Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals

Jacinta Holloway, Kerrie Mengersen, Kate Helmstedt

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 6.C., Day 3, 21/09, 15:30: Measuring sustainable development with satellite imagery data and spatial statistical machine learning methods

Jacinta Holloway
j1.holloway@qut.edu.au
Queensland University of Technology

Distinguished Professor Kerrie Mengersen
k.mengersen@qut.edu.au
Queensland University of Technology

Dr Kate Helmstedt
kate.helmstedt@qut.edu.au
Queensland University of Technology

Spatial and machine learning methods of satellite imagery analysis for Sustainable Development Goals

DRAFT VERSION 31/08/2018

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

ABSTRACT

The United Nations (UN) and World Bank have set Sustainable Development Goals (SDGs), with the aim for countries to reach targets related to important aspects of quality of life by 2030. An essential element of sustainable development is achieving social and economic aims to improve human quality of life, while conserving and managing natural resources. Earth observation data, such as satellite imagery data, are increasingly being used for monitoring the SDGs, and statistical machine learning methods are commonly used to analyse these types of data. However, current methods often exclude the spatial information inherent in earth observation data, which can provide useful insights. In this paper we review how spatial information is currently measured for remote sensing data, describe spatial machine learning methods in the literature and opportunities for further development of spatial methods. We also describe a minimum set of requirements to measure SDGs from satellite imagery data.

Keywords: satellite images, sustainable development goals, big data, machine learning, spatial data, decision trees, remote sensing

1. INTRODUCTION

The United Nations (UN) and World Bank have set Sustainable Development Goals (SDGs), with the aim for countries to reach targets related to important aspects of quality of life by 2030 (United Nations, 2018). For example, food security, sustainable cities and communities, clean water and life on land are all SDGs with targets. The United Nations has acknowledged the role big data can play in measuring and monitoring progress towards these SDGs, and supports a number of Task Teams which have investigated and implemented methods for using big data sources for official statistics, and continue to explore these applications (United Nations, 2018).

Earth observation data, including satellite images, are an example of a big data source which can be obtained at no cost, for a long time series, and used to produce statistics and indicators to measure sustainable development (United Nations, 2017). In 2017, the United Nations Task Team on Satellite Imagery and Geospatial Data published a report on the feasibility of using earth observation data to produce official statistics, including statistics relevant to SDGs such as agricultural indicators and land cover (United Nations, 2017). There are a number of useful resources on the topic of earth observation data for official statistics and SDGs, including Satellite earth observations in support of the SDGs (*CEOS EO HANDBOOK*, 2018), Earth Observation for Water Resources Management (García, Rodríguez, Wijnen, & Pakulski, 2016) and the United Nations Food and Agriculture Organisation’s Handbook on remote sensing for agricultural statistics (FAO, 2016).

Satellite images can be used to identify features of interest, such as agricultural land, forests, urban areas, roads and water based on how they appear in the images. This is called land cover classification (FAO, 2016). Identifying these features of interest is often viewed as a classification problem, which requires methods which label individual pixels as belonging to a class based on their spectral characteristics (Sharma, Ghosh, & Joshi, 2013). For example, pixels can be classified into different crop types, or a binary classification of forest or bare ground.

Statistical analyses of satellite images, such as land cover, can be applicable to indicators of SDGs. The Group on Earth Observation (GEO) has identified that a number of SDGs can be measured at some level using earth observation data (GEO, 2016, pp.5). In figure 1, blue boxes indicate the goal can be measured based on earth observation data. Examples include agricultural monitoring relevant to *SDG 2: No hunger* and biodiversity and ecosystem monitoring relevant to *SDG 15: Life on land*.

Figure 1 Sustainable Development Goals measurable by earth observation data

Source: (GEO, 2016, pp.5).

It is often much cheaper to perform statistical analyses on free satellite imagery data to measure SDGs compared with directly collecting data in the field or through surveys. However, some amount of directly collected data, or ‘ground truth’ data is required to validate results derived from statistical analysis of satellite images. To minimise the potential negative impacts of missing or limited ground truth data, methods which model spatial relationships in satellite images can be used.

