Study on Bhutan Census Grid Square Data Creation

Kazuyuki Neki

DOI: https://doi.org/10.17102/bjrd.rub.10.2.002

Abstract:

It is highly demanded to urgently maintain accurate GIS statistical data as the importance of planning a sustainable development plan due to widening disparities and inequalities; addressing new social challenges arising from rapid development and urbanization. As a former study of this paper proved the significance of creating the grid square statistics and the effectiveness of adopting the World Grid Square system rather than developing a dedicated system for Bhutan, in the context of the United Nations General Assembly adopted the UN resolution on Global Geodetic Reference Frame for Sustainable Development that was led by the Republic of Fiji in 2015.Currently in Bhutan, important GIS statistical data (e.g. population) is available only in large administrative levels, namely Gewog/Thromde (city-level), however GIS statistical data in smaller levels are required to conduct a detailed study on the new social challenges. This Grid Square system can be quickly adopted from open sources and provide granular information in smaller areas than cities and villages with a precise and globally adopted geodetic reference frame.There are two major objectives of this paper; (1) Creating practical GIS statistic data in the grid square format. (2) Clarifying the advantages, disadvantages and difficulties, in order to evaluate and improve the grid square format. Within the limitation of data availability in Bhutan, this study has assumed the population of each grid square covering the whole census area of Thimphu Thromde by an estimation model using sampled data in Enumeration Areas (EA) of ‘2017 Population & Housing Census of Bhutan’. Further, this paper discusses advantages and disadvantages between EA-units and grid squares on the presumption of taking suppression measurements and resampling from EA-units to grid squares.Consequently, we conclude that using the grid square format in practice could improve the efficiency of GIS analysis and data availability although there is one drawback in accuracy. Moreover, the population of each EA-unit has been predicted by the model comprising GIS-based variables (e.g. distance from facilities). The result has been validated by the whole population in Thimphu thromde. Accessibility to health and educational facilities is the most significant factor for the number of people living in the area.

Keywords; *GIS; GGRF; grid square statistics; standardization; Bhutan census*

Introduction

There is no doubt that information and communication technologies (ICT) have high potential value to make governance more efficient, transparent and inclusive across all sectors in developing countries such as Bhutan (BIPS 2004). Among ICT policy of Bhutan, transparency and accountability, effectiveness and efficiency, and citizen and business centricity in governance were highly prioritized (MoIC 2015). In recent years, Geographical Information System (GIS) is world-widely used for planning, decision making and advocating public in education, industries, and government offices, and in Bhutan, more importantly to provide important information to its citizens and customers.

In other countries such as the United Kingdom, geographically digitizing and providing census statistical data (Census Geography) has started in 1971 (Rhind, D. 1983). The census data are divided into Output Areas and Small Areas (OA) (23,296) and each OA contains 113.8 households on average (Openshaw, S. 1995). In Japan, providing census data has also started in 1969 and its data format was a grid square format. As of now, the smallest size of Japanese grid squares is 125m square (Statistics Japan 2019). On the other hand, thereafter 2013 when the charter of G8 regarding open data, privacy information policies on the GIS statistical data including sensitive data have been globally discussed (Geography Education Working Group 2014). This is because of a lack of policies and regulations for treating sensitive information, specifically in GIS formats as of now. Furthermore, in 2015, the United Nations General Assembly adopted the UN resolution on Global Geodetic Reference Frame (hereinafter referred to as GGRF) for Sustainable Development that was led by the Republic of Fiji. The UN Committee of Experts on Global Geospatial Information Management (UN-GGIM) recognized since its inception the growing demand for more precise positioning services, the economic importance of a global geodetic reference frame and the need to improve the global cooperation within geodesy.

Safeguarding the right to privacy of the individual is a fundamental element in the policy of statistical offices. Several countries have one or the other form of a legal framework dealing with this matter. In some instances, a law bears upon a particular survey, such as the Census of the Population (the United Kingdom). Relevant laws in the United States and the Netherlands have a wider scope (Willenborg 1996). Currently in Bhutan, important GIS statistical data (e.g. population) is available only in large administrative levels, namely Gewog/Thromde (city-level), however GIS statistical data in smaller levels are required to conduct a detailed study. There is not a privacy law and are only limited policies, guidelines and standards related to privacy information in Bhutan (Kuensel 2014).

Therefore, this study aims to contribute to creating and maintaining efficient GIS statistical data as the importance of planning a sustainable development plan due to widening disparities and inequalities; addressing new social challenges arising from rapid development and urbanization. Two major objectives of this paper are (1) Creating practical GIS statistic data in the grid square format and (2) Clarifying the advantages, disadvantages and challenges, in order to evaluate and improve the grid square format by using census data of Bhutan as an example under the limitation of current data availability.

Grid square format

The grid square format is one of GIS statistical formats which store a variety of statistical data such as demographical data (e.g. population), topographical data (e.g. elevation), meteorological and so on, with geographical information (longitude and latitude), inside uniformed size areas, called grid squares, which are divided into the same interval by longitude and latitude without any gap on the earth. Each grid square has an identification code to be immediately recognized where it locates and combine with other statistical data created by the grid square format.

