Journal of Terrestrial Observation

Volume 2, Issue 2 | *Spring 2010* | *Article 3*

Lessons Learned in the Collection of Surface Reference Information

Chris J. Johannsen

 ISSN 1946-1143.

Foreword

Lessons Learned from Remote Sensing Activities in Recent Natural Disasters

When I first started with the Laboratory for Applications of Remote Sensing (LARS) in 1966, I was assigned the responsibility of collecting "ground truth" on the agricultural fields that were being flown by the University of Michigan's 12-channel scanner. I was the only agronomist at LARS among a small group of engineers. Since I grew up on a farm, this was a fun and easy task, that is, until they started asking questions. The original task was to identify the crops, such as corn, soybeans, wheat, oats, and alfalfa. However, when two cornfields looked different on the scanner data, my credibility was questioned. It turned out that one of the fields had been recently cultivated and that caused the field to look darker because of the moist soil surface compared to the dry soil surface. I therefore added the category of cultivated corn to the corn category. Next came "weedy corn" and soon I had to add row direction since "north-south rows of corn" appeared different from "east-west rows of corn" because of sun angle and differences in reflection.

Within a few years, LARS was flying more flightlines and my tasks became larger, to the point where I had to train engineers to assist in collecting "ground truth." Engineers and even other disciplines can become quite visual in what they see and I started to get categories of "yellowish green" and "greenish yellow." I realized that we didn't have a Munsell Color Chart for crops like we have for describing soil colors. Also a person's background makes a difference on what he or she is observing and describing.

Through the years the accuracy of "ground truth" has meant many things, depending on the method used to collect the data, the experience of the person collecting the data, the purpose for the collection of the remote sensing data, and the time interval between when the remote sensing data were collected and when the ground truth data were collected. As McCoy (2005) explained, ground truth has been replaced by "reference information" to be more inclusive than "ground" and less absolute than "truth." Congalton and Green (2008) stated that the term "ground truth" was inappropriate and in some cases, misleading. Therefore these authors used the term "reference data" throughout their book. In a recently published book chapter, Craig Daughtry and I decided that the term "surface reference" would be a better term, because it is intended to mean any data or information collected to support the analysis and interpretation of remote sensing data obtained for studying air, land, and water resources (Johannsen and Daughtry, 2009). We sought to include the meteorology, hydrology, and oceanography research community in our descriptive term since "ground reference," which was already being used by some researchers, wouldn't include that community.

Surface reference data and information can serve three purposes: (1) to guide the analysis process by providing training sites for supervised classification, which is what I was doing when I started collecting information; (2) to assess and evaluate the accuracy of the results of the remote sensing analysis, which prompted me to explain why two crops were identified as the same crop (e.g., wheat and oats at an early stage of growth); and (3) to characterize and model the spectral behavior of radiation within the scene. The latter item worried me in the beginning, as I thought that the engineers were trying to work me out of a job. However, I begin to appreciate the importance of each perspective and I matured in my efforts.

Surface reference data includes attributes or measurements that describe surface conditions at a specific location and time. It is only as good as the experience of the individual collecting the data. One should always check on how it was collected, when it was collected, what instruments were used, what the calibration sources were, and similar factors before using someone else's reference data. Usually, reference data are collected with a specific purpose in mind and this may be different from your purpose. For example, if you want to follow the reflective changes of corn through the growing season and the person collecting the data may have collected the data from a different variety of corn than you are using, the results could be quite different. I learned the hard way about this issue when the person planting my Ph.D. research plots ran out of the corn variety he was planting and substituted another variety halfway through the plot. You could see to the row where the substitution was made as the second variety had wider leaves and therefore a different reflection measurement.

Surface reference data may be broadly classified as discrete or continuous. Discrete data are qualitative descriptions or nominal designations that convey basic land-use and land-cover differences among regions in the imagery; examples include water, urban, forest, rangeland, cropland, and wetland. Accurate and precise nominal labels concisely communicate significant information such as water temperature, types of residential areas, type and age of trees, types of grass and associated vegetation, specific crop types and types of wetlands. Continuous data or quantitative measurements of the physical and biological characteristics of surface features complement nominal labels with more specific data to document the precise meaning of the labels. Biophysical data may include slope, soil organic matter content, vegetation biomass, leaf area index, leaf chlorophyll content, and leaf angles.

All field observations and measurements that are used as reference must include a means of determining reliable locations for each sample site. Before GPS, land-surveying techniques using maps, compasses, and land features were used to locate sample sites. Today, with a relatively inexpensive GPS receiver, a researcher can determine coordinates of sample sites in the field or navigate to a point on the ground with coordinates derived from a map or geo-referenced image. Although all GPS coordinates contain errors, a researcher can often mitigate common GPS errors resulting from multi-path reflections, electrical interference, poor satellite geometry, and obstructions.

In addition to the collection of in situ data and recorded observations, there are many sources of reference information, such as soil maps, meteorology data, historical crop types, combine yield monitor data, digital terrain data, and sea surface temperature data. Each of these are references that can guide the analyst in making decisions with respect to selection of training samples and helping to interpret the results. However, all of these types of reference sources must be treated as general reference data and information, as they will vary in the scale, accuracy, and the purpose for which they were collected.

One of the more common types of surface reference information that is quoted in the literature is Normalized Difference Vegetation Index (NDVI). NDVI is misused and should never be called "ground truth." Researchers have noted that NDVI is not a perfect descriptor of the vegetation type or amount since they have sought to modify the VI as shown by Jensen (2007), where he describes examples such as Infrared Index (II), Perpendicular Vegetation Index (PVI), Greenness Above Bare Soil (GRABS), Soil Adjusted Vegetation Index (SAVI), and many others. The latter even has a Modified SAVI called MSAVI to show that not all VIs work satisfactory for different geographic regions. When researchers use any type of VI as reference information, they should describe the types of vegetation that are most common to their study area. This gives the reader a chance to decide if the Index provides the correction background information.

The major point of discussing this topic is that one should take the same care in the selection of surface reference information as one does in the collection of the data and the analysis techniques. It is important that one visits the location during the time of data collection; if at all possible, as it gives one a better view and understanding of the surface cover types and conditions. For temporal analysis, a visit to the location at the time of year that each dataset was collected would assist greatly in the choice of analysis techniques and interpretation of the results. After 50 years of collecting surface reference data and information, I am still learning how to improve the collection, assembly, and presentation of this important analysis tool.

Chris J. Johannsen
Professor Emeritus, Department of Agronomy
Purdue University

REFERENCES

Congalton, R.G., and K. Green. 2008. *Assessing the accuracy of remotely sensed data: Principles and practices.* CRC Press-Taylor & Francis Group.

Jensen, J.R. 2007. *Remote sensing of the environment: An earth resource perspective*. 2nd ed. Upper Saddle River, New Jersey, Prentice Hall.

Johannsen, C.J., and C.S.T. Daughtry. 2009. Surface reference data collection. Chapter 17 *in* Warner, T.A., M.D. Nellis, and G. Foody, eds. *The Sage handbook of remote sensing.* London, England, Sage Publications.

McCoy, R.M. 2005. *Field methods in remote sensing.* New York, Guilford Press.