Detecting air pollution clusters in Japan: A spatial analysis approach』

研究代表者	Alvaro Dominguez
報告年度	2022-03
URL	http://id.nii.ac.jp/1270/00000258/

調査報告書 21－06

Detecting air pollution clusters in Japan: A spatial analysis approach

令和 4（2022）年 3 月

公益財団法人　アジア成長研究所

Detecting air pollution clusters in Japan: A spatial analysis approach

Alvaro Domínguez

Abstract

We rely on satellite data to study the spatial distribution of air pollutants and economic activity for 1650 municipalities of all four main islands of Japan: Honshu, Kyushu, Hokkaido and Shikoku. Specifically, we analyze atmospheric particulate matter and ozone concentrations, as well as population density, accessibility to cities, and night lights for the above islands. We then make use of principal component analysis, spatial dependence analysis, and methods of regionalization to endogenously classify the municipalities based on their similarity in attributes and geographic location. The spatial dependence analysis provides results which show us the specific sites where the high-value clusters (hot spots) and low-value clusters (cold spots) are located. These show a high positive correlation between economic activity and air pollution. Additionally, we perform a regionalization analysis of the variables under consideration, which specifies how the four main islands can be regionalized into six to nine geographical regions or structures, each. The regionalization takes into consideration both pollution levels and economic activity. We then conclude by discussing how these different analyses can complement each other, and how they contribute in finding the locations where policies related to air quality can help in improving the quality of life of the population.

1 Introduction

The link between the environment and the economy is very important due to the implications it has both on economic growth and in the quality of life of people. Since the mid-twentieth century, with the rapid growth and development of many countries along with their industries, air pollutants became an increasing concern to policy makers. Many works have analyzed the economic consequences of air pollution and ways to deal with it (Carriazo, 2016; Crandall, 1983; Downing and Watson Jr, 1974; Kyriakopoulou et al., 2021; Luechinger, 2010; Wolozin, 1968). Additionally, there has been an increase in the population's awareness that in order to achieve sustainable lifestyles, it is necessary to redesign the socio-ecological regime at a system level (Beddoe et al., 2009).

We rely on spatial dependence analysis, principal component analysis (PCA), and regionalization methods to detect clusters of municipalities in Japan that possess similar air pollution levels and economic activity. To do so, we employ satellite data from *AidData* geoquery database (Goodman et al., 2019) to assess and analyze the spatial distribution, for each of the four main islands, of particulate matter ($PM_{2.5}$), ozone concentrations, population density, night lights, and accessibility to cities. We begin by implementing PCA to reduce the variables into two components. The first component (PC1), mostly sums up variables related to economic activity (population density, night lights and accessibility to cities), with the variables related to pollution (particulate matter concentration and ozone concentration) being included depending on the island under consideration. The second component (PC2) synthesizes the pollution variables. In the case of Kyushu and Shikoku, it includes both pollution variables. For Honshu it summarizes Ozone concentration while for Hokkaido it does so for particulate matter concentration.

We then follow up by applying two different spatial methods to spot geographically contiguous clusters, in each of the islands, for both components. The first one is the spatial dependence method of Anselin (1995) which lets us find so-called regional hot spots (i.e. high-value clusters), cold spots (i.e. low-value clusters), and also spatial outliers. The second one is a regionalization method by Duque et al. (2012), which endogenously derives the regional boundaries (in our case based on pollution levels and economic activity).

We find a positive and statistically significant level of spatial dependence for PC1 and PC2 over many municipalities. From the PC1 data, we observe clusters with higher levels of economic activity in the metropolitan areas of Tokyo, Nagoya and Osaka in the island of Honshu. Some other clusters with these characteristics are also present in the north of Kyushu, and to a lesser degree, to the northwest of Shikoku and Hokkaido. From the PC2 data, we see that there are also large clusters with high levels of pollution to the center and the south of Honshu (which could have been affected due to the changes in energy production after the 2011 earthquake). We also can see the presence of smaller clusters to the south of Kyushu, west of Shikoku, and north of Hokkaido.

The regionalization analysis, resting on the PC1 and PC2 data, implies that the islands conforming Japan can be divided into different geographical regions (ranging from 6 to 9, depending on the island) from similar constraints. The boundaries of these new regions are different from those present in the administrative regions of Japan. The design and control of policies aimed at improving air quality, therefore, would need to be coordinated across the different municipalities.

This article makes three contributions. The first one is that it restricts the detection of clusters to locational similarity among regions. The second one is that the results of the analysis are endogenously obtained, thus the number of clusters do not require to be previously defined. By relying on these methodologies, we differentiate ourselves from other studies of air pollution that make use of non-spatial clustering methods, therefore having to define the number of clusters "manually". Finally, we also contribute through the complementarity of spatial dependence analysis and regionalization methods. By combining both tools, we are able to find clusters that are endogenous, spatially contiguous, and robust. Through the information that these clusters provide, local and national governments may develop policies that improve the quality of life of its citizens, by targeting specifically at the regions that require more urgent action.

The literature on air pollution and its consequences is vast. Different studies explain that air pollution has severe and chronic effects on human health, going from minor upper respiratory irritation to lung cancer, aggravating pre-existing heart and lung disease, chronic bronchitis, and asthmatic attacks (Bernstein et al., 2004; Kampa and Castanas, 2008). An assessment of the costs of inaction of outdoor air pollution by Lanzi et al. (2018) shows that these costs slowly build up, reaching a whopping 1% of global GDP by the year 2060. According to Dechezleprêtre et al. (2019) and based on evidence from Europe, an increase of 1 $\mu g/m^3$ in particle matter concentration ($PM_{2.5}$) may reduce real GDP by 0.8% that same year. Some estimates show that the number of lost working days, which ultimately impact labour productivity, will increase to 3.7 billion US dollars by 2060 (Lanzi et al., 2016).

During the rapid industrialization process of the 1950s and 1960s, pollution levels in Japan increased substantially creating big environmental problems (Shoji and Miyamoto, 1977); at a point about 20% of the population were constantly exposed to air pollutants (Toyama, 1964). In the 1970s and 1980s, the Japanese manufacturing sector's energy intensity fell considerably, and through different measures such as air pollution control agreements and the environmental impact assessment law, pollution levels were substantially reduced (Kanada et al., 2013).

Different studies have been done examining the effects of air quality on the population in Japan. Yorifuji et al. (2015) rely on a nationwide population-based longitudinal survey and find that air pollution exposure during pregnancy increases the risk of low birth weight. Katanoda et al. (2011) study the effects of particulate matter, sulfur dioxide, and nitrogen dioxide on participants living in three different prefectures and find that long term exposure to air pollutants is associated to significant increases in lung cancer and respiratory diseases. On the economic side, Kim et al. (2019) analyze the spatial influence that economic growth has, at a national level, on trade and the emission of particulate matter $PM_{2.5}$ for Japan, China, and South Korea. They show that capital stock accumulation in China produce positive spillover effects in trade and emissions of particulate matter (at both production and consumption levels). Instead, Japan and South Korea reduce their emissions of pollutants through capital accumulation, creating positive shocks on trade.

In recent times, with more readily available spatial data and computational power, spatial approaches to the analysis of air quality have also become increasingly popular. Maantay (2007) examines the spatial correspondence of air pollution and asthma for the case of the Bronx in New York. She finds that individuals living within a certain distance of toxic land uses were significantly more likely to be hospitalized due to asthma, as well as being poor. Kumar et al. (2016) monitor the air quality in the city of Mumbai to analyze the health benefits of policy interventions aimed at reducing air pollution. Through an assessment based on air quality change from concentrations at a starting point and the population in that city, the authors' results indicate that reductions of pollutants bring enormous health benefits.

Kume et al. (2007) employ contour maps and characterize the spatial distribution of the monthly variation of environmental pollutants in Shizuoka, Japan from 2001 to 2002. Shimadera et al. (2009) estimate the contribution that transboundary air pollutants from neighboring Asian countries has on ionic concentrations in fog in the Kinki region in Japan. Their study concludes that transboundary air pollutants significantly affect concentrations in fog in that region. Araki et al. (2015) apply regression-kriging to air pollutants in Japan for 2009 and 2010. The authors observe that this methodology predicts with a high accuracy and spatial resolution the spatial distribution of air pollutants in this country.

Finding regions that are geographically contiguous and that share characteristics (i.e. economics, politics, etc.) can greatly contribute to regional planning. Regionalization as a problem has been widely studied in statistics, machine learning, and geography (Duque et al., 2007; Law and Neira, 2019; Wise et al., 1997). A homogeneous region is defined as a set or group of areas that are spatially contiguous, presenting a high level of similarity in certain characteristics or attributes (Fischer, 1980). In this paper, the characteristics under consideration are particulate matter concentration, ozone concentration, population density, night lights, and accessibility to cities.

