Formulation of Huge Lattice Spatial Adjacency Matrices With Non-rectangular Shape of Socio-economic Grid-Cell Data for the Analysis of Sustainable Economy With High Computational Efficiency

Gigih Fitrianto[1], Shojiro Tanaka[2] & Ryuei Nishii[3]

[1] Graduate School of Economics, Hiroshima University of Economics, Hiroshima, Japan

[2] Professor, Faculty of Economics, Hiroshima University of Economics, Hiroshima, Japan

[3] Professor, Institute of Mathematics for Industry, Kyushu University, Fukuoka, Japan

Correspondence: Gigih Fitrianto, Graduate School of Economics, Hiroshima University of Economics, 5-37-1, Gion, Asaminami, Hiroshima-shi, Hiroshima 731-0138, Japan.

Received: August 19, 2018 Accepted: September 4, 2018 Online Published: November 8, 2018

doi:10.5430/rwe.v9n2p1 URL: https://doi.org/10.5430/rwe.v9n2p1

Abstract

The advantage of using grid-cell data for socio-economic analysis should be the feasibility to incorporate satellite data that will enrich the regional analysis and has an important role to observe the relationship between socio-economics and nature. This advancement corresponds to the sustainable development goals that balance the socio-economic quality in harmony. In order to perform the analysis, formulation of a spatial adjacency matrix has an important role to project the spatial relationship within regions. However, no precedent research provided a practical formulation for the spatial adjacency matrix in grid-cell data structure (Fitrianto & Tanaka, 2017).

The general process that used shapefiles solely, which store geometry and attribute information for the spatial features (ESRI, 1998) to construct the adjacency matrix is not suitable. The problem arises due to the existence of NA cells that represent non-inhabitant areas such as water bodies, yet the shapefile does not contain this information inside the municipal body. The NA cells create a non-rectangular lattice and it is important to exclude them in the analysis to correctly project the real information.

This article provides a method to precisely project the real information by using Kronecker product to construct the adjacency matrix and applying a projection matrix to eliminate the NA cells (Tanaka & Nishii, 2009). It showed eminent efficiency compared with commonly used R package called ***spdep***. Experimental results verified that this method, even for huge dimension with a trillion elements, produces more than 2000 times faster elapsed time than the package.

Keywords: grid-cell data, Kronecker product, spatial adjacency matrix, sustainable development goals

1. Introduction

The establishment of sustainable development goals (SDG) and the targets by United Nations General Assembly on 25 September 2015 encouraged every nation to balance the improvement of socio-economic quality with the environmental sustainability (UNDP, 2015). Let us use some of the goals and targets of SDG (see details in Table 1) as an example of related analysis that can be achieved by the incorporated data. This approach enables us to comprehend not only for sustainable per capita economic growth by increasing economic productivity in the agricultural sector (Goal 8) but also for the impact of improper agricultural production causing deforestation and desertification (Goal 15).

In order to simultaneously analyze those two elements, the incorporation of grid-cell data for socio-economic variables and satellite data shall become a key role. The advantage of using this grid-cell data type of socio-economic variable enriches the regional analysis by incorporating satellite data (Tanaka & Nishii, 2015), such as night-time lights, carbon-dioxide concentration, and vegetation index. It becomes realized due to the availability of grid-cell database for socio-economic variables such as GPWv4, GRUMP, Landscan, Worldpop, and GeoStat (see examples in Table 2).

Table 1. Examples of related sustainable goals and targets for incorporated analysis

Goals	Targets
Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.	**8.1** Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7 per cent gross domestic product growth per annum in the least developed countries.
	8.2 Achieve higher levels of economic productivity through diversification, technological upgrading and innovation, including through a focus on high-value added and labor-intensive sectors.
Goal 15. Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss	**15.2** By 2020, promote the implementation of sustainable management of all types of forests, halt deforestation, restore degraded forests and substantially increase afforestation and reforestation globally.
	15.3 By 2030, combat desertification, restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land degradation-neutral world.

This incorporated data set should become more convenient due to the improvement of accessibility for earth observation satellite data provided by a database such as NASA ongoing project Open Data Cube (ODC) (ODC Documentation, 2017). The ODC allows storing rich information for full spatial and temporal coverage of earth observation (Lewis et al., 2017).

The objective of this project is to provide an open and freely accessible analysis ready data to increase developing countries capability for the usage space-based Earth observation technologies (Killough, 2017). The web-based architecture providing user-friendly features for data preparation, processing, and visualization. There are ten features available in the ODC such as cloud-free mosaic, NDVI anomaly, water quality, landslide, and urbanization (Open Data Cube, 2018). This recent improvement allows future research to analyze the interrelation between socio-economics and satellite data as a component of achieving the SDG goals and targets.