In the previous study (Kazuyuki 2018), the significance of creating the grid square format statistical data has been confirmed in comparison with each feature, advantage and difficulty, and using the World Grid Square System (Aki-Hiro Sato et al 2015 and 2017) was suggested rather than Bhutanese government creating or using their own GIS statistical data system and coordinate system. A digit code (grid square code) is used to identify the location of a specific grid cell. The system gives grid square codes by using from 6 digits (1st level) to 13 digits (6th level). The whole area of Bhutan is covered by 34,457 of the 3rd level grid, 137,828 of the 4th level grid squares, 551,312 of the 5th level grid squares and 2,205,248 of the 6th level grid squares (approximately 1 km^2, 0.26 km^2, 0.0659 km^2 and 0.0165 km^2, respectively).

The previous study[1)] has shown the preferable methodology to convert existing statistical data, such as Bhutan census and Agriculture census of Bhutan, to the grid square format data, with further necessary data and features to improve analysing and planning in private and public sectors in Bhutan. This study has created grid square format data in Thimphu Thromde (the core town area) as a sample region in order to clarify the advantages, disadvantages and difficulties of the grid square format. The original census data includes a variety of significant data such as the number of households, details of family members of each building which are very useful information for planning in public and private sectors. Nevertheless, these data were never disclosed except for the total administrative areas due to comprehending privacy information of respondents. Therefore, the following advantages of the grid square format allow creating population distribution data (i.e. census survey data with geographical information);

Advantage 1; Grid squares can treat sensitive privacy information data (e.g. asset value and the number of disabled) as impersonal statistical data in a specific area.)

Advantage 2; Grid squares will be never affected by administrative boundaries and other boundaries, and can be assembled to form areas/regions reflecting a specific purpose and study area (e.g. villages without any address systems, communities, water catchments and flood prone areas) by the smallest grid squares.

Census data in Bhutan

The first census, which is equivalent to civil registration in Bhutan, was conducted in the 18th century. The first-ever United Nations Standard Population and Housing Census in Bhutan was conducted in 2005 under the directives of the Royal Government of Bhutan that provided a much-needed comprehensive social, economic, and demographic information set at the lowest administrative level for planning and other purposes. The entire authority of the census data and mandates related to the conduct of the decennial population and housing census was transferred to the National Statistics Bureau (NSB). In 2017, the second census survey, the 2017 Population & Housing Census of Bhutan (PCHB), in a series for the country and the first to be conducted by NSB was conducted (RCSC 2012). The 2017 PHCB collected information on the following topics; i. General Demographic Characteristics and Migration; ii. Education and Employment; iii. Fertility of Women Age 15-49 Years; iv. Housing Conditions and Facilities; v. Particulars of the Deceased in Past 12 Months; vi. Household Land, Fruits Trees, Livestock & Poultry Ownership; vii. Household Debts and Income; viii. Psychological well-being. However, the survey data are available in only administrative boundaries such as Dzongkhags, Gewogs and Thromdes. This can mean that it is impossible to analyse detailed data which fall into smaller areas than the boundaries. Therefore, to develop a detailed development and management plan based on geographical features, detailed GIS statistical data is demanded.

This study focused on creating GIS data of population and households from the 2017 PHCB. The following sections will describe detailed methodologies of creating statistical GIS data in a grid square format, then present results of analysis to understand the current situation of dispersing of the population in cities in Bhutan.

Methodology of creating statistical GIS data in grid square format

The data collection and tabulation process of the 2017 PCHB can be separated into mainly two steps namely, EA-unit data tabulation and administrative data tabulation. EA (Enumeration Area) is defined as "a well-delineated territorial unit containing the prescribed number of households" in the 2017 PCHB. An EA-unit is divided so that each EA-unit consists of approximately 30 households which is enough large number for the suppression processing. The study used the EA-units tabulated data to create the grid square format data by the following methodology.

In order to convert from one format to another format, both formats must be tessellated (tiling). A tesselation is a plane filled with flat figures without gaps to cover the entire target area (**Figure 1**). The newly converted tiling format data (i.e. the World Grid Square system) will be estimated by the equation below (e.g. Population):

$$GSpop._i = \sum_{j=1}^{n} EApop._j \times \frac{A_{ij}}{A_j}$$

Where;

is population in Grid Square code
is population in EA-units code
is the area of EA-unit code j within Grid Square
is the area of EA-unit code j

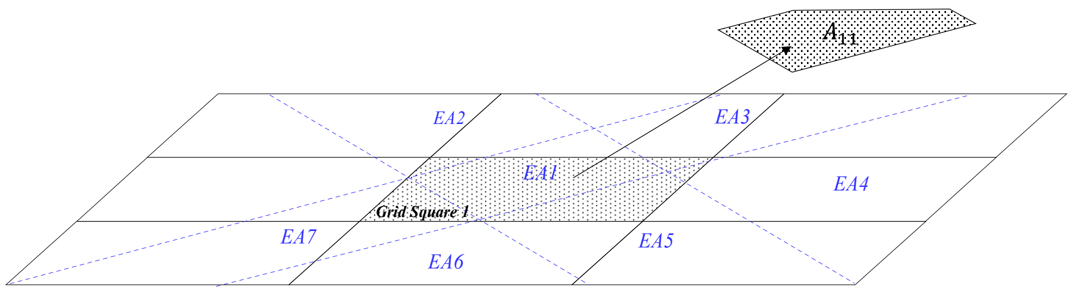

Figure 1: Conversion from EA-units to grid squares

The size of grid squares used in this study is the 6th level of the World Grid Square system (the length of its one side is approximately 143m and the area of one grid square is approximately 0.0165 km^2) which is smaller than EA-units and to include usually enough large of some houses in Thimphu Thromde in Bhutan to de-identify privacy information of each household. Both of the lengths and the areas are calculated based on a global standard coordination system, WGS 1984 UTM Zone 46N (WKID: 32646, Authority: EPSG).