Regionalization, which is defined as a procedure to add up or combine geographical areas into regions that are homogeneous has been called in various ways such as regional clustering (Maravalle and Simeone, 1995), conditional clustering (Lefkovitch, 1980), clustering under connectivity constraints (Hansen et al., 2003), and regionalization (Wise et al., 1997). Regional science practitioners use methods of spatial clustering to summarize information, detect the actual number of clusters, and as a way to sketch regions that are appropriate to analyze and monitor (Duque et al., 2007, 2011).

The rest of the paper is organized as follows. Section 2 describes the data we use along with providing descriptive statistics of the variables we analyze. Section 3 explains the methodology we rely on: PCA, spatial dependence, and regionalization (Max-p clusters). Section 4 presents the results obtained from each method. Section 5 discusses how these methods are complementary. Finally, section 6 concludes. For ease of exposition, all the graphs are presented in the appendix.

2 The Data

2.1 Description of the Database

Data on air pollution are from the *AidData* geoquery database (Goodman et al., 2019). From this novel database, we rely on five variables:

- **Particulate matter ($PM_{2.5}$) concentration:** This indicator shows the micrograms (one-millionth of a gram) of gaseous pollutants per cubic meter ($\mu g/m^3$) of ambient air, for 2013. These particles can be made of many different chemicals. The original source of the data comes from *Ambient air pollution exposure estimation for the Global Burden of Disease 2013* (Brauer et al., 2016).

- **Ozone concentration:** This indicator refers to the quantity of ozone molecules that accumulated in the air in 2013. Long term exposure to this gas can cause the development of asthma and other respiratory problems. The original source is the *Ambient air pollution exposure estimation for the Global Burden of Disease 2013* (Brauer et al., 2016).

- **Population density:** This indicator provides an estimate of the number of people per square kilometer in 2015. The original source of this data is the *CIESIN* (Columbia University, 2018).

- **Night lights:** This indicator conveys the lights from cities, towns, and other places with persistent lighting, including gas flares, measured in digital numbers (DNs) for 2013. This is used as an approximation to the economic activity that occurs in a given area or region. The original source of this data is *NOAA National Geophysical Data Center* (Lights, 2017).

- **Accessibility to cities:** This indicator measures the estimated time, in minutes, from a point to the nearest city in 2015. The original source is *The Malaria Atlas Project* (Weiss et al., 2018).

We rely on the indicators described above because:

1. They represent indicators related to the Sustainable Development Goals of the UN (to which Japan adheres). Specifically, particulate matter and ozone concentrations are related to SDG3 to "Ensure healthy lives and promote well-being for all at all ages". In the case of Night lights, it is related to SDG7 to "Ensure access to affordable, reliable, sustainable and modern energy for all" and SDG8 to "Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all". Finally, population density and accessibility to cities are related to SDG11 to "Make cities and human settlements inclusive, safe, resilient and sustainable".

2. These variables capture, without overlap, the concepts we want to study in this manuscript.

2.2 Descriptive Statistics and Maps

Table 1 below presents a summary of the indicators previously mentioned, for each of the main islands. The values are within what we would expect. Nevertheless, it is interesting to note that even though Kyushu has a lower mean population density than Honshu (523.46 vs 1434.03 people per square kilometer), it presents a higher concentration of particulate matter (17.56 $\mu g/m^3$ vs 14.27 $\mu g/m^3$) and has similar levels of ozone concentration. This may be due to pollution incoming from China and/or other nearby countries.[1] We also observe that Hokkaido has lower levels of particulate matter concentration and population density than the other islands.

The importance of night lights in economics rests on its strong correlation to economic activity. We use it as a way to analyze economic activity for different points in the map, with areas with higher economic activity, such as urban areas, presenting brighter lights than those with less activity (i.e. rural areas). For the case of the last variable, accessibility to cities, it quantifies the time it takes to reach the nearest urban center via surface transportation. This data helps understand how travel time differs across urban centers. Urban centers tend to concentrate more infrastructure and resources than rural areas.

Figure 1 gives a first glimpse at the spatial distribution of air pollution in Japan. In particular, figure 1a depicts the distribution of particulate matter concentration, whereas figure 1b does so for ozone concentration. We can observe that, for the case of particulate matter concentration, there are some large concentrations in the Tokyo area (Kantou region, island of Honshu) and in the Fukuoka area (island of Kyushu). Meanwhile, we can see that there are two big clusters of ozone concentration in the Kantou region and also in parts of the Chuubu and Kansai regions, to the center of the island of Honshu. It would appear that municipalities with high values tend to be neighbors with other municipalities with high values, and vice-versa. Nevertheless, we also notice that the clusters we show are not contiguous. An additional constraint is that the number of clusters shown is exogeneous, so it has to be decided a priori. Due to this, we later show the results of endogenous clusters that are also contiguous.

3 Methodology

3.1 Principal Component Analysis

The method of Principal Component Analysis (PCA) dates back to the work by Pearson (1901) and Hotelling (1933). In more recent times, with the advent of higher computing power from computers, it has become more popular. The purpose of using PCA is to maintain variability for a reduced number of linear combinations that serve as linear functions of the variables under study, helping to summarize the data (Jolliffe and Cadima, 2016). The new variables that are obtained do not correlate with each other and are expected to maximize variance. In order to obtain these new variables, the PCA method works around decomposing eigenvalues of a system.

Consider a matrix $\mathbf{X}$ made up of k variables, with n observations each. To circumvent possible differences in scale in each variable, we standardize each of them. We define $\mathbf{X}^T\mathbf{X}$ as a $k \times k$ cross-product correlation matrix. Through the PCA method we reduce the number of variables relying on the principal components, which explain a significant share of the variance

[1] Yoshino et al. (2016) explain that for the case of the city of Fukuoka in Kyushu, trans-boundary air pollution occurs not only during the winter-spring season but also in summer, thus significantly impacting the quality of air.

Table 1: Descriptive statistics for Honshu, Hokkaido, Kyushu, and Shikoku

(a) Honshu

Statistic	Mean	St. Dev.	Min	Q1	Median	Q3	Max	Obs.
Particulate Matter Concentration ($\mu g/m^3$ $PM_{2.5}$, 2013)	14.27	3.85	0	11.63	13.82	16.55	34.95	1179
Ozone concentration ($\mu g/m^3$, 2013)	60.68	4.10	0	59.98	61.24	62.37	63.88	1179
Population density (number of people, 2015)	1434.03	3121.68	2.50	93.36	298.21	1165.51	41654.99	1179
Night lights (DNs, 2013)	22.50	15.65	0	8.42	18.37	38.43	47	1179
Accessibility to cities (distance in minutes, 2015)	15.15	16.53	0	1.61	9.96	23.83	86.99	1179

(b) Hokkaido

Statistic	Mean	St. Dev.	Min	Q1	Median	Q3	Max	Obs.
Particulate Matter Concentration ($\mu g/m^3$ $PM_{2.5}$, 2013)	8.76	2.01	5.16	7.24	8.52	9.74	17.45	175
Ozone concentration ($\mu g/m^3$, 2013)	49.74	1.93	44.35	48.72	50.03	50.81	53.89	175
Population density (number of people, 2015)	70.42	170.49	2.61	11.20	20.45	54.45	1617.71	175
Night lights (DNs, 2013)	9.55	7.44	0.90	4.08	7.44	11.92	38.91	175
Accessibility to cities (distance in minutes, 2015)	49.46	29.88	2.35	28.20	43.06	66.26	142.26	175

(c) Kyushu

Statistic	Mean	St. Dev.	Min	Q1	Median	Q3	Max	Obs.
Particulate Matter Concentration ($\mu g/m^3$ $PM_{2.5}$, 2013)	17.56	3.28	10.95	14.96	17.30	17.33	36.51	205
Ozone concentration ($\mu g/m^3$, 2013)	59.92	0.58	58.01	59.69	59.85	60.02	62.35	205
Population density (number of people, 2015)	523.46	790.91	6.41	112.23	239.21	645.01	6674.94	205
Night lights (DNs, 2013)	17.34	12.55	0.08	7.11	17.34	26.92	46.68	205
Accessibility to cities (distance in minutes, 2015)	15.25	12.73	0	4.17	11.90	23.55	51.88	205

(d) Shikoku

Statistic	Mean	St. Dev.	Min	Q1	Median	Q3	Max	Obs.
Particulate Matter Concentration ($\mu g/m^3$ $PM_{2.5}$, 2013)	13.44	2.75	9.96	11.95	13.00	14.54	30.76	91
Ozone concentration ($\mu g/m^3$, 2013)	58.80	0.63	57.45	58.33	59.05	59.23	59.54	91
Population density (number of people, 2015)	324.71	455.42	7.55	53.45	116.48	371.71	2183.96	91
Night lights (DNs, 2013)	12.82	12.33	0	3.14	9.35	18.41	44.92	91
Accessibility to cities (distance in minutes, 2015)	20.81	19.73	0	5.04	16.16	30.51	80.16	91

in the original variables. The idea is that for every principal component y_j, the method finds the coefficients a_j for $j = 1, 2, ..., k$ so that:

$$y_j = a_1 x_1 + a_2 x_2 + ... + a_k x_k, \tag{1}$$

where y_j is a linear combination of the original variables $x_1, x_2, ..., x_k$.