Table 2. Examples of grid-cell data source for socio-economic variables

Database	Method	Data Sets	Data Availability
GPWv4 (source: CIESIN, Columbia University)	Population estimates are created by extrapolating the raw census estimates and proportionally allocated to raster cells using a uniform areal weighting approach to produce the population surfaces.	a. Population density. b. Population count. c. Land and water area d. National identifier grid.	All data contains 2000, 2005, 2010, 2015 with 1 km^2 grid resolutions.
LandScan (source: Oak Ridge National Laboratory)	The modeling process uses sub-national level census counts for each country and primary geospatial input or ancillary datasets. For each country, they calculate a “likelihood” coefficient for each cell and applies it to total population.	Global population distribution data.	There are data from 1998 to 2016 (except 1999 data) with 1 km^2 grid resolutions.
Worldpop (source: GeoData	The dataset produced by disaggregating census data for population mapping using random forest with remotely-sensed and	a. Population b. Birth c. Pregnancies	Dataset **a, d**, and **e** are created based on 100 m resolution data. Dataset **b** and **c** are 1 km

Institute, University of Southampton)	ancillary data (Stevens et al., 2015)	d. Urban change e. Age structure The data covers for Africa, Asia, Latin America and the Caribbean.	resolution data.

Let us consider an instance of a spatial analysis with socio-economic variables based on Anselin (1988) spatially lagged autoregressive model,

$$Y = \alpha i_N + \delta WY + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

where Y is the crime rate, x_1 and x_2 represent household income and housing value respectively, the data size N is the 49 areas in Columbus, Ohio, α is a constant, i_N is a column vector, δ is the spatial regression coefficient, and $\varepsilon \sim N(0, \sigma^2 I_N)$. All of the variable matrices have $N \times 1$ dimension. The spatial W matrix stores the neighborhood relationship for each area, which is an $N \times N$ adjacency matrix. Therefore, the WY variable represents the influence of neighborhood's crime rate.

The role of W matrix in the above model is to represent the impact of each neighborhood's crime rate to the neighboring locations. Anselin (1988) and Anselin (1992) showed that based on several estimation methods, the $\hat{\delta}$ values are always positive ($0.3 < \hat{\delta} < 0.5$) and significant. This implies the existence of strong spatial spillover that the crime rate for each location affected by the neighboring values. From those results, we see the importance of the W matrix to capture the spatial relationship of neighboring locations on the analysis.

Bivand, Pebesma, and Gómez-Rubio (2008), Arbia (2014), Dmowska and Stepinski (2017), and Baddeley, Turner, and Rubak (2018), proceeded to carry out the grid-cell data analysis with the extraction of base data by overlaying it with specific regional shapefile in general (ESRI, 1998). The output object and data created by this process then become the base information for the succeeding analysis, which is summarized in Figure 1. Based on those data types, we can analyze the neighborhood relation inside the data and store it in spatial adjacency matrix W.

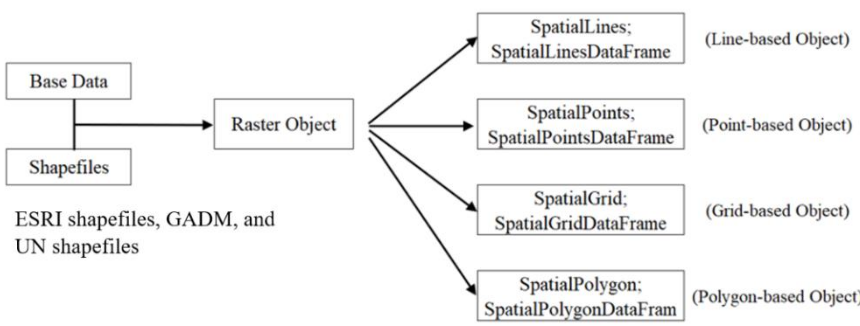

Figure 1. General shapefile handling process for spatial analysis

However, this general process based on shapefile is not suitable to derive the W matrix from grid-cell data because it does not exclude the NA cells inside the administrative area. To demonstrate this drawback, we choose the neighborhood area of Lake Apopka, Florida as shown in Figure 2.

Figure 2 map (a) shows that the shapefile does not reflect the surface information of the lake area in map (b). It is clear that there is no population in water body areas in map (b), which are denoted by NA cells. By using the shapefile, we have rectangular lattice matrix $D_{9\times9}$, without identifying the non-inhabitant area as seen on land indicator (a). Thus, we will misidentify that there is the possibility of a human settlement exists over non-inhabitant areas. By applying a projection matrix (Tanaka & Nishii, 2009) to exclude the NA cells of non-rectangular $D_{9\times9}$ lattice based on land indicator (b), then we will have the correct $W^*_{53\times53}$ instead of $W_{81\times81}$.

The importance of correctly formulate W^* are shown by several precedent researches such as: residential values analysis based on surrounding land use (Geoghegan, Wainger, & Bockstael, 1997), land use change impact to economic and ecological condition (Irwin & Geoghegan, 2001) or rural-urban interface (Bell & Irwin, 2002), and roads impact on deforestation (Nelson & Hellerstein, 1997). Their analysis, in particular, indicated the role of spatial

W matrix to represent windborne seeding effects that the neighboring locations are more likely to have the same vegetation type. By using the correct W^*, they exactly observed that effect for a vegetative cover area.