In this paper we review spatial approaches for analysing satellite imagery data, emerging spatial machine learning methods and outline a minimum set of requirements for measuring SDGs from satellite imagery data.

2. SPATIAL AUTOCORRELATION AND EARTH OBSERVATION DATA

Spatial autocorrelation occurs when a variable is correlated with itself in space, and this correlation exists after the effects of other variables have been accounted for (Hefley et al., 2017). Mathematically, spatial autocorrelation can be expressed as

$$\Gamma_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}\, Y_{ij}. \qquad (1)$$

The spatial autocorrelation, Γ_{ij}, between location i and all other sites j is given by two matrices, W and Y (Getis, 2010). W_{ij} is a matrix of the spatial relationship between location i and other sites j, and Y_{ij} is a matrix of non-spatial relationships between observation Y at site i and the value of that observation at other sites, j (Getis, 2010).

When spatial autocorrelation is present in data, the values of the dependent variable are influenced by their location and proximity to other data. Satellite images are an example of earth observation data that have spatial autocorrelation. For satellite images, generally pixels located close to each other are more likely to have similar values than pixels spaced further apart (Woodcock, Strahler, & Jupp, 1988).

Modelling spatial autocorrelation is important because, in cases where a spatial relationship exists between observations, this provides additional information to inform statistical models. Traditional statistical models treat observations as independent, and in cases where there is spatial autocorrelation in the data this is not the case; the observations are spatially dependent. By overlooking spatial dependence, traditional models exclude useful information from the analysis.

3. APPROACHES FOR SPATIAL ANALYSIS OF EARTH OBSERVATION DATA

The existence, strength and nature of the spatial relationship between observations can be identified and quantified using a number of measures. For satellite images there are global measures, which apply to autocorrelation across an entire image, and local measures, which assess spatial autocorrelation for individual observations or groups of observations (Getis, 2010; Spiker & Warner, 2007). Local measures of spatial autocorrelation can identify which observations are contributing to spatial relationships evident at the global (whole image) level (Spiker & Warner, 2007). Examples of global measures of spatial autocorrelation include semivariance, Geary's c and Moran's I, and local measures include local Geary's c and local Moran's I. These measures can be applied to other spatial data, but are described in the satellite imagery data context in this review.

These measures of spatial autocorrelation are defined as follows. Semivariance is the average variance in observation values separated by some distance (Cressie, 1989; Peterson & Hoef, 2010). Moran's I measures autocorrelation as a function of covariance between observations, and can take values from -1, indicating strong and negative autocorrelation, to +1, indicating strong and positive autocorrelation (Spiker & Warner, 2007). Geary's c is a measure of spatial autocorrelation, derived from the squared difference between pixels, divided by the image variance (Spiker & Warner, 2007). Geary's c takes values between 0 and 2, with 0 representing maximum positive spatial autocorrelation, 1 no spatial autocorrelation and 2 maximum negative spatial autocorrelation (Spiker & Warner, 2007).

These global and local measures of spatial autocorrelation have been implemented for satellite image analysis in a number of ways. Examples of these applications are described in terms of global or local measures.

Application: Global measures of spatial autocorrelation

For satellite image analysis, global spatial autocorrelation measures identify spatial relationships at the image level and have many applications. For example, Moran's I and semivariance have been applied to satellite imagery analysis to detect landslide events in Myanmar (Mondini & C., 2017) and identify regions with higher and lower social heat vulnerability based on derived surface temperature, normalized difference vegetation index (NDVI) and socioeconomic data (Sim, 2017).

Das & Ghosh, (2017) used an index based on estimated autocorrelations between neighbourhood pixels to classify regions in Landsat satellite imagery of the Kolkata region of India as changed or unchanged in terms of land cover. The authors use an extension of Moran's I to calculate the spatial autocorrelation between two Landsat images at time t and $t+t'$, then produce a binary change index. Gao, Tang, Jing, Li, & Ding, (2017) proposed an unsupervised evaluation method for segmenting high resolution remote sensing images for geographic object-based image analysis (GEOBIA), combine a measure of spatial stratified heterogeneity, the q statistic, with Global Moran's I statistic, a measure of spatial autocorrelation, to produce a global assessment metric.