Advantages and disadvantages of conversion

Using the grid square format in practice could improve the efficiency of GIS analysis and data availability, however there is one drawback in accuracy. Specifically in this study case, building-related information including the number of households and population may be distorted to some extent because the EA-unit format which the previous format before the conversion has already been tabulated from the raw data. This is because the raw data contains privacy information, therefore the individual household data has been aggregated so that the sensitive information of individual houses cannot be identified.

The general pros and cons of using an EA-unit and a grid square format are listed in **Table 1**. Although terms of data accuracy may include some distortion mainly due to the previous discussion by cases of specific conversion, the advantages including ease of analysis and improving detail data availability weigh the disadvantage. For example, the advantages of grid square in overlay analysis and data accuracy are described in **Figure 2**Figure 3**.** Due to its larger size and predefined form of areas, the EA-unit analysis will not be able to segment to fit the shape of the inundated areas while the grid square analysis will be able to resemble the shape of the analysis target by combining small grid squares, allowing for more detailed analysis.

Table 1: Comparison between an EA-unit and a grid square format

Features	**EA-unit**	**Grid square**
Overlay analysis	△ Conversion is needed	O Straight forward
Data accuracy	O Exactly but as EA-units only	△Depends on process
Data confidency	× None in small areas	O Can be processed (high potential)

The disadvantage of the inaccuracy of data can be beneficial when one single grid square cannot hide individual household information because there are only a few households in the grid square. In **Figure 2**, several grid squares have only one house. If the population data is created for these grid squares, the number of household members will be explicitly exposed. For that areas, suppression measures should be taken to numerically process to hide the individual household's information. These detailed methodologies will be discussed in the following sections.

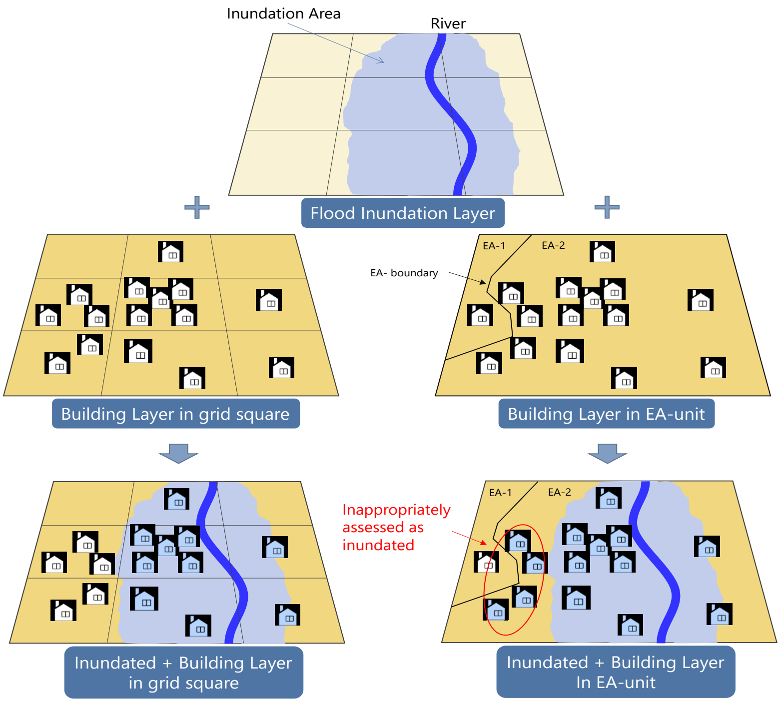

Figure 2: Example of Overlay Analysis and Data accuracy for flood risk analysis (Grid Square versus EA-unit)

Privacy information regulations

The importance of privacy information protection has been declared in *Bhutan Information Communications and Media Act 2018* that has been amended from the previous *Bhutan Information, Communications and Media Act 2006*, the privacy information collected by any kinds of agencies have to be deliberately prevented from misuse. An ICT and Media facility or service provider and vendor shall limit the collection, use and disclosure of personal information, to that which a reasonable person would consider appropriate in the circumstances. Also in the 2017 PCHB regulations, EA-unit tabulated data of the 2017 PCHB are unable to be disseminated and shared even with other Bhutanese government agencies. Similarly, the surveyors and analysts shall limit the collection and use of the data related to the 2017 PCBH to only the intended purposes of the census survey (RCSC 2012). This lack of standardisation (i.e. ambiguous or eliminating any possibility of sharing) and policies related to sensitive data sharing makes it more difficult to share the important data with relevant agencies and the private sector for decision making because it is safest not to share due to the difficulties of judgment.