Fitrianto and Tanaka (2017) found that very few precedent researches focused on the practical method of formulation W, especially for grid-cell data structure with big data size. No practical paper was found for the real-projected W^* for non-rectangular case. To remove the non-inhabitant cells, we overhauled the whole process of formulation and found the Kronecker product provides the best results (Note 1).

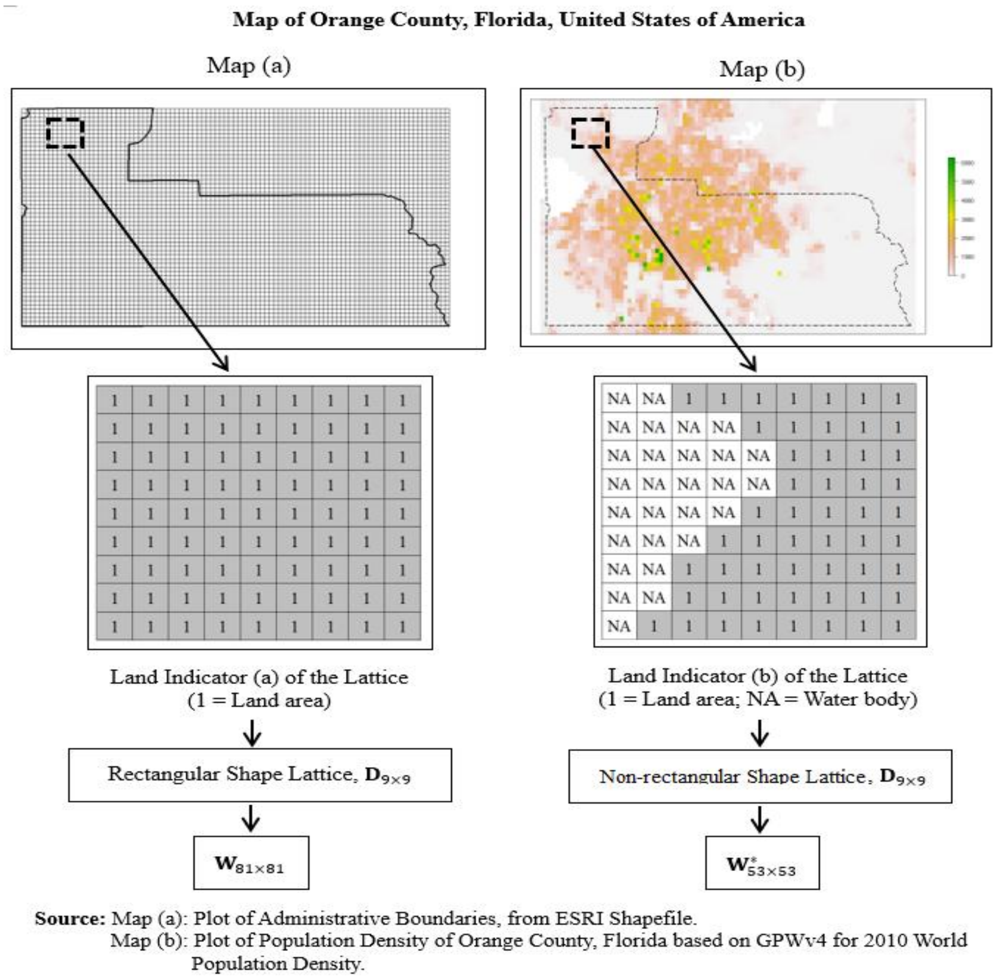

Figure 2. Comparison of formulation of w for the neighborhood of eastern area of Lake Apopka, Florida

2. Utilization of Kronecker Product: Formulation of W Matrices

Construction of spatial adjacency matrix based on a spatial lattice matrix, **D**, inherit a spatial information of a lattice cell denoted by s_i at (x_i, y_i) position

$$D = \{s_i = (x_i, y_i) : i = 1, \dots, \mathcal{Z}\} \tag{1}$$

where, $x_i = 1, \dots, c$ and $y_i = 1, \dots, r$, given that r and c are row and column of **D** respectively, and $\mathcal{Z} = rc$ is the data size on a given rectangular lattice. The neighborhood information for each s_i cell is defined as follows (Cressie 1991, pp. 384-385),

$$N_i = \{s_j = (x_j, y_j) : s_j \text{ is a neighbor cell of } s_i\},\ i \neq j,\ i, j = 1, \dots, \mathcal{Z} \tag{2}$$

N_i consists of horizontal-vertical and diagonal neighborhood of each cell.

