Supplementary Materials: Statistical Machine Learning Methods and Remote Sensing for Sustainable Development Goals: A Review

Jacinta Holloway and Kerrie Mengersen

Table S1. Earth observation and geospatial information resources for SDG monitoring.

SUSTAINABLE DEVELOPMENT GOALS	Population distribution	Cities and infrastructure mapping	Elevation and topography	Land cover and use mapping	Oceanographic observations	Hydrological and water quality observations	Atmospheric and air quality monitoring	Biodiversity and ecosystem observations	Agricultural monitoring	Hazards, disasters and environmental impact monitoring
1 No poverty	■	■		■					■	
2 Zero hunger	■	■	■	■	■	■	■	■	■	■
3 Good health and well-being	■	■		■		■	■		■	■
4 Quality education	■	■								
5 Gender equality	■	■								
6 Clean water and sanitation	■	■	■	■	■	■	■	■	■	■
7 Affordable and clean energy	■	■	■	■	■	■	■	■		■
8 Decent work and economic growth	■	■	■	■					■	
9 Industry, innovation and infrastructure	■	■	■	■			■			■
10 Reduced inequalities	■	■								
11 Sustainable cities and communities	■	■	■	■			■	■	■	■
12 Responsible consumption and production	■	■	■	■				■		■
13 Climate action	■	■	■	■	■	■	■	■	■	■
14 Life below water	■	■	■	■	■	■	■	■	■	■
15 Life on land	■	■	■	■		■	■	■	■	■
16 Peace, justice and strong institutions	■	■		■						
17 Partnerships for the goals	■	■	■	■	■	■	■	■	■	■

Source: [1]

1. The Three Step Process for Statistical Machine Learning Analyses of Remote Sensing Data

In this review, we focused on the analysis step of a three step process; pre-processing, analysis and evaluation. In the supplementary material we discuss the pre-processing and evaluation steps.

1.1. Pre-processing Considerations

We note here three general issues that impact strongly the analysis and evaluation steps.

1.1.1. Quality Assessment of Remote Sensing Data

There is a misconception that big data sources, such as remote sensing, will be statistically valid simply due to the magnitude of data available. Although increasing sample size can reduce sampling errors, it does not reduce other sources of statistical bias such as measurement error. Further, data sets derived from these sources are not necessarily random samples of the target population practitioners want to make inferences about, therefore increasing the size of the dataset may not reduce sampling error or improve the quality of the estimates. In addition, big data sources have their own specific issues which can impact on statistical validity. These include accumulation of errors (noise) and potential spurious correlations. The different sources of data, and their aggregation, can also lead to statistical biases[2].

Some of these challenges of using remote sensing data for statistical purposes can be addressed by using appropriate methodologies, depending on the data source and intended outputs. However, it should be noted that these methodologies will not counter issues related to data quality, data representativeness or other related issues.

1.1.2. Other Information Sources

Remote sensing data can be analysed on its own, or with other datasets. The choice of other datasets depends on the following considerations. The first is aim; different datasets will be required for different purposes, based on the required estimates, selected method for analysis, spatial and temporal resolution and scope of the analysis, quality of the available remote sensing data, and so on. The second is access; datasets can include existing resources and/or they can be obtained from external parties. Access considerations also include intellectual property, cost, and continuity of access. The third is content; a wide range of data can potentially complement remote sensing data in an analysis, such as censuses, household or agricultural surveys, administrative data, environmental and meteorological data.

The World Bank and FAO report; a Global Strategy for Improving Agricultural and Rural Statistics (2010) provides advice about constructing and maintaining a Minimum Set of Core Data (MSCD) that includes remote sensing and other spatial and non-spatial information in a Master Sample Frame (MSF) database[3]. The MSF is constructed by combining satellite images classified by geo-referenced land use and digitised administrative data, digitised enumeration areas, and overlaying this with population and agricultural census data, other relevant data such as a register of commercial farms, and an area frame [3].

1.1.3. Eliminating the Pre-Processing Step: Data Cubes

A potential alternative to performing your own pre-processing of satellite images is to access a Data Cube product if it is available for your region of interest. A Data Cube provides a time series of pre-processed satellite imagery data that is ready for analysis and free [4]. This reduces the data pre-processing time and makes the analysis of satellite imagery data more accessible to practitioners outside the earth science field. The processing in the Data Cubes is also performed consistently to the same standard[4], which makes it simpler to draw comparisons of data from multiple countries. Current coverage of the Data Cube products is shown in Figure S1.

