能である。ただし、 $\mathbf{Z}$  は、 $I \times Q$  の説明変数行列、 $\boldsymbol{\beta}_t = [\beta_{1,t}, \dots, \beta_{Q,t}]^{Tr}$  である。さらに、離散的な観測点  $t_1, \dots, t_M \in T$  を考えると、次式が得られる。

$$\mathbf{Y} = \mathbf{Z}\mathbf{B}\boldsymbol{\Theta}^{Tr} + \mathbf{E}, \quad (11)$$

ここで、 $\mathbf{Y}$  は、 $y_t$  から成る  $I \times t_M$  行列、 $\mathbf{B}$  は、 $\mathbf{b}_q$  から成る  $Q \times K$  行列、 $\boldsymbol{\Theta}^{Tr}$  は、 $\boldsymbol{\vartheta}_t$  から成る  $K \times t_M$  行列、 $\mathbf{E}$  は、 $\boldsymbol{\varepsilon}_t$  から成る  $I \times t_M$  行列である。

$\mathbf{B}$  は、次式の最小化により求められる。

$$\int_T \|\mathbf{A}\boldsymbol{\vartheta}_t - \mathbf{Z}\mathbf{B}\boldsymbol{\vartheta}_t\|^2 dt + \sum_{q=1}^Q \pi_q \int_T [L(\mathbf{b}_q^{Tr} \boldsymbol{\vartheta}_t)]^2 dt, \quad (12)$$

ここで、 $\|\bullet\|$  はノルムを示し、 $\mathbf{A} = (\mathbf{a}(s_1) \cdots \mathbf{a}(s_I))^{Tr}$  は、 $\mathbf{a}(s_i)$  から成る  $I \times K$  の回帰係数行列である。 $L$  は、線形微分演算子 (linear differential operator) であり、例えば二次微分演算子  $\int \beta_{k,t}''^2 dt$  等によって与える。 $\mathbf{B}$  の推定量とその分散は、次式により与えられる。

$$\begin{aligned} \text{vec}(\hat{\mathbf{B}}^{Tr}) = & \\ & [(\mathbf{Z} \otimes \mathbf{J}_{99}^{1/2})^{Tr} (\mathbf{Z} \otimes \mathbf{J}_{99}^{1/2}) + \mathbf{P}_A]^{-1} (\mathbf{Z} \otimes \mathbf{J}_{99}^{1/2})^{Tr} \text{vec}(\mathbf{J}_{99}^{1/2} \mathbf{A}^{Tr}), \\ \hat{\text{Var}}[\text{vec}(\hat{\mathbf{B}}^{Tr})] = & \\ & (\mathbf{Z}_J^{*Tr} \mathbf{Z}_J^* + \mathbf{P}_A)^{-1} \mathbf{Z}_J^{*Tr} (\mathbf{I}_{[I]} \otimes \hat{\boldsymbol{\Sigma}})^{Tr} \mathbf{Z}_J^* (\mathbf{Z}_J^{*Tr} \mathbf{Z}_J^* + \mathbf{P}_A)^{-1}, \quad (13) \end{aligned}$$

ここで、 $\text{vec}(\bullet)$  は vec operator,  $\otimes$  はクロネッカー積を示す。 $\mathbf{J}_{99}^{1/2}$  は、その  $p, q$  成分を  $\int \vartheta_{p,t} \vartheta_{q,t} dt$  で与える  $K \times K$  行列、 $\mathbf{I}_{[I]}$  は  $I \times I$  の単位行列、 $\hat{\boldsymbol{\Sigma}}$  は残差から計算する分散共分散行列、 $\mathbf{Z}_J^* = \mathbf{Z} \otimes \mathbf{J}_{99}^{1/2}$  である。

また、 $\mathbf{A} = \text{diag}(\pi_1, \dots, \pi_Q)$ 、 $\mathbf{R}$  をその  $p, q$  成分を  $\int (L\vartheta_{p,t})(L\vartheta_{q,t}) dt$  で与える  $K \times K$  行列としたとき、 $\mathbf{P}_A$  は、 $\mathbf{A} \otimes \mathbf{R}$  で定義される。 $\pi_1, \dots, \pi_Q = \pi$  のとき、 $\pi$  は、一般化クロスバリデーション (generalized cross-validation (GCV)) や、REMLによって推定可能である。本研究では、REMLを用いている (Reiss et al., 2010)。REMLによる推定方法の詳細については、Wood (2011) を参照されたい。