Application: Local measures of spatial autocorrelation

For satellite image analysis, local spatial autocorrelation measures apply to the observation level, such as a pixel or cluster of pixels within an image. For example, Greene, Robinson, & Millward, (2018) used a spatial autocorrelation indicator, bivariate Moran's I, to evaluate spatial relationships between median household income and access to urban tree canopy in Toronto, Canada based on data derived from QuickBird satellite imagery classified by the USDA Forestry Service. The method identified statistically significant spatial clusters of high and low urban tree canopy.

Based on analysis of data derived from QuickBird panchromatic satellite imagery of Morgantown, West Virginia, USA, Spiker & Warner (2007) concluded global autocorrelation measures; semivariance, Geary's c and Moran's I, are useful to quantify autocorrelation in remote sensing images, but local measures, including local Moran's I and local Geary's c, are necessary and useful to understand which are the dominant contributors to these global metrics.

Other methods for dealing with spatial autocorrelation

Besides these global and local measures of spatial autocorrelation, other approaches have been implemented to address the spatial autocorrelation in satellite images. Other approaches Wang et al (2016) performed a review of methods for incorporating spatial information in satellite image classification, covering methods under five categories: pre-classification, sample selection, classifiers, post-classification and accuracy assessment (Wang, Shi, Diao, Ji, & Yin, (2016)). An example of incorporating spatial information in classification of satellite imagery is object based image analysis (OBIA) which first segments an image into homogeneous sections based on some feature e.g. topography and then runs parametric, machine learning or fuzzy logic methods on these segments rather than individual pixels to classify them (Wang et al., 2016). Geographic object-based image analysis (GEOBIA) builds on OBIA by emphasizing the geographical space of the objects in a satellite image (Wang et al., 2016).

An example of addressing spatial autocorrelation at the sample selection stage is Sheffield, Morse-McNabb, Clark, Robson, & Lewis, (2015); the authors attempted to address autocorrelation in the data to map dominant land use over time in Victoria, Australia by selecting one pixel per parcel of land and selecting those pixels that were most representative of the land parcel. Their approach is in data selection rather than in specification of their model, which means it was developed specifically for these data and would not necessarily apply to other problems and data sources. For more examples of spatial autocorrelation approaches within each of the five categories see Wang et al (2016).

Evidently, there are a number of ways spatial autocorrelation is measured in statistical analysis of satellite imagery data. Another approach which is emerging is the adaptation of existing machine learning methods, which perform faster and often have higher accuracy than other methods, to model spatial relationships. This will be described in section 4.

4. SPATIAL MACHINE LEARNING FOR EARTH OBSERVATION DATA

Statistical machine learning methods, also called artificial intelligence methods, use algorithmic models to analyse data. These methods treat the way in which the data were generated or relationships between the variables as unknown (Breiman, 2001). However, the existence of spatial relationships in satellite images is known, and an emerging way of modelling these relationships is to adapt existing machine learning algorithms proven to be effective for analysing these types of data. In this paper, we focus on decision tree methods.

A decision tree is a classification method which can be used to categorise large numbers of observations, such as pixels in satellite images. The classification structure of a decision tree is estimated from training data using a statistical procedure (Sharma et al., 2013). The 'tree' is made of a root node, internal nodes and leaves. Nodes are located where trees branch or split the dataset and terminal nodes are called leaves. Leaves contain the most homogeneous classes, or final classifications (Sharma et al., 2013). Due to their computational simplicity, decision tree can be applied to large amounts of data, which is useful for satellite imagery analysis, that is often performed on thousands or millions of pixels. Examples of decision trees in the remote sensing literature include identifying land cover (Al-Obeidat, Al-Taani, Belacel, Feltrin & Banerjee, 2015); mapping forest change globally (Hansen et al., 2013); and identifying land use classes such as agriculture, built up areas and water in Surat city, India (Sharma et al., 2013).

