Ex-post identification of geographical extent of benefited area by a transportation project: Functional data analysis method

Seya, Hajime

Yoshida, Takahiro

Tsutsumi, Morito

Ex-post identification of geographical extent of benefited area by a transportation project: Functional data analysis method

Hajime Seya
Departments of Civil Engineering, Graduate School of Engineering Faculty of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan,
E-mail: hseya@people.kobe-u.ac.jp

Takahiro Yoshida
Graduate School of Systems and Information Engineering, University of Tsukuba,
1-1-1, Tennodai, Tsukuba, Ibaraki, 305-8573, Japan,
E-mail: yoshida.takahiro@sk.tsukuba.ac.jp

Morito Tsutsumi
Faculty of Engineering, Information and Systems, University of Tsukuba,
1-1-1, Tennodai, Tsukuba, Ibaraki, 305-8573, Japan,
E-mail: tsutsumi@sk.tsukuba.ac.jp

Abstract: This study proposes an approach for ex-post identification of the geographical extent of an area benefiting from a transportation project, using functional data analysis methods. Our approach focuses on real estate (land) price data for the ex-post spatial evaluation. First, we prepare a panel of land prices observed before, during, and after the project in the areas that are potentially impacted. Second, using functional data analysis, movements of land prices in each observed site during the target period are approximated by continuous functions. Third, using the functional ordinary Kriging technique, the functions for land price movements in each micro district are spatially predicted. Lastly, by employing the functional clustering (functional K-means) technique, potential areas of benefit may be identified. Different from exiting before-and-after methods, including difference-in-differences method, the proposed procedure based on functional data analysis can describe a map with a complex spatial distribution pattern of benefit rather than using distance bands (ring buffer) from transportation cores, such as railway stations, bus stops, and highway interchanges. Then, the proposed procedure is empirically applied to a large-scale Japanese heavy railway project. The obtained result shows so-called redistributive effect, that is, land prices decrease around exiting stations and increase around new stations. In addition, interestingly enough, the spatial distribution pattern of the identified areas of benefit using this procedure are fairly similar to that of ex-ante predicted areas of benefit by the hedonic approach. Thus, capitalization is observationally confirmed with regard to accessibility improvement.

Keywords: functional data analysis, functional clustering, functional ordinary Kriging, transportation project

1. Introduction

A transportation project may improve the accessibility of land (Zhu and Liu, 2004). The improvement of accessibility may be reflected in both the price of the land and the intensity of development (Medda, 2012). Typically, the effect of transportation on land use and development may take a relatively long time to observe, but the effect of transportation on property values is likely to occur sooner (Stokenberga, 2014). Hence, existing studies on transportation impact on land use tried to examine not only the former impact but also the latter impact. Because higher land prices discourage the development of low-density housing in railway station/bus stop areas, and encourage high-density, transportation-oriented development (Knaap et al., 2001), investigating capitalization impacts of transportation projects on land/housing prices is fairly important from the viewpoint of sustainable transportation and land use.

Existing studies on the relationship between transportation and land/housing prices typically employ the with-and-without hedonic approach in a cross-sectional sense (Rosen, 1974; Kanemoto, 1988) to analyze the effects of transportation on land/housing prices (or rent values)[1]. One empirical difficulty of such an approach includes possible omitted variable bias, which may be mitigated under some particular situation using spatial econometric models (Seya et al., 2013, 2015). To confront such a problem and to avoid possible omitted variable bias, other studies employ the "before-and-after" approach (e.g., Lin and Hwang, 2004; Mayor et al., 2012; Kuminoff and Pope, 2014), by focusing on the changes in land/housing prices before and after projects or their announcement[2]. When we implement the before-and-after approach using, for example, the well-known difference-in-differences method (Dubé et al., 2013, 2014), we need to define proximity effects. According to Dubé et al. (2013), existing measures can be categorized into [1] dummies based on distance (or travel time) cut-off criteria generating ring buffers around stations or along service routes; [2] continuous measures of distance (or travel time) from each house to the nearest station; and [3] mixed approaches combining both categories. However, because spatial distribution patterns of benefit could be complex, such an isotropic assumption is sometimes not realistic.

Thus, here, we attempt to propose a novel approach for ex-post identification of the geographical extent of a benefited area by a transportation project, using the functional data analysis methods. Our approach focuses on real estate (land) price data. First, we prepare a panel of land

---

[1] Findings in Western studies are mixed with respect to effects of magnitude and direction, ranging from a negative to an insignificant or a positive effect (Vessali, 1996; Debrezion et al., 2007; Dubé et al., 2013), and also mixed in Asia (Seya and Timmermans, 2016).

[2] Announcement effects, or anticipated effect, are actually found in the exiting studies. Examples include McDonald and Osuji (1995) and Grimes and Young (2013) for the US, Bae et al. (2003) for South Korea, Tsutsumi and Seya (2008) for Japan, and Jayantha et al. (2015) for Hong Kong.

prices observed before, during, and after the project in the areas that are potentially impacted. Second, using the FDA methods, movements of land prices in each observed site during the target period are approximated by continuous functions. Third, using the functional ordinary Kriging (FOK) technique developed by Giraldo et al. (2011), the functions for land price movements in each micro district (precisely divided region based on its seven-digit postcode, called *cho-cho-moku* in Japanese) are spatially predicted. Lastly, by employing the functional clustering (FC) technique, potential areas of benefit may be identified. Conventional data clustering methods, including the K-means method, implicitly assume that each observation is independent and ignore the observation order. By contrast, the FDA method explicitly considers the time-series nature of data, and therefore, may capture the trend of data more properly. In addition, when data is very noisy, the FDA method has merit because the method is robust to outliers in the sense that it uses a curve for clustering rather than noisy data itself.

The proposed procedure is empirically applied to a large-scale Japanese heavy railway, the Tsukuba Express (TX) line project. The TX line, which connects Akihabara district of Tokyo Prefecture to Tsukuba city[3], Ibaraki Prefecture, has a length of 58.3 km and commenced operations on August 24, 2005. Cervero and Landis (1995) noted for developed countries that, "although new transportation investments no longer shape urban form by themselves, they still play an important role in channeling growth and determining the spatial extent of metropolitan regions by acting in combination with policies such as supportive zoning and government-assisted land assembly." Ewing (2008) followed this discussion in a highway context, and suggested that even if the net economic impacts, which he called "aggregate impact" of highway investments, are small, highways may have important impacts on the geographic pattern of development within a region, which he called "localized impacts." We might say that the impact of the TX line is a mix of generative (aggregate) and redistributive (localized) impact because, although the line is located in a developed country, Japan, the TX line has reduced the travel time from Akihabara to Tsukuba from 85 min, using the existing Joban line or a highway bus, to 45 min.

Among the studies that have focused on land prices around the TX line, Tsutsumi and Seya (2008) showed the existence of so-called "announcement effects," that is, land prices around the train stations began to rise from 2001, before the start of operations. Pior et al. (1998), Pior and Shimizu (2001), and Tsutsumi and Seya (2009) evaluated the benefits of the TX line project based on the hedonic approach and Ahsan et al. (2001) evaluated the benefits using an integrated land-use transportation model. However, there is no study that has attempted an ex-post assessment of the TX line project focusing on land prices. The remainder of this paper is organized as follows. Section 2 introduces the FDA method, including the FOK and FC techniques. Section 3 presents an empirical demonstration. Finally, Section 4 concludes this paper with an outlook.

---

[3] Tsukuba city is one of the hub towns of research and science in Japan.

2. Functional data analysis and functional ordinary Kriging

FDA concerns the analysis of information on curves or functions, and was developed by Prof. Jim Ramsay and his colleagues in the 1990s (e.g., Ramsay and Silverman, 2002, 2005; Ramsay et al., 2009). Although FDA is now used widely in many fields (e.g., Ferraty and Vieu, 2006; Ramsay et al., 2009; Ferraty and Romain, 2011), we can find no application studies in the field of transportation[4], except for the work by Chiou et al. (2014) for missing value imputation and outlier detection for traffic flow data. Delicado et al. (2010) and Giraldo et al. (2010; 2011) proposed FOK for functional data as a natural extension of the Kriging technique to functional data; the former is one of the well-established methods for spatial prediction and interpolation of scalar or vector data. In FOK, the relationship among functions is modeled using the variogram, based on the notion that functions whose geographical locations are close to each other exhibit similar values. Here, we summarize the FOK technique based on Delicado et al. (2010) and Giraldo et al. (2011).