本研究では、まず関数回帰分析でトレンドを考慮し、適用後の残差関数  $\hat{\mathbf{E}}$  に対して、関数OKを適用して残差関数を内挿し、内挿地点のトレンド項と足し合わせるという方法をとる。この手法を、関数 Regression kriging (RK) と称することとする。ただし、本研究では関数回帰分析における  $\mathbf{B}$  の推定において、空間的自己相関を考慮していない。 $\mathbf{B}$  の推定にバリオグラムで構造化された分散共分散をフィードバックして組み入れ、FGLS (feasible generalized least squares) による繰り返し推定を行うことは今後の重

要な課題である。

## 4. 実証

### 4.1. 実証の方法：関数OK

本節では、関数OKの手法を用いて、つくばエクスプレス沿線の開業後の地価変動に関する分析を行う。用いるデータは、図2, 3に示したもので、茨城県、千葉県それぞれについて分析を行う。以下、具体的な手順について示す。

まず、4次B-スプライン基底関数を用いて、2003年~2010年の地価推移を近似する (ただし、相対的な推移パターンの差異のみに着目するため、図2, 3同様2002年価格との比率とする)。次に、関数OKを用いて、図1に示す町丁目の中心座標における関数を予測する。最後に、2003年1月, ..., 2010年1月の10時点における関数値のベクトルを属性とし (8属性)、 $k$ -mean法による空間クラスタリングを試みる。観測点の数は、茨城：42、千葉：99、予測点の数は、茨城433、千葉417である。

### 4.2. 分析の結果：関数OK

計算において、バリオグラムの形状は、一般的な Exponential型、Gaussian型、Spherical型の中から、式 (7) の残差平方和の観点で最も当てはまりの良かった Spherical型を採用した。また、関数のスプライン近似におけるFCVにおいて、 $\mu$  は、かなり小さい値から大きい値まで探索するために、

```
for(i in 1:30){
  mu[i]= exp(i-10)
}
```

から最良のものを探索することとし、 $L+4$  (基底関数の数) は、観測時点数が8時点であることを考慮し、5, 6, ..., 15の中から最良のものを探索することとした。したがって、組み合わせ数は  $30 \times 11 = 330$  となる。図4に、FCVの結果 ( $\mu$  と  $L$  は表1に示す値) に基づく最良ケースのRMSE値を示す。2010年の茨城県では、0.11程度の誤差が存在することとなる。表1に、パラメータの推定結果を示す。レンジパラメータは、茨城県、千葉県でそれぞれ15.6km, 11.1kmと、大小関係も含め直感的に妥当な値が得られた。

推定されたバリオグラムを用いて町丁目中心座標における関数を予測、 $k$ -mean法によって分類したのが図5である。なお、 $k$ -mean法ではカテゴリ数を事前に設定する必要があるが、本研究では、カテゴリの平均値や解釈のし易さ等を勘案して、5つのカテゴリに分類した。図5において、[IBARAKI: number], [CHIBA: number] は、茨城県、千葉県におけるクラスタ (以下、CL) 番号を示し、図6, 7に示したCL

毎の平均値、および図9、10に示したCL位置の図と対応している。似たような変動を示す町丁目は、概ね同一のクラスタに分類されていることが分かる。

まず、図5、図7の茨城県の結果を見ると、例えば、CL1、CL2では開業後2008年までに、それぞれ地価が1.4、1.2倍程度にまで上昇している一方で、CL5では一貫して下落していることが見てとれる。CL3は、地価が対象期間中概ね一定である町丁目と、2008年まで上昇し、その後下落した町丁目が混在しているが後者においても、その上昇幅はCL1、CL2と比較して小さい。CL4と、CL5は、全体的に下落傾向にある。しかしながら、特にCL4については、開業後数年間、下落率（曲線の傾き）が小さくなっており、例えば対象地域周辺の市区町村との相対値で見るという方法をとれば、TX事業の影響が見られると考えられる（Tsutsumi and Seya, 2008）。

千葉県（図5、図8）においても、概ね同様の傾向が見られるが、CL1の上昇幅は、茨城県に比べて小さい。また、2009年、2010年において、地価の下落が顕著である点も、特徴として指摘できる。

図9、図10は、クラスタをGIS上にマッピングしたものである。茨城県において、TX終着駅のつくば駅より、守谷駅のほうが地価の上昇幅が大きい点は興味深い（図6の駅名参照）。対象地域において、快速列車が停車する駅は、南流山、流山おおたかの森、守谷、つくばであるが、このうちの南流山、守谷、停車しない駅の中では、柏の葉キャンパスにおいて地価の上昇幅が大きい。また、全体的な傾向としては、CL4、CL5に分類されている町丁目が多く、TX事業の影響はかなり局所的であるといえる。