For this work, PCA is used on five variables: particulate matter concentration, ozone concentration, population density, night lights, and accessibility to cities. The idea is to obtain a more reduced set of variables (the principal components) that sum up a large part of the variance of these five variables related to pollution levels and economic activity.

3.2 Spatial Dependence

In order to combine the concepts of attribute similarity with locational similarity, we use spatial dependence analysis. More specifically, a test for global dependence checks whether we can find a pattern of clusters in the spatial distribution of a given attribute. The null hypothesis for such a test consists in showing if there is randomness or not in the spatial location under study (i.e. the regions under study are independent of each other and their location provides no relevant information). Rejection of the null hypothesis, thus, proposes the presence of structures or clusters that supply relevant information for our study. The most popular test for global spatial dependence is Moran's I (Cliff and Ord, 1981). We define this test in the following way:

$$I = \sum_i \sum_j w_{ij} (x_i - \mu) (x_j - \mu) / \sum_i (x_i - \mu)^2 \tag{2}$$

where w_{ij} represents the row-standardized element of the weighted matrix, summarizing the spatial structure of the data under study. The variable x_i denotes the level of pollution/economic activity in municipality i, x_j is the same but for municipality j, and μ shows the average level of the level of pollution/economic activity.

We briefly explain the concept of spatial weights, w_{ij}, due to its importance in spatial analysis. The spatial weights are represented through a weights matrix W, summarizing the spatial structure. Positive values in this matrix (i.e. $w_{ij} > 0$) denote a relationship between neighbors in a geographical area, while zero-values reflect a lack of it. Different ways exist in which these weights can be specified. One of these, referred to as Queen contiguity, considers two regions to be neighbors if they share a common border or vertex. In another, the Rook contiguity, regions are deemed neighbors when they share a common border (although this type of structure does not consider vertices, thus being a subset of Queen contiguity). Other neighbor structures can be defined based on distance thresholds, k-nearest neighbors, and inverse distances. In this paper, we make use of a Queen contiguity structure because it is easy to implement and interpret the results derived from it.

The Moran scatter plot was proposed by Anselin (1995) as a way to visually analyze spatial dependence. This plot displays the relationship between the variable x under study, and the spatially lagged variable W_x. In other words, the Moran scatter plot represents the relationship between the attribute of a location x and the weighted average, W_x, of this location's neighbors. The Moran's I statistic depicts the slope of the fitted line of the two variables. The Moran scatter plot provides a convenient way to classify spatial dependence. Positive spatial autocorrelation is indicated through a positive slope in the graph, and it implies the existence of a cluster in which the values taken in a particular place are surrounded by neighbors with similar values. If the graph shows a negative slope, it signifies the presence of negative spatial autocorrelation in which spatial outliers are ruling. In this case, the values in a particular location are surrounded by values of the neighbors with reversed signs.

Anselin (1995) also proposed other methods of spatial association, through local indicators of spatial association (LISA). The local Moran statistic evaluates local spatial patterns through so called "hot spots" (displaying relatively high patterns), "cold spots" (relatively low values) and spatial outliers (high values surrounded by low values or vice-versa).[2]. The local Moran's I is defined as:

$$I_i = \frac{(x_i - \mu)}{\sum (x_i - \mu)^2} \sum_j w_{ij} (x_j - \mu) \tag{3}$$

where the variables and the notation are those of Moran's I, presented in equation 2.

3.3 The Max-p Method for Regionalization

The Max-p algorithm is used to detect clusters that are spatially constrained, and rests on a mixed integer programming model. Duque et al. (2012) developed it as the solution to the constrained optimization problem of the form:

$$Min\ Z = \left(- \sum_{k=1}^{n} \sum_{i=1}^{n} x_i^{k0} \right) * 10^h + \sum_i \sum_{j|j>i} d_{ij} t_{ij}, \tag{4}$$

Subject to:

[2]Relying on local spatial dependence can complement the analysis performed with the global one. This is due to the latter identifying whether clustering patterns exist or not, while the local analysis detects the exact location of these clusters and spatial outliers

$$\sum_{i=1}^{n} x_i^{k0} \leq 1 \quad \forall k = 1, \ldots, n \tag{5}$$

$$\sum_{k=1}^{n} \sum_{c=0}^{q} x_i^{kc} = 1 \quad \forall i = 1, \ldots, n \tag{6}$$

$$x_i^{kc} \leq \sum_{j \in N_i} x_j^{k(c-1)} \quad \forall i = 1, \ldots, n; \forall k = 1, \ldots, n; \forall c = 1, \ldots, q \tag{7}$$

$$\sum_{i=1}^{n} \sum_{c=0}^{q} x_i^{kc} l_i \geq \text{ threshold } * \sum_{i=1}^{n} x_i^{k0} \quad \forall k = 1, \ldots, n \tag{8}$$

$$t_{ij} \geq \sum_{c=0}^{q} x_i^{kc} + \sum_{c=0}^{q} x_j^{kc} - 1 \quad \forall i, j = 1, \ldots, n \mid i < j; \forall k = 1, \ldots, n \tag{9}$$

$$x_i^{kc} \in \{0, 1\} \quad \forall i = 1, \ldots, n; \forall k = 1, \ldots, n; \forall c = 0, \ldots, q \tag{10}$$

$$t_{ij} \in \{0, 1\} \quad \forall i, j = 1, \ldots, n \mid i < j \tag{11}$$

Where the decision variables are:

$$t_{ij} = \begin{cases} 1, & \text{if areas } i \text{ and } j \text{ belong to the same region } k, \text{ with } i < j \\ 0, & \text{otherwise} \end{cases}$$

$$x_i^{kc} = \begin{cases} 1, & \text{if areas } i \text{ is assigned to region } k \text{ in order } c \\ 0, & \text{otherwise} \end{cases}$$

The parameters of the problem are:

$i, I =$ Index and set of areas, $I = \{1, \ldots, n\}$
$k =$ index of potential regions, $k = \{1, \ldots, n\}$
$c =$ index of contiguity order, $c = \{0, \ldots, q\}$, with $q = (n-1)$

$$w_{ij} = \begin{cases} 1, \text{ if areas } i \text{ and } j \text{ share a border, with } i, j \in I \text{ and } i \neq j \\ 0, \text{ otherwise} \end{cases}$$

$N_i = \left\{ j \mid w_{ij} = 1 \right\}$, the set of areas that are adjacent to area i
$d_{ij} =$ dissimilarity relationships between areas i and j, with $i, j \in I$ and $i < j$

$h = 1 + \left\lfloor \log \left(\sum_i \sum_{j|j>i} d_{ij} \right) \right\rfloor$, which is the number of digits of the floor function of $\sum_i \sum_{j|j>i} d_{ij}$, with $i, j \in I$

$l_i =$ spatially extensive attribute value of area i, with $i \in I$
threshold $=$ minimum value for attribute l at regional scale.

We now proceed to explain the equations that make up this method. Equation 4 represents the objective function, constituted by two terms. The first one defines the number of regions through an aggregation of the number of areas chosen as the root areas. The second term establishes the heterogeneity by adding the areas of a region based on their pairwise dissimilarities. Equation 5 bounds an aggregated region to having no more than a single core area. Equation 6 says that each area will only by allocated to a single region k and a single contiguity order c. Equation 7 tells us that area i is assigned to region k in order c if an area j exists and is allotted to that region k in order c. Equation 8 explains that when a new area is defined, it is done based on a threshold from a spatially intensive attribute. For this work, we set it as 10 percent of the population. Equation 9 expresses that total heterogeneity is obtained from pairwise dissimilarities. The last two equations, equations 10 and 11, point that the integrity of the variables should be maintained.

4 Results

4.1 Principal Component Analysis

Table 2 below summarizes the results of the principal component analysis. This table is separated into a) and b), summarizing two islands each, and is composed of three parts: the first part shows the proportion of total variance obtained from the principal components. PC1 (the first component) and PC2 (the second component) explain the variance for each island. In the case of Honshu, PC1 explains the 55 percent and PC2 the 21 percent. In the case of Hokkaido, PC1 and PC2 explain the 48 and the 21 percent respectively. For Kyushu, PC1 is 50 percent and PC2 is 23 percent. Finally, Shikoku's PC1 represents 49 percent and PC2, 26 percent. When considered cumulatively, these two components represent more than three quarters of the total variance (except for Hokkaido where it is above two thirds).

In the second part, we observe the squared correlations of the components with the original variables. This shows the size of the correlation, helping us understand the components in function of the original variables. As an example, the variables that chiefly compose PC1 for Honshu are Particulate matter, Population density, Night lights, and Accessibility to cities. For the case of PC2, Ozone concentration is the main variable of interest. Regarding the criterion to choose the number of components, we rely on that of Kaiser (1960). This criterion suggests utilizing only those eigenvalues that exceed a value of one. In our case, this would imply using two components, which explain a percentage higher than 70 percent of total variance for each island (once again, Hokkaido is a little bit lower, at 69 percent).