Suppose the functional random processes $\{\chi(s) : s \in D \subseteq \Re^2\}$, whose realizations (or values) are functions defined on $T = [a,b] \subseteq \Re$ belong to

$$L_2(T) = \left\{ f : T \to \Re, \ s.t. \int_T f(t)^2 \, dt < \infty \right\},$$

(1)

where $L_2(T)$ denotes Hilbert space with inner product $\langle f, g \rangle = \int_T f(t)g(t)dt$. In addition, for every fixed $t_0 \in T$, $\{\chi_{t_0}(s) : s \in D \subseteq \Re^2\}$ is a scalar-valued random process defined on $\Re^2$ with values in $\Re$. Our goal is to spatially predict a function $\chi_{[2003,2010]}(s_0) :\to \Re$ (where $a = 2003$ and $b = 2010$) based on the observed land price change functions, $\chi(s_i), i = 1, ..., I$, where s_i ($i = 1, ..., I$) are the observed sites.

Here, we assume that the functional random process satisfies the assumption of weak stationarity, that is, $E(\chi_{t_0}(\mathbf{s})) = m_{t_0}, V(\chi_{t_0}(\mathbf{s})) = \sigma_{t_0}^2$, $\forall t_0 \in T, s \in D$, where m_{t_0} and $\sigma_{t_0}^2$ are mean and variance functions, respectively and $Cov(\chi_{\tilde{t}_0}(\mathbf{s} + \boldsymbol{h}), \chi_{t_0}(\mathbf{s})) = C_{\tilde{t}_0, t_0}(\boldsymbol{h})$, $\forall t_0, \tilde{t}_0 \in T, \boldsymbol{h}$, where $C_{.,.}(\boldsymbol{h}) : T \times T \to \Re$ is a square-integrable function. In other words, a weak stationarity assumption here means that the mean and variance functions are constant and the covariance depends only on the distance between sampling points for any time period. With the weak stationarity assumption,

---

[4] Sadahiro and Kobayashi (2014) developed their original method for spatial analysis of temporal data. We find that their idea of visualizing temporal patterns from GIS is fairly similar to ours, although the objectives and methods are quite different.

we can obtain the following variogram function:

$$\gamma_{\tilde{t}_0,t_0}(\boldsymbol{h}) = \frac{1}{2}Var\left(\chi_{\tilde{t}_0}(\boldsymbol{s}+\boldsymbol{h}) - \chi_{t_0}(\boldsymbol{s})\right). \tag{2}$$

This is used to model the spatial autocorrelation relationships among functions, which are used for spatial prediction. The best linear unbiased predictor for $\chi(\boldsymbol{s}_0)$ can be obtained as

$$\hat{\chi}_t(\boldsymbol{s}_0) = \sum_{i=1}^{I} \lambda_i \chi_t(\boldsymbol{s}_i), \ \ \lambda_1,....,\lambda_I \in \Re, \tag{3}$$

The weight λ_i ($i=1,...,I$), assigned to the function for the observed site s_i ($i=1,...,I$) may be determined by minimizing

$$E\left[\int_T (\hat{\chi}_t(\boldsymbol{s}_0) - \chi_t(\boldsymbol{s}_0))^2 dt\right], \tag{4}$$

with the constraint $E(\hat{\chi}(\boldsymbol{s}_0) - \chi(\boldsymbol{s}_0)) = 0$. Now, our optimization problem can be formulated as

$$\min_{\lambda_1,...,\lambda_I} \int_T Var\left(\hat{\chi}_t(\boldsymbol{s}_0) - \chi_t(\boldsymbol{s}_0)\right)dt, \tag{5.a}$$

$$s.t. \sum_{i=1}^{I} \lambda_i = 1. \tag{5.b}$$

where (5.b) is the unbiasedness constraint.

Because observations are given for some discrete time periods, in order to evaluate $\int_T (\chi_t(\boldsymbol{s}_i) - \chi_t(\boldsymbol{s}_j))^2 dt$, where s_j ($j=1,...,I$) is the observed site, we need to approximate $\chi_t(\boldsymbol{s}_i)$ and $\chi_t(\mathbf{s}_j)$ by some continuous function. Unlike classical studies, which attempted to approximate $\chi_t(s)$ by a parametric function, Giraldo et al. (2011) employed cubic B-splines for smoothing[5]. Suppose that $\chi_t(\boldsymbol{s}_i)$ observed at site s_i can be expressed as

$$\tilde{\chi}_t(\boldsymbol{s}_i) = \sum_{k=1}^{L+4} a_k(\boldsymbol{s}_i)\vartheta_{k,t} = \boldsymbol{a}(\boldsymbol{s}_i)^{Tr}\boldsymbol{\vartheta}_t, \tag{6}$$

where $a_k(\boldsymbol{s}_i)$ denotes a regression coefficient, $\vartheta_{k,t}$ denotes a basis function of cubic B-splines, $\boldsymbol{a}(s_i)$ is a vector whose element is given by $a_k(\boldsymbol{s}_i)$, $\boldsymbol{\vartheta}_t$ is a vector whose element is given by $\vartheta_{k,t}$, and ($\bullet^{Tr}$) denotes the transpose of a matrix or a vector ($\bullet$). Let the discrete observed time period be denoted as $t_1,...,t_M \in T$. Then, the regression coefficient can be estimated by minimizing

---

[5] Other basis functions developed in non-/semi-parametric statistics also can be applied (Ruppert et al., 2003, 2009).

$$\min_{a \in \Re^{L+4}} \sum_{m=1}^{M} (\chi_{t_m}(s_i) - \tilde{\chi}_{t_m}(s_i))^2 + \mu \int_{T} (\tilde{\chi}''_t(s_i))^2 dt \quad , \tag{7}$$

where $(')$ denotes the differential operator, and μ is the parameter that controls the degree of smoothness. μ can be estimated by minimizing the following functional cross-validation (FCV) score.

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

where $\tilde{\chi}_{t_m}(\mathbf{s}_i)^{(i)}$ denotes the prediction on s_i evaluated at t_m, by excluding the site s_i from the sample.

After the curve on s_i is predicted as $\hat{\tilde{\chi}}_t(s_i)$, we classify the predicted curves using the FC approach. Here, we employ the most basic functional K-means approach, which can be implemented by minimizing

$$\min \sum_{i=1}^{I} \int_{T} (\hat{\tilde{\chi}}_t(\mathbf{s}_i) - h_{t,k})^2 dt \quad . \tag{9}$$

where $h_{t,k}$ denotes the k-th cluster center (e.g., Ferraty and Romain, 2011). As the conventional K-means algorithm, a priori specification of a number of clusters "k" is needed. We can use the "kmeans.fd" function of the "fda.usc" package in R to implement K-means clustering for functional data.

3. Empirical investigation

3.1. Study area

As explained in section 1, the proposed procedure is empirically applied to a large-scale Japanese heavy railway, the TX line project. In 1991, the Metropolitan Intercity Railway Company was established, and the TX project was formulated by Japan's Ministry of Land, Infrastructure and Transport in 1999. Construction commenced in 2001, and operations started in 2005.