また、他の興味深い点として、図10において、TXの路線と接続していない柏駅周辺において、地価がCL2に分類された変動を示している点が挙げられる。言うまでもなく、鉄道の開業の影響は、路線沿線地域でなく、他地域にも及ぶ（堤・瀬谷, 2010）。

さて、図9、10を見ると、茨城県、千葉県に共通して、地価上昇エリアは非常に狭い範囲にとどまっていることが見てとれる。この範囲が、堤・瀬谷 (2010) においてヘドニック・アプローチに基づき、アクセシビリティ（最寄駅距離と最寄駅から東京駅までの時間）の変化から試算した土地資産価値の上昇範囲と比べてかなり狭いことは興味深い。また、堤・瀬谷 (2010) では、守谷駅よりも、開業前後のアクセシビリティの向上が著しいつくば駅周辺における地価上昇が大きいと予測したが、実際は逆であった。このことは、東京への通勤鉄道としての重要性が、堤・瀬谷 (2010) のモデル構築時点より増加していると解釈できよう。いずれにしろ、図9、10の結果からは、TX事業の便益帰着範囲は、十分に

smallであるといつてよいと考えられる。

### 4.3. 分析の結果：関数RK

本節では、特に茨城県のデータに着目し、3.4で示した関数 RK の枠組みで用途地域による地価変動の差異の考慮を試みる。関数回帰分析における説明変数としては、切片 (Intercept) と、用途地域ダミー（第一種住居地域；第一種中高層住居専用地域；第一種低層住居専用地域；第二種住居地域；第二種中高層住居専用地域）を用いている。しかしながら、本実証において被説明変数としては「比率」を用いているため、回帰係数そのものに関する考察は難しく、あくまでトレンド除去のための分析となる点に注意されたい。

表2に、パラメータ推定結果、図11に回帰係数推定値の  $t$  値関数を示す。関数 FCV における最良ケースの FCV 得点で見ると、関数 OK の場合と比べて、関数 RK で悪化するという結果となった。これは、用途地域ごとの変動パターンの差異が必ずしも大きくないことと、本実証においては従属変数を比率としているため、導入した用途地域変数の回帰係数の多くが、有意でないことによると考えられる。

回帰係数は必ずしも有意でないものの、 $t$  値の変動パターンが用途によって異なる点は、実証的な知見として興味深い。例えば、第一種低層住居専用地域の回帰係数の  $t$  値は、他の用途と大きく異なる変化を見せている。このような結果を、用途ごとの沿線開発計画と照らし合わせながら、都市計画的な観点から考察することで、より詳細な便益帰着構造の把握を行うことができるといえる。したがって今後、関数 RL 等の FDA の技法を用いた多くの実証研究の蓄積が望まれるところである。

