Puttanapong, Nattapong et al.

Working Paper

Predicting poverty using geospatial data in Thailand

ADB Economics Working Paper Series, No. 630

Provided in Cooperation with:
Asian Development Bank (ADB), Manila

Suggested Citation: Puttanapong, Nattapong et al. (2020) : Predicting poverty using geospatial data in Thailand, ADB Economics Working Paper Series, No. 630, Asian Development Bank (ADB), Manila, https://doi.org/10.22617/WPS200434-2

This Version is available at:
https://hdl.handle.net/10419/246707

PREDICTING POVERTY USING GEOSPATIAL DATA IN THAILAND

Nattapong Puttanapong, Arturo M. Martinez Jr., Mildred Addawe, Joseph Bulan, Ron Lester Durante, and Marymell Martillan

NO. 630

December 2020

ADB ECONOMICS WORKING PAPER SERIES

ASIAN DEVELOPMENT BANK

ADB

ADB Economics Working Paper Series

Predicting Poverty Using Geospatial Data in Thailand

Nattapong Puttanapong, Arturo M. Martinez Jr., Mildred Addawe, Joseph Bulan, Ron Lester Durante, and Marymell Martillan

No. 630 | December 2020

Nattapong Puttanapong (nattapong@econ.tu.ac.th) is an assistant professor at Thammasat University Faculty of Economics. Arturo Martinez, Jr. (amartinezjr@adb.org) is a statistician. Joseph Albert Nino Bulan (jbulan@adb.org) is an associate statistics analyst. Mildred Addawe (maddawe.consultant@adb.org), Ron Lester Durante (rdurante1.consultant@adb.org), and Marymell Martillan (mmartillan.consultant@adb.org) are consultants at the Economic Research and Regional Cooperation Department of the Asian Development Bank.

This paper was prepared as background material for the report *Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific*. The authors thank colleagues from the National Statistical Office of Thailand who closely worked with the project team.

ASIAN DEVELOPMENT BANK

6 ADB Avenue, Mandaluyong City, 1550 Metro Manila, Philippines
Tel +63 2 8632 4444; Fax +63 2 8636 2444
www.adb.org

ISSN 2313-5867 (print), 2313-5875 (electronic)
Publication Stock No. WPS200434-2
DOI: http://dx.doi.org/10.22617/WPS200434-2

CONTENTS

TABLES AND FIGURES

TABLES

FIGURES

ABSTRACT

Poverty statistics are conventionally compiled using data from household income and expenditure survey or living standards survey. This study examines an alternative approach in estimating poverty by investigating whether readily available geospatial data can accurately predict the spatial distribution of poverty in Thailand. In particular, geospatial data examined in this study include night light intensity, land cover, vegetation index, land surface temperature, built-up areas, and points of interest. The study also compares the predictive performance of various econometric and machine learning methods such as generalized least squares, neural network, random forest, and support vector regression. Results suggest that intensity of night lights and other variables that approximate population density are highly associated with the proportion of an area's population who are living in poverty. The random forest technique yielded the highest level of prediction accuracy among the methods considered in this study, perhaps due to its capability to fit complex association structures even with small and medium-sized datasets. Moving forward, additional studies are needed to investigate whether the relationships observed here remain stable over time, and therefore, may be used to approximate the prevalence of poverty for years when household surveys on income and expenditures are not conducted, but data on geospatial correlates of poverty are available.

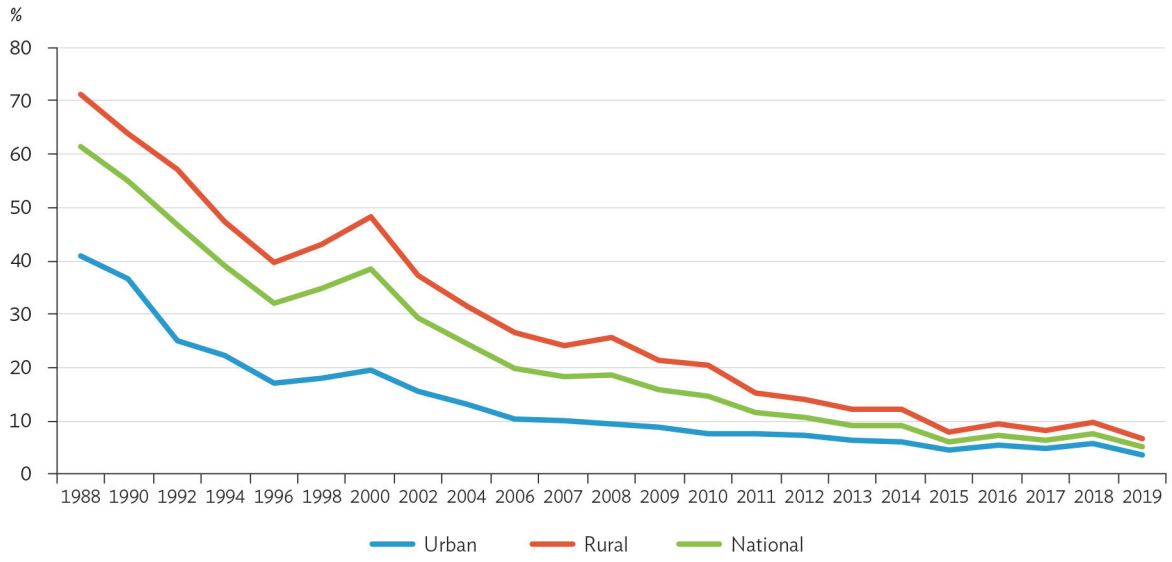
Keywords: big data, computer vision, data for development, machine learning algorithm, multidimensional poverty, official statistics, poverty, SDG, Thailand

JEL codes: C19, D31, I32, O15

I. INTRODUCTION

Over the past 3 decades, real gross domestic product per capita of Thailand has grown by more than twofold. The country's economic growth is accompanied by declining household poverty rates, dropping from 61.41% in 1988 to 5.04% in 2019 (Figure 1). However, there are areas in the country where significant pockets of poverty still exist. For instance, about 6.76% of households living in rural areas are still considered poor (NSO 2020). Furthermore, the pandemic brought by the coronavirus disease (COVID-19) may undermine some of the gains in poverty reduction in the country.

Figure 1: Proportion of Thailand's Population Living Below its National Poverty Line

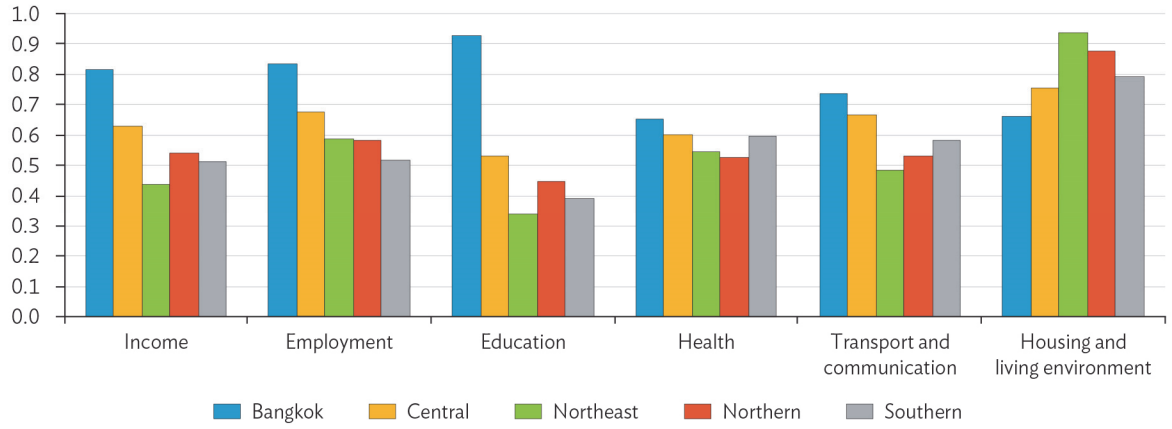


Notes: Thailand's poverty line is calculated based on the minimum standard required by an individual to fulfill one's basic food and nonfood commodities. Details are provided in the National Economic and Social Development Council (2015).

Source: National Statistical Office of Thailand.

The concentration of poverty in rural areas is possibly driven by Bangkok's high agglomeration force and the fact that most economic activities are concentrated in Bangkok and its suburbs (OECD 2018). Since rural provinces have a limited variety of economic activities, they have a constraint of creating nonagriculture jobs. Trends in nonpecuniary indicators of development are also concerning (Figure 2). For instance, half of the country's working population are still in precarious employment. There is also ample room for improvement in the education sector as rural migrants and urban poor generally lack the skills demanded by modern jobs.

Figure 2: The Human Achievement Index in 2017



Note: The Human Achievement Index value ranges from 0 (worst outcome) to 1 (best outcome).
Source: National Economic and Social Development Council. 2017. *Human Achievement Index Report 2017.* Bangkok.

Given that poverty is still prevalent in select areas of Thailand, particularly rural areas, poverty monitoring remains an important task for the country's development practitioners. At present, the National Economic and Social Development Council and National Statistical Offices are the government agencies responsible for compiling poverty statistics in Thailand. In particular, these agencies rely on the Household Socio-economic Survey (HSES), in which the household's income is surveyed every two years. In addition, HSES results are compiled at the national and provincial levels only, as the sample size of the aforementioned survey is not large enough to yield reliable estimates beyond provincial level, yet demand for granular data on poverty and other socioeconomic indicators continues to expand.

Given the high cost of increasing the frequency and scope of such surveys, an option which government agencies with limited resources may be unable to sustain in the long run, a number of initiatives within Thailand and elsewhere are exploring alternative data collection strategies which entail tapping other data sources. For instance, small area estimation techniques which combine survey with census and other types of administrative data have been widely used to facilitate estimation at levels more granular than what working with surveys alone can afford. More recently, efforts to use innovative data coming from call detail records, social media data, digital transactions, and remote sensing for compilation of development statistics are expanding too.

Mapping the spatial distribution of poverty is an area which could greatly benefit from the integration of multiple data sources. In this context, two types of analytical frameworks are worth pointing. First, by capitalizing on ongoing developments in computer vision techniques and satellite imagery, several researchers have shown that it is feasible to develop an algorithm that can automatically predict survey-based estimates of poverty with satisfactory levels of accuracy (Jean et al. 2016; Hofer et al. 2020; ADB 2020; Piaggesi et al. 2019). Such approach is quite attractive for instances wherein collecting survey data, particularly in remote and/or hard-to-reach areas, is onerous, and no other types of supplementary data are readily available. However, since the features extracted by computer vision techniques are relatively abstract (ADB 2020), it is difficult to manually pinpoint exactly which features are being picked up by the computer when predicting poverty. Consequently,

it is also difficult to validate what could have triggered an unexpectedly low or high estimate of poverty, if such instances arise. Alternatively, if structured geospatial data are readily available, one can develop a more tractable econometric model for predicting poverty. Whereas in the first approach, one is letting the computer extract abstract features or patterns from satellite images that are potentially correlated with poverty, in the second approach, one is leveraging on quantitative data that have already been precompiled. In principle, the former approach can draw vast number of potential predictors of poverty as a computer can directly extract numerous features from the satellite images whereas the latter approach is limited to structured precompiled data. However, presumably, it is easier to validate the results of the second approach since the data used for predicting poverty are more structured and tractable.

This study explores the second approach where poverty is predicted by identifying correlates from precompiled geospatial data. It contributes to the existing literature by assessing whether it is feasible to develop a model with satisfactory predictive performance even if we are solely depending on precompiled geospatial datasets which theoretically, can be considered as just a fraction of the number of covariates that the first approach can potentially generate, a feat that has not been explored thoroughly in the context of Thailand, in previous studies. Up to some extent, it may be considered as follow up to the ADB (2020) study which examined the first approach of using geospatial data to predict poverty. However, there are slight differences in research objectives. Whereas, the main objective of the ADB (2020) study is to examine the feasibility of providing poverty estimates that are more granular than government-published estimates by using artificial intelligence on data from survey, census, and satellite imagery, our focus here is to examine whether we can develop a reasonably good poverty prediction model even if we limit ourselves in using covariates from readily available or precompiled geospatial dataset(s). Furthermore, this study's objective is to briefly compare the performance of different machine learning techniques, a topic that has not been well explored in previous studies of poverty estimation using nontraditional data sources. By doing so, the study aims to contribute to the literature that explores other cost-effective methods of predicting poverty through integrating the use of innovative data with surveys and register-based data which in turn, could provide rich inputs as relevant government agencies aim to meet the growing data requirements of economic planners and policymakers.