Our target area includes municipalities that include the TX stations, as well as those that share boundaries with the stations. In the study area, two railway lines other than the TX line exist—the Joso line, which is a relatively small intra-prefecture non-electrified railway line passing through the west side of the study area[6], and the Joban line, which is a relatively large-scale

---

[6] At the Moriya Station, the TX and Joso lines are interchanged. However, there are no "trackage rights" for either line.

inter-prefecture railway line that connects Tokyo and the Tohoku region. In fact, there was no rail connectivity for most of the areas along the line (Figure 1), and in this regard, the impact of the TX line is expected to be significant. In addition, there is another reason of expectation about land price growth around the TX stations. Although exiting studies suggest that local development around the transportation cores do not always occur (e.g., Handy, 2005), in the TX line case, various urban improvements or development projects actually have been implemented near the new stations. Such developments are ensured by the new law, "Act on special measures concerning comprehensive advancement of housing development and railway construction in metropolitan areas." This law attempts to harmonize construction of railway lines and urban planning around stations, including the technique of land readjustment (e.g., Sorensen, 2001), and thus far, has been applied only to the TX line. Mizutani (2012) empirically verified land-use change around the TX station[7]. Figure 2 shows the names of TX stations and the elevation information derived from a 90-m resolution Shuttle Radar Topography Mission[8] digital elevation model (known as SRTM DEM). There is a fairly high elevation area in the north part of Tsukuba city (Figure 2), where Mount Tsukuba (877 m) is located.

[Figure 1: Location of Tsukuba Express (TX) line] around here

[Figure 2: Station names of TX line in the study area overlaid with digital elevation model] around here

3.2. Data

We use officially assessed land-price data for residential use located in Ibaraki prefecture around the TX line. The Land Appraisal Committee under the Ministry of Land, Infrastructure, Transport and Tourism (MLIT), Japan publishes market values per square meter of "standard sites" as of January 1 every year, which have been assessed by the Committee based on real property appraisal[9]. The land price data, in our study area, were collected annually from 2002 to 2010 (i.e., from 3 years before the commencement to 5 years after the commencement). Figure 3 represents the spatial distribution of observation sites for the land price data, where land prices are collected annually from 2002 to 2010; the data constitute panel data. The boundaries for our *cho-cho-moku*

---

[7] Kawada et al. (2010) investigated travel behavior change after the start of operations.

[8] http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1

[9] The Committee try to select standard lots in terms of current use, environment, site area, shape, etc. of sites in the neighborhood area as "standard sites" of Land Market Value Publication (MLIT, 2016). Needless to say, the market value of a standard site does not uniformly indicate the value of all sites within the neighborhood area, and therefore we cannot deny the possibility of selection bias. 25,270 standard sites were selected across the country by the Committee in 2016. For more details, see MLIT (2016).

micro districts are shown in Figure 3[10]. Figure 4 represents the changes in land prices at those sites, shown as a relative index to the prices in 2002. From Figure 4, we find that for some observations sites, land prices increased dramatically, especially during 2005 to 2008, although they decreased consistently for the majority of the other sites over the target period.

[Figure 3: Observed sites for officially assessed land prices where residential land prices are continuously observed from 2002 to 2010] around here

[Figure 4: Changes in land prices at the observed sites in the study area] around here

3.3. Procedure and settings

A combination of Figures 3 and 4 enables a geographical consideration of the changes of land prices at observed sites. However, from the viewpoint of identification of the extent of the areas of benefit, we are interested in not only the observed sites, but also the unobserved sites. Hence, this study first approximates the movements of land prices in each observed site during the target period by continuous functions, and subsequently, interpolates the functions of each observed site to the micro districts (*cho-cho-moku*) by the FOK technique of Giraldo et al. (2011). Finally, the functional K-means method is applied to summarize and visualize the movements of land prices in the study area. The detailed procedure is as follows.

First, by employing the cubic B-splines for smoothing, land price changes (index) from 2002 to 2010 are approximated. Second, using the FOK technique, functions at the center point of each district shown in Figure 3 are spatially predicted (interpolated). Lastly, the predicted functions at the micro districts are classified into five groups[11], using the functional K-means method. The smoothing parameter in FOK is calibrated by the FCV method for two parameters μ and L[12]. The optimum value for μ is selected from the candidate set generated from

---

[10] The mean size of the *cho-cho-moku* in km^2 is 0.675, and standard deviation is 0.996. Because the Japanese address system is based on this micro district and land prices are assessed at this level, it is natural to employ this analysis unit.

[11] This number has been determined by trial and error.

[12] The FCV for two parameters, μ and L, is computationally demanding. Therefore, it is important to find a more efficient way, including the restricted maximum likelihood (REML) method (Reiss et al., 2010).

$$\text{for}(i \text{ in } 1{:}30)\{$$
$$\text{mu}[i] = \exp(i{-}10)$$
$$\}.$$

Here, we can search for the optimum value for μ from a relatively large candidate set, including both very small and very large values. In addition, $L+4$ (the number of basis function) is searched from among 5, 6, ..., 15, considering the number of time periods. Thus, the number of parameters passed to the functional K-means is 30×11=330. There are 227 observed sites, and 1,488 prediction sites. Of the exponential, Gaussian, and spherical functions, we employ the spherical variogram function, which is the best performer in terms of the residual sum of squares in the FCV calculation.

One of the limitations of our approach is the validity of the assumption of spatial stationarity, which may not hold in general. One way to overcome this limitation is to develop a combined model of functional Kriging and functional regression to allow for non-stationarity (Caballero et al., 2013). However, a statistically sound parameter-estimation procedure for such a model is not yet available. Moreover, because our focus is the averaged movement in each micro district, local effects are beyond our consideration. This is why the non-stationarity problem is less severe in our case, which partially justifies use of the FOK technique.

3.4. Results

Table 1 represents the parameter estimation results. The number of basis functions ($L+4$) is estimated as 10, and the lambda is estimated as 0.04979. The range parameter, which represents the separation distance over which spatial dependence is apparent, is estimated as 9.891 km, which seems reasonable for our study area. Using these parameter estimates, we approximate land price changes by the cubic B-splines, and apply the FOK technique to predict the function in the centroid of every micro district. The movement in each of the 1,488 centroids is shown in the left-hand side of Figure 5 (each line on the left-hand side of Figure 5 denotes the land-price change index, in which the price in 2002 is set to 1.0). Then, using the FC technique, we categorize these functions into five groups. The names of groups are C1, C2, C3, C4, and C5 (Figure 5)[13]. The right-hand side of Figure 5 indicates that functions with similar movements are successfully classified into the same class/category. For comparison, we also apply the conventional K-means method evaluated at discrete observed years from 2003 to 2010.

[Table 1: Parameter estimation results] around here

[Figure 5: Assigned groups of each function by functional K-means (left) and estimated

---

[13] This value is set by trial and error by considering model fit and ease of interpretation.

centers (means) of each group] around here

The bottom parts of Figures 6 and 7 denote the representative movement (average) of land prices of each five classes/categories. The movements between Figures 6 (K-means) and 7 (functional K-means) are fairly different[14], especially in C2. More specifically, the increase of land prices in C2 by the functional K-means technique is rather mild when the K-means technique is used. Clearly, the conventional K-means technique, based on discrete years, does not capture the overall changing pattern compared to the functional K-means, which is based on the integral of a curve. Hence, the following discussion focuses on the result by the functional K-means technique.

In the functional K-means result, the price index in C1 changes up to around 1.43 in 2009. In category C2, the index goes up to 1.04 in 2008—fairly stable over the period, and we interpret this as the project mitigating the tendency of decreasing land prices, which is a typical characteristic of Japan's collapsed bubble economy, except for Tokyo and some other mega cities. In the categories C3 to C5, land prices decreased consistently over the period, in descending order. However, we notice that the ratio of decrease after the commencement of operations in August 2005 became small for these three categories.

It is interesting to note that in Figures 6 and 7, more districts around Moriya station are categorized in C1 compared to districts around Tsukuba station, which is the terminal station of the TX line (see Figure 2 for station names). These stations are rapid train stops. Because the area around Tsukuba station had no railway connectivity prior to the commencement of operations of the TX line, Tsutsumi and Seya (2009) predicted that large benefit incidence would occur around Tsukuba station[15]. In fact, Moriya station had been there before the start of commencement of the TX line, and thus far, significant onsite commercial development has not been conducted around the station, different from Tsukuba station. However, our FDA analysis results suggest that actually, benefit incidence around Moriya station would be greater in terms of spatial extent, suggesting that people put high value on the reduced time distance to Tokyo[16] (frequency of rail departure from Moriya station to Akihabara station on weekdays, except for peak hours, is 10 departures per hours, whereas that from Tsukuba station is 6 departures per hours). Of course, we cannot deny the possibility that other Moriya station-specific factors may produce such a trend.