After reducing the variables into a smaller number in PC1 and PC2, which summarize a large proportion of the variance, we can now study their spatial distribution. We proceed by finding the spatial dependence and the patterns of regionalization for each of these two components. With this, we can encounter clusters that are characterized by being spatially contiguous and that face similar characteristics or attributes.

4.2 Spatial Dependence

In figure 2 and figure 3 below, we show the plots of the local Moran scatter plot for autocorrelation. In these plots, the X axis presents the standardized value of PC1 and PC2, respectively. The Y axis shows the weighted mean of the neighbors (i.e the spatial lag). The regression line in the graphs explains the degree of spatial dependence in the region. The Moran scatter plot consists of four quadrants: Top-right, top-left, bottom-right, and bottom-left. The top-right and

Table 2: Principal Component Analysis

a) Total variance and cumulative proportion.										
	Honshu					Hokkaido				
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Proportion of variance	0.55	0.21	0.13	0.08	0.04	0.48	0.21	0.17	0.10	0.04
Cumulative proportion	0.55	0.76	0.89	0.96	1.00	0.48	0.69	0.86	0.96	1.00
Squared correlations between the components and the variables.										
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Particulate matter ($PM_{2.5}$) concentration	**0.64**	0.11	0.02	0.23	0.00	0.10	**0.58**	0.30	0.01	0.01
Ozone concentration	0.20	**0.71**	0.00	0.09	0.00	**0.36**	0.25	0.25	0.12	0.01
Population density	**0.49**	0.08	0.37	0.05	0.01	**0.49**	0.18	0.12	0.19	0.02
Night lights	**0.79**	0.08	0.02	0.00	0.11	**0.76**	0.05	0.05	0.06	0.08
Accessibility to cities	**0.63**	0.06	0.24	0.01	0.06	**0.68**	0.01	0.15	0.11	0.06
Criterion to choose of number of components										
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Eigenvalues	2.75	1.03	0.65	0.38	0.19	2.40	1.06	0.87	0.49	0.18
Kaiser Criterion	2					2				

b) Total variance and cumulative proportion.										
	Kyushu					Shikoku				
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Proportion of variance	0.50	0.23	0.15	0.09	0.03	0.49	0.26	0.14	0.10	0.01
Cumulative proportion	0.50	0.74	0.88	0.97	1.00	0.49	0.75	0.89	0.99	1.00
Squared correlations between the components and the variables.										
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Particulate matter ($PM_{2.5}$) concentration	0.16	**0.42**	0.42	0.00	0.00	0.00	**0.67**	0.30	0.02	0.00
Ozone concentration	0.08	**0.60**	0.29	0.02	0.00	0.04	**0.59**	0.36	0.00	0.00
Population density	**0.64**	0.10	0.00	0.24	0.02	**0.84**	0.01	0.00	0.12	0.03
Night lights	**0.88**	0.04	0.00	0.01	0.08	**0.93**	0.00	0.00	0.02	0.04
Accessibility to cities	**0.75**	0.00	0.03	0.18	0.03	**0.64**	0.00	0.04	0.32	0.00
Criterion to choose of number of components										
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**
Eigenvalues	2.51	1.17	0.74	0.45	0.13	2.46	1.28	0.71	0.49	0.07
Kaiser Criterion	2					2				

bottom-left parts allow us to single-out spatial clusters. These represent the cases where a municipality and its neighbors have similar high/low values, respectively, in PC1 and PC2. For the case of the top-left and bottom-right sections of the graph, these show the spatial outliers and their geographical position. These outliers are municipalities with high values in PC1 and PC2, while its neighbors' values are low, or vice-versa. The dots that are highlighted in figures 2a-2d as well as in figures 3a-3d show the observations that are statistically significant. We consider municipalities that are significant at a p-value of 0.01. By relying on a lower significance level, we can minimize any problems related to the multiple comparison problem. Anselin (1995) explains that this is a common problem when performing analysis of local indicators of spatial association (LISA).

Figures 4a-4d, (for PC1) and 5a-5d (for PC2) present the spatial distribution of municipalities that are statistically significant. Referring to the four sections in the Moran scatter plot, the municipalities are organized into hot spots (high-high) clusters, cold spots (low-low) clusters, and the remaining consist of spatial outliers. The idea is that a spatial cluster consists of a core group (the municipalities) and their neighbors. In figures 4a-4d and 5a-5d, the cold spots and hot spots depict the cores of the spatial clusters. Meanwhile, we can see that in figures 6a-6d and 7a-7d, the spatial clusters are composed of the core and its neighbors, for PC1 and PC2 respectively.

Figures 2, 4, and 6 show a positive and statistically significant amount of spatial dependence. For the case of Honshu, the level of particulate matter concentration, population density, night lights, as well as accessibility to cities (stipulated by PC1) encountered by the average municipality, is of a highly similar nature as that of its geographic neighbors. For the municipalities in Hokkaido, we observe that ozone concentration, population density, night lights and accessibility to cities are of a highly similar nature. Finally, Kyushu and Shikoku have a matching pattern: in PC1 only population density, night lights and accessibility to cities have a similar nature for the municipalities.

In these graphs, we can see the global Moran's I, which encapsulates this relationship. This coefficient's value is 0.860 for Honshu, 0.707 for Hokkaido, 0.765 for Kyushu and 0.683 for Shikoku respectively. In figure 6a we find clusters of hot spots located in the Tokyo, Nagoya, and Osaka areas, with the cores consisting of 162 municipalities. These hot spots consist of high levels of particulate matter concentration (19.68 $\mu g/m^3$),[3] population density (6593.28 people per square km), night lights (45.60 DNs), and accessibility to cities (0.28 minutes).[4] The cold spots are composed of 138 municipalities, with a large concentration of these towards the north of the Tohoku region. They are characterised by lower levels of particulate matter (10.10 $\mu g/m^3$), population density (76.17 people per square km), night lights (6.79 DNs), and a higher distance to nearby cities (42.18 minutes). Figure 6b shows the same, but for Hokkaido. We see that the hot spots are located towards the northeast and west, whereas the cold spots concentrate to the mid west. Figures 6c and 6d present the same, but for Kyushu and Shikoku respectively. The hot spots in Kyushu are located to the north, which include the cities of Kitakyushu and Fukuoka. Meanwhile, the cold spots are situated to the southwest.

Figures 3, 5, and 7 exhibit the spatial dependence for PC2. We find a positive and statistically significant amount of spatial dependence, with the slope of this dependency being 0.288 for Honshu, 0.596 for Hokkaido, 0.812 for Kyushu, and 0.584 for Shikoku. The PC2 hot spots for Honshu are located to the south (in parts of the Kansai region), center-north (parts of the Chuubu region), and to the east (encompassing parts of the Kantou and Touhoku regions), comprising 137 municipalities. The cold spots, instead, only consist of 6 municipalities fairly distanced from one another. For the case of Hokkaido, small pockets of hot spots are located to the north and center, whereas the cold spots can be found mostly to the south. It is of interest to mention that for both Kyushu and Shikoku, the main variables of interest in PC2 are concentrations of particulate matter and of ozone. There are only 15 municipalities in Kyushu and 10 in Shikoku classified as hot spots. These municipalities are characterised by higher levels of particle matter concentrations (14.32 $\mu g/m^3$ for Kyushu and 15.75 $\mu g/m^3$ for Shikoku) and ozone concentration (61.42 $\mu g/m^3$ for Kyushu and 57.61 $\mu g/m^3$ for Shikoku).

The Moran scatter plot supplies us with a way to recognize clusters that are bi-dimensional. In these clusters, the first dimension points out to the values of the principal components (economic activity and air pollution) while the second dimension explains the spatial contiguity (geographic proximity) of the municipalities. We would then be inclined to ask ourselves how we can classify the grey areas in the maps (figures 4 and 5), which are not statistically-significant. To do so, in the next section we rely on the Max-p clustering algorithm proposed by Duque et al. (2012)

[3]The numbers shown in parenthesis specify the mean value for the municipalities that are classified as hot spots. These values are in their original scale, as in the case of table 1.

[4]It is important to note that the lower this number is, the higher the accessibility is.

4.3 Regionalization through the Max-p algorithm

In figure 8 we can observe the regional division of Japan. These administrative regions are classified based on natural and historical reasons. At lower levels, Japan is also divided into 47 prefectures and into municipal levels. Simultaneously, different administrative areas tend to differ in the geographical scope of the economic activities being performed. Therefore, if we don't correctly take into consideration spatial autocorrelations, we may overestimate the effects of urban agglomeration (Otsuka, 2021).