The rest of this paper is structured as follows. The second section reviews related literature while the third and fourth sections introduce the data and research methodologies, respectively. The fifth section presents the key findings of the econometric and machine learning methods adopted in this study. The last section summarizes lessons learned and draws brief recommendations for future studies.

II. LITERATURE REVIEW

Using Precompiled Geospatial Data for Predicting Socioeconomic Indicators

The existing literature offers a wide range of case studies showcasing various applications of satellite imagery and geospatial data for development-related analyses. For instance, data on night light intensity compiled through Defense Meteorological Satellite Program Operational Linescan System and Suomi National Polar-Orbiting Partnership Visible and Infrared Imaging Radiometer Suite are widely used. Elvidge et al. (1997); Doll, Muller, and Elvidge (2000); Doll, Muller, and Morley (2006);

Sutton, Elvidge, and Tilottama (2007); Chen and Nordhaus (2011); Henderson, Storeygard, and Weil (2012); Bickenbach et al. (2013); Forbes (2013); Li et al. (2013); and Li, Zhao, and Li (2016) found statistically significant relationship between the density of night lights and various ground data such as gross domestic product, electricity consumption, inequality, and infant mortality rate.

In addition to night lights intensity, Landsat, National Oceanic and Atmospheric Administration Polar Orbiting Environmental Satellites and Terra-Moderate Resolution Imaging Spectroradiometer satellites have been scanning the Earth's surface with multi-spectrum sensors. These multi-spectrum data have been used by various researchers to compile a number of geospatial indicators such as the building density, water coverage, Normalized Difference Vegetation Index (NDVI), Land Surface Temperature (LST), Normalized Difference Water Index, Normalized Difference Snow Index, Normalized Difference Soil Index, and Normalized Difference Built-up Index. Specifically, NDVI represents the spatio-temporal pattern of forest and cultivated areas and is considered one of the conventional indices commonly used in remote-sensing analysis of vegetation. NDVI is calculated by measuring the difference between near infrared (which vegetation reflects) and red light (which vegetation absorbs). The studies of Sun et al. (2010), Li et al. (2015) and Jin et al. (2008) demonstrated correlation between urban expansion and decreasing NDVI. Similarly, the research of Kristjanson et al. (2005), Bhattacharya and Innes (2006), Morikawa (2014), and Aburas et al. (2015) showed the statistical relationship between NDVI and the spatial distribution of income inequality.

Data on land surface temperature is another type of precompiled geospatial information which researchers are using to predict income. For instance, Weng (2001); Buyantuyev and Wu (2010); Huang, Zhou, and Cadenasso (2011); Li et al. (2014); Ruthirako, Darnsawasdi, and Chatupote (2014); Youneszadeh, Kruger geb. Amiri, and Pilesjo (2015); Cooper et al. (2017); and Dissanayake et al. (2019) saw statistically significant relationship between land surface temperature and income. Richardson (2007); Maccini and Yang (2009); Barrios, Strobl, and Bertinelli (2010); Arezki and Bruckner (2012); Thiede (2014); Sarsons (2015); Shah and Steinberg (2017) Amare et al. (2018); and Gilmont et al. (2018) also found statistically significant correlation between rainfall on income, human capital, and economic activity in developing countries. In addition, Leroux et al. (2016) and Sruthi and Aslam (2015) documented the formulation of forecast models using both temperature and NDVI for predicting drought and in turn, forecasting the loss of agricultural output and its effect on farmers' incomes.

Efforts to crowd source geospatial data are also expanding. A good example is OpenStreetMap (OSM), a collaborative project producing a crowdsourced geographic database, and one of the major platforms promoting the use of geospatial data in the fields of global humanitarian action and community development. OSM database also features other types of geospatial data as presence of road, river, built-up area and point of interest (POI), enabling the investigation on the association among the geographical characteristics and socioeconomic conditions. Studies such as those by Hu et al. (2016), Ye et al. (2019), and Deng et al. (2019) demonstrate that OSM can provide details of spatial distribution of population and economic activities.

Poverty Mapping in Thailand

As mentioned earlier, official poverty statistics in Thailand are based on the HSES which provides reliable estimates from national down to provincial levels. However, recognizing the importance of having more geographically disaggregated poverty data as inputs for policy targeting, National Statistical Office (NSO) of Thailand started compiling small area (tambon or subdistrict level) poverty

estimates in 2003 in collaboration with other development partners like the then National Economic and Social Development Board (NESDB), Thailand Development Research Institute (TDRI),[1] and the World Bank. Since then, small area poverty estimates in the country have been compiled for the following years: 2005, 2007, 2008, 2011, 2012, 2015 and 2017.[2] The outputs in 2003 and 2005 were jointly prepared by three local institutions, namely, NESDB, NSO, and TDRI, together with the technical advisory from the World Bank. In 2015, the World Bank provided further technical assistance to NSO, to build capacity to implement small area estimation among more NSO staff. Additional technical details on the process of compiling poverty maps are documented by Jitsuchon (2004), Healy and Jitsuchon (2007), and Jitsuchon and Richter (2007).

However, despite availability of analytical tools for compiling granular estimates of poverty, it is important to identify alternative methods due to limitations associated with the conventional poverty mapping technique which heavily relies on availability of census data. For instance, since censuses are usually conducted every 5 to 10 years only, poverty mapping models that use covariates derived from census have restrictively strong assumptions (Bedi, Coudouel, and Simlaer 2007).

In a study published recently, researchers from ADB extended the conventional small area poverty estimation framework by tapping geospatial data extracted from daytime and nighttime imagery through machine learning algorithms to create granular poverty maps of the Philippines and Thailand (ADB 2020, Hofer et al. 2020). The adopted method was inspired specifically by Jean et al. (2016) which was further used and/or enhanced in subsequent studies (e.g., Babenko et al. 2017; Tingzon et al. 2019; Heitmann and Buri 2019; Yeh et al. 2020). These studies fall under the strand of literature that broadly aim to explore applications of artificial intelligence and computer vision techniques for estimating poverty. However, as hinted earlier, this methodology has several technical issues. First, validating aberrant or unexpected predictions becomes challenging because of the fact that the features being used to correlate poverty are abstract. Second, instead of directly predicting poverty, the method employs an intermediate step wherein an algorithm is first trained to predict the intensity of night lights. The intermediate step is necessary in this context because sources of night light data, particularly satellite imagery, are readily accessible and can cost-effectively provide large volumes of labelled images on which to train a computer vision algorithm, something that cannot be easily done if we were to predict poverty outright since readily available poverty data are not quite granular. Using data on night lights as a proxy for poverty during the intermediate step is arguably valid if it is assumed that places that are brighter at night are less poor than those places that are less well lit. However, if there are places that are equally lit but show varying levels of poverty on the ground, such an intermediate step could potentially lead to loss of vital information by not predicting poverty outright.

This study contributes to the existing literature of poverty measurement in Thailand by developing a prediction model whose correlates were derived from precompiled geospatial data. By doing so, we aim to assess whether it is feasible to develop a model with satisfactory predictive performance even if we are solely depending on precompiled geospatial dataset(s) instead of applying computer vision techniques to automatically extract satellite image features that are potentially correlated with poverty, a feat that has not been thoroughly explored in previous studies.

---

[1] In 2019, NESDB has been renamed as the National Economic and Social Development Council. TDRI is an independent think tank foundation.

[2] For years when census was not conducted, small area poverty estimates were calculated using data from the latest census (See http://www.nso.go.th/sites/2014/lists/ebook_thailand_poverty_maps/allitems.aspx).

III. DATA

Satellite Data

Google Earth Engine

Google Earth Engine (GEE) is an open cloud-based data storage and computing platform which provides access to satellite imagery for free. In this study, we extracted the following information from GEE: Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)'s rainfall data.

- Land Surface Temperature (LST)
- Normalized Difference Vegetation Index (NDVI). NDVI is widely used as an indicator representing the land cover of forest and agricultural activity.
- Intensity of Night Lights

Table 1 summarizes the range of data that can be obtained from GEE.

Table 1: Technical Specification of Data Obtained from Google Earth Engine

	Data	Satellite	Data Full Name	Resolution	Area (approximate)	Data availability	Duration	Frequency
1	Rainfall		CHIRPS Pentad: Climate Hazards Group InfraRed Precipitation with Station Data (version 2.0 final)	0.05 arc degree	110 m^2/Pixel	1 Jan 1981–26 Jul 2019	38y	Monthly
2	Nighttime Light (Old)	DMSP OLS	DMSP OLS: Nighttime Lights Time Series Version 4, Defense Meteorological Program Operational Linescan System	30 arc seconds	1 km^2/Pixel	1 Jan 1992–1 Jan 2014	27y	Yearly
3	Nighttime Light (New)	VIIRS	VIIRS Nighttime Day/Night Band Composites Version 1	15 arc seconds	375 m^2/Pixel	1 Apr 2012–1 Jun 2019	8y	Monthly
4	Land Surface Temperature (day)	MODIS	MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1 km	30 arc seconds	1 km^2/Pixel	5 Mar 2000–22 Aug 2019	20y	Daily
5	Land Surface Temperature (Night)	MODIS	MOD11A2.006 Terra Land Surface Temperature and Emissivity 8-Day Global 1 km	30 arc seconds	1 km^2/Pixel	5 Mar 2000–13 Aug 2019	20y	8 day
6	NDVI	MODIS	MODIS Combined 16-Day NDVI	15 arc seconds	375 m^2/Pixel	18 Feb 2000–14 Mar 2017	18y	16 day

CHIRPS = Climate Hazards Group InfraRed Precipitation with Station, DMSP OLS = Defense Meteorological Satellite Program Operational Linescan System, km = kilometer, m = meter, MODIS = Terra-Moderate Resolution Imaging Spectroradiometer, NDVI = Normalized Difference Vegetation Index, VIIRS = Visible and Infrared Imaging/Radiometer Suite.

Source: Google Earth Engine.

Global Urban Footprint

The Global Urban Footprint project by the German Remote Sensing Data Center of the German Aerospace Center compiles geocoded data which identify urban areas, land surface, and water bodies. Geocoded data on built-up and nonbuilt-up areas are also available from the Global Urban Footprint.

Global Human Settlement Layer

Mainly supported and supervised by the Directorate General Joint Research Centre of the European Commission, the Global Human Settlement Layer project has produced a fully open and free geospatial spatial dataset. The generated geospatial database provides informative evidence and the broadened insight of global human presence.

United States Geological Survey

This geospatial dataset has been generated based on the ten years (2001–2010) collection of Terra-Moderate Resolution Imaging Spectroradiometer-based Global Land Cover maps (MCD12Q1 land cover type data). There are 16 classifications for each pixel, identifying the type of land cover based on the method of highest confidence during 2001-2010, as described in Broxton et al. (2014).

European Space Agency Land Cover

Initially, the main objective of the European Space Agency (ESA)'s Climate Change Initiative is to produce an accurate land-cover classification that can serve the climate modeling community. This project has developed the Essential Climate Variable spatial dataset based on the extensive archives of remote-sensing data (ESA 2017). The database covers time series from 1992 to 2017 and contains 38 land cover classes, which are based on the United Nations Land Cover Classification System.

Crowd-Sourced Geospatial Database

Open Street Map

OpenStreetMap features crowd-sourced data on locations of infrastructures, human settlements and economic activities. In this study, we extract the following information from OSM: road count, road length, POI, and built-up area. We categorized POIs into 16 types based on its economic activity matched to the official classifications of 16 production and service sectors published by the NESDB.