Another interesting finding includes that the areas around Mount Tsukuba (see also the DEM in the Figure 2) and Tsukuba station are classified in the same cluster, C2. After the start of the

---

[14] The differences are summarized in Table 2.

[15] Tsutsumi and Seya (2009) evaluated the benefit of the TX line, and applied the spatial hedonic approach (Kim et al., 2003) based on changes of distance to the nearest station and time required to travel by train from the nearest station to central Tokyo.

[16] The numbers of average passengers in 2014 were 17,131 for Tsukuba station and 23,575 for Moriya station.

operation of the TX line, Mount Tsukuba received attention as a new tourism spot, and benefits seems to be generated in that area. This is one of the interesting empirical findings of our study, and is difficult to capture using the distance bands-based DID method.

The spatial hedonic analysis of Tsutsumi and Seya (2009) suggested that the benefits of the TX project would be capitalized into land prices not only along the TX line, but also along the Joso line, because the time required to travel from the nearest station to central Tokyo by train would be reduced. In addition, the authors suggested that positive benefits might not be generated along the Joban line because the time to Tokyo would not be reduced (Figure 8). Then, if we consider categories C1, C2, and C3 in Figure 7 to be the areas of benefit, then we could observe similarities of the spatial distribution pattern of identified areas of benefit by the FDA procedure (Figure 7) with the predicted areas of benefit by the hedonic approach (Figure 8). More specifically, land prices around the Joso line were fairly constant during the target period (a kind of complementary effect), but land prices around the Joban line decreased (a kind of substitution effect). Thus, we observe the redistributive effect (Handy, 2005). Of course, we cannot conclusively state that the effect of the TX line project is purely redistributive. Rather, the impact may be a mix of generative and redistributive effects because although the line is located in a developed country, Japan, the TX line has reduced the travel time from Akihabara to Tsukuba dramatically, as explained. Such a finding can be added to the discussion of transportation impact on land use, of which the knowledge is comparatively limited compared to the opposite direction effects (e.g., Ewing and Cervero, 2001, 2010).

[Figure 6: Visualizing classification results by K-means in the target area] around here

[Figure 7: Visualizing classification results by Functional K-means in the target area] around here

[Table 2: Cross aggregation of K-means and Functional K-means results] around here

[Figure 8: Benefit evaluation results by the hedonic approach in Tsutsumi and Seya (2009) (The figure is newly created for this paper)**] around here**

11

4. Concluding remarks

This study proposed a novel approach for ex-post identification of the geographical extent of areas benefiting from a transportation project, using the functional data analysis methods for a case study of Japan's Tsukuba Express (TX) line. First, a panel of land prices observed before (2002–2004), during (2005), and after (2006–2010) the TX project in the areas that are potentially impacted were prepared. Second, using the FDA method, movements of land prices in each observed site during the target period were approximated by the cubic B-splines function. Third, using the functional ordinary Kriging technique, functions at the center point of each district shown in Figure 3 were spatially predicted (interpolated). Lastly, by employing the FC (functional K-means) technique, the predicted functions at micro districts were classified into five groups. The clustering results by the functional K-means technique was compared to the results produced by the conventional K-means technique, where the time-series nature of data is ignored.

The results with the former method suggest that land prices rose around TX stations with rapid train stops (i.e., Moriya and Tsukuba stations) during 2005 to 2009 (or 2008), and began to decrease from 2009. In addition, the results suggest that land prices around the Joso line were fairly constant during the target period (a kind of complementary effect) but land prices decreased around the Joban line (a kind of substitution effect). Thus, we found a so-called redistributive effect (Handy, 2005). It is interesting to observe that such a spatial distribution pattern was successfully captured by applying the with-and-without hedonic approach (Figures 7 vs. 8).

In future research, it is important to extend our model to include sites where land prices are not observed continuously (and where prices are missing for some years), because some of the observed sites of officially assessed land prices are replaced every year. In addition, because land price trends may differ depending on local attributes, such as land-use zoning, it would be interesting to consider spatial non-stationarity by using a functional Kriging-combined regression (Caballero et al., 2013) or a functional geographically weighted regression (Yamanishi and Tanaka, 2003). Moreover, although our analysis focused on absolute change of land price, relative change to the other sites may be important (Tsutsumi and Seya, 2008).

12

References

- 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.

- Bae, C.H.C., Jun, M.J. and Park, H. (2003) The impact of Seoul's subway Line 5 on residential property values, *Transport Policy*, 10 (2), 85–94.

- Caballero, W., Giraldo, R. and Mateu, J. (2013) A universal kriging approach for spatial functional data, *Stochastic Environmental Research and Risk Assessment*, 27 (7), 1553–1563.

- Cervero, R. and Landis, J. (1995) The transportation-land use connection still matters. Access, 7, 2–11.

- Chiou, J.M., Zhang, Y.C., Chen, W.H. and Chang, C.W. (2014) A functional data approach to missing value imputation and outlier detection for traffic flow data, *Transportmetrica B*, 2 (2), 106–129.

- Delicado, P., Giraldo, R. Comas, C. and Mateu, J. (2010) Statistics for spatial functional data: Some recent contributions, *Environmetrics*, 21, (3-4), 224–239.

- Debrezion, G., Pels, E. and Rietveld, P. (2007) The impact of railway stations on residential and commercial property value: A meta-analysis, *The Journal of Real Estate Finance and Economics*, 35 (2), 161–180.

- Dubé, J., Legros, D., Thériault, M. and Des Rosiers, F. (2014) A spatial difference-in-differences estimator to evaluate the effect of change in public mass transit systems on house prices, *Transportation Research Part B*, 64, 24–40.

- Dubé, J., Thériault, M. and Des Rosiers, F. (2013) Commuter rail accessibility and house values: The case of the Montreal South Shore, Canada, 1992–2009, *Transportation Research Part A*, 54, 49–66.

- Ewing, R. (2008) Highway-induced development: Research results for metropolitan areas, *Transportation Research Record*, 2067, 101–109.

- Ewing, R. and Cervero, R. (2001) Travel and the built environment: A synthesis, *Transportation Research Record*, 1780, 87–114.

- Ewing, R. and Cervero, R. (2010) Travel and the built environment, *Journal of the American Planning Association*, 76(3), 265–294.

- Ferraty, F. and Romain, Y. (2011) *The Oxford Handbook of Functional Data Analysis (Oxford Handbooks in Mathematics)*, Oxford University Press, Oxford.

- Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory And Practice (Springer Series in Statistics)*, Springer, New York.

- 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.

- Giraldo, R., Delicado, P. and Mateu, J. (2011) Ordinary kriging for function-valued spatial data, *Environmental and Ecological Statistics*, 18 (3), 411–426.

- Grimes, A. and Young, C. (2013) Spatial effects of urban rail upgrades, *Journal of Transport Geography*, 30, 1–6.

- Handy, S. (2005) Smart growth and the transportation-land use connection: What does the research tell us?. *International Regional Science Review*, 28 (2), 146–167.

- Jayantha, W.M., Lam, T.I. and Chong, M.L. (2015) The impact of anticipated transport improvement on property prices: A case study in Hong Kong, *Habitat International*, 49, 148–156.

- Kanemoto, Y. (1988) Hedonic prices and the benefits of public projects, *Econometrica*, 56 (4), 981–989.

- Kim, C.W., Phipps, T.T. and Anselin, L. (2003) Measuring the benefits of air quality improvement: A spatial hedonic approach, *Journal of Environmental Economics and Management*, 45 (1), 24–39.

- Knaap, G.J., Ding, C. and Hopkins, L.D. (2001) Do plans matter? The effects of light rail plans on land values in station areas, *Journal of Planning Education and Research*, 21 (1), 32–39.

