SATELLITE IMAGE AUTHORIZATION WITH STATISTICAL METHODS

Vootla Sridhar
Research Scholar
Shri JJT University
Rajasthan

Dr. Anupama A Deshpande
Associate Professor
Shri JJT University
Rajasthan

Dr. T. SRINIVAS
Professor
ECE Department
Principal
Mother Theressa
College Of Engineering
And Technology.

ABSTRACT

This study presents a new method of statistical differencing for the enhancement of contrast images. Our approach controls the sharpening effect using two constants in such a way that enhancement occurs in intensity edges areas and very little in uniform areas. It has been proven that our method is superior to similar existent and can be applied to pre-process satellite images. Image processing on satellite imagery is an important challenge. In this area, the identification of objects in a satellite image is a pivotal task. Machine learning methods can be employed for this purpose as there has been significant amount of research in the area of image processing through machine learning.

Keywords: satellite, statistical differencing, Image processing.

Introduction

Environmental protection has been becoming more significant in our lives it is becoming increasingly important to understand the necessity to monitor many environmental aspects. But it's not possible to observe and analyze the some of the most fascinating aspects of environment just relying on experts from human perspective. Experts cannot be found in all areas of interest simultaneously. Furthermore, some subjects are not accessible to experts by human at reasonable prices. This makes remote sensing a necessity in large-scale environmental monitoring. Remote sensing focuses on the methods and information that are used to analyze, understand, and monitor environmental changes by employing microwave and optical images generated by various sensors. The control of natural resources as well as growing challenges associated with environmental monitoring makes remote sensing one of most important technologies for strategic planning in coming century. For instance, airborne and satellite remote sensing has already played crucial roles in predicting and monitoring weather conditions. It is now recognized that climate models that rely on trace gas forcing by them cannot accurately reflect changes in climate. To deal with climate change, variables that are highly variable in time and space should be included in models. These include cloud, aerosols, ocean productivity land usage, and their interactions. The most effective tools to determine these parameters are sophisticated satellites that can remotely sense. The most efficient computing and sensing tools are available in increasing numbers and there are many high-resolution satellites with greater coverage of many areas. This means that remote sensors will play an even more significant role in the management of resources and environment in the next 10 years.

The development of Statistical Machine Learning (SML) methods and computational algorithms to analyse remote sensing data has been expanding for over half a century. An early example is the Laboratory for Applications of Remote Sensing (LARS), which was established in 1960. This centre produced crop identification based on the Apollo satellite spectral data and automated analysis as early as 1969, and machine implemented multispectral analysis in 1970. Since then, machine learning techniques applicable to remote sensing data have continued to develop and, as more data from quality sensors are becoming freely available, additional applications of these data are being explored. A current key focus of remote sensing data analysis in the statistical research community is deriving environmental and agricultural statistics. Land use, land cover change, crop identification, deforestation, and water quality are some examples of statistics that are currently being derived from remote sensing data analysis. The use of remote sensing data for deriving these types of statistics and metrics is also topical internationally, as it conforms to the United Nations 2030 Agenda for Sustainable Development. Use of remote sensing data for monitoring and supporting implementation of the Sustainable Development Goals, targets, and indicators is being explored and encouraged by the United Nations through global working groups, and National Statistical Organisations, such as Statistics Canada, are also producing these types of analyses. There is also growing interest internationally about using remote sensing data and other big data sources to examine the relationship between environmental and socio-economic indicators, such as poverty, and between environmental and economic indicators through ecosystem accounting, with the latter using the United Nations System of Environmental-Economic Accounting framework as the global standard.

Furthermore, the process of automatically interpreting sensor images requires large amounts of image information and different types and types of manipulation, based on the type of data that is being extracted. This presents a variety of difficult problems that need to be solved, for instance, creating efficient multisource data fusion strategies to gain additional information; constructing algorithms that are able to automatically decide on the kind of images should be combined, or the types of operations suitable for extracting particular data; and constructing algorithms that combine the spectral and contextual information in analysis, just as humans use to interpret.