Tables A.1 and A.2 of the Appendix provide the list of variables obtained from geospatial data of 2015 and 2017, respectively.

Poverty Data

Income-Based Poverty

As mentioned earlier, poverty mapping is a regular initiative conducted by Thai government. In this study, the ratio of the population living below the national poverty line per total population in each tambon (i.e., subdistrict) is used as one of the dependent variables in our computations.

Multidimensional Poverty

As an alternative metric of poverty, the Office of the National Economic and Social Development Committee and National Electronics and Computer Technology Center also compile statistics on prevalence of multidimensional poverty starting 2017. The data are based on:

(i) a census-based Basic Minimum Need data, supervised by the Community Development Department, Ministry of Interior, which includes approximately 36 million population.

(ii) a register-based data source of approximately 11.4 million individuals gathered by the Ministry of Finance through the national welfare card program.

The criteria used in identifying a multidimensionally poor person is inspired by the Multidimensional Poverty Index method developed by the Oxford Poverty and Human Development Initiative and United Nations Development Programme.

Reference Period

Our target reference period coincides with two most recent years where tambon-level estimates of poverty in Thailand are available: 2015 and 2017 for income poverty, and 2017 for multidimensional poverty index.

IV. METHODS

In this study, we consider Generalized Least Squares (GLS) method, and three other widely used machine learning algorithms: neural network, random forest estimation, and support vector regression (SVR). To assess whether the models have satisfactory generalization performance, 50% of the data were allocated for training while the remaining 50% constituted the validation set. Based on this allocation, we resampled the data 100 times. The values of metrics used to compare machine learning algorithms are based on averages from these 100 datasets.

Generalized Least Squares

GLS is considered a modification of the Ordinary Least Squares (OLS) as it relaxes the assumption that the variance of an observation is homogeneous regardless of the explanatory variables associated with it.

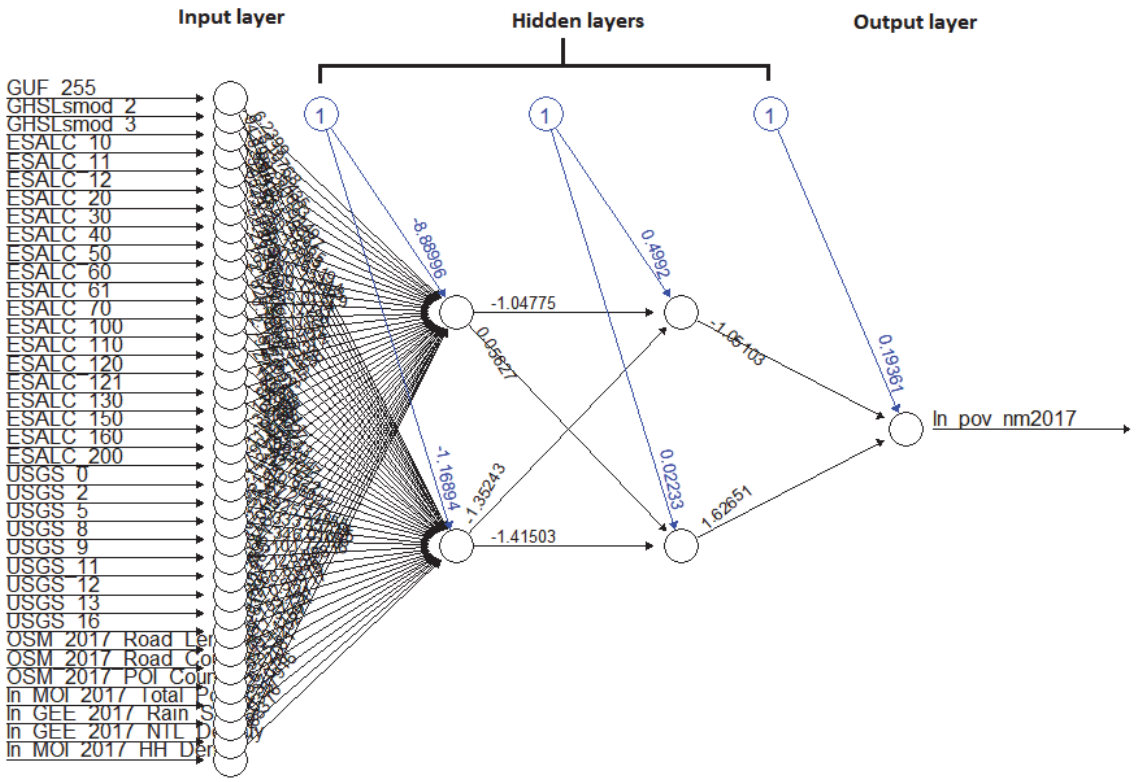
Neural Network

A neural network is an example of a machine-learning model inspired by the biological neural network that constitutes the human brain. As with other types of machine-learning models, a neural network can learn to perform different tasks without being explicitly programmed to do so (ADB 2020).

Structurally, a neural network is composed of numerous nodes and edges. A node can be a variable or a mathematical function connected by edges. These nodes combine together to form different layers within the neural network. The input layer takes in the raw data. In the hidden layers,

each node or neuron serves as filter and is activated each time it detects a specific pattern or feature. The output layer simply organizes the identified features into an appropriate category. The best way to represent these connections is through computational graphs as shown in Figure 3 (ADB 2020).

Figure 3: Illustration of a Sample Neural Network



Source: Graphics generated by the study team.

Random Forest

Random forests are an ensemble method based on decision trees, with each tree building on a random subset of the training data and a random subset of the independent variables. It can perform classification and/or prediction-related tasks and by using averaging, it can improve a model's predictive accuracy and control overfitting.

In this study, variable importance (VIMP) and minimal depth (MD) were used to conduct further analyses. These metrics use the main features obtained from all decision trees to assess the relative significance of explanatory variables in selecting the final predictors in the model.

Support Vector Regression

Typically, the main objective in a linear regression framework is to minimize a specific loss function. For instance, OLS method aims to minimize the sum of squared errors. Methods like lasso or ridge regression extend this framework by introducing additional penalty parameters to minimize complexity and/or reduce the number of covariates that marginally contribute to the model's predictive performance.

On the other hand, a method like the support vector regression provides an alternative framework wherein instead of minimizing a specific loss function, one is only concerned about reducing it to a certain degree. This gives greater flexibility in the estimation and helps in dealing with the limitations pertaining to distributional properties of the variables included in the analyses. In general, flexibility with allowable error renders SVR superior than other conventional estimation methods that are fixated on minimizing a loss function.

In this study, neural network, random forest estimation, and support vector regression were implemented using R software. Table 2 lists the details of relevant R packages, including links to the main sources of technical references.

Table 2: R Packages Used in Machine Learning Computations

Method	Package's name	Technical reference
Neural Network (NN)	Nnet	https://cran.r-project.org/web/packages/nnet/nnet.pdf
Random Forest (RF)	randomForestSRC	https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf
Support Vector Regression (SVR)	e1071	https://cran.r-project.org/web/packages/e1071/e1071.pdf

Source: The Comprehensive R Archive Network.

V. ANALYTICAL RESULTS

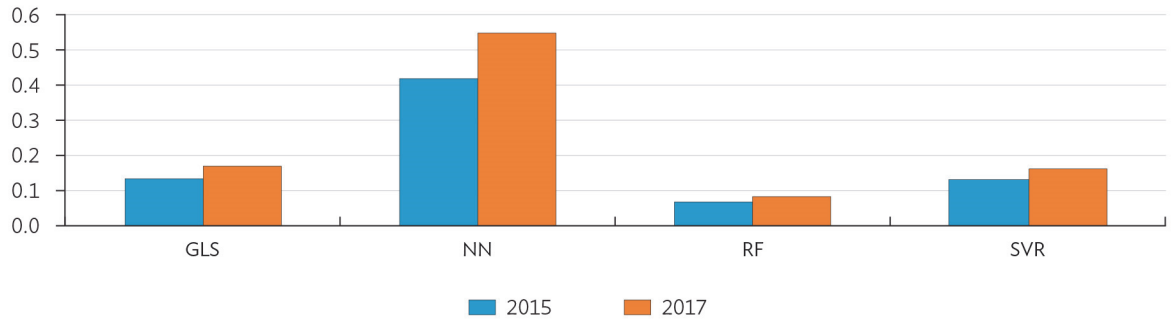
Preliminary Analysis

As preliminary estimation tools, we first estimated a full model and various model specifications using OLS and stepwise regression. In general, we found that the proportion of people living below income-based poverty line and the value of the multidimensional poverty index are negatively associated with geospatial indicators that represent the degree of an area's urbanization, i.e., intensity of night lights, building density, and number of points of interest which are associated with manufacturing and utility sectors. On the other hand, poverty outcomes are positively correlated with rainfall, NDVI, and other land cover classes that are typically associated in rural areas. While the directions of these correlations align with our expectations, the resulting adjusted R2 values are relatively low, ranging from 0.13 to 0.33.

Using Machine Learning Algorithms to Predict Income-Based Poverty Rate

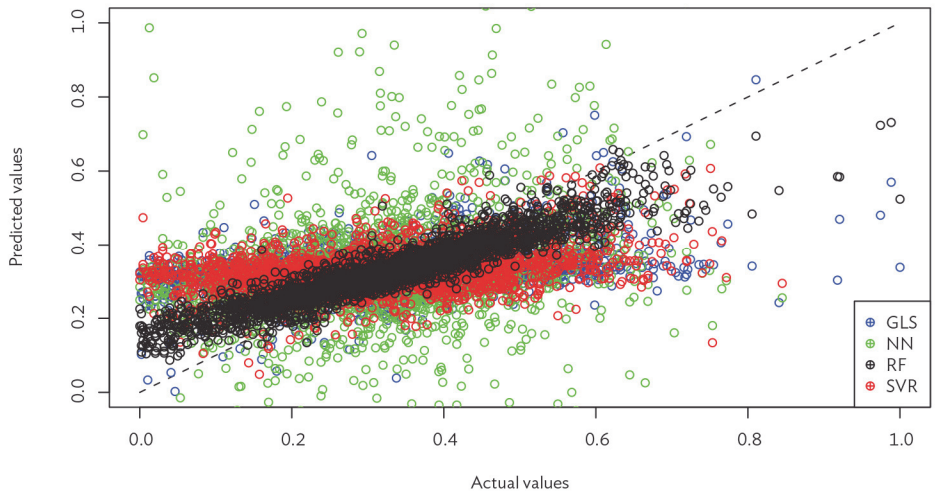
Comparison of the RMSE (averaged across 100 trials) from the four computational methods (Figure 4) shows that the random forest method yielded the lowest RMSE value. Graphical illustration of the goodness-of-fit (Figures 5 and 6) also confirms that the Random Forest (RF) has the best predictive performance among the four methods that we have considered—generating predicted values that are closest to the actual ones. SVR and GLS performed second and third under the same criteria. Notably, neural network generated the highest RMSE.

Figure 4: Comparison of Average Root Mean Squared Error Obtained from Four Machine Learning Algorithms



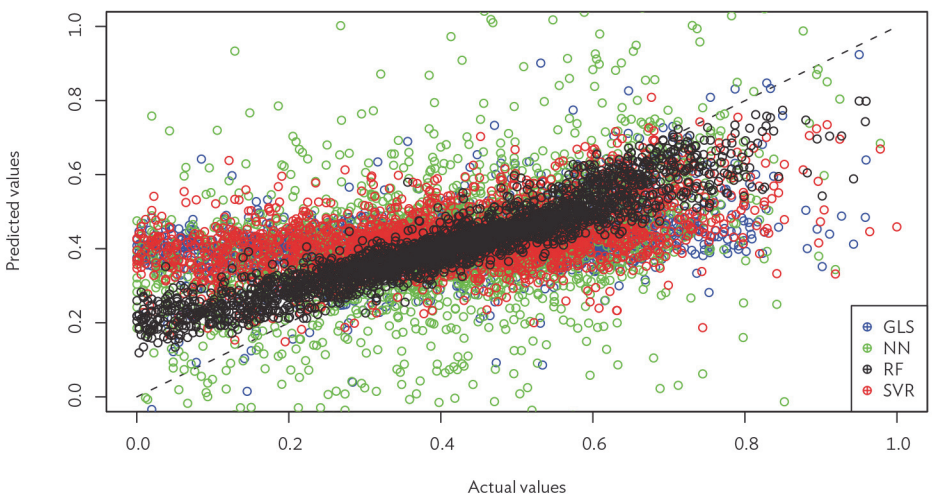
GLS = Generalized Least Squares, NN = Neural Network, RF = Random Forest, SVR = Support Vector Regression.
Source: Calculation and graphics generated by the study team.