- Kuminoff, N.V. and Pope, J.C. (2014) Do "Capitalization effects" for public goods reveal the public's willingness to pay?, *International Economic Review*, 55 (4), 1227–1250.

- Lin, J.J. and Hwang, C.H. (2004) Analysis of property prices before and after the opening of the Taipei subway system, *The Annals of Regional Science*, 38(4), 687–704.

- Mayor, K., Lyons, S., Duffy, D. and Tol, R.S. (2012) A hedonic analysis of the value of rail transport in the Greater Dublin Area, *Journal of Transport Economics and Policy*, 46 (2), 239–261.

- McDonald, J.F. and Osuji, C.I. (1995) The effect of anticipated transportation improvement on residential land values, *Regional Science and Urban Economics*, 25 (3), 261–278.

- Medda, F. (2012) Land value capture finance for transport accessibility: A review, *Journal of Transport Geography*, 25, 154–161.

- Mizutani, C. (2012) Construction of an analytical framework for polygon-based land use transition analyses, *Computers, Environment and Urban Systems*, 36(3), 270–280.

- Ministry of Land, Infrastructure, Transport and Tourism (2016) *How to look at Land Market Value Publication*, (http://tochi.mlit.go.jp/english/land-prices/land-market-value-publication)

- 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.

- 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.

- Ramsay, J.O., Hooker, G. and Graves, S. (2009) *Functional Data Analysis with R and Matlab*, Springer, New York.

- Ramsay, J.O. and Silverman, B.W. (2002) *Applied Functional Data Analysis: Method and Case Studies*, Springer, New York.

- Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis, Second Edition*. Springer-Verlag, New York.

- 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.

- Rosen, S. (1974) Hedonic prices and implicit market, product differentiation in pure competition, *Journal of Political Economy*, 82, 34–55.

- Ruppert, D., Wand, M.P. and Carroll, R.J. (2003) *Semiparametric Regression (Cambridge Series in Statistical and Probabilistic Mathematics)*, Cambridge University Press, Cambridge.

- Ruppert, D., Wand, M.P. and Carroll, R.J. (2009) Semiparametric regression during 2003–2007, *Electronic Journal of Statistics*, 3, 1193–1256.

- Sadahiro, Y. and T. Kobayashi (2014) Exploratory analysis of time series data: Detection of partial similarities, clustering, and visualization, *Computers, Environment and Urban Systems*, 45, 24–33.

- Seya, H. and Timmermans, H. (2016) An overview of Asian studies on transport and land use, in Zhang, J. and Feng, C-M. (eds.) *Asian Transport Studies*, Chapter 25, forthcoming, Routledge, London.

- Seya, H., Yamagata, Y. and 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.

- Seya, H., Murakami, D., Tsutsumi, M. and Yamagata, Y. (2015) Application of LASSO to the eigenvector selection problem in eigenvector based spatial filtering, *Geographical Analysis*, 47 (3), 284–299.

- Stokenberga, A. (2014) Does Bus rapid transit influence urban land development and property values: A review of the literature, *Transport Reviews*, 34(3), 276–296.

- 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.

- Tsutsumi, M. and Seya, H. (2009) Hedonic approaches based on spatial econometrics and spatial statistics: Application to evaluation of project benefits, *Journal of Geographical Systems*, 11 (4), 357–380.

- Vessali, K.V. (1996) Land use impacts of rapid transit: A review of the empirical literature,

Berkeley Planning Journal, 11 (1), 71–105.

· 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.