Spatial decision trees, then, are a group of statistical machine learning methods which extend on existing decision tree algorithms to model spatial relationships. These methods are suited to estimating spatial autocorrelation because they can effectively handle very large numbers of observations and variables, which has been a computational barrier to modelling spatial relationships in satellite imagery data (Griffith & Chun, 2016). Some examples of spatial decision tree methods are as follows.

Liu, Cao, Zhao, Mulligan, & Ye, (2018) use a geostatistical random forest approach, which applies regression kriging. Regression kriging is a hybrid approach that combines linear regression term of the primary variable on auxiliary variables to estimate trend with a kriging term of the regression residuals. The authors apply the Random Forest Regression Kriging (RFRK) model created by (Hengl et al., 2015) to measure particulate matter concentration derived from satellite imagery and ground measurements. The RFRK is a two-step procedure where the random forest is used to simulate non-linear trend between the dependent variable and covariates, and at the second step kriging is used to estimate residuals of the random forest trend. The authors found the RFRK approach was able to capture non-linear relationships between variables well, as is expected with a random forest model, and explicitly model spatial dependence of the particular concentrations.

Hengl et al. (2018) created a random forest for spatial predictions framework (RFsp) which includes euclidean buffer distances of observations as covariates to model spatial autocorrelation. The RFsp was found to produce estimates as accurate and unbiased as common geostatistical method, kriging, however the authors caution using this approach for greater than 1000 observations due to computational difficulties. Typical analysis of remote sensing data is performed on thousands or millions of pixels, so this could be a challenge for applying the RFsp method to these data. Li, Heap, Potter, & Daniell (2011) investigated Support Vector Machine (SVM) and random forest algorithms for spatial modelling of environmental variables, in this case Australian seabed mud content. The machine learning methods were compared with a number of geostatistical methods, including kriging, co-kriging and universal kriging, in addition to other statistical methods such as generalised linear models. The authors found random forest models combined with ordinary kriging and random forest combined with inverse distance squared outperformed all other methods in terms of predictive accuracy.

These examples demonstrate the implementation of spatial approaches to machine learning methods have been explored in the literature and found to have value for modelling environmental variables which have underlying spatial autocorrelation. A limitation of spatial random forest approaches is the high computational demand. Through evaluations not reported here, spatial implementation of a boosted regression tree method, called gradient boosted machine, was found to be more computationally feasible than the same spatial implementation of a random forest method.

Spatial decision tree models combine key benefits of statistical machine learning methods for large spatial data; ability to fit non-linear relationships, effectively cope with significantly large amounts of data and high predictive accuracy, and can further improve this predictive accuracy by modelling the spatial dependence that exists between the pixels. However, existing spatial decision tree models have a high computational cost, and there is opportunity to explore models that can approximate spatial relationships and efficiently handle larger numbers of pixels.

5. REQUIREMENTS FOR MEASURING SUSTAINABLE DEVELOPMENT GOALS FROM SATELLITE IMAGERY DATA

Some consideration should be given to the minimum resources required in order to measure Sustainable Development Goal (SDG) indicators using satellite imagery data. This section describes recommendations on some minimum requirements for measuring SDGs using satellite imagery data, based on research, including contributing chapters to the United Nations Satellite Imagery and Geospatial Data Task Team report (United Nations, 2017) and practical experience of the authors.

In order to measure sustainable development indicators from satellite images, at the simplest level there are three requirements. Firstly, the indicator needs to be identifiable in a satellite image. Secondly, satellite images need to be available for the location of interest. Thirdly, some form of validation data is needed to verify statistical outputs produced by the model. These criteria are expanded on here.

1. **The indicator is measurable from satellite imagery data.** At the simplest level, this means the indicator can be seen or extracted from a satellite image. For example, turbidity in water is visible from the spectral bands of a satellite image, and this is also an indicator of water quality (Phinn, 2005). This statistical output links to SDG 6: *Clean water and sanitation* and indicator 6.3.2 *Percentage of bodies of water with good ambient water quality* (United Nations, 2017). Some additional examples of sustainable development indicators measurable from remote sensing data are in Table 1.

Table 1. Remote sensing data, as used to measure UN Sustainable Development Goals.