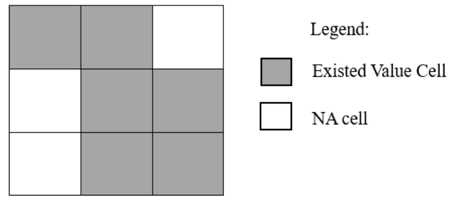

Figure 3. Nonrectangular Lattice $\mathrm{D}_{3\times 3}$

We focused on nearest neighbor case for each cell, which divided into two groups, primary and secondary neighborhood. The vertical and horizontal neighbors called as primary neighbors and the diagonal neighbors called as the secondary. Let us use Figure 3 as an example, we construct the neighborhood of cell $(1,1)$, $(2,2)$, and $(3,2)$ on non-rectangular $\mathrm{D}_{3\times 3}$ as shown in Figure 4.

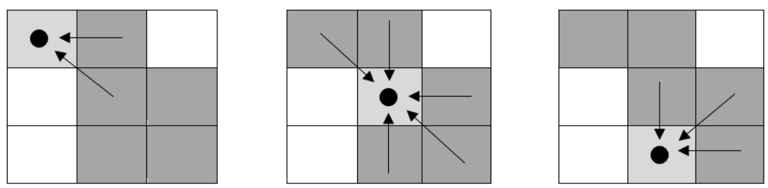

Figure 4. Neighborhood Examples for cell $(1,1)$, $(2,2)$, and $(3,2)$ on $\mathrm{D}_{3\times 3}$

All the neighborhood information is stored on spatial adjacency matrix, $\mathrm{W}_{\mathcal{Z}\times\mathcal{Z}}$ which includes NA cells. However, as shown in Figure 4, NA cells should be excluded from the analysis. Therefore, we apply the projection matrix on a rectangular case W based on Tanaka and Nishii (2009) to obtain the real projected W^*.

We consider the formulation of W as the basis neighborhood information prior to the projection matrix. In nearest neighbor case, based on Cressie and Wikle (2011, p. 167), the $\mathrm{W} \equiv (w_{ij})$ as a $\mathcal{Z}\times\mathcal{Z}$ matrix with $w_{ii} = 0$ and $w_{ij} = 1$ if cell s_j is the neighbor of s_i as defined on equation (2). The formulation utilizes Kronecker product of an identity matrix I_m and an A_m matrix which constructed as follows,

$$\mathbf{A}_m = \begin{bmatrix} 0 & 1 & 0 & & 0 & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 \\ 0 & 1 & 0 & & 0 & 0 \\ & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}_{(m\times m)} \tag{3}$$

where A_m has a $m\times m$ dimension which is either $r\times r$, or $c\times c$ given that r and c is the number of row and column of D respectively. Thus, we can formulate W matrix as follows,

$$\mathrm{W}_{\mathrm{P},\mathcal{Z}\times\mathcal{Z}} = \mathrm{A}_r \otimes \mathrm{I}_c + \mathrm{I}_r \otimes \mathrm{A}_c, \quad \mathrm{W}_{\mathrm{S},\mathcal{Z}\times\mathcal{Z}} = \mathrm{A}_r \otimes \mathrm{A}_c, \quad \text{and} \quad \mathrm{W}_{\mathcal{Z}\times\mathcal{Z}} = \mathrm{W}_{\mathrm{P},\mathcal{Z}\times\mathcal{Z}} + \mathrm{W}_{\mathrm{S},\mathcal{Z}\times\mathcal{Z}} \tag{4}$$

The $\mathrm{W}_{\mathcal{Z}\times\mathcal{Z}}$ consists of two components, primary ($\mathrm{W}_{\mathrm{P},\mathcal{Z}\times\mathcal{Z}}$) and secondary neighborhood ($\mathrm{W}_{\mathrm{S},\mathcal{Z}\times\mathcal{Z}}$). Practically, in R language program, equation (4) can be run by Script 1 which applies sparse matrix library by using *Matrix* package (Bates & Maechler, 2017).

```
## Constructing the Function* ##
A.mat <- function(dims){
   B <- as(diag(1, dims-1, dims-1), "CsparseMatrix")
   D <- as(matrix(rep(0, dims), nrow = dims, ncol = 1), "CsparseMatrix")
   E <- cbind(rbind(t(D[1:(dims-1),]),B), D)
   return(t(E) + E)
}

library(Matrix)
r = r; c = c
ic  <- as(diag(1, ncol = c, nrow = c), "CsparseMatrix")
ir  <- as(diag(1, ncol = r, nrow = r), "CsparseMatrix")
ac <- A.mat(c); ar <- A.mat(r)

W  <- kronecker(ar, ic) + kronecker(ir, ac) + kronecker(ar, ac)
```

Script 1. Kronecker Product to Construct W

Note: *) r and c is the number of rows and columns of **D** and 'dims' is referred to the value of r or c.