· Zhu, X. and Liu, S. (2004) Analysis of the impact of the MRT system on accessibility in Singapore using an integrated GIS tool, *Journal of Transport Geography*, 12 (2), 89–101.
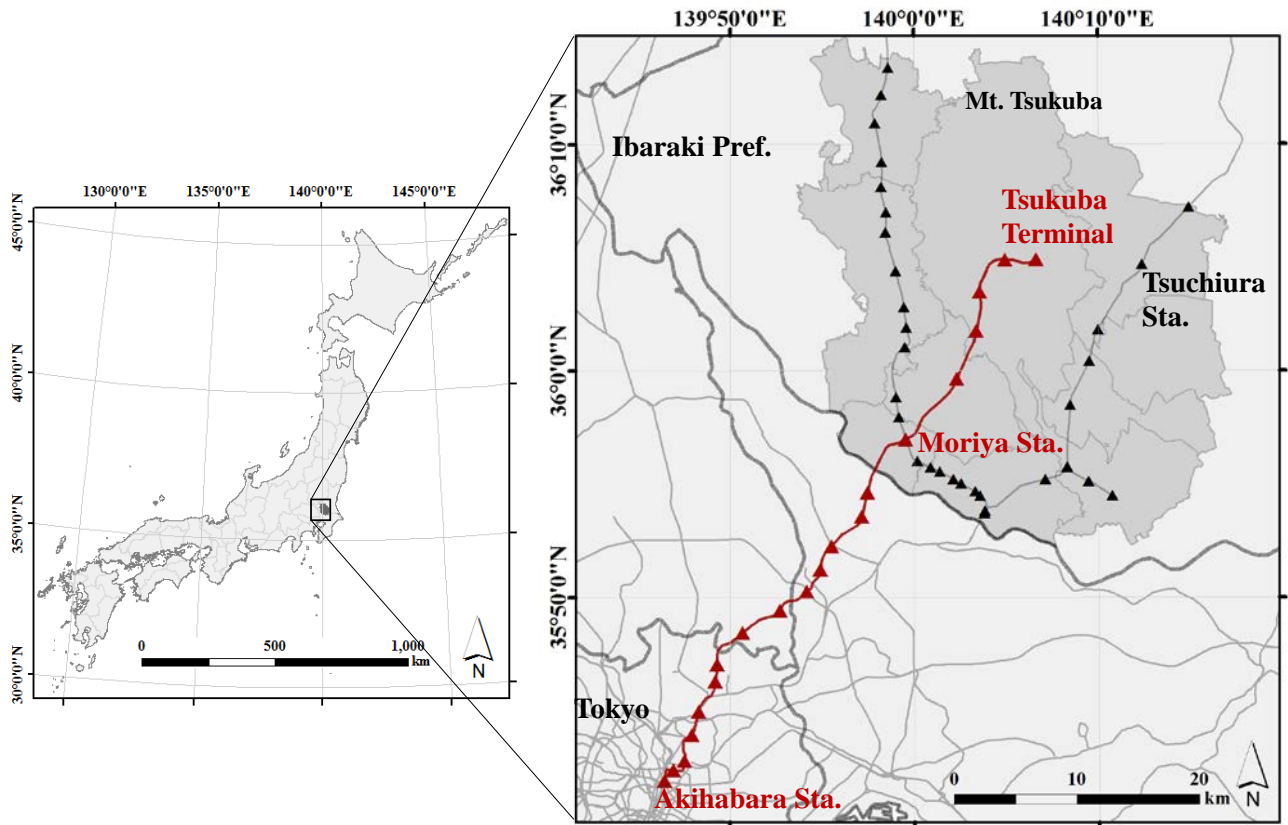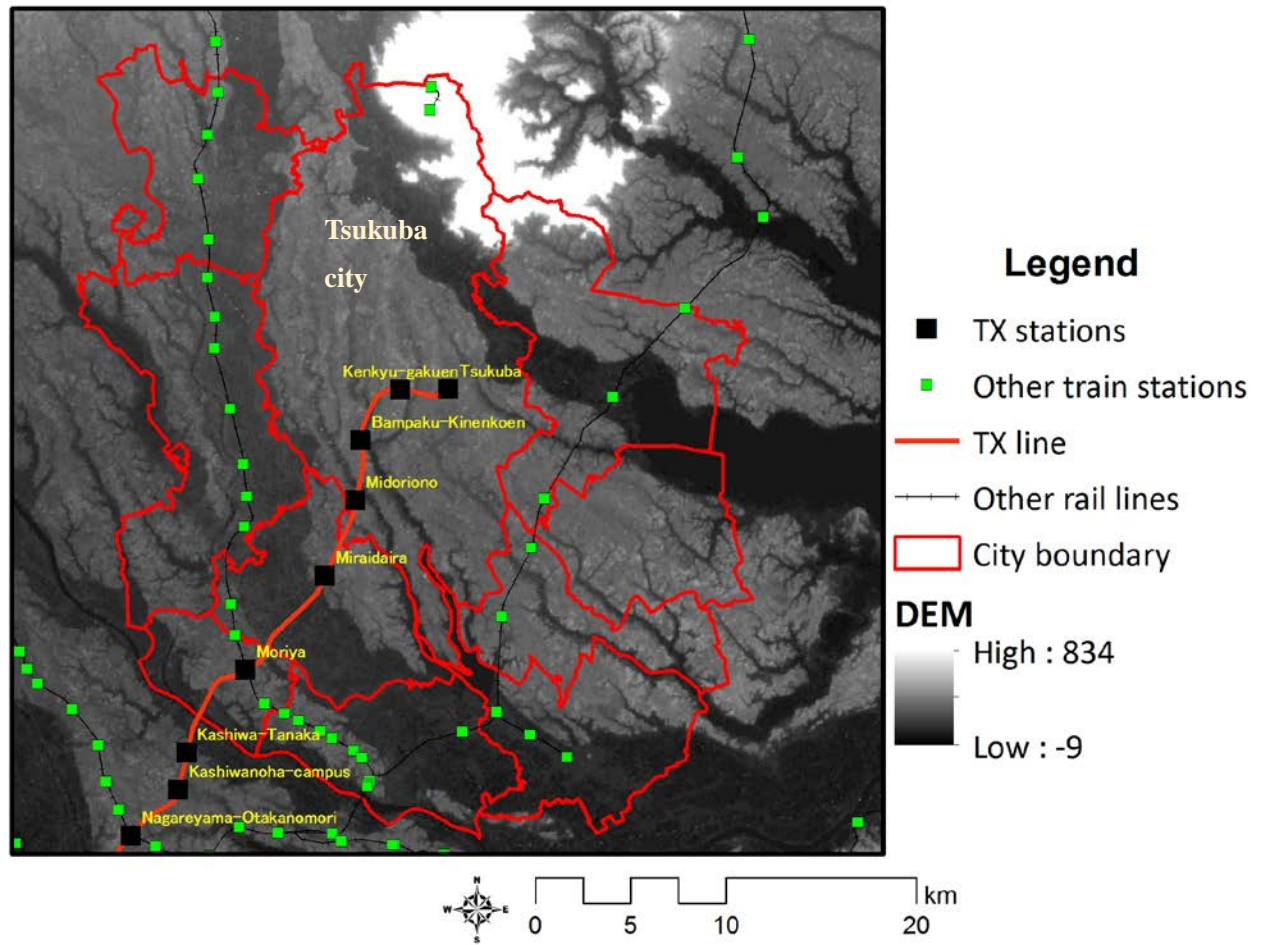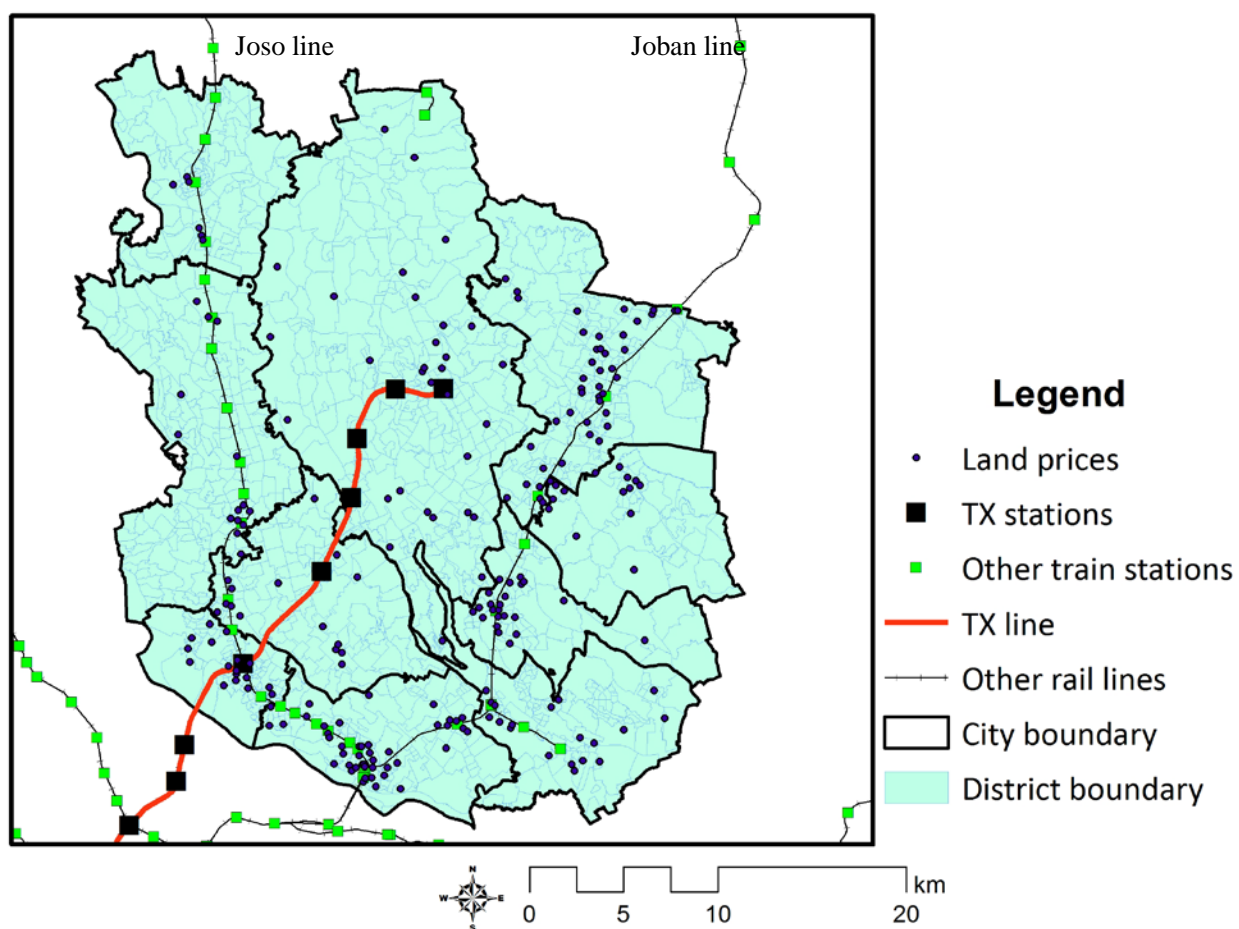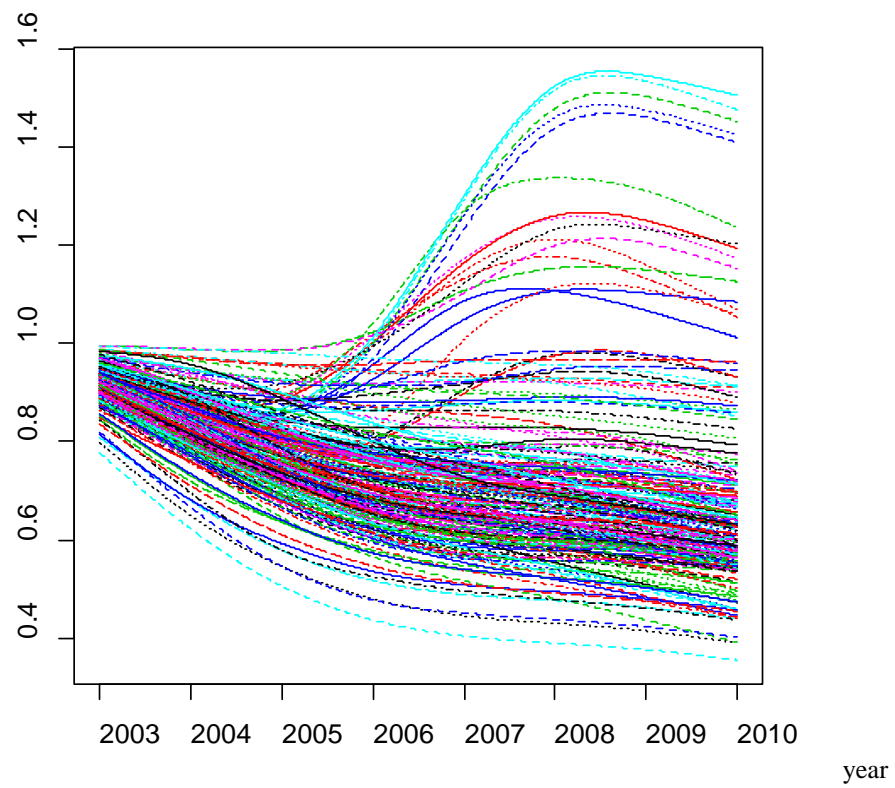
**Figure 1: Location of Tsukuba Express (TX) line**