Figure 5: Scatter Plot of Published and Predicted Income Poverty Rates, 2015



GLS = Generalized Least Squares, NN = Neural Network, RF = Random Forest, SVR = Support Vector Regression.
Source: Calculation and graphics generated by the study team.

Figure 6: Scatter Plot of Published and Predicted Income Poverty Rates, 2017



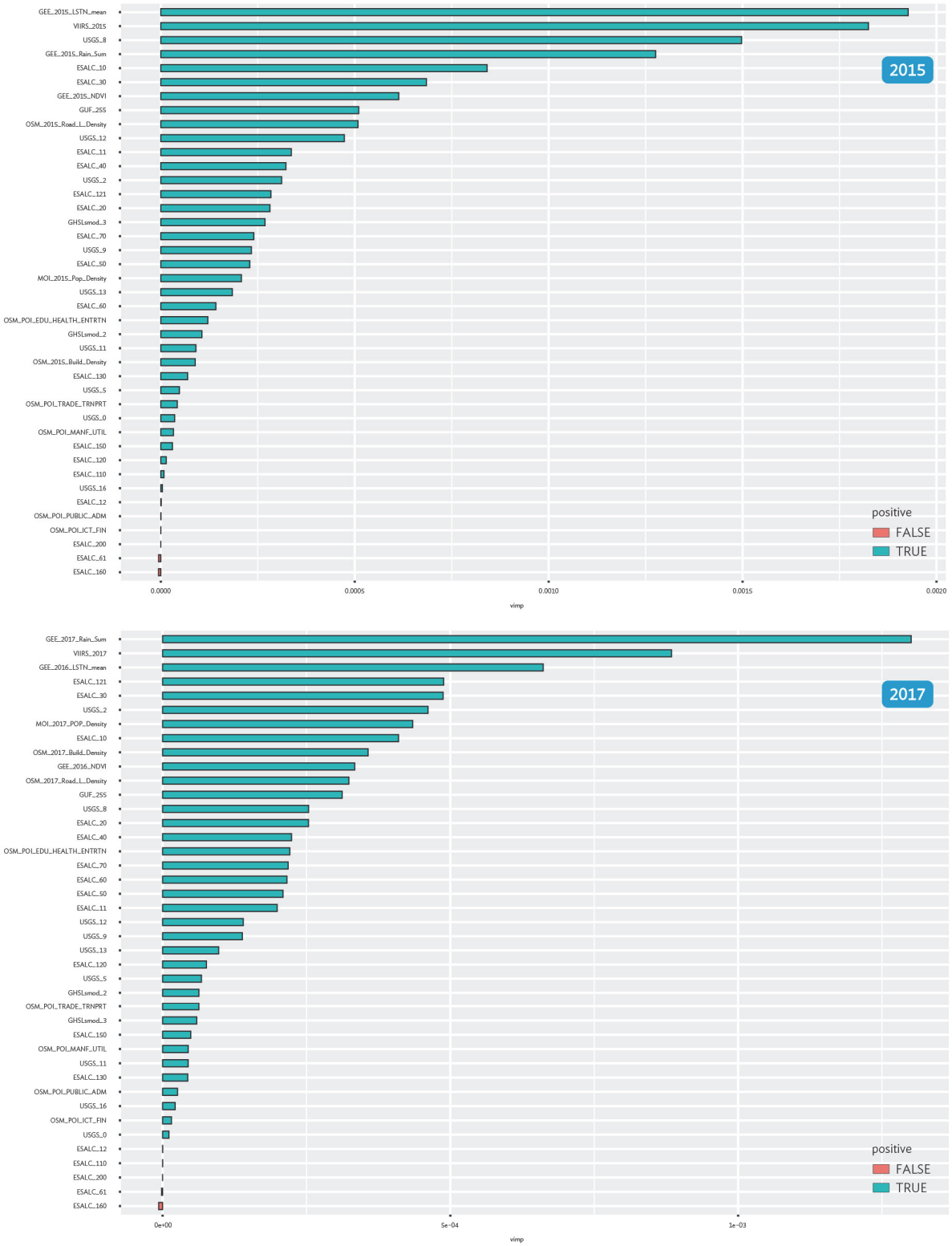
GLS = Generalized Least Squares, NN = Neural Network, RF = Random Forest, SVR = Support Vector Regression.
Source: Calculation and graphics generated by the study team.

VIMP and MD were further conducted to identify the significance of each variable. Figure 7 shows the result of VIMP for 2015 and 2017, while Figure 8 illustrates the outcomes of computing MD.

VIMP identified intensity of night lights and population density-related variables as the biggest contributors to the model. Meanwhile, five variables were identified as false positive in the VIMP's results for 2015 and 2017, indicating the irrelevance of these variables in predicting the poverty headcount rate. Alternatively, it is also possible that the information provided by these variables is already captured by other variables. The 'unimportant' variables are the area covered by tree or shrub (ESALC_12), the area covered by tree, broadleaved and deciduous more than 40% (ESALC_61), the area covered by mosaic herbaceous more than 50% (ESALC_110), the area covered by tree, flooded, fresh or brackish water (ESALC_160) and the bare areas (ESALC_200).

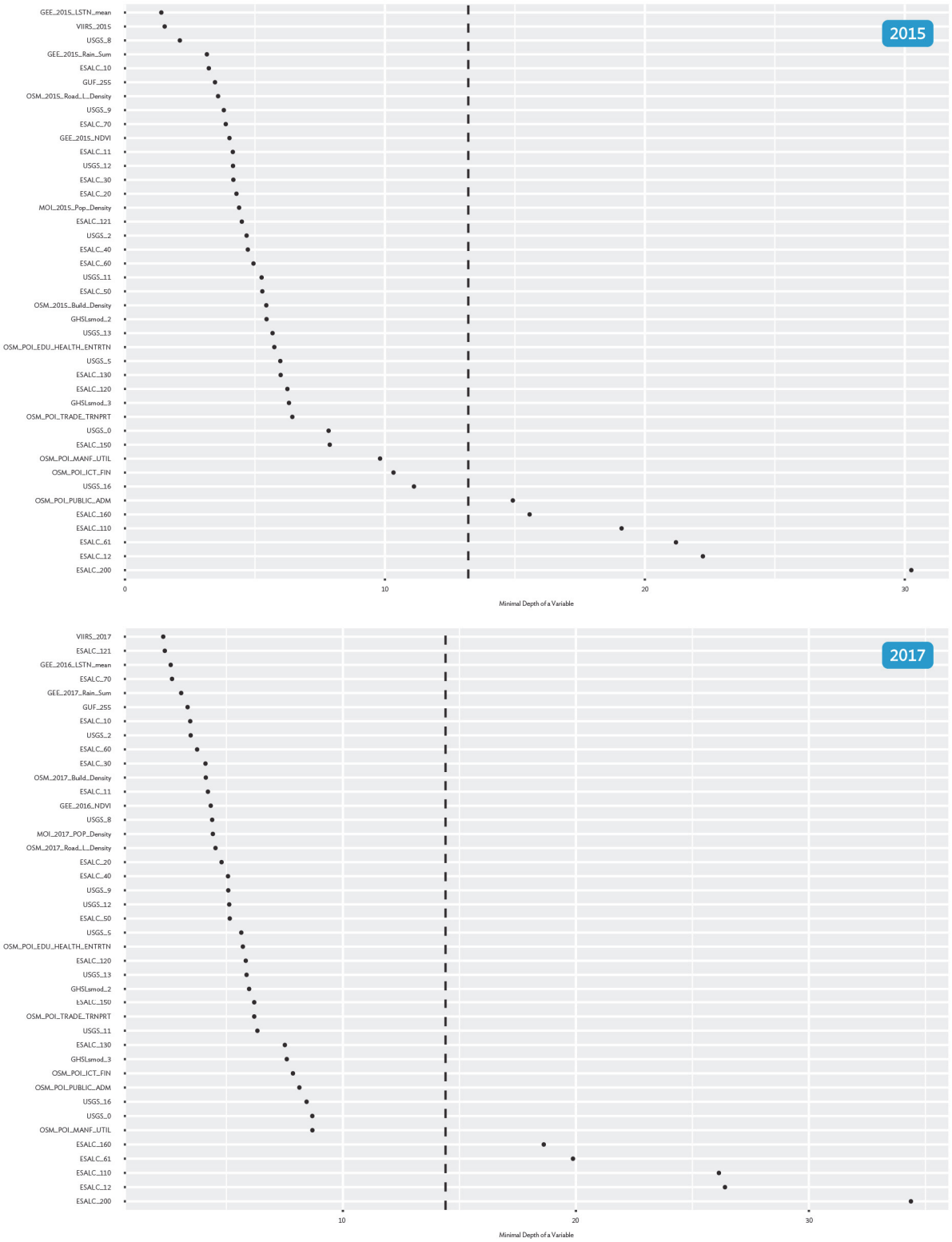
The results obtained from MD calculation generated similar outcomes, confirming that intensity of night lights and population density-related variables are highly associated with poverty headcount. Similarly, the results show that five variables possess very low predictive power—the same five variables identified by VIMP results as irrelevant to the model. These variables can therefore be excluded from the model in further analysis.

Figure 7: Results of Variable Importance Computation



Source: Calculation and graphics generated by the study team.

Figure 8: Results of Minimal Depth Computation



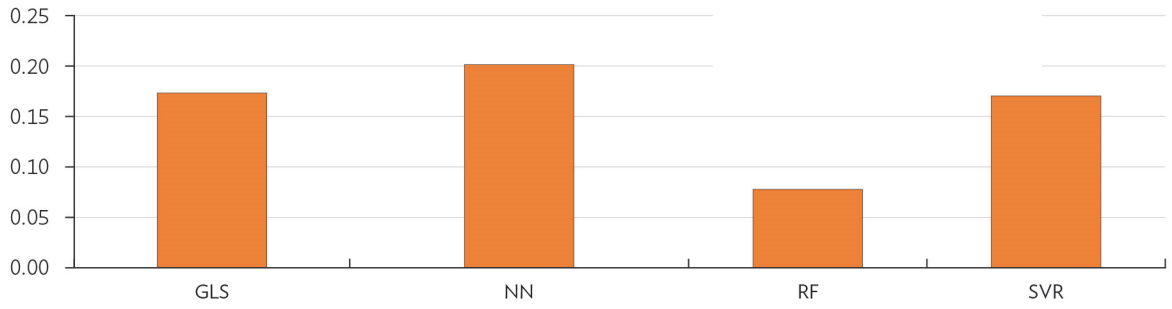
Source: Calculation and graphics generated by the study team.

Using Machine Learning Algorithms to Predict Multidimensional Poverty Index

In addition to income-based poverty rate, we also applied GLS, neural network, random forest, and support vector machine to predict the multidimensional poverty index (MPI).