Through the Max-p algorithm, we are able to cluster different regions into analytical regions based on economic activity/pollution levels. Contrary to the traditional regional divisions, which only provide information on locational similarity, by dividing each of the four main islands by analytical regions we obtain important information on both locational and attribute similarity. On the other hand, we see that the municipalities in figure 8 only have a similar location (i.e. locational similarity), but they do not necessarily share a common level of pollution or economic activity (i.e. attribute similarity). The analytical regions shown in figures 9 and 10 allow us to single out clusters of spatially contiguous municipalities that: (a) have similar levels of pollution/economic activity, and (b) to an extent have spatial heterogeneity in each administrative region. We can appreciate this in the region of Chuubu in figure 8. When we consider the economic activity and pollution, we detect four different clusters within its administrative boundaries (figure 9a and 10a).

The major findings from the regionalization process can be broken into two. In the first one, the economic activity and the pollution levels (expressed through PC1 and PC2 respectively) span various administrative regions. This implies that there is a necessity to coordinate throughout all these regions in order to successfully tackle this issue. The second one concerns the fact that some regions are spatially more uneven and heterogeneous than others. As an example, for PC1, Hokkaido and Kyushu include six and nine clusters respectively, while Touhoku only includes two clusters.

Through a deeper understanding of how multiple variables interact spatially, we can better communicate to policy-makers and any interested party how to deal with these issues. Authors such as Gaspar et al. (2019) explain that low and middle income countries may have to spend an additional 4 percent of GDP each year in order to attain their SDG goals. A country such as Japan, with more resources and technology at its disposal, is thus better equipped to achieve such goals. By better grasping how pollution is spatially distributed throughout the various regions, policy-makers may develop and target more effectively different policies.

5 Discussion

From the two previous sections, we can observe that both the local Moran and the Max-p clustering method may be used as complementary analyses. First we may rely on the local Moran analysis to detect hot spot and cold spot clusters. After that, the Max-p method can help in classifying into clusters the municipalities that were left. Relying on these methods does not necessarily imply that there will be a one to one superposition of clusters. Nevertheless, performing such an analysis can facilitate the detection of spatial clusters that are robust and worthy of deeper studies by the academic community. These clusters also can serve local, regional, and national governments in the development of policies prioritizing areas where improving air pollution is more urgent.[5]

[5]Policies that are especially desirable are those geared towards municipalities where activities that generate air pollution strengthens or reinforces that of their neighbors and vice-versa.

There are two points to consider when contrasting the spatial clusters produced by these methodologies. The first one is that the local Moran method only considers the observations from the high-high and low-low quadrants, whereas the Max-p algorithm uses all of the observations. The second one is that the size of the clusters obtained through the Max-p algorithm is usually bigger. The definition of spatial contiguity that the local Moran analysis considers is that of first order neighbors (i.e. the ones that are more proximate or direct neighbors). Instead, the analysis performed by the Max-p algorithm can also be of higher order. Therefore, these methodologies work well as complements and not as substitutes. With the variation of spatial contiguity and observations, both of these methodologies can help us appreciate different things through the respective information they provide.

Regarding policies to improve air quality, those aimed at generating incentives for populations to move into less populated areas can help in the reduction of concentrations in certain areas with higher pollution levels. Additionally, even though air pollution mostly has consequences at the local and regional level, many pollutants and small particles such as $PM_{2.5}$ can be carried across large distances by wind, affecting other countries and regions. Because of this, it is important to coordinate with other countries and regions to find solutions that tackle this effectively. A global transformation of the energy system, along with less polluting and more environmentally friendly technologies is a fundamental part of any cost-effective policy response.

6 Conclusion

In this article we detect clusters of regions in Japan with similar levels of pollution and of economic activity. To do so, we rely on principal component analysis and different methods of spatial data analysis. Through satellite data of 1650 municipalities in the four main islands, we assess differences, at the regional level, of particulate matter concentration, ozone concentration, population density, night lights, and accessibility to cities. By applying principal component analysis, we reduce these variables into two components: PC1 (the first principal component) and PC2 (the second principal component). PC1 mostly represents the municipal variation in population density, night lights and accessibility to cities. The variables related to pollution (particulate matter concentration and ozone concentration) are included depending on the island under consideration. PC2 mostly represents the municipal variation in the pollution variables. We then identify groups of municipalities that present spatial contiguity (location similarity) and similar economic activity and levels of air pollution (attribute similarity).

We make use of two methods to find clusters that are geographically contiguous. The first one consists in studying spatial dependence through the local Moran analysis of Anselin (1995). This analysis shows the high-value clusters (hot spots), low-value clusters (cold spots), and spatial outliers. The second method is the Max-p algorithm of Duque et al. (2012). This algorithm outlines a new map of Japan where the regional borders are endogenously obtained based on differences in air pollution and economic activity.

From the local Moran analysis we find that there is a positive and statistically significant level of spatial dependence throughout the regions. In other words, the economic activity and level of air pollution (depicted by PC1 and PC2) that the average Japanese municipality experiences is highly similar to that of its geographically proximate neighbors. Noteworthy are the PC1 clusters of hot spots for metropolitan areas of Tokyo, Nagoya, and Osaka, in the island of Honshu. These show high levels of particulate matter concentration, population density, and night lights. On the flip side, the north of the Tohoku region presents a cold spot for these same variables. Municipalities in this latter cluster, thus, have lower levels of particulate matter concentration, population density, and night lights.

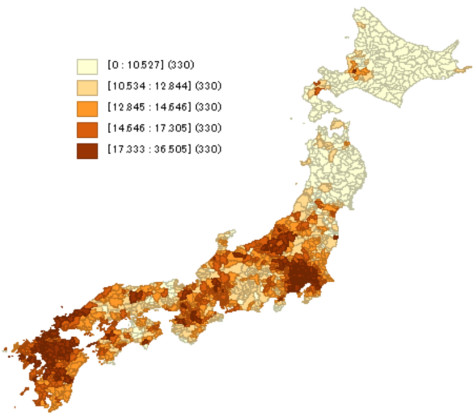
The Max-p algorithm generates results that are consistent to a large extent with those of the local Moran analysis. The Max-p algorithm shows that Japan can be subdivided into different regions in each island, from the results of the PC1 and PC2 data. The borders obtained by the algorithm are different from those of the administrative regions, suggesting that pollution and economic activity cross administrative borders. Through these new boundaries, it can be more clear how different municipalities can coordinate in the design and implementation of policies aimed at improving air pollution levels.

Future research may focus on extending the analysis through the sensitivity of the Max-p algorithm using different size and initialization parameters. Additionally, implementing alternative clustering methods such as She et al. (2017) for regionalization of Japanese municipalities could be promising.

7 Appendix

Figure 1: Air and Ozone pollution in Japan

(a) Particle matter ($PM_{2.5}$) concentration



(b) Ozone concentration

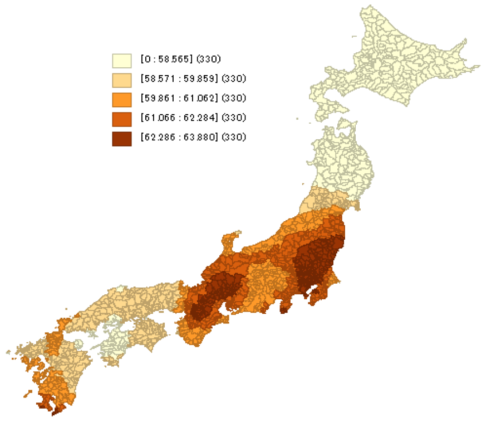
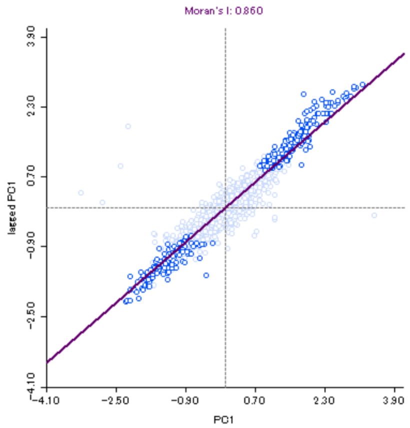
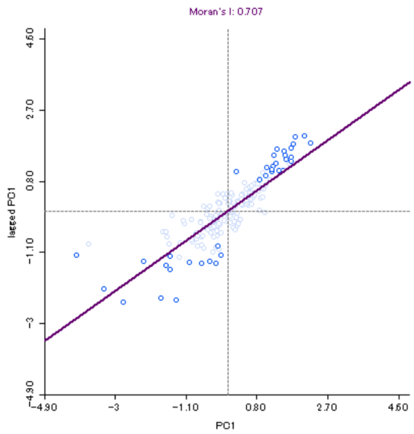


Figure 2: Spatial Autocorrelation (PC1).