Figure 2: Station names of TX line in the study area overlaid with digital elevation model (DEM)

Figure 3: Observed sites for officially assessed land price where residential land prices are continuously observed from 2002 to 2010

Figure 4: Changes in land prices at the observed sites in the study area

(Shown as an index relative to the prices in 2002)
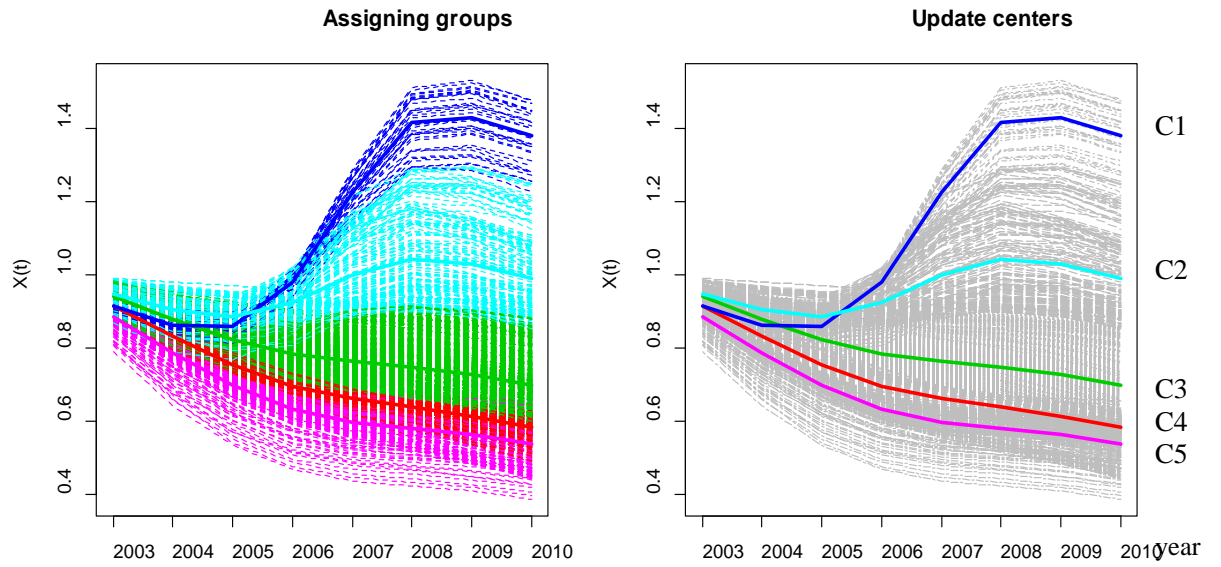
Figure 5: Assigned groups of each function by functional k-means (left) and estimated centers (means) of each group. Each line in the left figure denotes the land price change index (price in 2002 is set to 1.0).
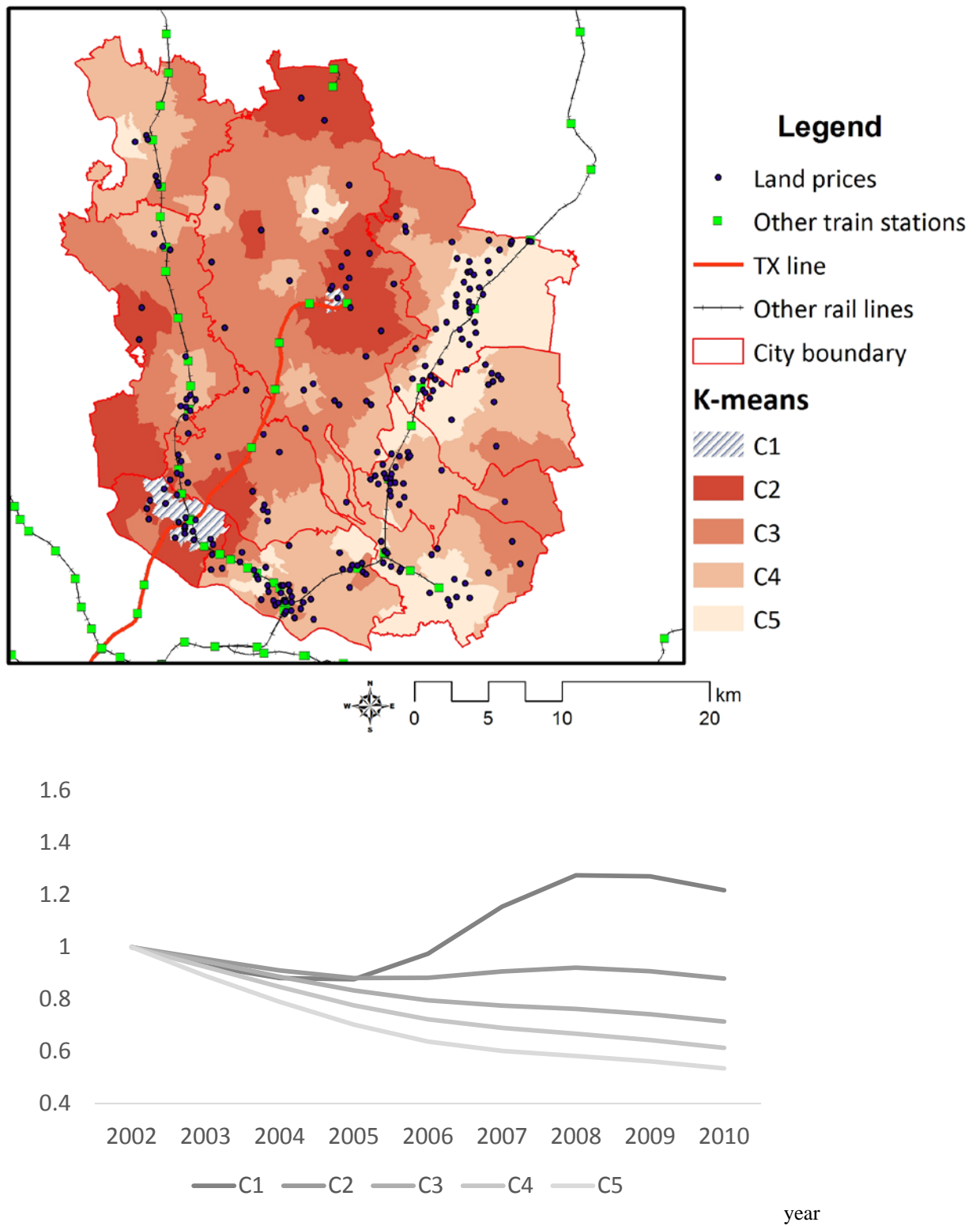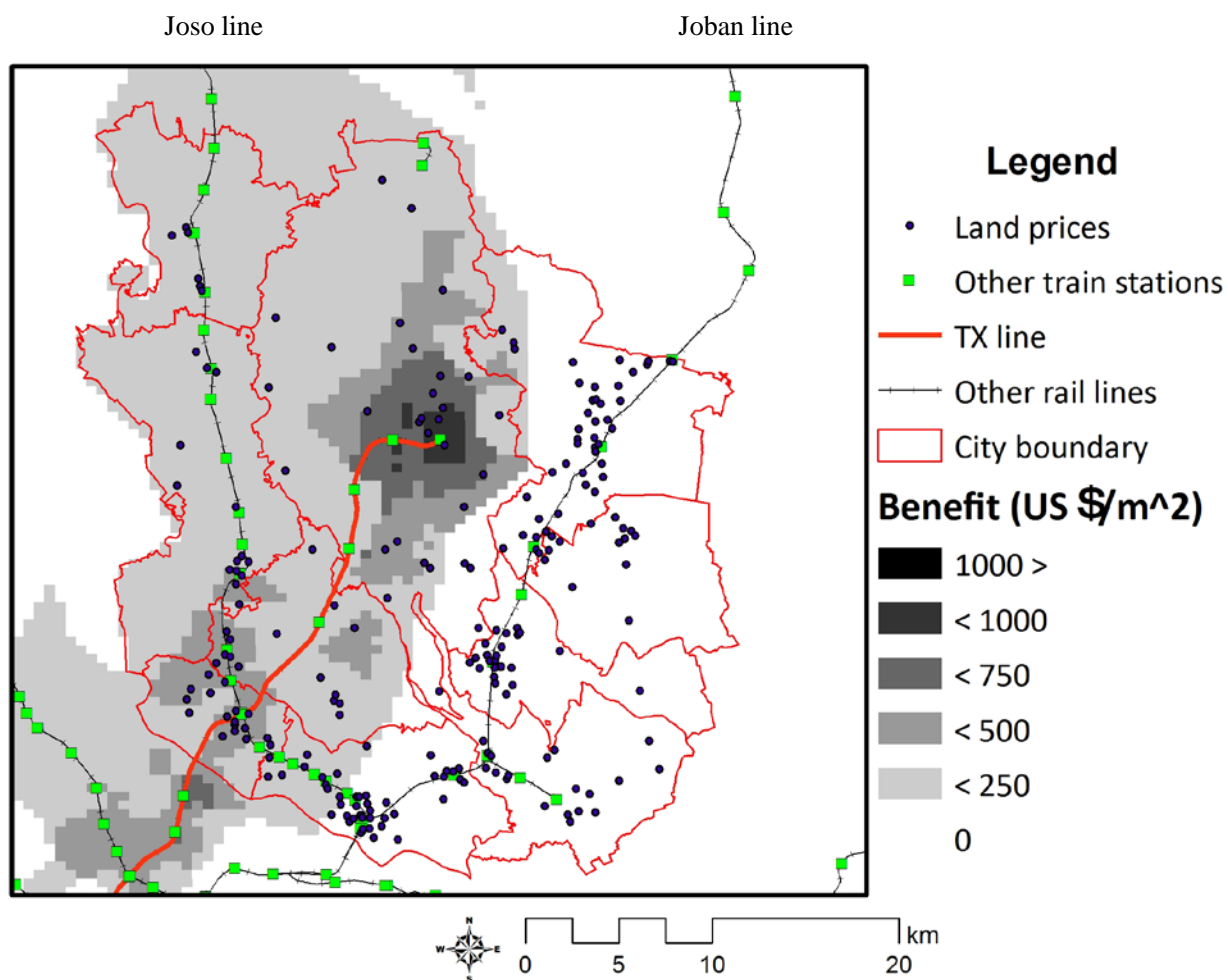
Figure 6: Visualizing classification results by K-means in the target area