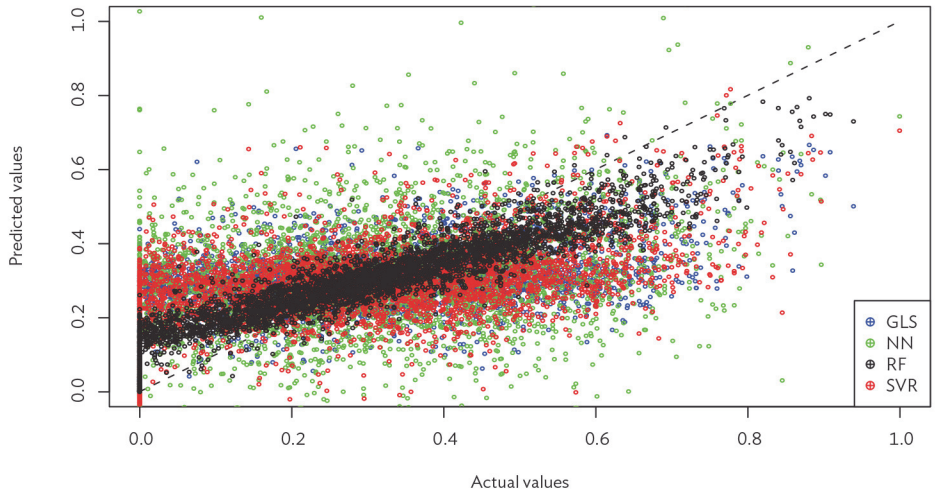
Figure 9 depicts the comparison of RMSE obtained from four Machine Learning methods. Similar to the case of income poverty rates, the random forest method yielded the lowest RMSE. The scatterplot in Figure 10 compares the actual MPI and the predicted values. It shows that most predicted values generated by random forest are located closest to the 45-degree line, suggesting that it has the best fit among the four methods considered in this study.

Figure 9: Comparison of Average Root Mean Squared Error Obtained from Four Machine Learning Algorithms, 2017



GLS = Generalized Least Squares, NN = Neural Network, RF = Random Forest, SVR = Support Vector Regression.
Source: Calculation and graphics generated by the study team.

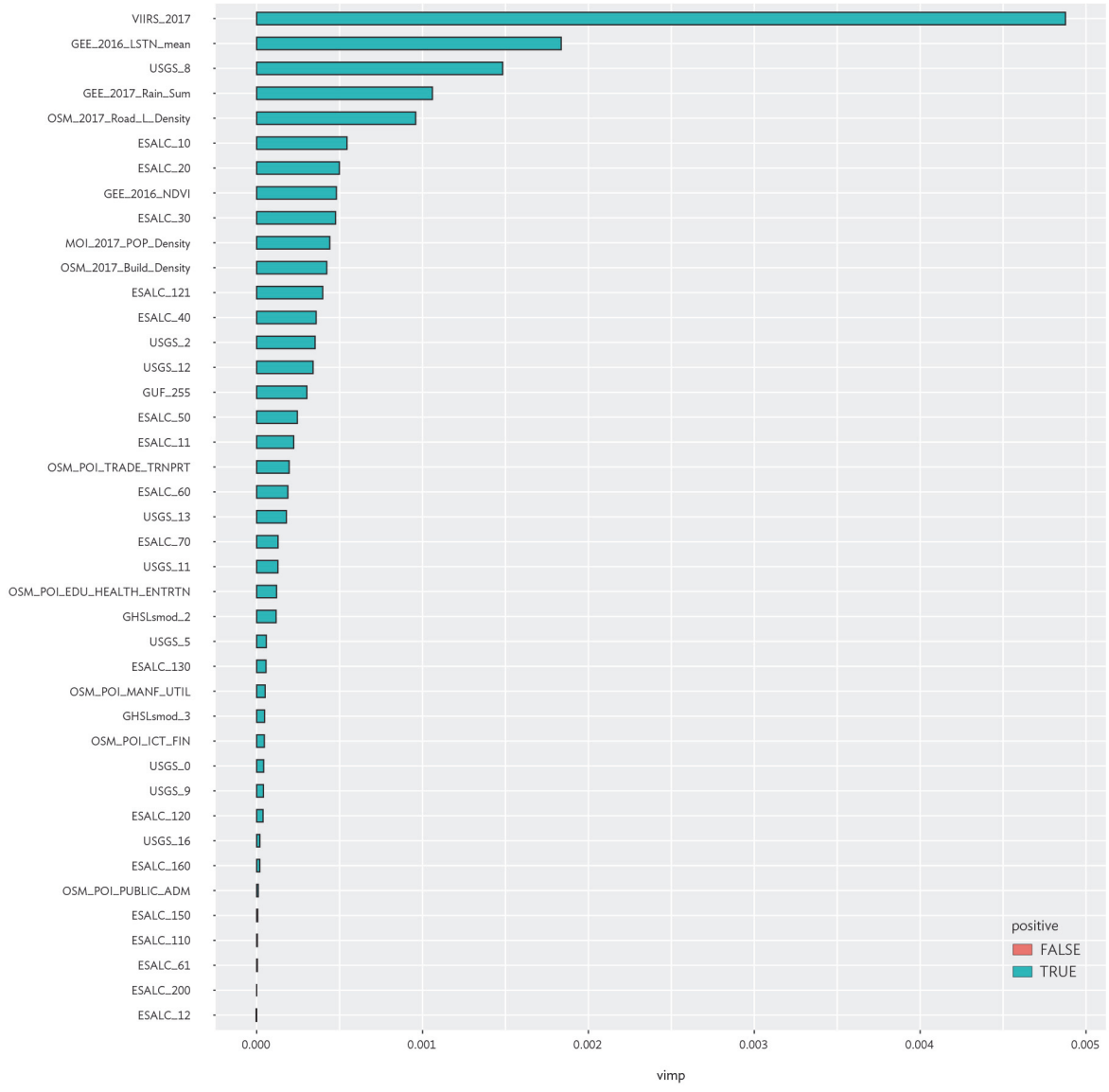
Figure 10: Scatter Plot of Published and Predicted Multidimensional Poverty Index, 2017



GLS = Generalized Least Squares, NN = Neural Network, RF = Random Forest, SVR = Support Vector Regression, TPMAP = Thai People Map and Analytics Platform.
Source: Calculation and graphics generated by the study team.

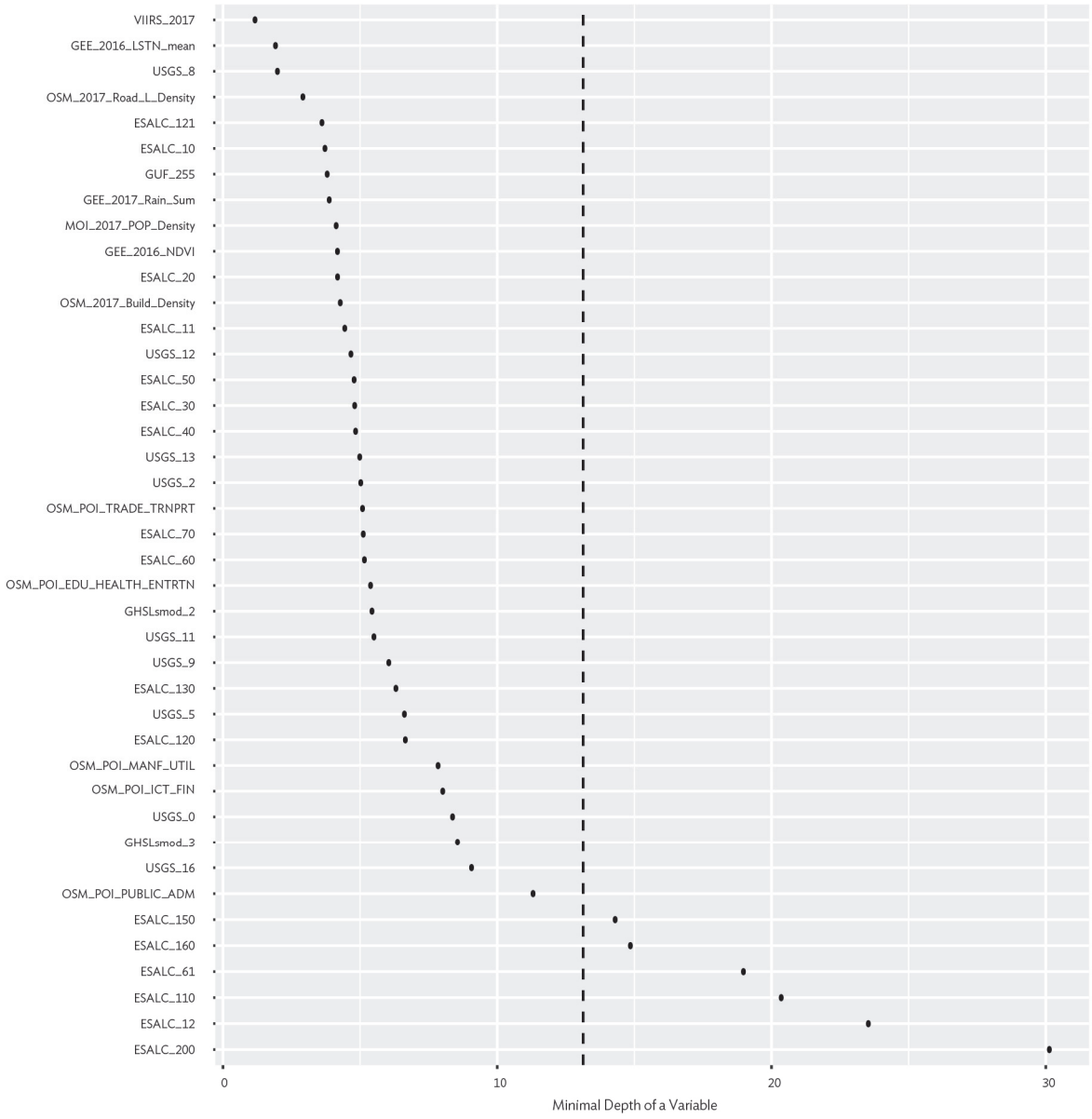
Again, we examined the degree of explanatory power of each variable by calculating VIMP and MD. Figure 11 exhibits that variables related to population density such as nighttime light, LST, and road density have a high degree of contribution to predict the variation in poverty rate, based on VIMP. The result obtained from MD, as illustrated in Figure 12, also show qualitatively similar results, revealing that nighttime light, LST, rainfall, road density, and the area covered by woody Savannas (USGS8) are key geographical features associated with the value of MPI.

Figure 11: Result of Variable Importance Computation



Source: Calculation and graphics generated by the study team.

Figure 12: Result of Minimal Depth Computation



Source: Calculation and graphics generated by the study team.

In summary, among the methods applied in this study, the random forest technique yielded the highest level of accuracy when predicting both income poverty rate and multidimensional poverty index. Furthermore, the resulting random forest models fit the datasets well, as suggested by the adjusted R-square values presented in Table 3.

Table 3: Adjusted R^2 Values for Random Forest-Based Models

Variable being predicted	Adjusted R^2
NSO's poverty headcount of 2015	0.8526
NSO's poverty headcount of 2017	0.8459
TPMAP's poverty rate of 2017	0.8332

NSO = National Statistical Office, TPMAP = Thai People Map and Analytics Platform.

Source: Calculation generated by the study team.

As stated in Wang, Aggarwal, and Liu (2018), the random forest algorithm tends to outperform other machine learning methods due to its capability to fit complex association structures even with small datasets.[3]

VI. CONCLUSION

The contribution of this study is twofold. Firstly, it introduces the integration of data, composed of the nationwide survey, register-based data, geospatial information, and the satellite imagery. In addition, since most of these are open data, data acquisition cost is minimal. This is potentially attractive for national statistical offices with scarce resources but wish to explore geospatial data can be used to enhance the compilation of poverty statistics. Secondly, this paper has applied computational techniques to examine the relationship between geospatial features such as intensity of night lights, land cover, land use, etc. and proportion of people living below poverty line as measured using conventional method of estimating poverty. It is shown that the Random Forest is the best prediction method, yielding the accuracy of more than 80%. These contributions suggest the potential of applying the open data and open-source computational tools to analyze the spatial distribution of poverty. Furthermore, the results obtained from VIMP and MD reveal the associations between geospatial covariates such as intensity of night lights, population density, and poverty rates. Moving forward, if it can be proven that such relationships remain stable over time, it might be possible to apply these techniques to predict poverty for years when household surveys on income and expenditures are not conducted, but data on geospatial correlates of poverty are available.