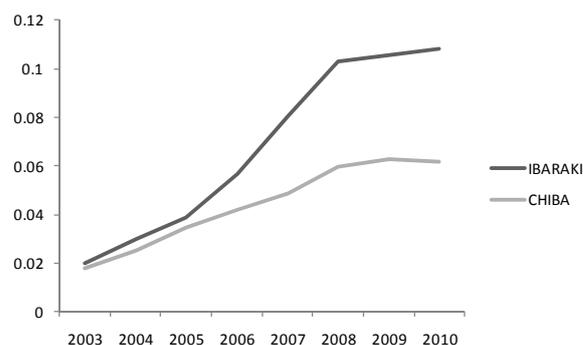
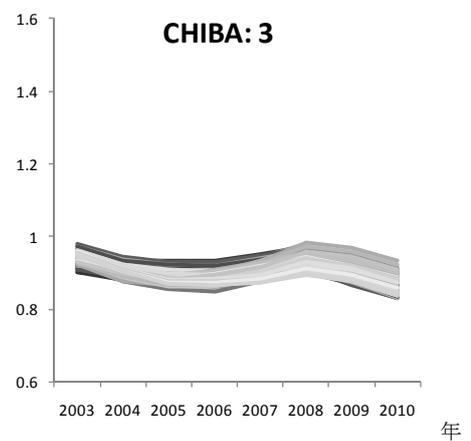
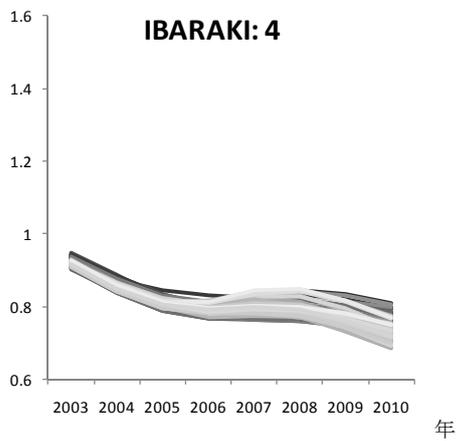
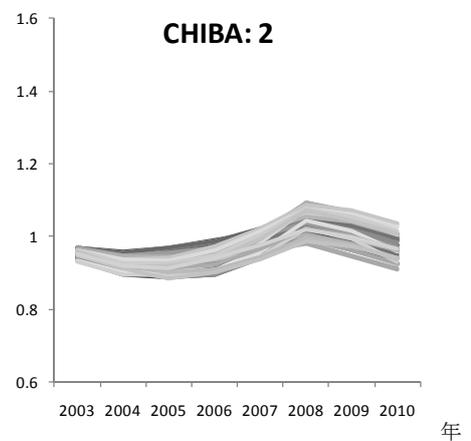
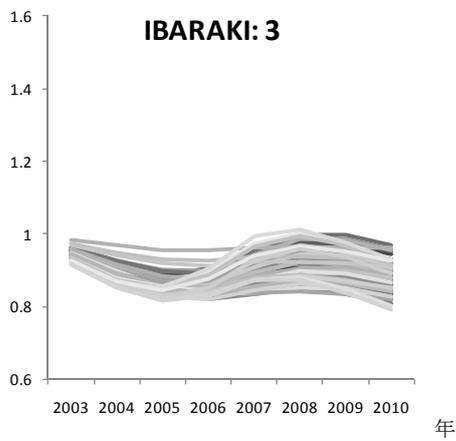
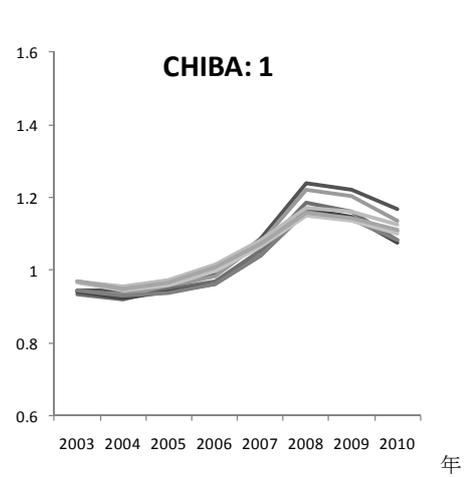
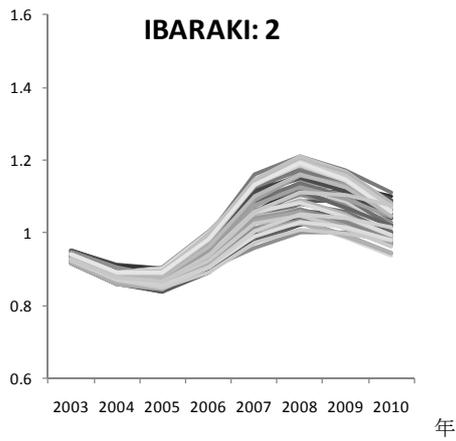
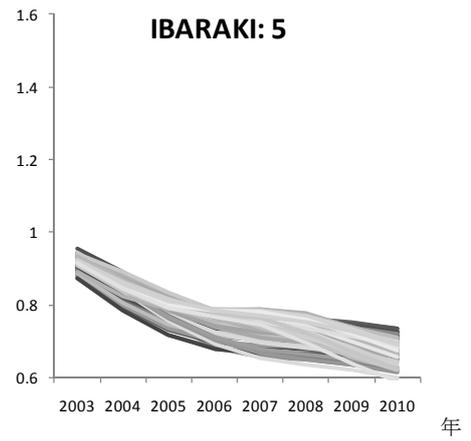
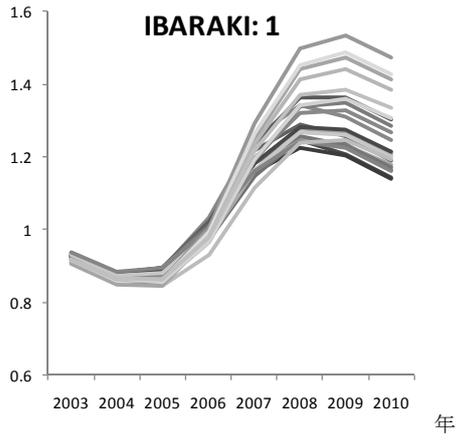


図4 CVに基づくRMSE値

表1 パラメータ推定結果

	IBARAKI	CHIBA
Nugget	0.316	0.0400
Partial-sill	0.319	0.0718
Range	15.6	11.1
Number of basis	10	11
Mu	0.04979	0.0009119
Functional-CV score	1.68	1.55