As examples of Japan which is one of the countries originally adopting the grid square format to census data, in the proposed guideline for treating GIS data in Japan, the precedents of violating individual privacy information policies are shown as follows: i, Personal name; ii, Date of birth, age, contacts (address, contact number and e-mail address), permanent domicile, educational background, stamp seal, place of employment and so forth; iii, Personal mobile number; iv, Medical record; v, Employment management record; vi, Personal image (NTT docomo 2019). Thus, the sharing of data is facilitated by defining common and distinct items of personal information in the law in Japan.

Since such detailed GIS statistical data have not been publicised in Bhutan, any specific regulation related to GIS statistical data does not exist at present. To propagate the GIS statistical data in Bhutan, these regulations treat sensitive information as the GIS statistical data will be required.

Anonymisation measurement for the grid square format

The privacy information must be protected otherwise the data cannot be used for any purposes in evidence-based planning when the grid square format treats sensitive information (e.g. family members and so forth). Taking anonymisation measures, which process GIS statistical data so that individuals can no longer be identified in the resulting data by numerical treatments, have to be taken into consideration so that the useful GIS statistical data (i.e. census) will be utilised in the grid square format in a wide range of circumstances and works without breaching data protection principles (Peter 1995).

To treat sensitive information, there are mainly two ways: pre-processing and post-processing. Obtaining the consent of a person before interviewing and conducting surveys is a primary example of pre-processing. On the other hand, post-processing comprises three steps; namely de-identification, aggregation and then suppression processing (Wellenborg 1996). For instance, in Japan where its census data are maintained in the grid square format, some of the suppression processing have been taken for suppressing the number of population in the case that the size of samples was less than a certain amount. The subtracted number of population in the grid squares was added to the greater size of grid squares that contains the smaller grid squares whose population was subtracted.

In post-processing, firstly, de-identification, which is also known as data anonymization or classification, means the process to prevent data including personal identities from being linked to statistical data. With keeping each record of data, privacy information (e.g. personal name) will be removed and some sensitive information (e.g. age) will be classified into clusters (e.g. 30-40 years old). Secondly, aggregation means to tabulate the de-identified data by the clusters and the areas (i.e. EA-unit or grid squares) and makes the GIS statistical data easy to analyse. However, in this stage, there are possibilities that the individual's data can be explicitly identified because there are only a few original geocoded data inside the area. Thirdly, in such a case, an analyst has to take suppression processing.

Compared to tabular formats, the types of suppression processing are limited in GIS statistical formats due to less flexibility of geographical information which is able to be identified with other sources (i.e. the number of households by satellite images). Possible measures for suppression in the GIS statistics are listed in **Table 2. Figure 3** presents an example of suppression processing (aggregation and segregation) that when one area has only a few households. In the suppression process, at first peripheral areas will be aggregated as one large size area then the large area will be segregated into the original areas with the average number of households. Another example is simply to be replaced with no data in the area which has only a few samples. In this context, the accuracy of the GIS statistical data will be more or less distorted. Therefore, the deducted value shall be added to an adjacent grid square or an upper-level grid square so that the total number of data can be maintained as the original. In recent years, these processes can be used in data collection of a person's locations and movements by mobile phones with GPS functions (Open Knowledge International 2019).

Table 2: Possible suppression processing in the GIS statistical data

Name	Procedure	Remarks
Exclusion	To hide the value of the grid square from the statistics.	To prevent the hided value from being re-identified by back calculating the total, few more values have to be hidden. (secondary suppression/ complementary suppression)
Aggregation and segregation	To aggregate the value with its peripheral grid square and then redistribute to each area.	Peripheral areas will be modified as well.

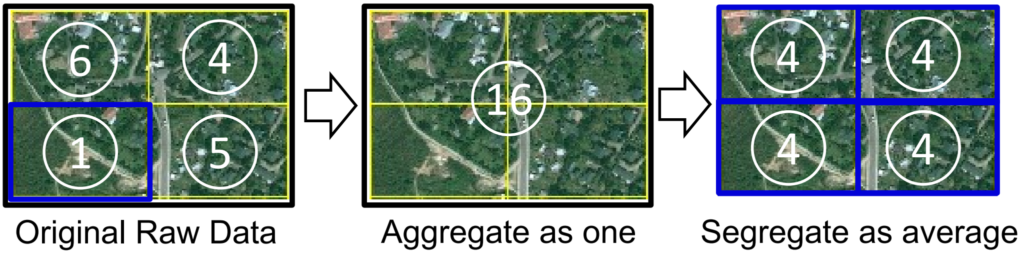

Figure 3: Example of suppression processing

On the other hand, other ordinal suppression processing such as replacing the values with range values and rounding the values are unsuitable to such the number of population in the GIS statistical data because these methods will affect to the total number of the statistics. For the same reasons, data swapping and imputation which are a way of completely changing the total value is not applicable to the GIS statistical data as well.