[3] Other studies that applied the random forest technique in other contexts also noted qualitatively similar results. For example, Fernández-Delgado et al. (2014), using the entire University of California Irvine dataset, have demonstrated that the random forest outperformed 179 classifiers from 17 families. Similarly, Díaz-Uriarte and Alvarez de Andrés (2006) have stated that random forest is the best method for gene selection and classification, and Ali et al. (2012) have shown that random forest yields the highest accuracy in predicting breast cancer. Nevertheless, the robustness of the results of random forest technique to the size of the training and validation data warrants further investigation.

REFERENCES

Aburas, Maher Milad, Sabrina Ho Abdullah, Mohammad Firuz Ramli, and Zulfa Hanan Ash'aari. 2015. "Measuring Land Cover Changes in Seremban, Malaysia Using NDVI Index." *Procedia Environmental Sciences* 30: 238–243. doi: https://doi.org/10.1016/j.proenv.2015.10.043.

Ali, Jehad, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. 2012. "Random Forests and Decision Trees." *International Journal of Computer Science Issues* 9 (5): 272–278.

Alpaydin, Ethem. 2010. *Adaptive Computation and Machine Learning (Adaptive Computation and Machine Learning Series)*. Cambridge: MA.

Amare, Mulubrhan, Nathaniel D. Jensen, Bekele Shiferawb, and Jennifer Denno Cissé. 2018. "Rainfall Shocks and Agricultural Productivity: Implication for Rural Household Consumption." *Agricultural Systems* 166: 79–89.

Arezki, Rabah, and Markus Brückner. 2012. "Rainfall, Financial Development, and Remittances: Evidence from Sub-Saharan Africa." *Journal of International Economics* 87 (2): 377–385.

Asian Development Bank (ADB). 2020. Mapping Poverty through Data Integration and Artificial Intelligence. A Special Supplement of the Key Indicators for Asia and the Pacific. Manila. https://www.adb.org/publications/mapping-poverty-data-integration-ai.

———. 2015. *Thailand: Industrialization and Economic Catch-Up.* Manila.

Babenko, Boris, Jonathan Hersh, David Newhouse, Anusha Ramakrishnan, and Tom Swartz. 2017. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, with an Application in Mexico." Proceedings from NIPS 2017: Neural Information Processing Systems Workshop on Machine Learning for the Developing World. Long Beach, CA.

Balli, Serkan, Ensar Arif Sagbas, and Musa Peker. 2018. "Human Activity Recognition from Smart Watch Sensor Data Using a Hybrid of Principal Component Analysis and Random Forest Algorithm." *Measurement and Control* 52 (3): 002029401881369.

Barrios, Salvador, Eric Strobl, and Luisito Bertinelli. 2010. "Trends in Rainfall and Economic Growth in Africa: A Neglected Cause of the African Growth Tragedy." *The Review of Economics and Statistics* 92 (2): 350–366.

Bedi, Tara, Aline Coudouel, and Kenneth Simler. 2007. *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington, DC: World Bank.

Bhattacharya, Haimanti and Robert Innes. 2006. "Is There a Nexus between Poverty and Environment in Rural India." *Food Security and Poverty*. doi:10.22004/ag.econ.21201.

Bickenbach, Frank, Eckhardt Bode, Mareike Lange, and Peter Nunnenkamp. 2013. "Night Lights and Regional GDP." Kiel Working Paper No. 1888.

Brown, Casey and Upmanu Lall. 2006. "Water and Economic Development: The Role of Variability and a Framework for Resilience." *Natural Resources Forum*, 306–317.

Broxton, Patrick, Xubin Zeng, Damien Sulla-Menashe, and Peter Troch. 2014. "A Global Land Cover Climatology Using MODIS Data." *Journal of Applied Meteorology and Climatology* 53 (6): 1593–1605.

Buyantuyev, Alexander and Jianguo Wu. 2010. "Urban Heat Islands and Landscape Heterogeneity: Linking Spatiotemporal Variations in Surface Temperatures to Land-Cover and Socioeconomic Patterns." *Landscape Ecology* 25 (1): 17–33.

Chen, Xi and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Science* 108 (21): 8589–8594. https://doi.org/10.1073/pnas.1017031108.

Cooper, L. Annie, Ashley P. Ballantyne, Zachary Holden, and Erin L. Landguth. 2017. "Disturbance Impacts on Land Surface Temperature and Gross Primary Productivity in the Western United States." *Journal of Geophysical Research: Biogeosciences* 122 (4): 930–946.

Deng, Yue, Jiping Liu, Yang Liu, and An Luo. 2019. "Detecting Urban Polycentric Structure from POI Data." *ISPRS International Journal of Geo-Information* 8 (6): 283.

Díaz-Uriarte, Ramon and Sarah Alvarez de Andrés. 2006. "Gene Selection and Classification of Microarray Data Using Random Forest." *BMC Bioinformatics* 7 (1): 3.

Dissanayake, DMSLB, Takehiro Morimoto, Yuji Murayama, Manjula Ranagalage, and Hepi H. Handayani. 2019. "Impact of Urban Surface Characteristics and Socio-economic Variables on the Spatial Variation of Land Surface Temperature in Lagos City, Nigeria." *Sustainability* 11 (1): 25.

Doll, Christopher N.H., Jan-Peter Muller, and Christopher D. Elvidge. 2000. "Night-Time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas Emissions." *Ambio* 29 (3): 157–62. DOI:10.2307/4315020.

Doll, Christopher N.H., Jan-Peter Muller, and Jeremy G. Morley. 2006. "Mapping Regional Economic Activity from Night-Time Light Satellite Imagery." *Ecological Economics* 57 (1): 75–92. DOI: 10.1016/j.ecolecon.2005.03.007.

Efroymson, M.A. 1960. "Multiple Regression Analysis." In *Mathematical Methods for Digital Computers*, edited by A. Ralston, and H.S. Wilf. New York.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, Ethan R. Davis, and Clay W. Davis. 1997. "Relation between Satellites Observed Visible-Near Infrared Emissions, Population, Economic Activity and Electric Power Consumption." *International Journal of Remote Sensing* 18 (6): 1373–1379. DOI: 10.1080/014311697218485.

European Space Agency (ESA). 2017. Land Cover CCI Product User Guide Version 2. https://www.esa-landcover-cci.org/?q=node/199.

Fernández-Delgado, Manuel, Eva Cernadas, Senen Barro, and Dinani Amorim. 2014. "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?" *Journal of Machine Learning Research* 15: 3133–3181.

Forbes, Dolores Jane. 2013. "Multi-scale Analysis of the Relationship between Economic Statistics and DMSP-OLS Night Light Images." *GIS Science & Remote Sensing* 50 (5): 483–499. DOI:10.1080/15481603.2013.823732.

Gilmont, Michael, Jim W. Hall, David Grey, Simon James Dadson, Simon Abele, and Michael Simpson. 2018. "Analysis of the Relationship between Rainfall and Economic Growth in Indian States." *Global Environmental Change* 49: 56–72.

Healy, Andrew and Somchai Jitsuchon. 2007. "Finding the Poor in Thailand." *Journal of Asian Economics* 18 (5): 739–759.

Heitmann, Soren and Sinja Buri. 2019. *Poverty Estimation with Satellite Imagery at Neighborhood Levels– Results and Lessons for Financial Inclusion from Ghana and Uganda.* International Finance Corporation. World Bank Group.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028. DOI: 10.1257/aer.102.2.994.

Hendrix, Cullen S., and Idean Salehyan. 2012. "Climate Change, Rainfall, and Social Conflict in Africa." *Journal of Peace Research* 49 (1): 35–50.

Hoddinott, John and Bill Kinsey. 2001. "Child Growth in the Time of Drought." *Oxford Bulletin of Economics and Statistics* 63 (4): 409–436.

Hofer, Martin, Tomas Sako, Arturo Martinez, Mildred Addawe, Joseph Bulan, Ron Durante, and Marymell Martillan. 2020. "Applying Artificial Intelligence on Satellite Imagery to Compile Granular Poverty Statistics." ADB Economics Working Paper Series No. 629.

Hu, Tengyun, Jun Yang, Xuecao Li, and Peng Gong. 2016. "Mapping Urban Land Use by Using Landsat Images and Open Social Data." *Remote Sensing* 8: 151.

Huang, Ganlin, Weiqi Zhou, and M.L. Cadenasso. 2011. "Is Everyone Hot in the City? Spatial Pattern of Land Surface Temperatures, Land Cover and Neighborhood Socioeconomic Characteristics in Baltimore, MD." *Journal of Environmental Management* 92 (7): 1753–1759.

Jean, Neal, M. Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *Science* 353 (6301): 790–794.

Jenerette, G. Darrel, Sharon L. Harlan, Anthony Brazel, Nancy Jones, Larissa Larsen, and William L. Stefanov. 2007. "Regional Relationships between Surface Temperature, Vegetation, and Human Settlement in a Rapidly Urbanizing Ecosystem." *Landscape Ecology* 22 (3): 353–365.

Jin, Xiaomei, L. Wan, You-Kuan Zhang, and Michael Schaepman. 2008. "Impact of Economic Growth on Vegetation Health in China based on GIMMS NDVI." *International Journal of Remote Sensing* 29 (13): 3715–3726.

Jitsuchon, Somchai. 2004. "Small Area Estimation Poverty Map for Thailand." Paper presented at the SMERU Research Institute and Ford Foundation International Seminar, Mapping Poverty in Southeast Asia. Jakarta.

Jitsuchon, Somchai and Kaspar Richter. 2007. "Thailand's Poverty Maps from Construction to Application." In *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*, edited by T. Bedi, A. Coudouel, and K. Simler. Washington, DC: World Bank.

Klyuev, Vladimir. 2015. "Structural Transformation—How Does Thailand Compare?" IMF Working Paper No. WP/15/51.

Kristjanson, Patricia, Maren Radeny, Isabelle Baltenweck, Joseph Ogutu, and An Notenbaert. 2005. "Livelihood Mapping and Poverty Correlates at a Meso-level in Kenya." *Food Policy* 30: 568–583.

Lanzafame, Matteo. 2014. "Temperature, Rainfall and Economic Growth in Africa." *Empirical Economics* 46: 1–18.

Leroux, Louise, Christian Baron, Bernardin Zoungrana, Seydou B. Traoré, Danny Lo Seen, and Agnes Bégué. 2016. "Crop Monitoring Using Vegetation and Thermal Indices for Yield Estimates: Case Study of a Rainfed Cereal in Semi-arid West Africa." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (1): 347–362.

Li, Deren, Xia Zhao, and Xi. Li. 2016. "Remote Sensing of Human Beings—A Perspective from Nighttime Light." *Geo-spatial Information Science* 19 (1): 69–79. DOI: 10.1080/10095020.2016 .1159389.

Li, Guang Yu, Shuang S. Chen, Yue Yan, and Cheng Yu. 2015. "Effects of Urbanization on Vegetation Degradation in the Yangtze River Delta of China: Assessment Based on SPOT-VGT NDVI." *Journal of Urban Planning and Development* 141 (4): 05014026.

Li, Lin, Tan Yongbin, Shen Ying, Zhonghai Yu, Zhen. Li, and Honghao Lan. 2014. "Impact of Land Cover and Population Density on Land Surface Temperature: Case Study in Wuhan, China." *Journal of Applied Remote Sensing* 8 (1): 084993.

Li, Xi, Huimin Xu, Xiaoling Chen, and Chang Li. 2013. "Potential of NPP-VIIRS Nighttime Light Imagery for Modeling the Regional Economy of China." *Remote Sensing* 5 (6): 3057–3081. DOI:10.3390/rs5063057.