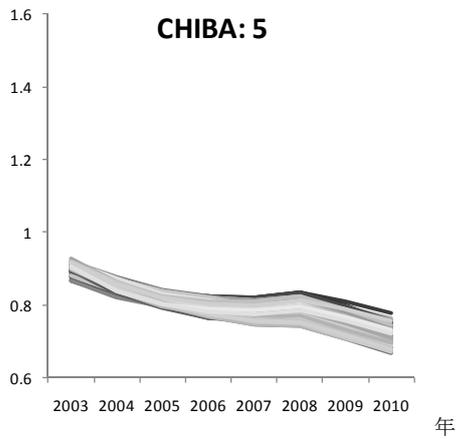
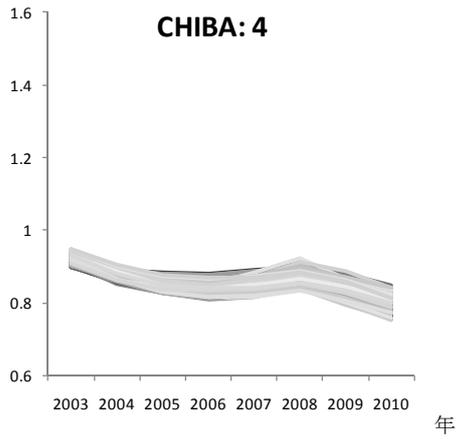