Sustainable Development Goal	Sustainable Development Target Description	Remote Sensing Application and Indicator
Goal 2: End Hunger	End hunger, achieve food security and improved nutrition, and promote sustainable agriculture. By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding, and other disasters, and that progressively improve land and soil quality.	Crop area estimation, crop yield, land cover classification, extreme weather event detection. Indicator 2.4.1 Proportion of agricultural area under productive and sustainable agriculture.
Goal 6: Clean water and sanitation	Ensure availability and sustainable management of water and sanitation for all. By 2020, protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers, and lakes.	Water quality monitoring. Indicator 6.6.1 Change in the extent of water-related ecosystems over time. Indicator 6.3.2 Proportion of bodies of water with good ambient water quality.
Goal 15: Life on land	Protect, restore, and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, halt and reverse land degradation, and halt biodiversity loss. By 2020, ensure the conservation, restoration, and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains, and drylands, in line with obligations under international agreements.	Forest cover monitoring, deforestation detection. Indicator 15.1.1 Forest area as a proportion of total land area. Indicator 15.3.1 Proportion of land that is degraded over total land area. Indicator 15.4.2 Mountain Green Cover Index.

Source: (Holloway & Mengersen, 2018)

More extensive description of SDGs measurable from remote sensing data are provided by the Group on Earth Observation (GEO, 2018).

2. **Analysis-ready satellite images are available.** Before satellite images can be statistically analysed, some pre-processing is required to correct for atmospheric effects such as cloud shadows and topographic adjustments (United Nations, 2017). Pre-processing can be performed by the practitioner, however this requires earth science domain specific knowledge and skills which are not always available. Alternative sources of free analysis ready satellite imagery data include the Digital Earth Australia Data Cube (Geoscience Australia, 2018), CEOS Data Cube products (Killough, 2017) and United States Geological Survey (USGS, 2018).

3. **Access to validation data and/or spatial information.** Access to validation data, also referred to as 'ground truth' is auxiliary data that can be used to validate statistical outputs produced by modelling satellite imagery data. Examples of ground truth include directly collected field measurements, survey results and weather data. In general, directly collected data is thought to have higher accuracy than modelled data, however statistics produced from satellite images are useful because they can be produced at a larger spatial scale, at lower cost and more frequently than other methods such as surveys and field measurements (United Nations, 2017).

Lack of access to directly collected data, often due to prohibitively high costs, is a key motivation to explore free big data sources such as satellite images. While ideally there would be large amounts of ground truth data available to validate statistical models, in practice when this is not possible modelled estimates with some uncertainty around them are useful as an alternative to no measurement of sustainable development indicators. The addition of spatial information will only be beneficial if there is a spatial relationship in the data, and this is often the case for satellite imagery data (Spiker & Warner, 2007). The global and local measures of spatial autocorrelation described in section 3 can be used to determine the existence, strength and nature of a spatial relationship in satellite data. When there is a lack of validation data to verify model derived statistics, other methods of improving model accuracy can be useful. This is an example of where incorporating spatial information can improve predictive accuracy of existing statistical models, such as the decision trees described in section 4.

There are many other considerations for producing SDG indicators and official statistics based on satellite imagery data. Useful criteria for determining whether earth observation data, such as satellite images, is suitable for producing a particular statistical output are provided by the United Nations Satellite Imagery and Geospatial Data Task Team report (United Nations, 2017, pp.118-119). Further considerations in terms of a cost-benefit analysis for a national statistical office considering using earth observation data for statistical production are beyond the scope of this review, but are defined by Tam & Clarke (2015). The cost benefit criteria described by Tam & Clarke (2015) include reduction in provider load, sustainability of the data source, timeliness and accuracy, and are also discussed in chapter 5 of the United Nations Satellite Imagery and Geospatial Data Task Team report (United Nations, 2017).

In terms of statistical methods, there are many which can be applied to remote sensing data to measure environmental features relevant to SDGs to produce accurate results (Holloway & Mengersen 2018). There is also opportunity to go beyond identifying spatial autocorrelation using local and global measures, and further develop existing statistical and machine learning methods to explicitly model spatial autocorrelation in satellite imagery data.