Liaqut, Anum, Isma Younes, Rakhshanda Sadaf, and Haris Zafar. 2019. "Impact of Urbanization Growth on Land Surface Temperature Using Remote Sensing and GIS: A Case Study of Gujranwala City, Punjab, Pakistan." *International Journal of Economic and Environmental Geology*, 44–49.

Maccini, Sharon, and Dean Yang. 2009. "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall." *American Economic Review* 99 (3): 1006–1026.

Morikawa, Robert. 2014. "Remote Sensing Tools for Evaluating Poverty Alleviation Projects: A Case Study in Tanzania." *Procedia Engineering* 78: 178–187.

National Economic and Social Development Council. 2015. *Poverty and Inequalities Report 2013.* Bangkok.

———. 2017. *Human Achievement Index Report 2017.* Bangkok.

National Statistical Office (NSO). 2020. Key Statistical Data. http://statbbi.nso.go.th/staticreport/page/sector/en/08.aspx (accessed 1 December 2020).

Organisation for Economic Co-operation and Development (OECD). 2018. *Multi-dimensional Review of Thailand.* Paris.

Piaggesi, Simone, Laetitia Gauvin, Michele Tizzoni, Ciro Cattuto, Natalia Adler, Stefaan Verhulst, Andrew Young, Rhiannan Price, Leo Ferres, and Andre Panisson. 2019. "Predicting City Poverty Using Satellite Imagery." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 90–96.

Richardson, Craig J. 2007. "How Much Did Droughts Matter? Linking Rainfall and GDP Growth in Zimbabwe." *African Affairs* 106 (424): 463–478.

Ruthirako, Poonyanuch, Rotchanatch Darnsawasdi, and Wichien Chatupote. 2014. "Intensity and Pattern of Land Surface Temperature in Hat Yai City, Thailand." *Walailak Journal of Science and Technology (WJST)* 12 (1): 83–94.

Sarsons, Heather. 2015. "Rainfall and Conflict: A Cautionary Tale." *Journal of Development Economics* 115: 62–72.

Shah, Manisha, and Bryce Millet Steinberg. 2017. "Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital." *Journal of Political Economy* 125 (2): 527–561.

Silhavy, Radek, Petr Silhavy, and Zdenka Prokopova. 2018. "Evaluating Subset Selection Methods for Use Case Points Estimation." *Information and Software Technology* 97: 1–9.

Sims, Daniel, Faiz Rahman, Vicente Cordova, Bassil El-Masri, Dennis Baldocchi, Paul V. Bolstad, Lawrence Flanagan, Allen H. Goldstein, David Hollinger, Laurent Misson, Russel K. Monson, Walter C. Oechel, Hans P. Schmid, Steven C. Wofsy, and Liukang Xu. 2008. "A New Model of Gross Primary Productivity for North American Ecosystems Based Solely on the Enhanced Vegetation Index and Land Surface Temperature from MODIS." *Remote Sensing of Environment* 112 (4): 1633–1646.

Sruthi, Swathandran and M.A. Mohammed Aslam. 2015. "Agricultural Drought Analysis Using the NDVI and Land Surface Temperature Data: A Case Study of Raichur District." *Aquatic Procedia* 4: 1258–1264.

Sun, Jinyu, Xuhui Wang, Anping Chen, Yuecun Ma, Mengdi Cui, and Shilong Piao. 2010. "NDVI Indicated Characteristics of Vegetation Cover Change in China's Metropolises over the Last Three Decades." *Environmental Monitoring and Assessment* 179: 1–14. doi:https://doi.org/10.1007/s10661-010-1715-x.

Sutton, Paul C., Christopher D. Elvidge, and Ghosh Tilottama. 2007. "Estimation of Gross Domestic Product at Sub-national Scales Using Nighttime Satellite Imagery." *International Journal of Ecological Economics & Statistics* 8: 5–21.

Thiede, Brian C. 2014. "Rainfall Shocks and Within-Community Wealth Inequality: Evidence from Rural Ethiopia." *World Development* 64: 181–193.

Tingzon, Isabelle, Ardie Orden, Stephanie Sy, Vedran Sekara, Ingmar Weber, Masoomali Fatehkia, Manuel Garcia-Herranz, and Dohyung Kim. 2019. "Mapping Poverty in the Philippines Using Satellite Imagery, and Crowd-Sourced Geospatial Information." *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-4/W19.

Wang, Suhang, Charu Aggarwal, and Huan Liu. 2018. "Random-Forest-Inspired Neural Networks." *ACM Transactions on Intelligent Systems and Technology* 69: 1–25.

Weng, Qihao. 2001. "A Remote Sensing? GIS Evaluation of Urban Expansion and Its Impact on Surface Temperature in the Zhujiang Delta, China." *International Journal of Remote Sensing* 22 (10): 1999–2014.

World Bank. 2020. “The World Bank in Thailand: Overview.” https://www.worldbank.org/en/country/thailand/overview.

Ye, Tingting, Naizhuo Zhao, Xuchao Yang, Zutao Ouyang, Xiaoping Liu, Quian Chen, Kejia Hu, Wenze Yue, Jiaguo Qi, Zhansheng Li, and Peng Jia. 2019. “Improved Population Mapping for China Using Remotely Sensed and Points-of-Interest Data within a Random Forests Model.” *Science of the Total Environment* 658: 936–946. https://doi.org/10.1016/j.scitotenv.2018.12.276.

Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. 2020. “Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa.” *Nature Communications* 11: 2583. https://doi.org/10.1038/s41467-020-16185-w.

Youneszadeh, S., Nina Kruger geb. Amiri, and Peter Pilesjo. 2015. “The Effect of Land Use Change on Land Surface Temperature in the Netherlands.” *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 40 (1): 745.

APPENDIX

Table A.1: List of Variables Obtained from Geospatial Data of 2015

Variable	Definition
VNTL2015f_sum	VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, year 2015
log_VIIRS_2015	Logarithm (based 10) of VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, 2015
log_VIIRS_density_2015	Logarithm (based 10) of VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, 2015, per area
log_POP_2018	Logarithm (based 10) of population size, 2018
log_Total_Pop_density	Logarithm (based 10) of population density, year 2015
log_LST_2015	Logarithm (based 10) of land surface temperature, 2015
log_Rain_2015	Logarithm (based 10) of amount of rainfall, 2015
SYNMAP_46	Synergetic Land Cover, pixel count of Urban, 2000
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
GUF_255	Global Urban Footprint, pixel count of built-up areas (2011–12?)
log_GUF_255	Logarithm (based 10) of Global Urban Footprint, pixel count of built-up areas (2011–12)
GHSLsmod2015_2	Global Human Settlement Layer, pixel count of "urban clusters" or low-density clusters, 2015
GHSLmod2015_3	Global Human Settlement Layer, pixel count of "urban centres" or high-density clusters, 2015
log_NDVI_2015	Logarithm (based 10) of normalized difference of vegetation index, 2015
log_NDVI_density_2015	Logarithm (based 10) of normalized difference of vegetation index, 2015, per area
USGS_0	USGS Land Cover, pixel count of Water (2001–2010 data)
USGS_1	USGS Land Cover, pixel count of Evergreen Needle Leaf Forest (2001–2010 data)
USGS_2	USGS Land Cover, pixel count of Evergreen Broadleaf Forest (2001–2010 data)
USGS_3	USGS Land Cover, pixel count of Deciduous Needle Leaf Forest (2001–2010 data)
USGS_4	USGS Land Cover, pixel count of Deciduous Broadleaf Forest (2001–2010 data)
USGS_5	USGS Land Cover, pixel count of Mixed Forests (2001–2010 data)
USGS_6	USGS Land Cover, pixel count of Closed Shrublands (2001–2010 data)
USGS_7	USGS Land Cover, pixel count of Open Shrublands (2001–2010 data)
USGS_8	USGS Land Cover, pixel count of Woody Savannas (2001–2010 data)
USGS_9	USGS Land Cover, pixel count of Savannas (2001–2010 data)
USGS_10	USGS Land Cover, pixel count of Grasslands (2001–2010 data)
USGS_11	USGS Land Cover, pixel count of Permanent Wetland (2001–2010 data)
USGS_12	USGS Land Cover, pixel count of Croplands (2001–2010 data)
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
USGS_14	USGS Land Cover, pixel count of Cropland/Natural Vegetation Mosaic (2001–2010 data)
USGS_16	USGS Land Cover, pixel count of Barren or Sparsely Vegetated (2001–2010 data)
USGS_PCA	USGS Land Cover, First Principal Component of USGS_0–USGS_16
ESALC2015_10	ESA Land Cover, pixel count of Cropland, rainfed, 2015
ESALC2015_11	ESA Land Cover, pixel count of Herbaceous cover, 2015
ESALC2015_12	ESA Land Cover, pixel count of Tree or shrub cover, 2015
ESALC2015_20	ESA Land Cover, pixel count of Cropland, irrigated or post-flooding, 2015

continued on next page

Table A.1: ***Continued***

Variable	Definition
ESALC2015_30	ESA Land Cover, pixel count of Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%), 2015
ESALC2015_40	ESA Land Cover, pixel count of Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%), 2015
ESALC2015_50	ESA Land Cover, pixel count of Tree cover, broadleaved, evergreen, closed to open (>15%), 2015
ESALC2015_60	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed to open (>15%), 2015
ESALC2015_61	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed (>40%), 2015
ESALC2015_70	ESA Land Cover, pixel count of Tree cover, needleleaved, evergreen, closed to open (>15%), 2015
ESALC2015_80	ESA Land Cover, pixel count of Tree cover, needleleaved, deciduous, closed to open (>15%), 2015
ESALC2015_100	ESA Land Cover, pixel count of Mosaic tree and shrub (>50%) / herbaceous cover (<50%), 2015
ESALC2015_110	ESA Land Cover, pixel count of Mosaic herbaceous cover (>50%) / tree and shrub (<50%), 2015
ESALC2015_120	ESA Land Cover, pixel count of Shrubland, 2015
ESALC2015_121	ESA Land Cover, pixel count of Evergreen shrubland, 2015
ESALC2015_122	ESA Land Cover, pixel count of Deciduous shrubland, 2015
ESALC2015_130	ESA Land Cover, pixel count of Grassland, 2015
ESALC2015_150	ESA Land Cover, pixel count of Sparse vegetation (tree, shrub, herbaceous cover) (<15%), 2015
ESALC2015_160	ESA Land Cover, pixel count of Tree cover, flooded, fresh or brackish water, 2015
ESALC2015_170	ESA Land Cover, pixel count of Tree cover, flooded, saline water, 2015
ESALC2015_180	ESA Land Cover, pixel count of Shrub or herbaceous cover, flooded, fresh/saline/brackish water, 2015
ESALC2015_190	ESA Land Cover, pixel count of Urban areas, 2015
ESALC2015_200	ESA Land Cover, pixel count of Bare areas, 2015
ESALC2015_210	ESA Land Cover, pixel count of Water bodies, 2015
Density_2015_Road_Count	Number of road paths per area, 2015
Density_2015_Road_Length	Total length of road paths per area, 2015
Density_2015_POI	Number of Point of Interest (POI) per area, 2015
NESDB7_3	Number of POIs in 2015 of this type: manufacturing
NESDB7_7	Number of POIs in 2015 of this type: wholesale and retail trade and repair of motor vehicles
NESDB7_8	Number of POIs in 2015 of this type: transportation and storage
NESDB7_9	Number of POIs in 2015 of this type: accommodation and food service activities
NESDB7_10	Number of POIs in 2015 of this type: information and communication
NESDB7_11	Number of POIs in 2015 of this type: financial and insurance activities
NESDB7_13	Number of POIs in 2015 of this type: professional, scientific, and technical activities
NESDB7_14	Number of POIs in 2015 of this type: administrative and support service activities
NESDB7_15	Number of POIs in 2015 of this type: public administration and defense; compulsory social security
NESDB7_16	Number of POIs in 2015 of this type: education
NESDB7_17	Number of POIs in 2015 of this type: human health activities

continued on next page

Table A.1: ***Continued***

Variable	Definition
NESDB7_industry	Number of POIs in 2015 of this type: mining and quarrying / manufacturing / electricity, gas, steam, and air conditioning supply / water supply, sewerage, waste management and remediation activities / construction
NESDB7_svcs1	Number of POIs in 2015 of this type: wholesale and retail trade and repair of motor vehicles / transportation and storage / accommodation and food service activities
NESDB7_svcs2	Number of POIs in 2015 of this type: information and communication / financial and insurance activities / real estate activities / professional, scientific, and technical activities / administrative and support service activities
NESDB7_svcs3	Number of POIs in 2015 of this type: public administration and defense; compulsory social security / education / human health activities / arts, entertainment and recreation / other service activities
Density_POI_Area_2015	Total area of Point of Interest per area, 2015
Density_NESDB7_industry	Number of POIs per sq km in 2015 of this type: mining and quarrying / manufacturing / electricity, gas, steam, and air conditioning supply / water supply, sewerage, waste management and remediation activities / construction
Density_NESDB7_svcs1	Number of POIs per sq km in 2015 of this type: wholesale and retail trade and repair of motor vehicles / transportation and storage / accommodation and food service activities
Density_NESDB7_svcs2	Number of POIs per sq km in 2015 of this type: information and communication / financial and insurance activities / real estate activities / professional, scientific, and technical activities / administrative and support service activities
Density_NESDB7_svcs3	Number of POIs per sq km in 2015 of this type: public administration and defense; compulsory social security / education / human health activities / arts, entertainment and recreation / other service activities
log_House_density_2015	Logarithm (based 10) of registered house, 2015, per area
Density_Building_Area	Total sq. meter of building per area, year 2015 and 2015
log_F2015_Buil_Density	Logarithm (based 10) of build-up (square meter), 2015, per area