図5 2003-2010年までの  
クラスター毎の地価変動

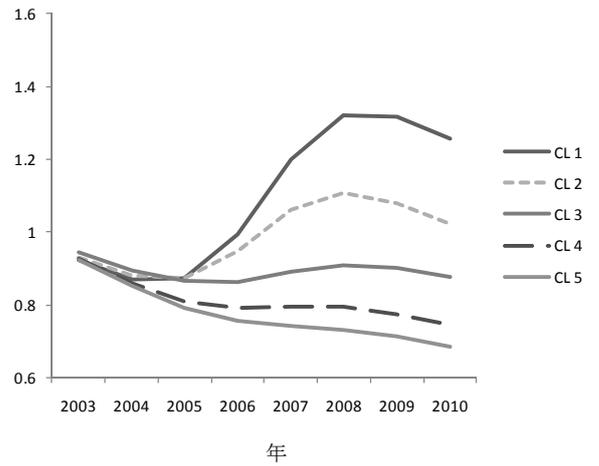


図7 クラスタリングごとの平均値（茨城）

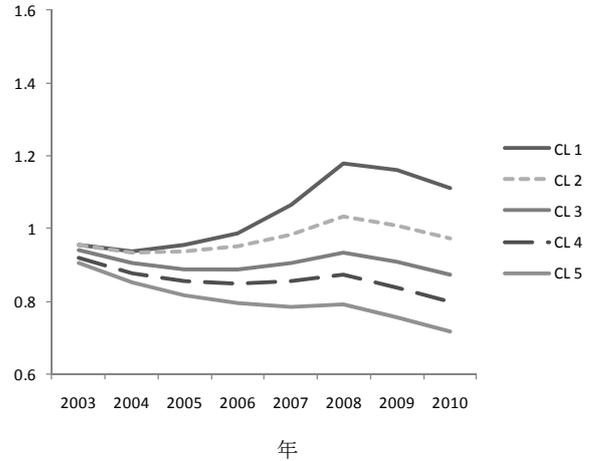


図8 クラスタリングごとの平均値（千葉）



図6 TX 駅名（対象地域のみ）

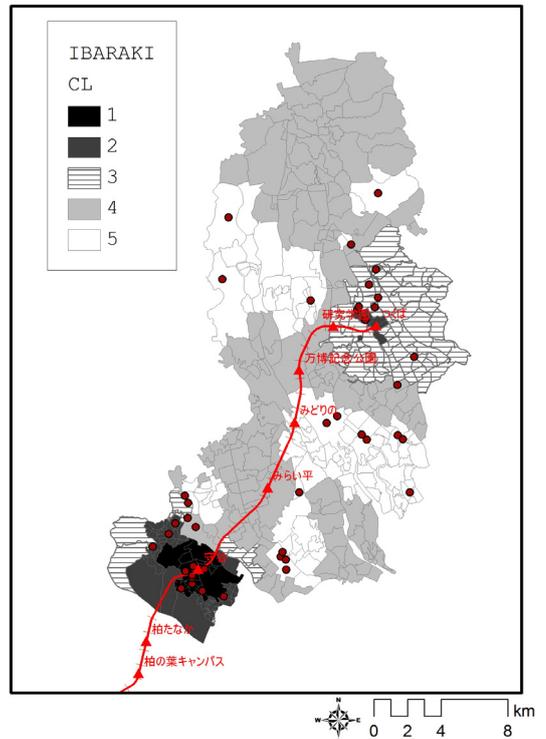


図9 クラスタリング結果の視覚化（茨城）

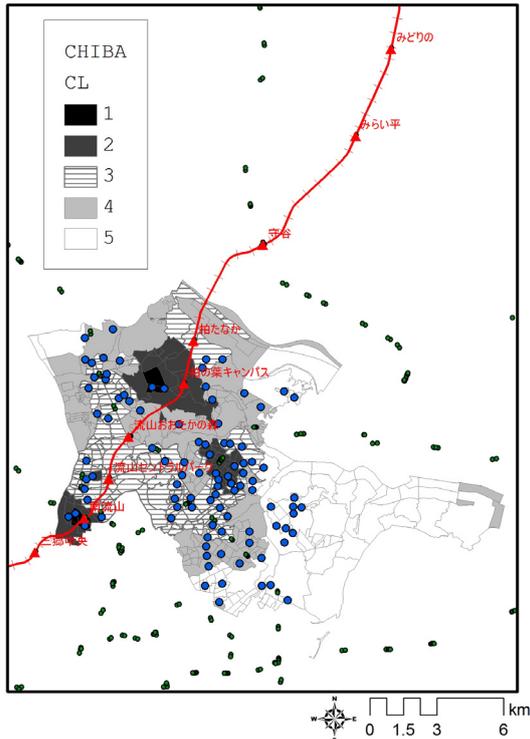


図 10 クラスタリング結果の視覚化 (千葉)

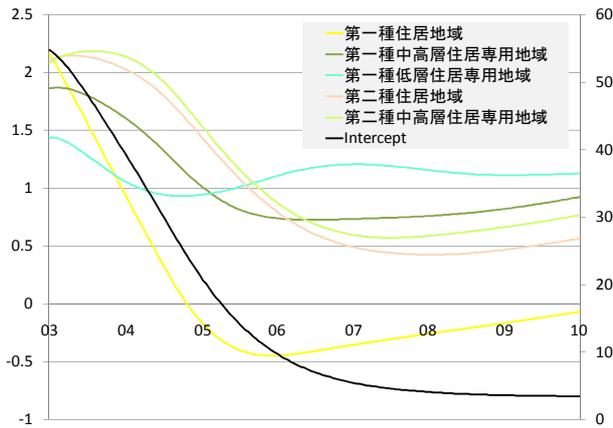


図 11  $t$  値と各用途地域 (茨城)  
(右軸は切片, 左軸はその他の変数)

表2 パラメータ推定結果

	IBARAKI
Nugget	0.000
Partial-sill	0.584
Range	3.97
Number of basis	10
Mu	0.04979
Pi	4.048
Functional-CV score	2.53

## 5. おわりに

本研究では, FDAの手法に着目し, 各公示地点での対象期間の地価変動を関数として表現し, その関

数を用いて対象地域の時系列的な地価変動を分析する方法論について検討し, 実際に適用した. 具体的には, 公示地点における地価変動を4次B-スプライン関数を用いて近似し, その関数を関数OKの手法を用いて町丁目中心座標に内挿し,  $k$ -mean法を用いて5クラスに分類した. 結果として, 開業後2005–2008年にかけて快速列車が停車する駅を中心に地価が上昇し, 2009年, 2010年においては対象地域全体で, 下落する傾向が示された.

今後, 対象地域を広げるとともに, さらに継続的に分析を行ってゆくことで, クラスターの安定性について考察してゆきたい. また, 関数データである点を明示的に考慮したクラスタリング手法もいくつか開発されており (例えば, Chiou and Li, 2007), これらの適用についても課題としたい.

また, 本研究では公示地価が連続的に観測されている地点のみを対象としたが, FDAでは, 関数で近似するステップを挟むことで, 欠損値を含む, すなわち連続観測されていない地点を同じフレームで分析に加えることも可能であると考えられる. しかしながら, このことは関数の近似精度を悪化させるため, 慎重な実証研究が必要となる.

さらに, 回帰モデルとの組み合わせによる方法 (関数RK) を精緻化し, 予測精度を上昇させるためのモデル開発を行い, 地理的加重回帰分析 (Geographically weighted regression (GWR)) に基づくYamanishi and Tanaka (2003) のアプローチと比較することも重要な課題である.