3. Non-rectangular Lattice Case

To eliminate NA cells, we apply a projection matrix (P) based on Tanaka and Nishii (2009) as follows,

$$\mathrm{W}^{*}_{\mathcal{V}\times\mathcal{V}} = \mathrm{P}^{T}_{\mathcal{Z}\times\mathcal{V}}\,\mathrm{W}_{\mathcal{Z}\times\mathcal{Z}}\mathrm{P}_{\mathcal{Z}\times\mathcal{V}} \tag{5}$$

where $\mathcal{V}$ represents the number of valid value cells. By using Figure 3 as an example of non-rectangular case, we will get a correctly projected W^{*} matrix as shown in Appendix (see Script 2 to run equation (5)).

```
# Let us assume the spatial lattice matrix denotes as D #
# Identify the valid value cells on D and store the information in 'no.na' object* #
# The 'NA' values are originated from grid-cells such as GPWv4 #
no.na <- which(as.numeric(t(D)) != 'NA', arr.ind = TRUE)

library(Matrix)
P <- as(matrix (0, nrow = (ncol(D)*nrow(D)), ncol = length(no.na)), "CsparseMatrix")

count=0
for (i in 1:nrow(P)){
   if (is.element(i, no.na)){
      count = count+1;
      P[i, count] <- 1;
   }
}
```

Script 2. R Script for projection matrix

Note: *) The transpose of D matrix is necessary due to the default system of cell-indexing matrix in R which started from the row and the cell-indexing format for equation (4) is inserted by column.

4. Computational Efficiency Comparison: A Simulation

In this section, we compare the performance of our Kronecker Product method with ‘cell2nb’ and ‘nb2mat’ functions in *spdep* package. Note that ‘poly2nb’ function is commonly used based on the polygon object, but that is not suitable for the purposes of this article focusing only on grid-cell data. Based on several D matrices, we measured their actual computational space and time. All simulations were conducted using the computational environment in Table 3 and a basic assumption that:

Assumption: There are no NA cells for all D.

The assumption is used to simplify the simulation process since the processes of equation (5) are common and easily applied to the system.

Table 3. Computational environments for simulations

Computational Environment		
Processor		Intel core i7-6700K, 4 GHz, 8 MB
Memory		DDR4-2133 64 GB (16 GB x 4 slots)
R version	Main Body	3.4.2 (64-bit)
	Library (*Matrix*)	1.2-11
	Library (*spdep*)	0.7-7
	Library (*raster*)	2.5-8

In Table 4, our program based on Script 1 showed more than 2000 times faster for the largest W. All the results obviously indicate that our method produces the best result, especially for larger dimension cases. In Table 5, we have actual object memory size comparison of both methods. The results also show that our method consumed 1/3 times less memory compared to ‘cell2nb’ and ‘nb2mat’. Figure 5 and Figure 6 provides the graphical comparison for those simulation results.

Table 4. Total elapsed time comparison for all methods*

Spatial Lattice Matrix ($\mathbf{D}_{r\times c}$)	W Elements ($\mathcal{Z}^2$)	Script 1	‘cell2nb’ and ‘nb2mat’
$\mathbf{D}_{3\times3}$	81	0.005	0.012
$\mathbf{D}_{60\times60}$	1.296×10^7	0.006	1.223
$\mathbf{D}_{120\times120}$	2.074×10^8	0.008	4.838
$\mathbf{D}_{180\times180}$	1.050×10^9	0.011	10.932
$\mathbf{D}_{240\times240}$	3.318×10^9	0.018	19.722
$\mathbf{D}_{300\times300}$	8.100×10^9	0.025	31.117
$\mathbf{D}_{330\times330}$	1.186×10^{10}	0.052	37.790
$\mathbf{D}_{500\times500}$	6.250×10^{10}	0.137	108.191
$\mathbf{D}_{1000\times1000}$	1.000×10^{12}	0.234	492.589

Note: *) Each elapsed time for our generic function in Script 1 is an average of 100 times iteration. On the other hand, ‘cell2nb’ and ‘nb2mat’ method is measured once by considering the amount of time needed to run.

The main reason why ‘cell2nb’ and ‘nb2mat’ function consumed a lot of space during the formulation process is that there is no option inside the function to utilize sparse matrix. However, there is a function to convert the produced matrix to sparse matrix. Even though we applied the function, we still found that ‘cell2nb’ and ‘nb2mat’ required more space compared to Script 1 in the process due to the intermediate list-class presence by ‘cell2nb.’ The ‘cell2nb’ function create the neighbor list object from the information for grid-cell dimension. Thus, ‘nb2mat’ function use this neighbor list object to construct the W matrices.