6. CONCLUSION

The United Nations and World Bank have identified Sustainable Development Goals (SDGs) and set priorities for countries to achieve these goals by 2030. Freely available satellite images are a useful data source that can be analysed to measure indicators of progress towards the SDGs. One issue with SDG indicators produced from satellite images is typically some validation data is needed to verify the accuracy of the statistical model results, and this validation data is often expensive to collect or unavailable. By including spatial information in analysis of satellite images, it is possible to achieve more accurate estimates without additional directly collected data. In this paper, we reviewed how spatial information is currently measured for satellite imagery data, described spatial machine learning methods currently in the literature and opportunities for further development of spatial methods. Spatial machine learning methods are continually being developed and have been used to produce accurate environmental statistics. There is scope for further development of spatial machine learning methods.

We also described a minimum set of three requirements to measure SDGs from satellite images; the indicator needs to be identifiable in a satellite image, these images need to be available for the location of interest, and access to validation data and/or spatial information. We also provided references to other useful sources of criteria when considering whether it is advisable to produce official statistics and SDG indicators from earth observation data, such as satellite images.

7. References

Al-Obeidat Feras; Al-Taani, Ahmad T.; Belacel, Nabil; Feltrin, Leo; Banerjee, N. (2015). A Fuzzy Decision Tree for Processing Satellite Images and Landsat Data. *Procedia Computer Science*, *52*, 1192–1197. https://doi.org/10.1016/J.PROCS.2015.05.157

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–231. Retrieved from https://projecteuclid.org/download/pdf_1/euclid.ss/1009213726

CEOS EO HANDBOOK. (2018). Retrieved from http://eohandbook.com/sdg/

Cressie, N. (1989). Geostatistics. *The American Statistician*, *43*(4), 197. https://doi.org/10.2307/2685361

Das, M., & Ghosh, S. K. (2017). Spatio-temporal Autocorrelation Analysis for Regional Land-cover Change Detection from Remote Sensing Data. In *Proceedings of the Fourth ACM IKDD Conferences on Data Sciences - CODS '17* (pp. 1–10). New York, New York, USA: ACM Press. https://doi.org/10.1145/3041823.3041835

FAO. (2016). *Handbook on remote sensing for agricultural statistics*. Retrieved from http://gsars.org/wp-content/uploads/2017/09/GS-REMOTE-SENSING-HANDBOOK-FINAL-04.pdf

Gao, H., Tang, Y., Jing, L., Li, H., & Ding, H. (2017). A Novel Unsupervised Segmentation Quality Evaluation Method for Remote Sensing Images. *Sensors (Basel, Switzerland)*, *17*(10). https://doi.org/10.3390/s17102427

García, L., Rodríguez, D., Wijnen, M., & Pakulski, I. (Eds.). (2016). *Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector*. The World Bank. https://doi.org/10.1596/978-1-4648-0475-5

GEO. (2016). *Earth Observations and Geospatial Information: Supporting Official Statistics in Monitoring the SDGs*. Retrieved from http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

Geoscience Australia. (2018). Open Data Cube. Retrieved from http://www.ga.gov.au/dea/odc

Getis, A. (2010). Spatial Autocorrelation. In M. M. Fischer & A. Getis (Eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications* (pp. 255–278). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_14

Greene, C. S., Robinson, P. J., & Millward, A. A. (2018). Canopy of advantage: Who benefits most from city trees? *Journal of Environmental Management*, *208*, 24–35. https://doi.org/10.1016/J.JENVMAN.2017.12.015

Hansen, M. C., Potapov, P. V, Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science (New York, N.Y.)*, *342*(6160), 850–3. https://doi.org/10.1126/science.1244693

Hefley, T. J., Broms, K. M., Brost, B. M., Buderman, F. E., Kay, S. L., Scharf, H. R., … Hooten, M. B. (2017). The basis function approach for modeling autocorrelation in ecological data. *Ecology*, *98*(3), 632–646. https://doi.org/10.1002/ecy.1674/suppinfo