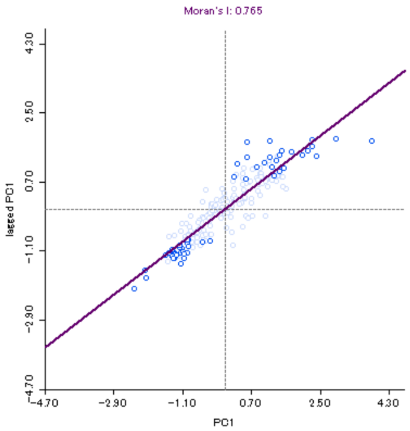
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

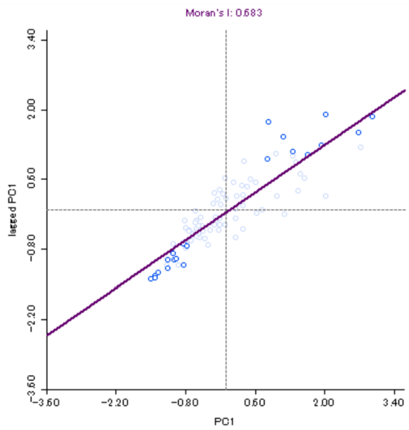
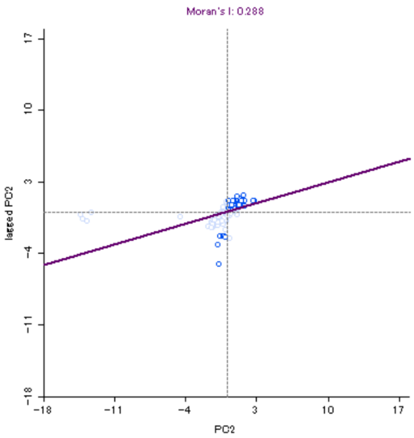
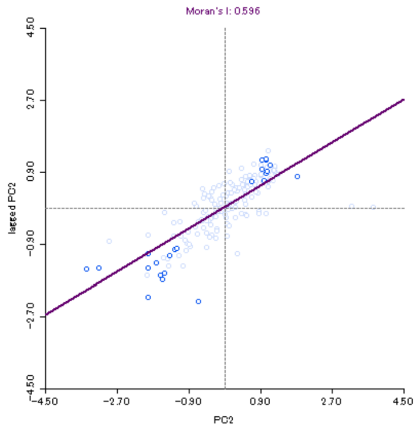


Figure 3: Spatial Autocorrelation (PC2).

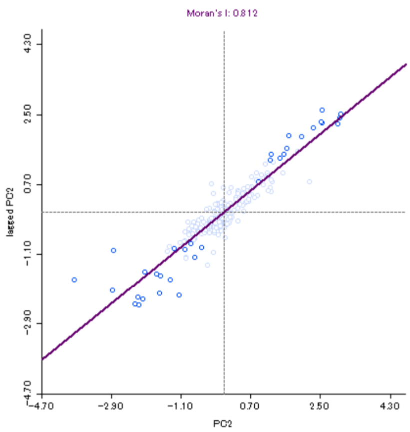
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

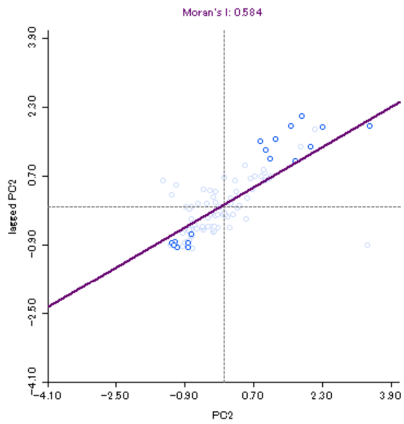
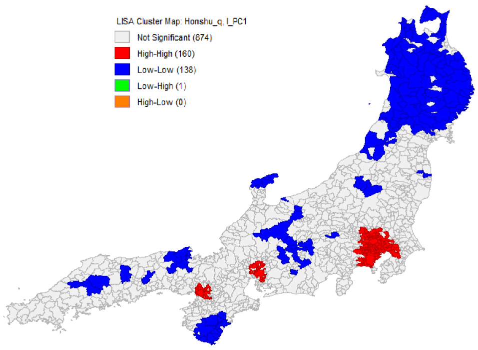
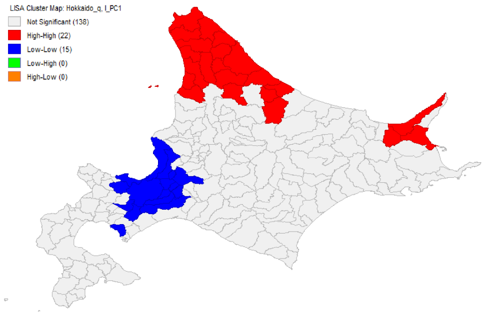


Figure 4: Hot spots and Cold spots (PC1).

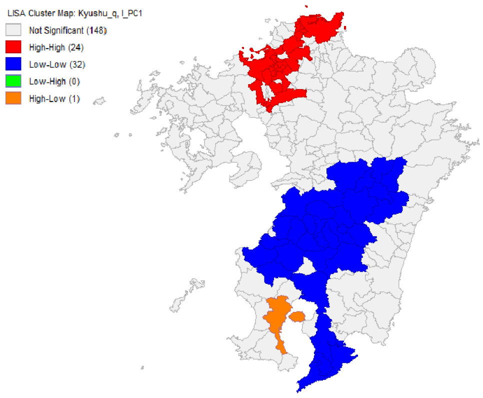
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

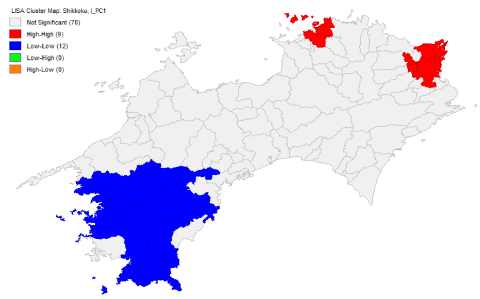
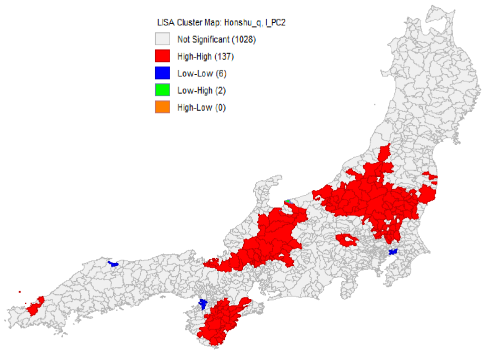
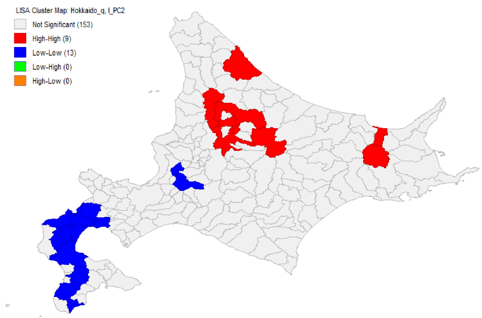


Figure 5: Hot spots and Cold spots (PC2).

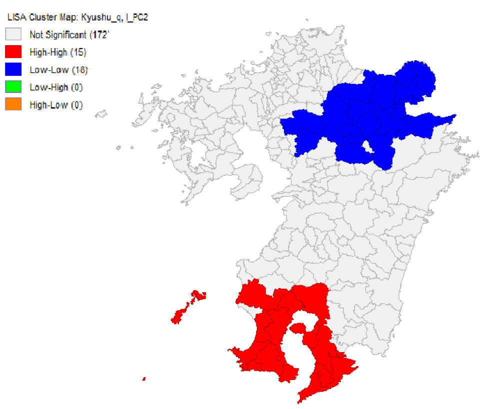
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