Criteria for anonymisation measurement

In this section, we discuss the criteria to determine which a gird square has to be suppressed in terms of no other data (e.g. commercial) but census data. Generally, linear sensitivity measures, which mean that the union of some safe grid squares will be also safe, are used.

The prior posterior or n-k dominance rule is most commonly used to assess whether the values have to be suppressed and other common measures for the sensitivity assessment are p-q rule, the p-percent rule. The prior posterior rule and the p-percent rule define a cell as unsafe if a lower or an upper bound of one of its contributions can be calculated lying within p percent of the true value. According to the dominance or (n, k)-rule, a cell is primarily unsafe when the total of the n largest contributions is greater than k percent of the cell value. This rule is based on the idea that a coalition of n-1 respondents in a cell may not be able to estimate an n-th contribution with certain accuracy (Atsuhiro 2003 and Loeve 2001).

However, in the context of current Bhutanese society and especially the census data, taking the sensitivity measures to suppress population from neighbour respondents can be negligible. Therefore, only the following criteria can be considered; the grid square may be defined as unsafe when there are only one or two respondents (households) contributing to the grid square value. The idea is that these few respondents' contributions can be estimated fairly accurately if the value of the grid square is known.

In GIS statistics data, there is the same problem with table forms, called holding problems, which is caused by respondents contributing to different areas (i.e. grid squares). For example, a business owner with different branches in different areas may contribute to different areas. However, in a marginal area, these contributions can be seen as coming from the same business owner, so they have to be aggregated and counted as one contribution in other statistics. This will become important when dealing with such data in a GIS statistics format. Most rules for determining which areas are unsafe will use information about the number or the size of the contributions in an area.

Estimated population distribution in Thimphu Thromde

Converting the 2017 PCHB data from the undisclosed EA-unit format to the grid square format, it is necessary to obtain a certain amount of data sets. However, because of the policy of the 2017 PHCB, the full EA-unit tabulated data are inaccessible for any other purpose. As the results of the request for the 2017 PHCB data in the EA-unit format to NSB, a hundred EA-unit data randomly sampled from Thimphu Thromde comprising of the total number of households and population in each EA-unit area from the 2017 PHCB have been provided for only the single purpose of this study. The sampled data includes information as follows: i, Unique EA code; ii, Number of households in each EA-unit; iii, Total Population in each EA-unit; iv, School age population (6-18 years) in each EA-unit. By combining the sampled data and other GIS-based data such as the locations of educational institution, historical places and so forth, a population estimation model has been developed in order to fill in other EA-unit area's information. Furthermore, we consider the relationships between the GIS-based variables of urban facilities and the population distribution in the urban area of Bhutan.

A multivariate regression model has been selected because this model can clarify relationships between population and quantity and qualitative variables. The general equation of regression model is shown below:

$$y = ax_1 + bx_2 + cx_3 \cdots$$

To develop a multivariate regression model for population distribution in Thimphu Thromde, Variables in each EA-unit and grid square have been collected with GIS processing based on limited data availability and listed in **Table 3**. Regarding the MoWHS data, we requested the ministry to provide the latest data. As per the data records, the latest year of construction of facilities is 2008 (e.g. Babesa Middle Secondary School and so forth). Different data sources can pose different conditions of the data in terms of time and accuracy. Based on these facilities ’ locations, two types of variables have been created. One is if a facility is located within an EA-unit/a grid square. The other is the distance between the nearest facility and the weighted centre point of an EA-unit/a grid square. And dummy variables of Core town, Tashichoe Dzong area, Urban Village precincts and each district based on 2017 PCHB as well as grouped areas such as busy or peripheral areas have been created.

The final model has been developed as an objective variable is the population of an EA-unit/a grid square because population showed the best fitness among the number of households and school age population (6-18 years) and above. An intercept of models is defined as 0 because when the area or the number of houses is 0, the population has to be 0. A variable of the bus stops has been removed from the result because of its low relevance.

Table 3: Variables and data sources for the population distribution model

Sl.	Name	source
1	Number of houses	2017 PHCB EA maps (NSB)
2	Area of EA-unit/grid square	2017 PHCB EA maps (NSB)/Author*
3	Historical places	MoWHS
4	Hospitals/clinics	MoWHS
5	Education institutions	MoWHS
6	Bus stops	Bhutan Post Office (April 2019)
7	Average slopes	5m resolution DEM (Thimphu Thromde)

*Measured by ArcMap as the WGS1984 UTM 46N coordinate system

Table 4 presents the results fo the models. The variables of History, Health and Education mean whether historical places, hospitals/clinics and educational institutions are located inside EA-units or grid squares. In the regression model result, Significance F, which stands for the probability that the null hypothesis in the regression model cannot be rejected, is far smaller, which means the model is significant. The necessary sample number is defined as the number of factors for the model multiplied into 10 (Peter 1995). In this study, the number of samples is 100, therefore the number of variables can be of utmost 10. Model 3 and 4 can be considered as acceptable. The adjusted R square indicates more than 0.98 in all four models and the estimated whole population in Thimphu Thromde is 112,773 which is nearly equivalent to 114,551 of the result of the PCHB 2017.

Accessibility to health and educational facilities is the most significant factor for the number of people living in the area aside from physical factors such as the areas and number building of EA-units.