DNB = Day Night Band, ESA = European Space Agency, POI = point of interest, USGS = United States Geological Survey, VIIRS = Visible and Infrared Imaging/Radiometer Suite.

Sources: Directorate General Joint Research Centre of the European Commission, European Space Agency, German Aerospace Center, Google Earth Engine, National Economic and Social Development Committee of Thailand, National Statistical Office of Thailand, Open Street Map, Thailand Development Research Institute, and United States Geological Survey.

Table A.2: List of Variables Obtained from Geospatial Data of 2017

Variable	Definition
VNTL2017f_sum	VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, year 2017
log_VIIRS_2017	Logarithm (based 10) of VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, 2017
log_VIIRS_density_2017	Logarithm (based 10) of VIIRS cloud mask—outlier removed—nighttime lights average DNB radiance, 2017, per area
log_POP_2018	Logarithm (based 10) of population size, 2018
log_Total_Pop_density	Logarithm (based 10) of population density, year 2017
log_LST_2017	Logarithm (based 10) of land surface temperature, 2017
log_Rain_2017	Logarithm (based 10) of amount of rainfall, 2017
SYNMAP_46	Synergetic Land Cover, pixel count of Urban, 2000
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
GUF_255	Global Urban Footprint, pixel count of built-up areas (2011–12?)
log_GUF_255	Logarithm (based 10) of Global Urban Footprint, pixel count of built-up areas (2011–12?)
GHSLsmod2017_2	Global Human Settlement Layer, pixel count of "urban clusters" or low-density clusters, 2017
GHSLmod2017_3	Global Human Settlement Layer, pixel count of "urban centres" or high-density clusters, 2017
log_NDVI_2017	Logarithm (based 10) of normalized difference of vegetation index, 2017
log_NDVI_density_2017	Logarithm (based 10) of normalized difference of vegetation index, 2017, per area
USGS_0	USGS Land Cover, pixel count of Water (2001–2010 data)
USGS_1	USGS Land Cover, pixel count of Evergreen Needle Leaf Forest (2001–2010 data)
USGS_2	USGS Land Cover, pixel count of Evergreen Broadleaf Forest (2001–2010 data)
USGS_3	USGS Land Cover, pixel count of Deciduous Needle Leaf Forest (2001–2010 data)
USGS_4	USGS Land Cover, pixel count of Deciduous Broadleaf Forest (2001–2010 data)
USGS_5	USGS Land Cover, pixel count of Mixed Forests (2001–2010 data)
USGS_6	USGS Land Cover, pixel count of Closed Shrublands (2001–2010 data)
USGS_7	USGS Land Cover, pixel count of Open Shrublands (2001–2010 data)
USGS_8	USGS Land Cover, pixel count of Woody Savannas (2001–2010 data)
USGS_9	USGS Land Cover, pixel count of Savannas (2001–2010 data)
USGS_10	USGS Land Cover, pixel count of Grasslands (2001–2010 data)
USGS_11	USGS Land Cover, pixel count of Permanent Wetland (2001–2010 data)
USGS_12	USGS Land Cover, pixel count of Croplands (2001–2010 data)
USGS_13	USGS Land Cover, pixel count of Urban and Built-up (2001–2010 data)
USGS_14	USGS Land Cover, pixel count of Cropland/Natural Vegetation Mosaic (2001–2010 data)
USGS_16	USGS Land Cover, pixel count of Barren or Sparsely Vegetated (2001–2010 data)
USGS_PCA	USGS Land Cover, First Principal Component of USGS_0–USGS_16
ESALC2017_10	ESA Land Cover, pixel count of Cropland, rainfed, 2017
ESALC2017_11	ESA Land Cover, pixel count of Herbaceous cover, 2017
ESALC2017_12	ESA Land Cover, pixel count of Tree or shrub cover, 2017
ESALC2017_20	ESA Land Cover, pixel count of Cropland, irrigated or post-flooding, 2017
ESALC2017_30	ESA Land Cover, pixel count of Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover) (<50%), 2017
ESALC2017_40	ESA Land Cover, pixel count of Mosaic natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%), 2017

continued on next page

Table A.2: ***Continued***

Variable	Definition
ESALC2017_50	ESA Land Cover, pixel count of Tree cover, broadleaved, evergreen, closed to open (>15%), 2017
ESALC2017_60	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed to open (>15%), 2017
ESALC2017_61	ESA Land Cover, pixel count of Tree cover, broadleaved, deciduous, closed (>40%), 2017
ESALC2017_70	ESA Land Cover, pixel count of Tree cover, needleleaved, evergreen, closed to open (>15%), 2017
ESALC2017_80	ESA Land Cover, pixel count of Tree cover, needleleaved, deciduous, closed to open (>15%), 2017
ESALC2017_100	ESA Land Cover, pixel count of Mosaic tree and shrub (>50%) / herbaceous cover (<50%), 2017
ESALC2017_110	ESA Land Cover, pixel count of Mosaic herbaceous cover (>50%) / tree and shrub (<50%), 2017
ESALC2017_120	ESA Land Cover, pixel count of Shrubland, 2017
ESALC2017_121	ESA Land Cover, pixel count of Evergreen shrubland, 2017
ESALC2017_122	ESA Land Cover, pixel count of Deciduous shrubland, 2017
ESALC2017_130	ESA Land Cover, pixel count of Grassland, 2017
ESALC2017_150	ESA Land Cover, pixel count of Sparse vegetation (tree, shrub, herbaceous cover) (<15%), 2017
ESALC2017_160	ESA Land Cover, pixel count of Tree cover, flooded, fresh or brackish water, 2017
ESALC2017_170	ESA Land Cover, pixel count of Tree cover, flooded, saline water, 2017
ESALC2017_180	ESA Land Cover, pixel count of Shrub or herbaceous cover, flooded, fresh/saline/brackish water, 2017
ESALC2017_190	ESA Land Cover, pixel count of Urban areas, 2017
ESALC2017_200	ESA Land Cover, pixel count of Bare areas, 2017
ESALC2017_210	ESA Land Cover, pixel count of Water bodies, 2017
Density_2017_Road_Count	Number of road paths per area, 2017
Density_2017_Road_Length	Total length of road paths per area, 2017
Density_2017_POI	Number of Point of Interest (POI) per area, 2017
NESDB7_3	Number of POIs in 2017 of this type: manufacturing
NESDB7_7	Number of POIs in 2017 of this type: wholesale and retail trade and repair of motor vehicles
NESDB7_8	Number of POIs in 2017 of this type: transportation and storage
NESDB7_9	Number of POIs in 2017 of this type: accommodation and food service activities
NESDB7_10	Number of POIs in 2017 of this type: information and communication
NESDB7_11	Number of POIs in 2017 of this type: financial and insurance activities
NESDB7_13	Number of POIs in 2017 of this type: professional, scientific, and technical activities
NESDB7_14	Number of POIs in 2017 of this type: administrative and support service activities
NESDB7_15	Number of POIs in 2017 of this type: public administration and defense; compulsory social security
NESDB7_16	Number of POIs in 2017 of this type: education
NESDB7_17	Number of POIs in 2017 of this type: human health activities
NESDB7_industry	Number of POIs in 2017 of this type: mining and quarrying / manufacturing / electricity, gas, steam, and air conditioning supply / water supply, sewerage, waste management and remediation activities / construction
NESDB7_svcs1	Number of POIs in 2017 of this type: wholesale and retail trade and repair of motor vehicles / transportation and storage / accommodation and food service activities

continued on next page

Table A.2: ***Continued***

Variable	Definition
NESDB7_svcs2	Number of POIs in 2017 of this type: information and communication / financial and insurance activities / real estate activities / professional, scientific, and technical activities / administrative and support service activities
NESDB7_svcs3	Number of POIs in 2017 of this type: public administration and defense; compulsory social security / education / human health activities / arts, entertainment and recreation / other service activities
Density_POI_Area_2017	Total area of Point of Interest per area, 2017
Density_NESDB7_industry	Number of POIs per sq km in 2017 of this type: mining and quarrying / manufacturing / electricity, gas, steam, and air conditioning supply / water supply, sewerage, waste management and remediation activities / construction
Density_NESDB7_svcs1	Number of POIs per sq km in 2017 of this type: wholesale and retail trade and repair of motor vehicles / transportation and storage / accommodation and food service activities
Density_NESDB7_svcs2	Number of POIs per sq km in 2017 of this type: information and communication / financial and insurance activities / real estate activities / professional, scientific, and technical activities / administrative and support service activities
Density_NESDB7_svcs3	Number of POIs per sq km in 2017 of this type: public administration and defense; compulsory social security / education / human health activities / arts, entertainment and recreation / other service activities
log_House_density_2017	Logarithm (based 10) of registered house, 2017, per area
Density_Building_Area	Total sq. meter of building per area, year 2017 and 2017
log_F2017_Buil_Density	Logarithm (based 10) of build-up (square meter), 2017, per area

DNB = Day Night Band, ESA = European Space Agency, POI = point of interest, USGS = United States Geological Survey, VIIRS = Visible and Infrared Imaging/Radiometer Suite.

Sources: Directorate General Joint Research Centre of the European Commission, European Space Agency, German Aerospace Center, Google Earth Engine, National Economic and Social Development Committee of Thailand, National Statistical Office of Thailand, Open Street Map, Thailand Development Research Institute, and United States Geological Survey.

Predicting Poverty Using Geospatial Data in Thailand

This study examines an alternative approach in estimating poverty by investigating whether readily available geospatial data can accurately predict the spatial distribution of poverty in Thailand. It also compares the predictive performance of various econometric and machine learning methods such as generalized least squares, neural network, random forest, and support vector regression. Results suggest that intensity of night lights and other variables that approximate population density are highly associated with the proportion of population living in poverty. The random forest technique yielded the highest level of prediction accuracy among the methods considered, perhaps due to its capability to fit complex association structures even with small and medium-sized datasets.

About the Asian Development Bank

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68 members —49 from the region. Its main instruments for helping its developing member countries are policy dialogue, loans, equity investments, guarantees, grants, and technical assistance.

ADB

ASIAN DEVELOPMENT BANK
6 ADB Avenue, Mandaluyong City
1550 Metro Manila, Philippines
www.adb.org