This interest in measuring natural resources and official statistics from freely available big data sources has led to an interface between the statistical and remote sensing disciplines, and a growing number of useful resources and frameworks are being produced as a result. Examples include the Committee on Earth Observation Satellites (CEOS) 2018 handbook; Satellite earth observations in support of the sustainable development goals; and the United Nations Satellite Imagery Task Team 2017 handbook on Earth Observations for Official Statistics. Other useful guides for practitioners tailored to specific applications include the World Bank's book; Earth Observation for Water Resources Management; and the United Nations Food and Agriculture Organisation's handbook on remote sensing for agricultural statistics. There is a growing body of literature from global institutes and in the research community from this multidisciplinary perspective, as we will describe further in this paper. Given the established and increasing interest in these types of analyses and the wide range of tools and techniques available, it is timely to review the SML methods that have emerged as the most popular and appropriate approaches for remote sensing applications. Noting that there is usually no one true or correct approach, it is also of interest to consider approaches to selecting an appropriate method for a given problem. In this paper, we address these needs by providing the remote sensing and statistical practitioner with an overview of methods for analysing remote sensing data, guidance on how to select methods for particular problems, and how to evaluate the results of these analyses. We also provide references and case studies for further reading.

Statistical Machine Learning Methods

Statistical machine learning methods, also referred to as empirical methods, can be defined as cases where a statistical relationship is established between the spectral bands or frequencies used and the variable measured (field-based) without there necessarily being a causal relationship. This relationship can be parametric, semi-parametric, or nonparametric.

The main advantages of SML approaches are that they provide a mathematically rigorous way of describing sampling and model error, estimating and predicting outcomes of interest and relationships between variables, and quantifying the uncertainty associated with these estimates and predictions. They can also be used to test hypotheses under specified assumptions, and some models, such as decision trees, have few assumptions. Disadvantages are the required ground truth data to train the model or verify model results (unlike the physics-based models). This reliance on ground truth data also means they may be difficult to extrapolate or transfer to other contexts.

There are many statistical machine learning algorithms that perform different tasks. Some of the algorithms that are relevant to remote sensing data as applied to Sustainable Development Goal targets are grouped according to four main analytic aims: Classification, clustering, regression, and dimension reduction (Figure 1). An overview of these methods and their applications, including references for further reading, is provided in Table.

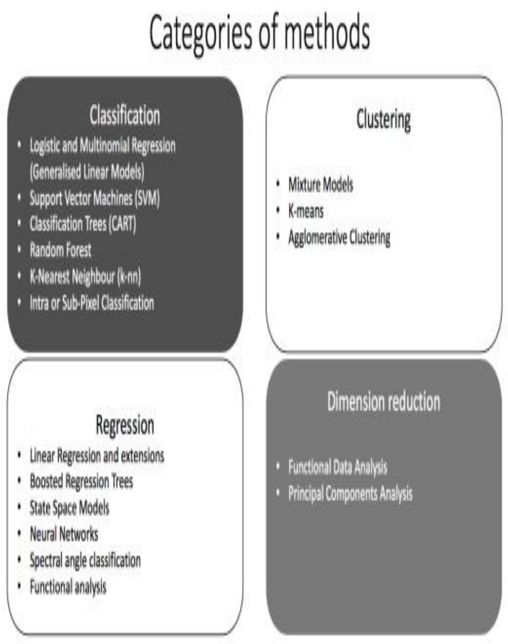


Figure 1. Statistical machine learning methods for analysis of remote sensing and other data, collated according to the analytic aim.

Categories of Statistical Machine Learning Methods

As indicated in Figure 1, four of the most common aims of remote sensing data analyses are classification, clustering, regression, and dimension reduction. The four categories of methods, as applied to remote sensing data, can be described as follows.

Classification: A classification method is applicable if the overall aim is to accurately allocate objects to a discrete (usually small) set of known classes or groups. This allocation is based on a set of input variables. In the literature, these are also called explanatory variables, factors, predictor variables, independent variables, covariates, or attributes.

Clustering: A clustering method is applicable if the aim is to combine objects into groups or classes based on a set of input variables. Clustering is an unsupervised learning method, which does not require a training data set. Unlike classification, we do not know the output variable or classes.

Regression: A regression method is applicable if the aim is estimation or prediction of a response variable based on a set of covariates. This is similar to classification methods, but the response is continuous instead of categorical. Like classification methods, the regression model is developed or trained based on a set of input variables for which the response is known.

Dimension Reduction: A dimension reduction method is applicable if there are many variables that can be extracted from remote sensing data (and other data sources), and the aim is to construct a small set of new variables that contain all (or most) of the information contained in the original (large) set of input variables. These new variables can be used as inputs into other analyses or they can be end products in their own right. For example, they may be inspected to gain a better understanding of important variables or interpreted as 'features' or 'indices' (e.g., two satellite reflectance variables are combined to give a single vegetation index variable, VI). This is an unsupervised learning process since there is no response variable to estimate.