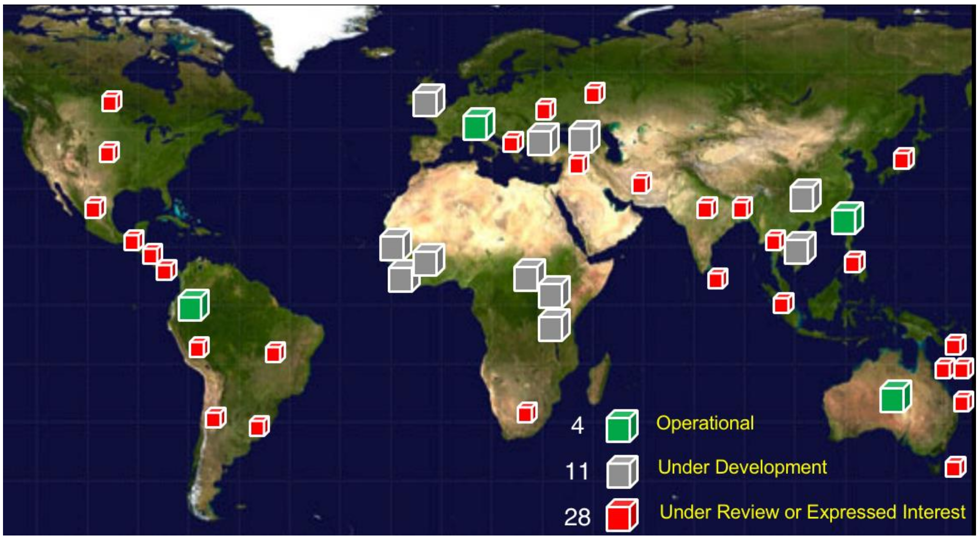

Figure S1. CEOS Data Cube products by status [5].

The Committee on Earth Observation Satellites (CEOS) currently has four operational Data Cubes in Australia, Colombia, Switzerland and Africa, with an additional eleven Data Cubes in development and 28 under review (43 total)[5]. As these Data Cubes become more prevalent, this will create more opportunities to perform analyses and make comparisons between countries based on satellite imagery data.

2. Evaluation Step

After the pre-processing step is completed or an analysis ready product has been chosen, the analysis step is completed as described in the body of the review. We now turn to the third and final step in the process; evaluation. The considerations to be addressed in the Evaluation step include critical assessment of the results and their accuracy, assessment of model accuracy and assessment of other potential biases or concerns with the results or their interpretation.

2.1. Critical Assessment of Results

As with any statistical analysis, it is essential to critically review the results with respect to whether they are sensible from both statistical and domain-knowledge perspectives. Evaluations can be performed on the model itself to determine whether the assumptions are met. Evaluations can also be performed on the statistical results; including critical consideration of the overall magnitude of the results, identification of results that are outside expected ranges, precision of the estimates, the size of the prediction intervals and other measures of result variability[6]. Once the results have been critically assessed, they can be compared with other relevant results and statistics to determine whether the results make sense in context[6]. Model and accuracy assessments are described in more detail in the following sections.

2.2. Model Assessment

Regardless of the type of methodological approach (SML, informed SML, physics based or object based), the goodness of fit of the model is typically assessed by comparing the estimated or predicted values obtained from the model with the observed values, if available. Statistical models are typically subjected to further goodness of fit evaluations using criteria such as AIC, BIC [7,8] and traditional hypothesis tests (if available and applicable). Where possible, estimates obtained from the model are typically compared with the observed data, using measures such as misclassification rates for categorical responses and mean squared error of prediction for continuous responses.

Depending on the available data, a typical statistical practice is to define complementary training and testing datasets by randomly partitioning the original data (for example, randomly splitting the dataset into a 75% training set and a 25% testing set). A model is then built using the training data, which is then assessed for accuracy in various ways on the test dataset. For example, for classification models, accumulative prediction error (APE) [7] could be used to assess how well the training data predict the test data, and hence how well the trained model may be expected to perform on new, unknown samples.

Cross-validation is another useful method of model assessment, which can be applied in almost any algorithm in most frameworks such as regression, classification and many others [8]. The idea of cross-validation is to repeat the validation procedure described above, using different data subsets each time. The measured fit (for example, the misclassification rate) is then averaged across these repeats to provide a more accurate measure of the predictive capabilities of the model under investigation [8]. Examples of cross-validation approaches include leave-one-out, leave-k-out and v-fold cross-validation [8]. As cross-validation techniques are so simple to apply and have minimal assumptions[8], the specific approach selected often simply comes down to computational burden.

2.3. Accuracy Assessment

Accuracy "is a relative measure of the exactness of an estimate and accounts for systematic errors also referred to as bias. Therefore, an accurate estimate does not systematically over or underestimate the true value" ([2], pp.1). When considering accuracy assessment, it is important to recognise that sampling effort (accuracy and representativeness of sampled data used for training) as well as the analytical technique(s) utilised play important roles. These issues have been described above.

Good practice involves reporting details of these processes, as well as details of the model fit, the accuracy of the estimates or predictions obtained from the model, and quantification of uncertainty of these estimates and predictions [9]. These issues are now discussed in more detail in the following sections.

2.4. Accuracy Assessment of Map Data

An example of accuracy assessment practices for remote sensing data analyses is given by the FAO Map Accuracy Assessment and Area Estimation Practical Guide [2] and we describe it here. In an accuracy assessment of map data, the map is compared with higher quality data. The higher quality data, called reference data, is collected through a sample-based approach, allowing for a more careful interpretation of specific areas of the map. The reference data is collected in a consistent manner and is harmonised with the map data, in order to compare the two classifications. The comparison results in accuracy measures and adjusted area estimates for each map category. This process is broken down into four major components: (i) a map, (ii) the sampling design (iii) the response design and (iv) the analysis. Here, we add an additional step, interpretation, in which the practitioner considers the implications of the accuracy assessment for their results and data collection and analysis processes.