## 参考文献

- 1) 荒木由布子・小西貞則 (2004) 動径基底関数展開に基づく関数回帰モデリング. 「応用統計学」, 33 (3), 243–256.
- 2) 五十嵐立青 (2007) 公共事業における計画化—事業化—実現化—プロセスの政治リソース分析—つくばエクスプレス開業のダイナミズム—. 「筑波大学大学院人文社会科学学位論文」.
- 3) 井上亮・清水英範・吉田雄太郎・李勇鶴 (2009) 時空間クリギングによる東京 23 区・全用途地域を対象とした公示地価の分布と変遷の視覚化. 「GIS—理論と応用」, 17 (1), 13–24.
- 4) 井元清哉・小西貞則 (1999) B-スプラインによる非線形回帰モデルと情報量規準. 「統計数理」, 47 (2), 359–373.
- 5) 上田孝行・堤盛人・武藤慎一・山崎清 (2009) わが国における応用都市経済モデル—特徴と発展経緯—. 「計画・交通研究会, ワーキングペーパーシリーズ」, WP09-04.
- 6) 王尾和寿・桑原祐史・黒崎健司・海老沢雅之・佐伯隆・村山祐司 (2005) つくばエクスプレス沿線市町村を対象とした土地利用変化の解析. 「地理情報システム学会講演論文集」, 14, 189–194.
- 7) 大村 謙二郎 (2007) つくばエクスプレス TX 沿線のまちづくりの課題と展望. 「Joyo ARC」,