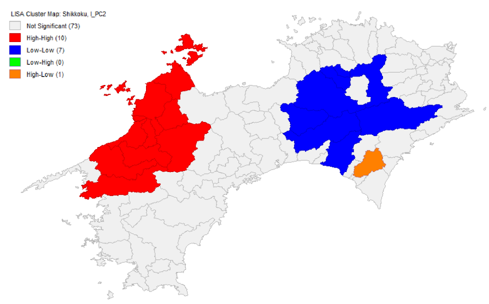
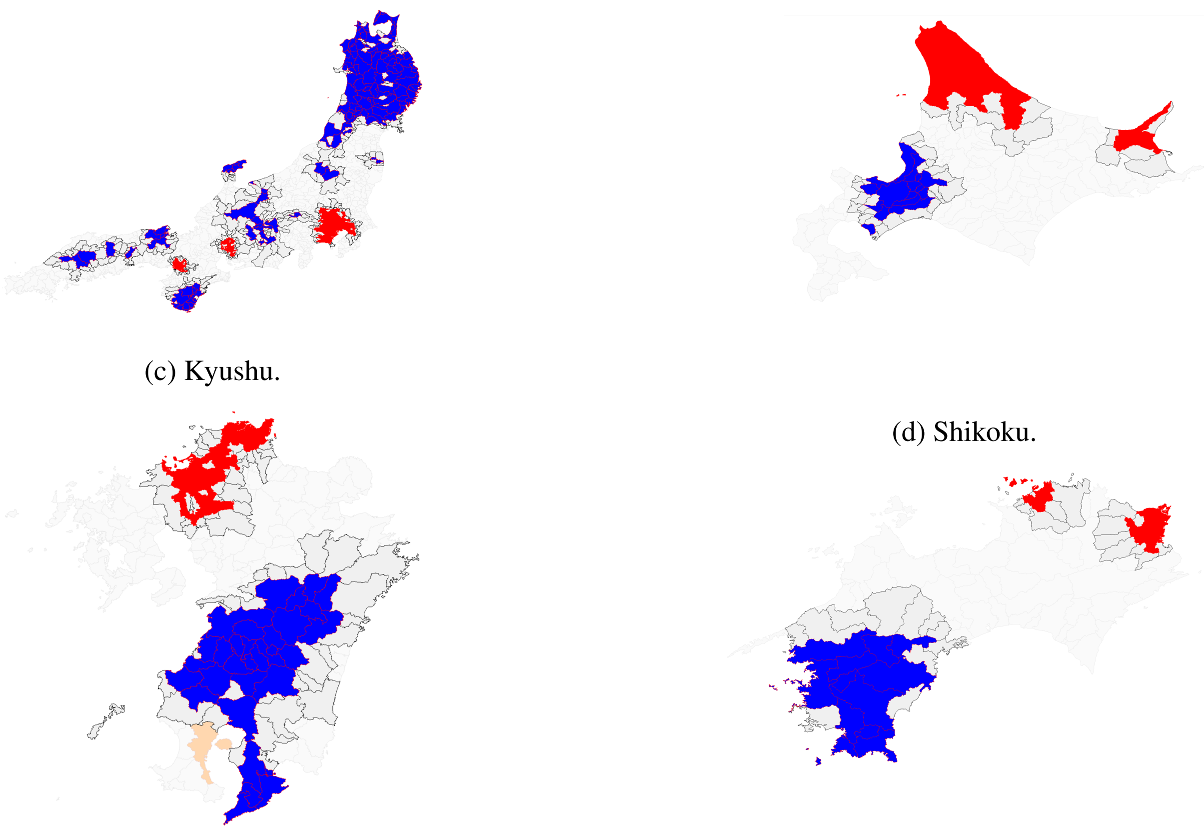


Figure 6: High & Low PC1 Clusters.



(a) Honshu.

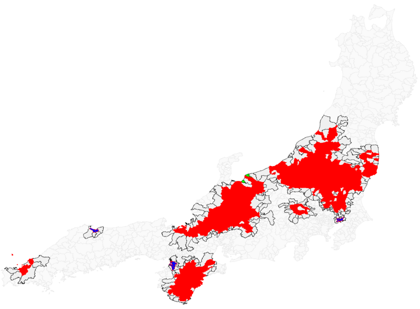
(b) Hokkaido.

(c) Kyushu.

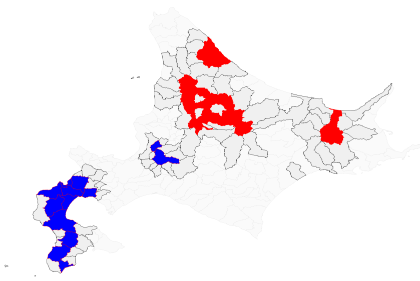
(d) Shikoku.

Figure 7: High & Low PC2 Clusters.

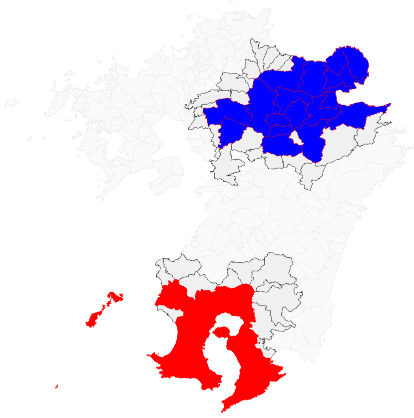
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

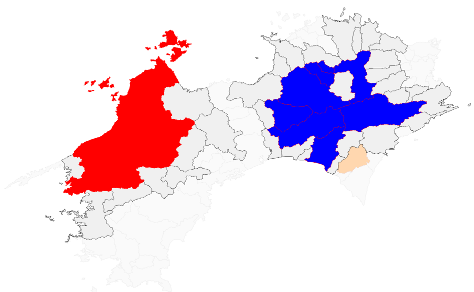


Figure 8: Administrative regions.

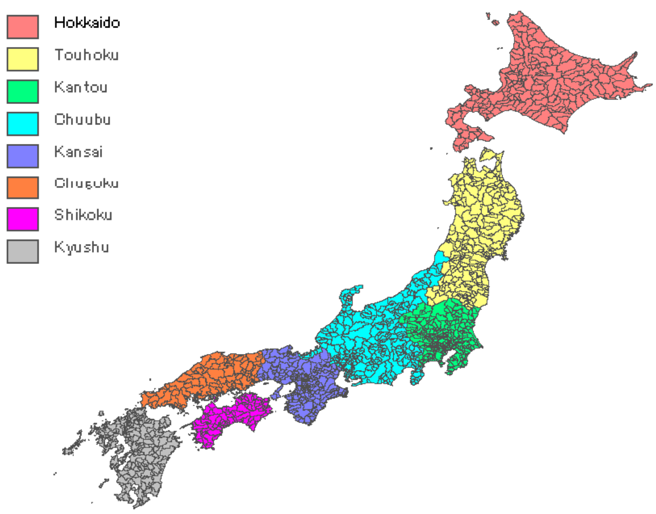
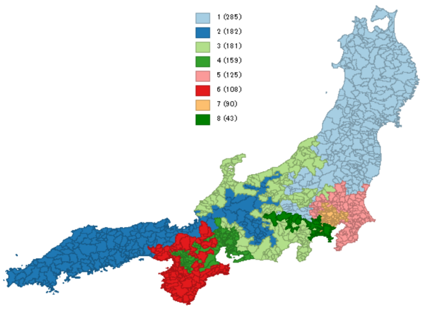
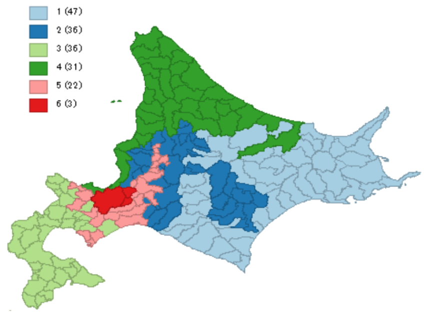


Figure 9: Regionalization: Analytical regions for PC1.

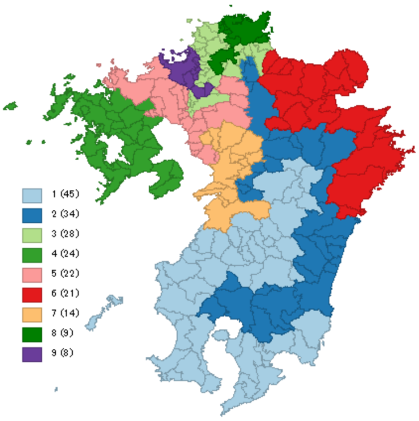
(a) Honshu.



(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.

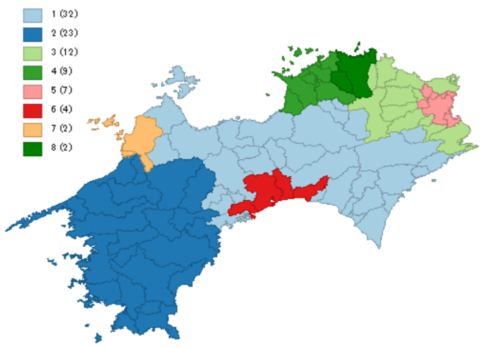
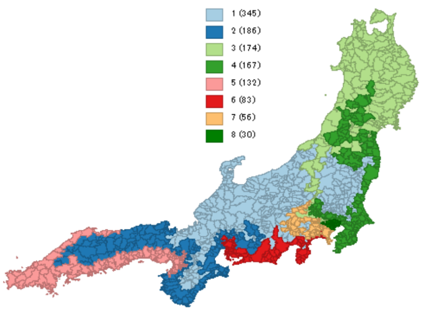
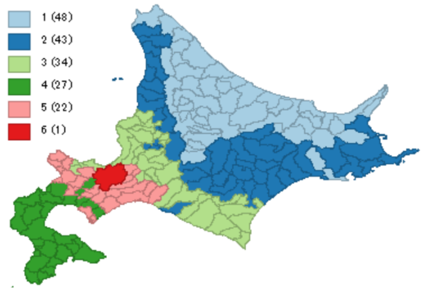


Figure 10: Regionalization: Analytical regions for PC2.

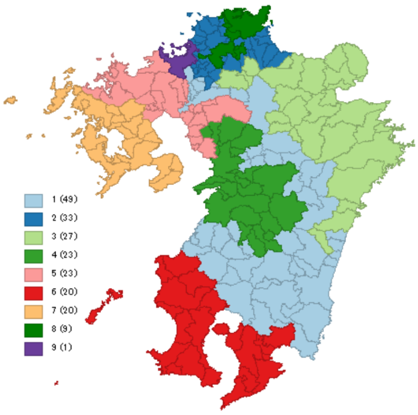
(a) Honshu.