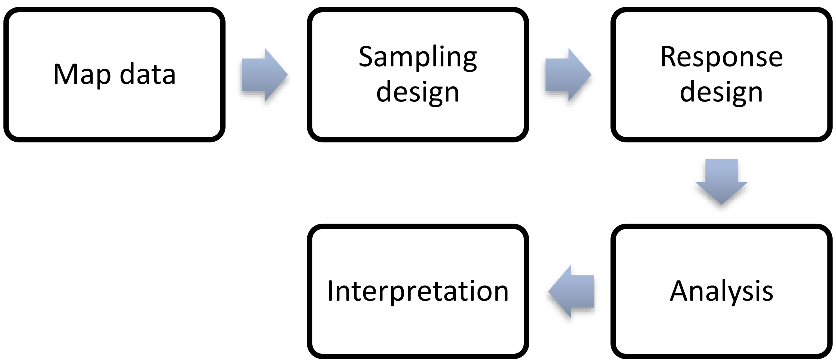

Figure S2. Adapted process for processing map data from raw form to interpreting statistical outputs.

At the map data stage, all map classes are defined, obvious errors are identified and corrected, strata are defined and their size is calculated. During sampling design, the approach, sample size, distribution of sample by strata, spatial unit of reference data and number of samples within the map data are determined. During the response design stage, the map class definitions are translated to the reference data classes and reference data is collected. Accuracy and area estimates and their confidence intervals are produced at the analysis stage [2]. After these four steps are completed, we propose an interpretation stage, when the accuracy estimates are considered in terms of the current study, and in terms of whether changes should be made at any of the previous stages (map data, sampling design, response design) to improve the accuracy of the final estimates.

2.5. Reporting Accuracy Assessment Results

FAO recommend reports of accuracy assessment results should include the estimates, adjusted area and their respective confidence intervals and relevant assumptions ([10], pp.23). Examples of assumptions that can influence level of accuracy include, but are not limited to, the minimum mapping unit and spatial assessment unit, the sampling design, the source of reference data and the confidence level used for calculating the confidence intervals (typically 95%). Presenting the error matrix in terms of estimated area proportions instead of absolute sample is also recommended [2].

FAO has applied these accuracy assessment steps to measure forest area and forest area change. These measures are important for countries with reporting requirements to access results-based payments for reducing emissions from deforestation and forest degradation. For details of this practical application of accuracy assessment, see the Global Forest Change (GFC) example in the FAO Map Accuracy Assessment and Area Estimation Practical Guide 2016 [2]. For further detailed information about using remote sensing data for crop identification and crop yield, refer to the FAO Handbook on Remote Sensing for Agriculture Statistics [11].

References

1. GEO *Earth Observations and Geospatial Information: Supporting Official Statistics in Monitoring the SDGs*. Available online: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E (accessed on 7 August 2018).

2. Food and Agriculture Organization of the United Nations *Map Accuracy Assessment and Area Estimation: A Practical Guide*; Rome, 2016. Available online: http://www.fao.org/3/a-i5601e.pdf (accessed on 16 April 2018).

3. World Bank, Food and Agricultural Organisation (FAO) of the U. N. 2010. *Global Strategy to Improve Agricultural and Rural Statistics*. Available online: http://www.fao.org/docrep/015/am082e/am082e00.pdf (accessed on 10 April 2018).

4. Committee on Earth Observation Satellites Open Data Cube. Available online: https://www.opendatacube.org/ceos (accessed on Apr 26, 2018).

5. Killough, B. Open Data Cube Background and Vision 2017.

6. Kass, R. E.; Caffo, B. S.; Davidian, M.; Meng, X.-L.; Yu, B.; Reid, N. Ten Simple Rules for Effective Statistical Practice. *PLOS Comput. Biol.* **2016**, *12*, e1004961, doi:10.1371/journal.pcbi.1004961.

7. Myung, J. I.; Tang, Y.; Pitt, M. A. Evaluation and comparison of computational models. *Methods Enzymol.* **2009**, *454*, 287–304, doi:10.1016/S0076-6879(08)03811-1.

8. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection *. *Stat. Surv.*

2010, *4*, 40–79, doi:10.1214/09-SS054.

9. Olofsson, P.; Foody, G. M.; Herold, M.; Stehman, S. V.; Woodcock, C. E.; Wulder, M. A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57, doi:10.1016/J.RSE.2014.02.015.

10. Srivastava, P. K.; Han, D.; Rico-Ramirez, M. A.; Bray, M.; Islam, T. Selection of classification techniques for land use/land cover change investigation. *Adv. Sp. Res.* **2012**, *50*, 1250–1265, doi:10.1016/J.ASR.2012.06.032.

11. FAO. 2016. *Handbook on Remote Sensing for Agricultural Statistics*; Available online: http://gsars.org/wp-content/uploads/2017/09/GS-REMOTE-SENSING-HANDBOOK-FINAL-04.pdf (accessed on 10 April 2018).