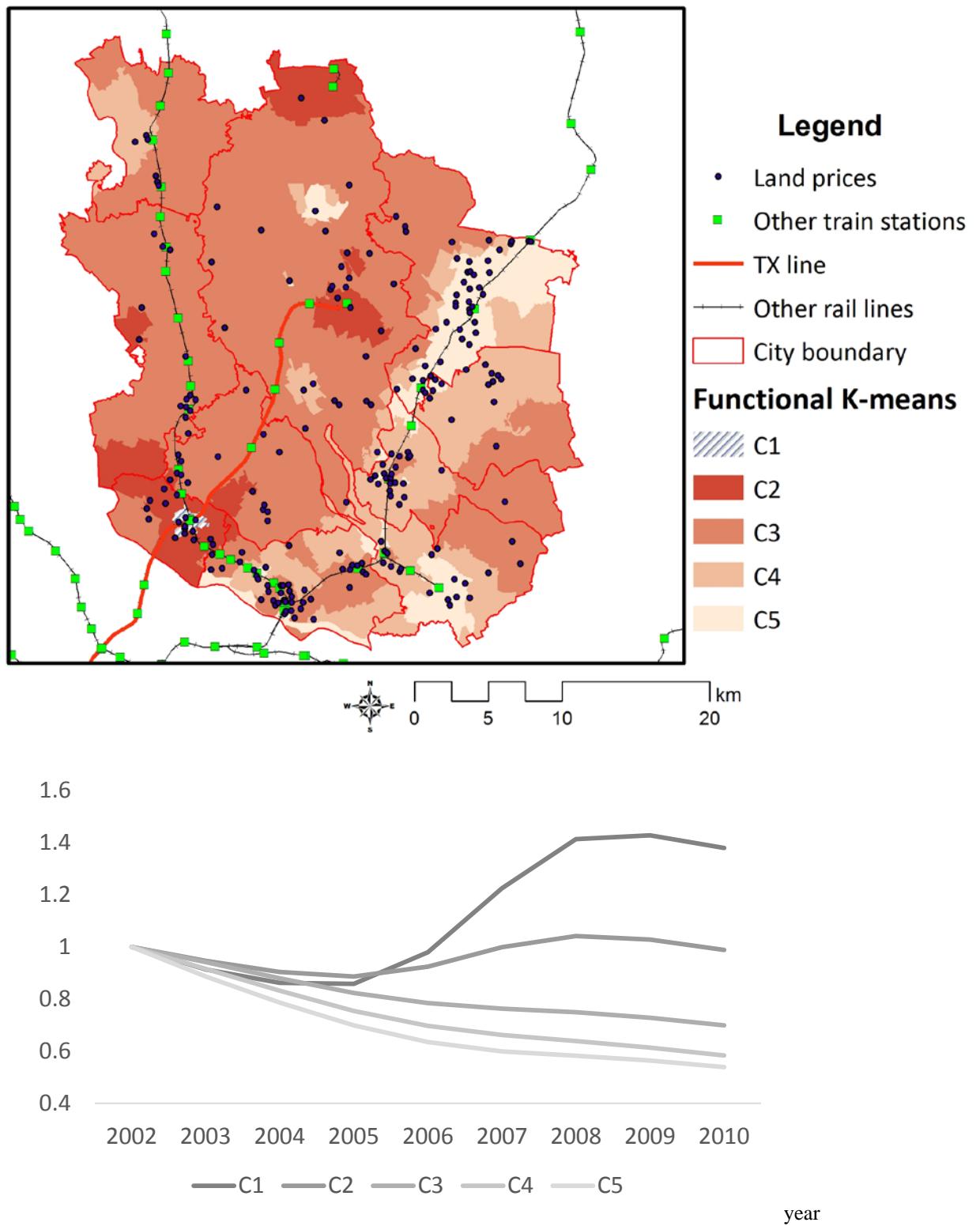
(Bottom figure denotes the estimated center (mean) of each group)

Figure 7: Visualizing classification results by functional K-means in the target area

(Bottom figure denotes the estimated center (mean) of each group)

Joso line

Joban line

Figure 8: Benefit evaluation results by the hedonic approach in Tsutsumi and Seya (2009)

(The figure is newly created for this paper)

Table 1: Parameter estimation result

Nugget	0.06970
Partial-sill	0.3160
Range	9.891
Number of basis	10
Lambda	0.04979

Table 2: Cross aggregation table of K-means and Functional K-means results

| | | Functional K-means | | | | | |
		C1	C2	C3	C4	C5	Total
	C1	35	57				92
	C2		107	100			207
K-means	C3			454			454
	C4			245	213	5	463
	C5				56	216	272
	Total	35	164	799	269	221	1488