- 39 (454), 10–17.
- 8) 岡本直久・川田真理絵・石田東生・堤盛人・谷口綾子・諸田恵士 (2008) つくばエクスプレス開業が周辺住民の交通行動に与えた影響. 「土木計画学研究・論文集」, 25 (3), 801–806.
  - 9) 小川茂男 (2007) ALOS 等を用いたつくば周辺の土地被覆・土地利用変化の把握. 「写真測量とリモートセンシング」, 46 (4), 26–31.
  - 10) 金子雄一郎・伊東誠 (2007) 鉄道整備事業の事後評価手法に関する諸検討—実際の評価の経験を踏まえて—. 「運輸政策研究」, 10 (3), 19–29.
  - 11) 金本良嗣 (1992) ヘドニック・アプローチによる便益評価の理論的基礎. 「土木学会論文集」, 449, IV-17, 47–56, 1992.
  - 12) 日下博幸・大庭雅道・鈴木智恵子・林陽生・水谷千亜紀 (2009) 冬季晴天日におけるつくば市のヒートアイランド:予備観測の結果. 「日本ヒートアイランド学会論文集」, 4, 10–14.
  - 13) 小荒井衛・中埜貴元 (2010) 時空間データセットを用いた TX 沿線の土地利用変化の特徴—つくば市 TX 研究学園駅周辺の事例—. 「CSIS DAYS 2010」, C05, 38.
  - 14) 篠崎寿夫・吉田正広・高橋宣明・富山薫順・松浦武信 (1994) 『現代工学のための数ベクトルの空間からヒルベルト空間へ (応用数学ガイドシリーズ)』. 現代工学社.
  - 15) 白柳博章・青山吉隆・中川大・松中亮治・野村友哉 (2000) ヘドニック法による都市間交通プロジェクトの便益計測システムの構築と Small 条件の検証. 「土木計画学研究・論文集」, 17, 57–66.
  - 16) 鈴木章裕・中井検裕 (2002) 宅鉄法における一体型土地区画整理事業の特徴とその効果に関する基礎的研究. 「都市計画論文集」, 37, 421–426.
  - 17) 瀬谷創・堤盛人 (2007) Kriging と GWR によるつくばエクスプレス開発効果の抽出. 「日本写真測量学会 平成 19 年度 年次学術講演会 発表論文集」, 213–217.
  - 18) 高津俊司・堀川淳・橋本浩史・佐藤馨一 (2006) つくばエクスプレス線の建設における鉄道と都市との一体整備に関する考察. 「土木計画学研究・論文集」, 23 (3), 635–640.
  - 19) 谷口綾子・浅見知秀・藤井聡・石田東生 (2009) 公共交通配慮型居住地選択に向けた説得的コミュニケーションの効果分析. 「土木学会論文集 D」, 65 (4), 441–448.
  - 20) 淡野寧彦・小島大輔・花島裕樹・亀川星二 (2009) つくば市におけるつくばエクスプレス開通による沿線地域の変容. 「地理空間」, 2 (1), 51–62.
  - 21) 堤盛人 (1997) 代表的な便益計測手法. 森杉壽芳編著 『社会資本整備の便益評価 一般均衡理論によるアプローチ』. 勁草書房, 第 4 章, 53–62.
  - 22) 堤盛人・瀬谷創 (2010) 便益計測への空間ヘドニック・アプローチの適用. 「土木学会論文集」, 66 (2), 178–196.
  - 23) 肥田野登 (1997) 『環境と社会資本の経済評価—ヘドニック・アプローチの理論と実際』. 勁草書房.
  - 24) 肥田野登・中村英夫・荒津有紀 (1986) 資産価値法に基づいた都市近郊鉄道の整備効果の計測. 「土木学会論文集」, 365, IV-4, 135–144.
  - 25) 肥田野登・林山泰久・山村能郎 (1992) ヘドニック・アプローチによる社会資本整備の評価/都市間交通施設整備がもたらす便益と地価変動. 「土木学会論文集」, 449, IV-17, 67–76.
  - 26) 藤谷匡 (2006) 衛星データを用いた近年のつくば市における土地被覆と地表面輝度温度の変化の解析. 「筑波大学大学院システム情報工学研究科修士論文」.
  - 27) 眞浦靖久・塚田幸広・井坪慎二 (2005) PP データによる TX 開通に伴う交通行動への影響に関する基礎的分析. 「土木計画学研究・講演集」, 33, CD-ROM, 300.
  - 28) 松本渉 (2010) 開業 5 周年を迎える、つくばエクスプレスの現状と今後の沿線の発展性について—つくばエクスプレス沿線地区の開発動向調査—. 「ちば経済季報」, 81, 1–13.
  - 29) 水田正弘 (2005) 関数データとその解析法. 「知能と情報」, 17 (4), 413–417.
  - 30) 柳原宏和・大瀧慈 (2004) B-スプラインノンパラメトリック回帰モデルにおける過剰適合の回避について. 「応用統計学」, 33 (1), 51–70.
  - 31) Ahsan, H.M., Nakamura, H. and Ueda, T. (2001) A land use-transport model: The structure and applications. *Journal of Civil Engineering*, 29 (2), 219–235.
  - 32) Chiou, J-M., and Li, P-L. (2007) Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69 (4), 679–699.
  - 33) Delicado, P., Giraldo, R. Comas, C. and Mateu, J. (2010) Statistics for spatial functional data: Some recent contributions. *Environmetrics*, 21, (3-4), 224–239.
  - 34) Ferraty, F. and Romain, Y. (2011) *The Oxford Handbook of Functional Data Analysis (Oxford Handbooks in Mathematics)*. Oxford University Press.
  - 35) Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory And Practice (Springer Series in Statistics)*. Springer-Verlag.
  - 36) Giraldo, R., Delicado, P. and Mateu, J. (2010) Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics*, 15 (1), 66–82.
  - 37) Giraldo, R., Delicado, P. and Mateu, J. (2011) Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, *in press*.
  - 38) Goulard, M. and Voltz, M. (1993) Geostatistical interpolation of curves: A case study in soil science. In Soares, A. (ed), *Geostatistics Tróia '92*. Kluwer, 805–816.
  - 39) Kawada, M., Okamoto, N., Ishida, H. and Tsutsumi, M. (2010) Effects of Tsukuba Express project on the residents' travel behavior. *Journal of the Eastern Asia Society for Transportation Studies*, 8, *in press*.
  - 40) Pior, M.Y. and Shimizu, E. (2001) GIS-aided evaluation system for infrastructure improvements: focusing on simple hedonic and Rosen's two-step approaches. *Computers, Environment and Urban Systems*, 25 (2), 223–246.
  - 41) Pior, M.Y., Shimizu, E. and Nakamura, H. (1998) GIS-aided benefit evaluation system for urban railway improvements: Focusing on hedonic approach. *Theory and Applications of GIS*, 6 (2),

- 11–22.
- 42) Ramsay, J.O., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and Matlab*. Springer-Verlag.
- 43) Ramsay, J.O. and Silverman, B.W. (2002) *Applied Functional Data Analysis: Method and Case Studies*. Springer-Verlag.
- 44) Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis, Second Edition*. Springer-Verlag.
- 45) Reiss, P.T., Huang, L. and Mennes, M. (2010) Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6 (1), Article 28.
- 46) Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.
- 47) Ruppert, D., Wand, M.P. and Carroll, R.J. (2009) Semiparametric regression during 2003–2007, *Electronic Journal of Statistics*, 3, 1193–1256.
- 48) Tsutsumi, M. and Seya, H. (2008) Measuring the impact of large-scale transportation project on land price using spatial statistical models. *Paper in Regional Science*, 87 (3), 385–401.
- 49) Wood, S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73 (1), 3–36.
- 50) Yamanishi, Y. and Tanaka, Y. (2003) Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of the Japanese Society of Computational Statistics*, 15 (2), 307–317.

(2011年5月6日 講演集原稿CD-ROM提出)