Table 5. Actual object memory size comparison between all methods* (in MB)

Spatial Lattice Matrix ($\mathbf{D}_{r\times c}$)	**W Elements** ($\mathcal{Z}^2$)	**Script 1**	**'cell2nb' and 'nb2mat'**
$\mathbf{D}_{3\times3}$	81	0.008	0.006
$\mathbf{D}_{60\times60}$	1.296×10^{7}	0.462	1.247
$\mathbf{D}_{120\times120}$	2.074×10^{8}	1.838	4.997
$\mathbf{D}_{180\times180}$	1.050×10^{9}	4.136	11.252
$\mathbf{D}_{240\times240}$	3.318×10^{9}	7.355	20.013
$\mathbf{D}_{300\times300}$	8.100×10^{9}	11.496	31.280
$\mathbf{D}_{330\times330}$	1.186×10^{10}	13.912	37.853
$\mathbf{D}_{500\times500}$	6.250×10^{10}	31.955	86.931
$\mathbf{D}_{1000\times1000}$	1.000×10^{12}	127.903	347.902

Note: *) All the simulations were done by one-shot run due to the same results produced by any iteration.

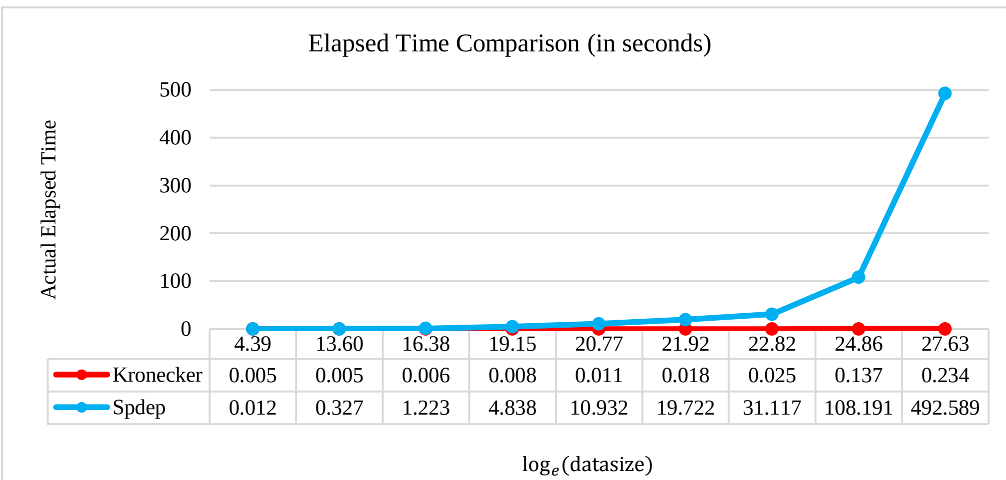

Figure 5. Comparison of elapsed time between Script 1 and *spdep* ('cell2nb' and 'nb2mat') Method

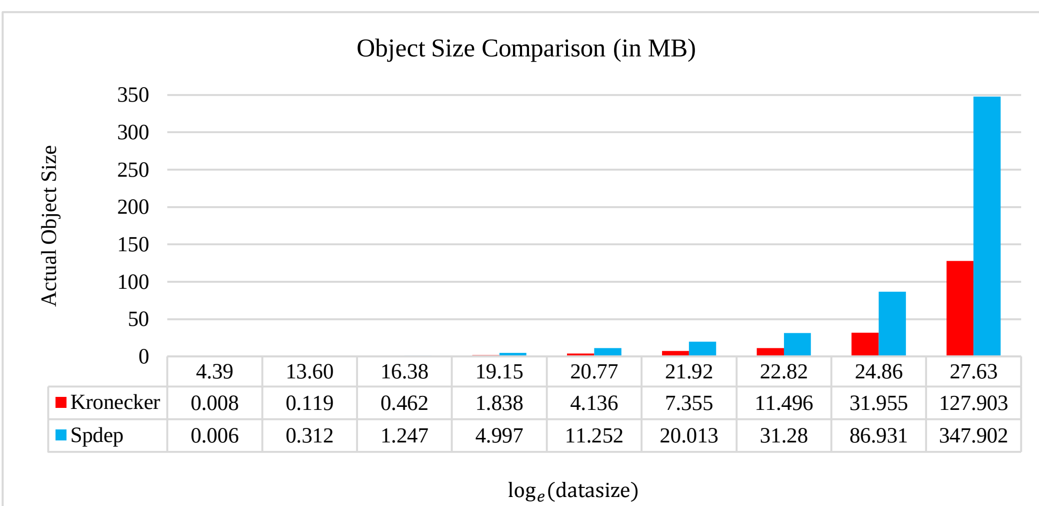

Figure 6. Object size comparison between Script 1 and *spdep* ('cell2nb' and 'nb2mat') Method

5. Kronecker Product Method Performance for Very Large W Matrices

By using Script 1 as the best method, we formulate larger W matrices. The simulation is essential due to the attempt to incorporate satellite images for higher resolution. Note that Landsat 8 images have the size of 4650×2571 pixels and generate a W matrix with more than 1.42×10^{15} elements.

We observed the performance to handle larger spatial lattice dimension. Table 6 summarizes our experiment of the elapsed time and object memory size. The results show the Kronecker product's handy availability to the formulation of huge W. Even for 5.063×10^{16} (more than 50 quadrillion) elements, it can produce the matrix in less than one minute.