Informed Statistical Machine Learning Methods

Informed statistical machine learning methods, also referred to as semi-empirical methods, combine knowledge about the process with SML or empirical models. This knowledge can be about the process itself or about the variables that are included in the process. These methods have been used for remote sensing data analysis for over a decade. For example, used Landsat 7 Enhanced Thematic Mapper image data to map selected water quality and substrate cover type parameters. The semi-empirical approach to water quality detection, in which knowledge about the spectral characteristics of some of the parameters is used to refine the statistical model. In their example, only a subset of variables is used in the model, with well-chosen spectral areas and appropriate wavebands or combinations of wavebands. The authors highlight the popularity and utility of these approaches, but also caution they still require ground truth data. While semi-empirical methods are arguably more transferable than purely empirical models (since they contain information about the process), they can still be limited by the generality of the data used to build them. The semi-empirical method, which incorporates physiological measures, spectral measures, and spatial features for estimating wheat yield. The authors note that while spectral (empirical) and physics-based (mechanistic) models based on vegetation indices are widely used, they are respectively limited by being data-intensive and complex. Their semi-empirical approach is proposed as an intermediate method.

Physics Based Method

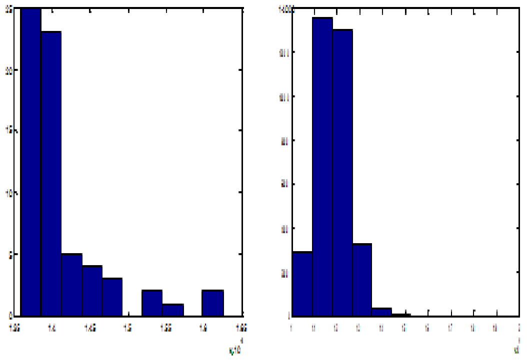
Physics based models are based on detailed knowledge of the system that is being modelled. They can be built without data, or data can be used to calibrate the model parameters. This method is suitable for automation across large areas provided that the model is appropriately and accurately parameterized.

Methodology

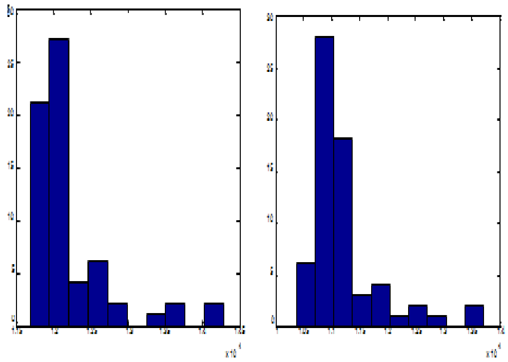
Image Dereferencing for the image is carried out by associating image with its latitude and longitude details provided with header file of the image concerned. Preprocessing of the image is an essential component to improve the quality of input image. Image filtration is applied on the input image to reduce and remove noise.

Classification technique can be applied to input data and output are recorded to choose an optimal one. The classification results employing Decision tree method are recorded. Accuracy assessment is carried out using Root mean square of error.