Table 4: Results of the population density model

	Log Total Population			
	1	2	3	4
Intercept	0	0	0	0
Number of houses	-0.008	-0.005		
	(0.006)	(0.006)		
	-1.379	-0.839		
	(0.171)	(0.403)		
Area(Log)	0.234**	0.191*	0.088**	0.093***
	(0.106)	(0.107)	(0.036)	(0.029)
	2.205	1.794	2.476	3.171
	(0.030)	(0.076)	(0.015)	(0.002)
Density(Log)	-0.155	-0.126		
	(0.110)	(0.112)		
	-1.407	-1.125		
	(0.163)	(0.263)		
History	-0.203	-0.187	-0.247	
	(0.297)	(0.303)	(0.291)	
	-0.682	-0.616	-0.851	
	(0.497)	(0.539)	(0.397)	
Health	0.249	0.307	0.312	
	(0.296)	(0.301)	(0.294)	
	0.840	1.022	1.064	
	(0.403)	(0.309)	(0.290)	
Education	0.279	0.345*	0.357*	0.343*
	(0.203)	(0.205)	(0.192)	(0.186)
	1.377	1.681	1.855	1.841
	(0.172)	(0.096)	(0.067)	(0.069)
Distace from historical places(Log)	0.112**	0.146***	0.138***	0.150***
	(0.055)	(0.054)	(0.050)	(0.046)
	2.039	2.724	2.764	3.226
	(0.045)	(0.008)	(0.007)	(0.002)
Distance from hospital(Log)	0.035	0.043	0.043	
	(0.056)	(0.056)	(0.054)	
	0.629	0.764	0.799	
	(0.531)	(0.447)	(0.426)	
Distance from educational institutions(Log)	0.078	0.097*	0.095*	0.108**
	(0.054)	(0.054)	(0.051)	(0.048)
	1.437	1.780	1.852	2.251
	(0.154)	(0.078)	(0.067)	(0.027)
Total length of roads within 250m(Log)	0.039			
	(0.038)			
	1.024			
	(0.308)			
Total length of roads within 375m(Log)	-0.202**			
	(0.086)			
	-2.356			
	(0.021)			
Total length of roads within 500m(Log)	0.511***	0.318***	0.326***	0.338***
	(0.095)	(0.044)	(0.039)	(0.036)
	5.373	7.270	8.272	9.314
	(0.000)	(0.000)	(0.000)	(0.000)
Average slope of the area(Log)	-0.029	-0.008		
	(0.088)	(0.088)		
	-0.333	-0.095		
	(0.740)	(0.925)		
Multiple R	0.998348259	0.998238575	0.998213219	0.998179629
R Square	0.996699246	0.996480254	0.996429631	0.996362572
Adjusted R Square	0.984749716	0.984848821	0.985288408	0.985683101
Standard Error	0.320768266	0.327495323	0.324419555	0.322240289
Observations	100	100	100	100
df	13	11	8	5
SS	2703.043005	2702.4491	2702.311813	2702.129947
MS	207.926385	245.6771909	337.7889766	540.4259894
F	2020.816174	2290.628365	3209.455993	5204.470468
Significance F	5.9578E-101	4.1602E-103	4.7915E-108	3.4787E-113

Note: Std. errors in parentheses *p<0.1; **p<0.05; ***p<0.01

Advantages, disadvantages and challenges of using a grid square format

The result of that the estimated population distribution in EA-units has been converted to the 6th level grid square (approximately 143m × 143m) is shown in **Figure 4**. The criteria of anonymisation measures will be less than 9 people which is equivalent to less than 3 households because the number of people per household in 2017 is 4.19 (The number of persons and households in Thimphu Thromde in 2017 is 106,487 and 25,408 (PHCB 2017)). The ratio of areas where only less than 9 people existed is going to be 22.6% which can require anonymisation measures. The grid squares may require anonymisation measures (green colour grid squares in **Figure 5**) are located only at peripheral areas and Tashichoe Dzong area in Thimphu Thromde.

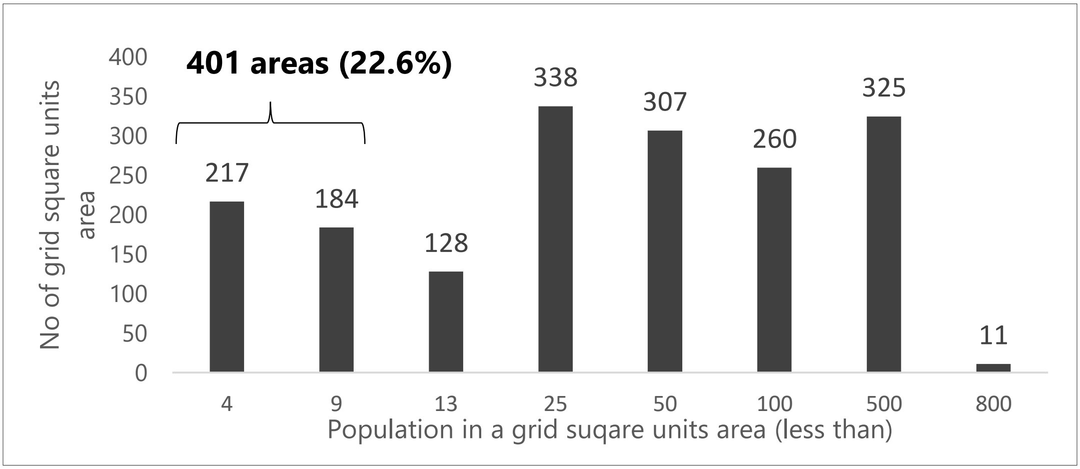

Figure 4: The number of population in a grid square (6th level)