Table 6. Kronecker product method performances for larger W Matrices*

Spatial Lattice Matrix ($\mathbf{D}_{r\times c}$)	W Elements ($\mathcal{Z}^2$)	Elapsed Time (in seconds)	Object Memory Size
$\mathbf{D}_{2500\times2500}$	3.906×10^{13}	1.45404	799.75 MB
$\mathbf{D}_{5000\times5000}$	6.250×10^{14}	5.91074	3.20 GB
$\mathbf{D}_{7500\times7500}$	3.164×10^{15}	13.49425	7.20 GB
$\mathbf{D}_{10000\times10000}$	1.000×10^{16}	25.14439	12.80 GB
$\mathbf{D}_{15000\times15000}$	5.063×10^{16}	55.90863	28.80 GB

Note: *) Each elapsed time is an average of 100 iteration times, but for the memory size run once.

6. Concluding Remarks and Future Works

To achieve the SDGs, we should focus not only on the increasing economic growth, but also on the environmental sustainability. To simultaneously observe them, the incorporation of grid-cell data for socio-economic variables and satellite data plays the key role.

Nonetheless, there were still few precedent researches that utilize those data for the analysis, which counts the adjacency interaction, such as Nelson and Hellerstein (1997) and Tanaka and Nishii (2015). It should be more desirable to carry out multi-dimensional quantitative analysis, such as dynamic spatial panel analysis (Elhorst, 2014). By taking spatio-temporal changes into account, we can generate more sophisticated analysis for sustainable development based on SDG indicators (UNSTATS, 2018).

For example, we can analyze the temporal relation between volume of production per labor unit by classes of farming/pastoral/forestry enterprise size and output growth of agricultural sector (indicators of **Goal 2.3** and **8.2**) and proportion of land that is degraded over total land area (indicator of **Goal 15.3**). This analysis can monitor the objectives of promoting sustainable agriculture that goes hand in hand with responsible production.

To capture the interrelations between the two indicators, the use of spatial adjacency matrices (W) provides much more precise and sophisticated analysis as stated in the Introduction. The general approach by using shapefiles is failed to project NA cells inside the municipal body, then we provided the utilization of Kronecker product to formulate a rectangular W and to apply the projection matrix to correctly construct the non-rectangular W^*.

Our method provides an eminent efficiency to construct the W. We can generate them more than 2000 times faster and with three times less space than the commonly used R package, 'cell2nb' and 'nb2mat' functions of *spdep*. This efficient process is important to handle huge data size, due to the increase of resolution, especially to deal with the satellite data. This approach would give practical insights for statistical imputation such as NA cell case of the neighbors and spatial interpolation.

Acknowledgements

The authors would like to thank Prof. Koichi Maekawa of Hiroshima University of Economics for the continuous supports during the writing process of this article. This work is supported by JSPS KAKENHI Grant Number JP17K400064.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.

Anselin, L. (2002). *SpaceStat Tutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data*. Technical Report S-92-1 of the National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Arbia, G. (2014). *A Primer for Spatial Econometrics*. Palgrave Texts in Econometrics. Springer.

Baddeley, A., Turner, R., & Rubak, E. (2018). *Handling shapefiles in the spatstat package*. R Vignette Document of *spatstat* Package. Retrieved from https://cran.r-project.org/web/packages/spatstat/vignettes/shapefiles.pdf

Bates, D., & Maechler, M. (2017). *Package Matrix. Sparse and Dense Matrix Classes and Methods*. R Documentation. Retrieved from https://cran.r-project.org/web/packages/Matrix/Matrix.pdf

Bell, K. P., & Irwin, E. G. (2002). Spatially explicit micro-level modelling of land use change at the rural–urban interface. *Agricultural Economics*, *27*(3), 217-232. https://doi.org/10.1016/S0169-5150(02)00079-8

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer.

Center for International Earth Science Information Network - CIESIN - Columbia University. (2017). *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 10*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4DZ068D

Cressie, N. A. C. (1991). *Statistics for Spatial Data*. New York: John Willey & Sons, Inc.

Cressie, N. A. C., & Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. New Jersey: John Willey & Sons, Inc.

Dmowska, A., & Stepinski, T. F. (2017). A High Resolution Population Grid for the Conterminous United States: The 2010 Edition. *Computers, Environment and Urban System*, *61*, 13-23. https://doi.org/10.1016/j.compenvurbsys.2016.08.006

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer-Verlag Berlin Heidelberg, Netherland: Dordrecht.