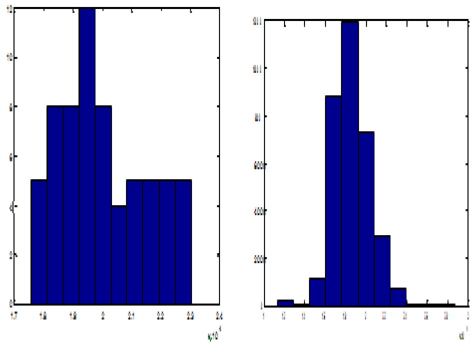
RESULTS



a. Band 1 2014 b. Band 2 2014



c. Band 3 d. Band 4



e. Band 5 f. Band 6

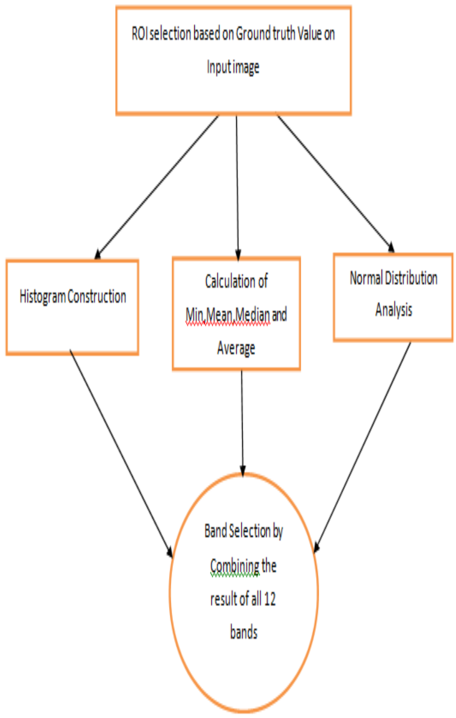


Figure: Workflow Diagram for Band Selection

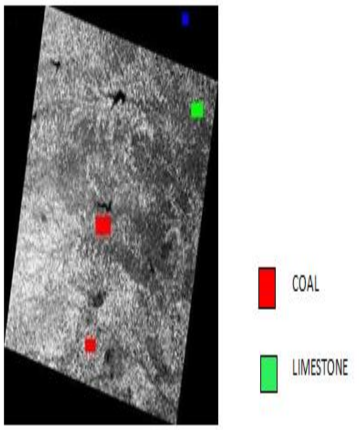


Figure: Sample ROI Selection for 2013 data

ROI chosen image is converted to ASCII values which is used for analyzing the data in histogram. ROI selection in the input image of 2013 data is shown in Figure. Bokaro Coalfield covers an area of 259 square kilometers and has total coal reserves of 4,246.30 million tones. Ranchi presently has 62 operative mines. The geological formation of the Palamau district comprises mainly of Calc-silicate rocks and crystalline limestone's. Most of the limestone dealers available in Hazaribagh as it produces major amount of limestone in Jharkhand state. These regions latitude and longitude are taken for training ROI's and distribution of minerals with histogram is analyzed.

Conclusions

This study has shown there is clearly an interface between the earth science and statistical domains, as remote sensing data continues to become more freely available and interest in deriving key environmental, social, and agricultural metrics continues to grow at the researcher, institute, and country level. We have described four key categories of statistical machine learning methods for analysing remote sensing data; namely, regression, classification, clustering, and dimension reduction. Following a discussion about the type of estimates that can be obtained from remote sensing data, the focus turned to two of the three broad steps in analysing remote sensing data, techniques for analysing the data, and the critical evaluation of the analysis results. A range of references has been provided to demonstrate the relevance of the methods described and their practical application to these data.

References

1. *Agterberg, FP, Bonham-Carter, GF, Cheng, Q, Wright, DF, Davis, JC & Herzfeld, HC, P 1993, 'Weights of evidence modeling and weighted logistic regression in mineral potential mapping', Computers in Geology, New York, Oxford University Press, pp. 13-32.*
2. *Akhavi, MS, Webster, TL & Raymond, DAP 2001, 'RADARSAT—1 imagery and GIS modeling for mineral exploration in Nova Scotia', Canada, Geocarto International, vol. 16, no. 2, pp. 55-61.*
3. *Akhil Kumar Das, Debasis Mandal, DW 2015, A New Statistical approach for Image Fusion Technique, International Journal of Computer Science and Engineering Technology, vol. 6, no. 1, pp. 5-9.*
4. *Albertson, JD & Parlange, MB, DW 1999, 'Natural integration of scalar fluxes from complex terrain," Advances in Water Resources Journal, vol. 23, no. 3, pp. 239-252.*
5. *Amin Beiranvand Pour & Mazlan Hashim, DW 2015, 'Hydrothermal altertaion mapping from Landsat-8 Data, Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran', Journal of TaibahUniversity for Science 9, vol. 9, no. 2, pp. 155-166.*
6. *Ammar Al Muhairi, Hosni Ghedira, Hussain Al-Ahmad & Ali Dawood, D 2011,'Using Remote Sensing Satellites for Water Quality monitoring in the UAE', IEEE GCC Conference and Exhibition (GCC), pp. 67-68.*
7. *Andrew James Elmore & John Fraser Mustard, DW 2003, 'Precision and Accuracy of EO-1 Advanced Land Imager (ALI) Data for Semiarid Vegetation Studies', IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 6, pp. 1311-1320.*
8. *Basabi Chakraborty & Atsushi Kawamura, DW 2018, 'A new penalty- based wrapper fitness function for feature subset selection with evolutionary algorithms', Journal of Information and Tele Communication, vol. 2, no. 2, pp. 163-180.*
9. *Bastin, LS 1997,'Comparison of fuzzy c-means classification, linear mixture modeling and MLC probabilities as tools for unmixing coarse pixels', International Journal of Remote Sensing, vol. 18, no.17, pp. 3629-3648.*
10. *Bernard, E, Hubbard, James, K, Crowley & David, R, Zimbelman, P 2003, 'Comparative Alteration Mineral Mapping using Visible to Shortwave Infrared (0.4–2.4_m) Hyperion, ALI, and ASTER Imagery', IEEE Transactions on Geoscience and Remote Sensing, vol. 41, no. 6, pp. 1401-1410.*