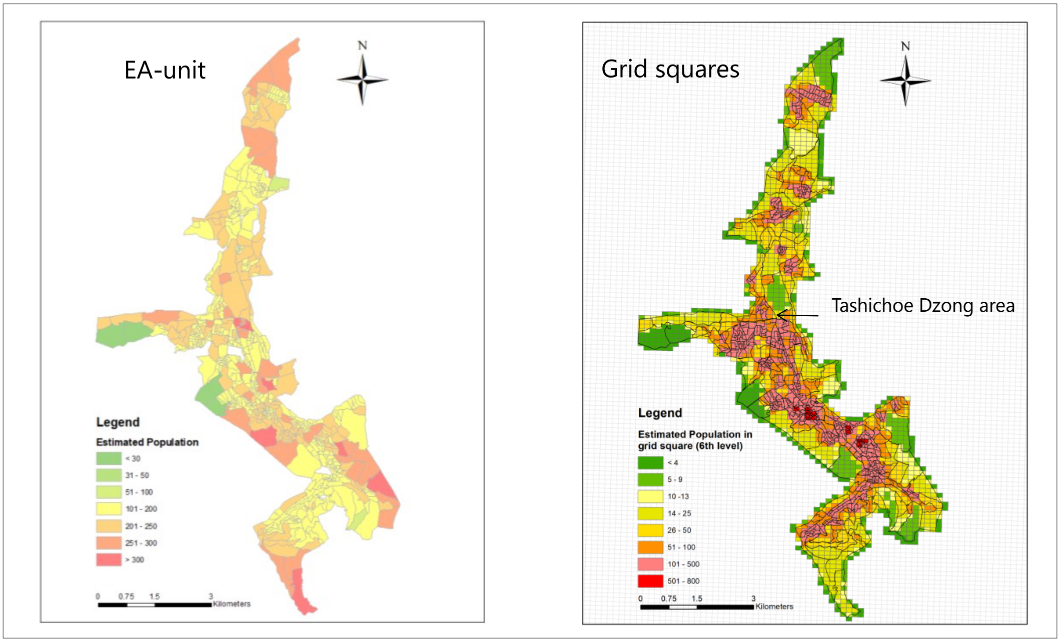

Figure 5: The PHCB 2017 population in EA-units and grid squares

There are pros and cons of converting an EA-unit format to a grid square format in the PHCB 2017. The advantages are, as already mentioned, primarily confidential data can be treated and making overlay and chronological analysis easier to integrate with other statistical data, which are uniform and of the same grid square size, in GIS processing. Furthermore, the grid squares displayed a more detailed population distribution and directly described the density of population in Thimphu Thormde.

The disadvantages are that anonymisation measures will be required on approximately one fifth of the grid squares and this can cause distortion of data especially in 6th level of grid squares. In reality, the distance between buildings in Bhutan is relatively wider specifically in peripheral areas even in Bhutan, which means that the size of grid squares has to be adjusted depends on the density of households in areas. Consequently, although in terms of judging if anonymisation measures are required and if the size of grid squares has to be adjusted, certain amount of manual work will be required, the methodology of converting the PHCB 2017 data from EA-units to grid squares can be adopted and bring benefits.

Conclusion

We discussed a course of action to incorporate policies to Bhutan based on current trends regarding GIS statistical data including sensitive information in the United Kingdom, Netherlands and Japan. Consequently, we conclude that using the grid square format in practice could improve the efficiency of GIS analysis and data availability although there is a possibility to distort the accuracy of data especially in peripheral areas of a town or rural areas because anonymisation measures have to be taken to suppress sensitive data.

As per the multivariate regression model for estimating the distribution of population in Thimphu Thromde by using the PHCB 2017 data and GIS-based variables, accessibility to health and educational facilities is the most significant factor for the number of people living in the area except for physical factors such as the areas and number building of EA-units. The results of the model analyses can be used for constructing similar analyses to identify the key elements for reducing inequality of development, and for examining how establishing or relocating facilities would impact peripheral demographic characteristics, by layering various GIS data.

This study has suggested the issues involved in not only creating GIS grid data but also promoting public data sharing in Bhutan. Establishing standardized data sharing policy and methodologies will allow collecting cross-disciplinary data to be collected over a longer period of time and on a broader scale. When these data are cumulated, it will be possible to plan from a more multifaceted and detailed perspective and achieve reducing inequality and efficient planning.

References

Kazuyuki N. (2018). Significance and methodology of creating grid square GIS statistical data in Bhutan. *Knowledge for Development in Bhutan – Messages of Japanese Volunteers from the Ground*, 20-33

Aki-Hiro Sato, Shoki Nishimura and Hiroe Tsubaki (2017). World Grid Square Codes: Definition and an example of world grid square data. *2017 IEEE International Conference on Big Data (Big Data)*, 10-14 Dec. 2017, pp. 4156-4165.