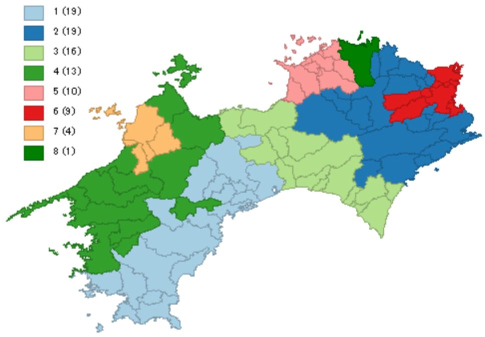
(b) Hokkaido.



(c) Kyushu.



(d) Shikoku.



References

Anselin, L. (1995). Local indicators of spatial association—lisa. *Geographical analysis*, 27(2):93–115.

Araki, S., Yamamoto, K., Kondo, A., et al. (2015). Application of regression kriging to air pollutant concentrations in japan with high spatial resolution. *Aerosol and Air Quality Research*, 15(1):234–241.

Beddoe, R., Costanza, R., Farley, J., Garza, E., Kent, J., Kubiszewski, I., Martinez, L., McCowen, T., Murphy, K., Myers, N., et al. (2009). Overcoming systemic roadblocks to sustainability: The evolutionary redesign of worldviews, institutions, and technologies. *Proceedings of the National Academy of Sciences*, 106(8):2483–2489.

Bernstein, J. A., Alexis, N., Barnes, C., Bernstein, I. L., Nel, A., Peden, D., Diaz-Sanchez, D., Tarlo, S. M., and Williams, P. B. (2004). Health effects of air pollution. *Journal of allergy and clinical immunology*, 114(5):1116–1123.

Brauer, M., Freedman, G., Frostad, J., Van Donkelaar, A., Martin, R. V., Dentener, F., Dingenen, R. v., Estep, K., Amini, H., Apte, J. S., et al. (2016). Ambient air pollution exposure estimation for the global burden of disease 2013. *Environmental science & technology*, 50(1):79–88.

Carriazo, F. (2016). Economics and air pollution. *Air Quality-Measurement and Modeling*.

Cliff, A. D. and Ord, J. K. (1981). *Spatial processes: models & applications*. Taylor & Francis.

Columbia University, C. f. I. E. S. I. N. . C. (2018). Gridded population of the world, version 4 (gpwv4): Population density adjusted to match 2015 revision un wpp country totals, revision 11.

Crandall, R. W. (1983). *Controlling industrial pollution: The economics and politics of clean air*. Brookings Institution, Washington, DC, USA.

Dechezleprêtre, A., Rivers, N., and Stadler, B. (2019). The economic cost of air pollution: Evidence from europe. *OECD Economics Department Working Papers*.

Downing, P. B. and Watson Jr, W. D. (1974). The economics of enforcing air pollution controls. *Journal of Environmental Economics and Management*, 1(3):219–236.

Duque, J. C., Anselin, L., and Rey, S. J. (2012). The max-p-regions problem. *Journal of Regional Science*, 52(3):397–419.

Duque, J. C., Church, R. L., and Middleton, R. S. (2011). The p-regions problem. *Geographical Analysis*, 43(1):104–126.

Duque, J. C., Ramos, R., and Suriñach, J. (2007). Supervised regionalization methods: A survey. *International Regional Science Review*, 30(3):195–220.

Fischer, M. M. (1980). Regional taxonomy: a comparison of some hierarchic and non-hierarchic strategies. *Regional Science and Urban Economics*, 10(4):503–537.

Gaspar, V., Amaglobeli, M. D., Garcia-Escribano, M. M., Prady, D., and Soto, M. (2019). *Fiscal policy and development: Human, social, and physical investments for the SDGs*. International Monetary Fund.

Goodman, S., BenYishay, A., Lv, Z., and Runfola, D. (2019). Geoquery: Integrating hpc systems and public web-based geospatial data tools. *Computers & geosciences*, 122:103–112.

Hansen, P., Jaumard, B., Meyer, C., Simeone, B., and Doring, V. (2003). Maximum split clustering under connectivity constraints. *Journal of Classification*, 20(2):143–180.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417.

Jolliffe, I. T. and Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151.

Kampa, M. and Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2):362–367.

Kanada, M., Fujita, T., Fujii, M., and Ohnishi, S. (2013). The long-term impacts of air pollution control policy: historical links between municipal actions and industrial energy efficiency in kawasaki city, japan. *Journal of cleaner production*, 58:92–101.

Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., Takezaki, T., Nakayama, T., Nitta, H., Tanabe, K., et al. (2011). An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in japan. *Journal of epidemiology*, pages 1102090211–1102090211.

Kim, E., Moon, S.-W., and Kagawa, S. (2019). Spatial economic linkages of economic growth and air pollution: developing an air pollution-multinational cge model of china, japan, and korea. *The Annals of Regional Science*, 63(2):255–268.

Kumar, A., Gupta, I., Brandt, J., Kumar, R., Dikshit, A. K., and Patil, R. S. (2016). Air quality mapping using gis and economic evaluation of health impact for mumbai city, india. *Journal of the Air & Waste Management Association*, 66(5):470–481.

Kume, K., Ohura, T., Noda, T., Amagai, T., and Fusaya, M. (2007). Seasonal and spatial trends of suspended-particle associated polycyclic aromatic hydrocarbons in urban shizuoka, japan. *Journal of hazardous materials*, 144(1-2):513–521.

Kyriakopoulou, E. et al. (2021). Cities and the environment: The case of air pollution. *International Review of Environmental and Resource Economics*, 15(4):319–358.

Lanzi, E., Dellink, R., and Chateau, J. (2018). The sectoral and regional economic consequences of outdoor air pollution to 2060. *Energy Economics*, 71:89–113.

Lanzi, E. et al. (2016). The economic consequences of outdoor air pollution. *Organization for Economic Cooperation and Development. Retrieved from https://www.oecd.org/environment/indicators-modelling-outlooks/Policy-Highlights-Economic-consequences-of-outdoor-air-pollution-web. pdf.*

Law, S. and Neira, M. (2019). An unsupervised approach to geographical knowledge discovery using street level and street network images. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pages 56–65.

Lefkovitch, L. P. (1980). Conditional clustering. *Biometrics*, pages 43–58.

Lights, D.-O. N. (2017). Image and data processing by noaa's national geophysical data center. dmsp data collected by the us air force weather agency.

Luechinger, S. (2010). Life satisfaction and transboundary air pollution. *Economics Letters*, 107(1):4–6.

Maantay, J. (2007). Asthma and air pollution in the bronx: methodological and data considerations in using gis for environmental justice and health research. *Health & place*, 13(1):32–56.

Maravalle, M. and Simeone, B. (1995). A spanning tree heuristic for regional clustering. *Communications in statistics-theory and methods*, 24(3):625–639.

Otsuka, A. (2021). *A New Driver of Regional Sustainability in Japan: Inter-regional Network Economies*, volume 54. Springer Nature.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572.

She, B., Duque, J. C., and Ye, X. (2017). The network-max-p-regions model. *International Journal of Geographical Information Science*, 31(5):962–981.

Shimadera, H., Kondo, A., Kaga, A., Shrestha, K. L., and Inoue, Y. (2009). Contribution of transboundary air pollution to ionic concentrations in fog in the kinki region of japan. *Atmospheric Environment*, 43(37):5894–5907.

Shoji, H. and Miyamoto, K. (1977). Environmental pollution in japan. *Japanese Economic Studies*, 5(4):3–40.

Toyama, T. (1964). Air pollution and its health effects in japan. *Archives of Environmental Health: An International Journal*, 8(1):153–173.

Weiss, D. J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., et al. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, 553(7688):333–336.

Wise, S., Haining, R., and Ma, J. (1997). Regionalisation tools for the exploratory spatial analysis of health data. In *Recent developments in spatial analysis*, pages 83–100. Springer.

Wolozin, H. (1968). The economics of air pollution: Central problems. *Law and Contemporary Problems*, 33(2):227–238.

Yorifuji, T., Kashima, S., and Doi, H. (2015). Outdoor air pollution and term low birth weight in japan. *Environment international*, 74:106–111.

Yoshino, A., Takami, A., Sato, K., Shimizu, A., Kaneyasu, N., Hatakeyama, S., Hara, K., and Hayashi, M. (2016). Influence of trans-boundary air pollution on the urban atmosphere in fukuoka, japan. *Atmosphere*, 7(4):51.

Detecting air pollution clusters in Japan: A spatial analysis approach

令和４年３月発行

発行所　　公益財団法人アジア成長研究所
〒803-0814　北九州市小倉北区大手町 11 番 4 号
Tel：093-583-6202／Fax：093-583-6576
URL：http://www.agi.or.jp
E-mail：office@agi.or.jp