Environmental System Research Institute (ESRI). (1998). *ESRI Shapefile Technical Description*. ESRI White Paper, New York. Retrieved from https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Fitrianto, G., & Tanaka, S. (2017). Fundamental Review on the Formulation of Large Lattice Spatial Neighbor Matrices. *HUE Journal: Humanities, Social and Science*, *40*(3), 41-88. http://dx.doi.org/10.18996/kenkyu2017400303

Geoghegan, J., Lisa A. W., & Bockstael, N. E. (1997). Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics*, *23*(3), 251-264. https://doi.org/10.1016/S0921-8009(97)00583-1

Irwin, E. G., & Geoghegan, J. (2001). Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment*, *85*(1-3), 7-24. https://doi.org/10.1016/S0167-8809(01)00200-6

Killough, B. (2017). *Open Data Cubes: A Big Data Solution for Global Capacity Building.* Handout for NASA GSFC for “How to Cloud for Earth Scientists” Series. Retrieved from https://earthdata.nasa.gov/files/GSFC%20Open%20Data%20Cube%2012-2017.pdf

Lewis, A., *et al.*. (2017). The Australian Geoscience Data Cube—Foundations and lessons learned. *Remote Sensing of Environment*, *202*, 276-292. https://doi.org/10.1016/j.rse.2017.03.015

Nelson, G. C., & Hellerstein, D. (1997). Do Roads Cause Deforestation? Using Satellite Images in Econometric Analysis of Land Use. *American Journal of Agricultural Economics*, *79*(1), 80-88. https://doi.org/10.2307/1243944

Oak Ridge National Laboratory. (n. a). *Landscan Documentation*. Retrieved from https://landscan.ornl.gov/documentation

ODC Documentation. (2017). ODC Governance Framework: The Open Data Cube Project. *ODC Documentation*. Retrieved from https:/docs.wixstatic.com/ugd/f9d4ea_80d4af32636f45d9a4a9695cec8d29ec.pdf

Open Data Cube. (2018). *CEOS Open Data Cube Whitepaper*. Retrieved from https://www.opendatacube.org/

Stevens, F. R., *et al.*. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. *Plos one*.

Tanaka, S., & Nishii, R. (2009). Nonlinear Regression Models to Identify Functional Forms of Deforestation in East Asia. *IEEE Transactions on Geoscience and Remote Sensing, 47*(8), 2617-2626. https://doi.org/10.1109/TGRS.2009.2015659

Tanaka, S., & Nishii, R. (2015). Incorporation of Gridded Data into the Analysis of Remotely-Sensed Image: Basic Quantitative Strategy to Analyze Deforestation by Population Growth. *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS),* 2548-2551. https://doi.org/10.1109/IGARSS.2015.7326331

United Nations Development Programme (UNDP). (2015). *Transforming our World: The 2030 Agenda for Sustainable Development*. Retrieved from https://sustainabledevelopment.un.org/content/documents/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

United Nations Statistics Division (UNSTATS). (2018). *Global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development*. Retrieved from https://unstats.un.org/sdgs/indicators/Global%20Indicator%20Framework%20after%20refinement_Eng.pdf

Worldpop. (n.d.). *Worldpop: Get Data*. GeoData Institute, University of Southampton. Retrieved from http://www.worldpop.org.uk/data/

Note

Note 1. All the source programs and the **W** and $\mathbf{W}^*$ matrices for the example **Figure 2** are available at: https://github.com/G-Fitrianto/2018-RWE-Article

Appendix. W Formulation for Non-rectangular $\mathrm{D}_{3\times3}$.

The $\mathrm{W}_{9\times9}$ based on rectangular $\mathrm{D}_{3\times3}$ can be constructed by applying equation (4) as follows,

$$\mathrm{W}_{9\times9} = \mathrm{A}_3 \otimes \mathrm{I}_3 + \mathrm{I}_3 \otimes \mathrm{A}_3 + \mathrm{A}_3 \otimes \mathrm{A}_3 = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & & & \\ 1 & 0 & 1 & 1 & 1 & 1 & & & \\ 0 & 1 & 0 & 0 & 1 & 1 & & & \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ & & & 1 & 1 & 0 & 0 & 1 & 0 \\ & & & 1 & 1 & 1 & 1 & 0 & 1 \\ & & & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{A1}$$

The non-rectangular $\mathrm{D}_{3\times3}$ in Figure 3 consist of NA cells and we should exclude them by calculating $\mathrm{W}^*_{6\times6}$ as follows,

$$\mathrm{W}^*_{6\times6} = \mathrm{P}^{\mathrm{T}}_{9\times6}\mathrm{W}_{9\times9}\mathrm{P}_{9\times6} = \begin{bmatrix} 0 & 1 & 1 & 0 & & \\ 1 & 0 & 1 & 1 & & \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ & & 1 & 1 & 0 & 1 \\ & & 1 & 1 & 1 & 0 \end{bmatrix}, \quad \mathrm{P}_{9\times6} = \begin{bmatrix} 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ 0 & 0 & 0 & & & \\ & 0 & 0 & 0 & 0 & \\ & 0 & 1 & 0 & 0 & \\ & 0 & 0 & 1 & 0 & \\ & & & 0 & 0 & 0 \\ & & & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \end{bmatrix} \tag{A2}$$