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., … Tondoh, J. E. (2015). Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLOS ONE*, *10*(6), e0125814. https://doi.org/10.1371/journal.pone.0125814

Holloway, J., & Mengersen, K. (2018). Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review. *Remote Sensing*, *10*(9), 1365. https://doi.org/10.3390/rs10091365

Killough, B. (2017). Status of the Open Data Cube. Retrieved from http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-44/3. Wednesday/2017.09.27_10.15_OpenDataCube_WGISS44.pdf

Li, J., Heap, A. D., Potter, A., & Daniell, J. J. (2011). Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software*, *26*(12), 1647–1659. https://doi.org/10.1016/J.ENVSOFT.2011.07.004

Liu, Y., Cao, G., Zhao, N., Mulligan, K., & Ye, X. (2018). Improve ground-level PM2.5 concentration mapping using a random forests-based geostatistical approach. *Environmental Pollution*, *235*, 272–

282. https://doi.org/10.1016/J.ENVPOL.2017.12.070
Mondini, A., & C., A. (2017). Measures of Spatial Autocorrelation Changes in Multitemporal SAR Images for Event Landslides Detection. *Remote Sensing*, *9*(12), 554. https://doi.org/10.3390/rs9060554
Nations, U. (2018). Sustainable Development Goals .:. Sustainable Development Knowledge Platform. Retrieved June 29, 2018, from https://sustainabledevelopment.un.org/?menu=1300
Peterson, E. E., & Hoef, J. M. Ver. (2010). A mixed-model moving-average approach to geostatistical modeling in stream networks. *Ecology*. WileyEcological Society of America. https://doi.org/10.2307/25661098
Sharma, R., Ghosh, A., & Joshi, P. K. (2013). Decision tree approach for classification of remotely sensed satellite data using open source support. *Journal of Earth System Science*, *122*(5), 1237–1247. https://doi.org/10.1007/s12040-013-0339-2
Sheffield, K., Morse-McNabb, E., Clark, R., Robson, S., & Lewis, H. (2015). Mapping dominant annual land cover from 2009 to 2013 across Victoria, Australia using satellite imagery. *Scientific Data*, *2*, 150069. https://doi.org/10.1038/sdata.2015.69
Sim, S. (2017). Social vulnerability to heat in Greater Atlanta, USA: spatial pattern of heat, NDVI, socioeconomics and household composition. In W. Heldens, N. Chrysoulakis, T. Erbertseder, & Y. Zhang (Eds.), *Remote Sensing Technologies and Applications in Urban Environments II* (Vol. 10431, p. 4). SPIE. https://doi.org/10.1117/12.2278678
Spiker, J. S., & Warner, T. A. (2007). Scale and Spatial Autocorrelation From A Remote Sensing Perspective. *Geo-Spatial Technologies in Urban Environments: Policy, Practice, and Pixels*, 197–213. https://doi.org/10.1007/978-3-540-69417-5_10
Tam, S.-M. & Clarke, F. (2015). Big Data, Statistical Inference and Official Statistics. International Statistical Review, 83(3), 436-448.
United Nations United Nations Global Working Group on Big Data for Official Statistics Available online: https://unstats.un.org/bigdata/ (accessed on Apr 10, 2018).
United Nations. (2017). Earth Observations for Official Statistics: Satellite Imagery and Geospatial Data Task Team report, (December). Retrieved from https://unstats.un.org/bigdata/taskteams/satellite/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf
USGS. (2018). GloVis - The USGS Global Visualization Viewer. Retrieved July 25, 2018, from https://glovis.usgs.gov/
Wang, L., Shi, C., Diao, C., Ji, W., & Yin, D. (2016). A survey of methods incorporating spatial information in image classification and spectral unmixing. *International Journal of Remote Sensing*, *37*(16), 3870–3910. https://doi.org/10.1080/01431161.2016.1204032
Woodcock, C. E., Strahler, A. H., & Jupp, D. L. B. (1988). The use of variograms in remote sensing: I. Scene models and simulated images. *Remote Sensing of Environment*, *25*(3), 323–348. https://doi.org/10.1016/0034-4257(88)90108-3