Aki-Hiro Sato and Hiroe Tsubaki (2015). Application of grid square statistics and its expansion to world grid square statistics, *6th conference of Transdisciplinary Federation of Science and technology*, pp. 266-272. (In Japanese)

National Statistics Bureau (2018). *Population & Housing Census of Bhutan 2017*.

Office of the Census Commissioner Royal Government of Bhutan (2006). *Population & Housing Census of Bhutan 2005*.

Kuensel (2015).*Digital Privacy: Issues and Challenges in Bhutan*. Retrieved from http://www.kuenselonline.com/digital-privacy-issues-and-challenges-in-bhutan/

Royal Government of Bhutan (2004). *Penal Code of Bhutan*. Retrieved from https://www.rbp.gov.bt/Forms/PCB%202004.pdf

National Statistics Bureau of Bhutan (2018). *2017 Population & Housing Census of Bhutan*. Retrieved from http://www.nsb.gov.bt/publication/files/PHCB2017_national.pdf

Royal Civil Service Commission, Royal Government of Bhutan (2012). *Bhutan Civil Service and Regulations 2012*. Retrieved from http://www.nationalcouncil.bt/assets/uploads/docs/download/2015/BCSR2012.pdf

Cabinet Secretariat of Japan (2010). *'Chirikuukanjouhou no katsuyou ni okeru kojinjohou no toriatsukai ni kansuru guideline'*. (In Japanese) Retrieved from http://www.mlit.go.jp/common/000123331.pdf

Rhind, D. (1983). *A census user's handbook*. Methuen, 1983.

Openshaw, S. ed. (1995). *Census Users' Handbook*. Wiley, 1995.

Statistics Japan. Historical Background. *Grid Square Statistics*. Retrieved April 19, 2019 from https://www.stat.go.jp/english/data/mesh/index.html

Peter Peduzzi, John Concato, Alvan R. Feinstein, Theodore R. Holford (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology, Volume 48*, 1503-1510.

Statistical Information Institute for Consulting and Analysis (2017). *'2017 Kokuzeichousa ni kansuru tiiki mesh toukei wo riyoushita tiikibunseki'*. (In Japanese) Retrieved from http://www.sinfonica.or.jp/topic/h27houkoku/keisanhouhou.pdf

Wellenborg, L. and De Waal, T. (1996). *Statistical Disclosure Control in Practice*. Lecture Notes in Statist., 111, Spring-Verlag, New York.

Loeve, J. A.(2001). Notes on sensitivity measures and protection levels, Division of Technology and Facilities, Methods and informatics department, Statistics Netherlands, Voorburg. Retrieved from http://neon.vb.cbs.nl/casc/related/marges.pdf

Atsuhiro Taki（2003）. 'Syukeihyo ni okeru cell hitokumondai to sono kennkyuu doukou'. Proceedings of the Institute of Statistical Mathematics, Volume 51-2 (2003), 337-350. Retrieved from

https://www.ism.ac.jp/editsec/toukei/pdf/51-2-337.pdf

UN-GGIM http://ggim.un.org/, Accessed on April 15, 2019.

Research Institute for World Grid Squares http://www.fttsus.jp/worldgrids/en/top/, Accessed on April 15, 2019.

Grid Square Statistics, Statistics Bureau under Ministry of Internal affairs and Communications http://www.stat.go.jp/english/data/mesh/index.html, Accessed on April 15, 2019.

NTT docomo, inc. *'Mobile geospatial statistics guideline'*. (In Japanese) Retrieved April 15, 2019 from

https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/guideline/index.html

Open Knowledge International. *Open Data Handbook*. Retrieved April 15, 2019 from

http://opendatahandbook.org/

Ministry of Works and Human Settlement, Royal Government of Bhutan. (2008). *Thimphu City Development Strategy*.

Royal Government of Bhutan (2018). *Bhutan Information Communications and Media Act of Bhutan 2018*.

Royal Government of Bhutan (2004). *Bhutan Information and Communications technology policy and Strategies*. Retrieved from

http://www.dit.gov.bt/sites/default/files/BIPS%20Final%20Report%20-%20v5.1.pdf

Ministry of Information and Communications (2015). *Revised Bhutan ICT Roadmap*. Retrieved from

https://www.moic.gov.bt/wp-content/uploads/2016/05/bhutan_ict_roadmap_2015_pdf_85407.pdf

Geography Education Working Group (2014). 'Tirikyoiku niokeru open data rikatsuyou to tizuryoku/GIS ginou no ikusei'. Retrieved from

http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t199-3.pdf

About the Author

Kazuyuki Neki has about 6 years of experience in the areas of infrastructure planning and policy development. He started his professional career as a consultant engineer in Japan and has developed many infrastructural policies. He was a Japan Overseas Cooperation Volunteer assigned to the Ministry of Works and Human Settlement from July 2017 to July 2019. He holds a MEng. in civil engineering and Professional Engineer of Japan in civil engineering.