

# 2010 CEOS FIELD REFLECTANCE INTERCOMPARISONS LESSONS LEARNED

*Kurtis Thome<sup>1</sup>, Nigel Fox<sup>2</sup>*

<sup>1</sup>NASA, Goddard Space Flight Center; <sup>2</sup>National Physical Laboratory

## ABSTRACT

This paper summarizes lessons learned from the 2009 and 2010 joint field campaigns to Tuz Golu, Turkey. Emphasis is placed on the 2010 campaign related to understanding the equipment and measurement protocols, processing schemes, and traceability to SI quantities. Participants in both 2009 and 2010 used an array of measurement approaches to determine surface reflectance. One lesson learned is that even with all of the differences in collection between groups, the differences in reflectance are currently dominated by instrumental artifacts including knowledge of the white reference. Processing methodology plays a limited role once the bi-directional reflectance of the white reference is used rather than a hemispheric-directional value. The lack of a basic set of measurement protocols, or best practices, limits a group's ability to ensure SI traceability and the development of proper error budgets. Finally, rigorous attention to sampling methodology and its impact on instrument behavior is needed. The results of the 2009 and 2010 joint campaigns clearly demonstrate both the need and utility of such campaigns and such comparisons must continue in the future to ensure a coherent set of data that can span multiple sensor types and multiple decades.

**Index Terms**— Vicarious calibration, SI-traceability, radiometric calibration

## 1. INTRODUCTION

This paper summarizes the lessons learned during the 2009 and 2010 joint field campaigns to Tuz Golu, Turkey for the comparison of techniques and instrumentation used for the vicarious calibration of optical imagers. The campaign marked an important step for the Committee of Earth Observations Systems (CEOS) to inform the community about the sources of uncertainties associated with the reflectance-based method [1] and start the process of attributing a Quality Indicator to satellite data and further to the end products derived from satellite imagery as recommended by the new international Quality Assurance framework for Earth Observation (QA4EO) [2].

The objectives of 2010 CEOS Key comparison were:

- 1) Determine biases between field instrumentation using a series of laboratory and in situ cross-comparisons of participants' radiometers and reference panels
- 2) Estimate a range of values for reflectance uncertainties associated with the reflectance-based method for vicarious calibration of optical sensors
- 3) Evaluate differences in sampling methods used to associate a "reflectance value" or a "radiometric value" for both a moderately-sized (0.03 km<sup>2</sup>) and large-sized area (1.0 km<sup>2</sup>)
- 4) Document "best practices" used by the participants in the 2010 CEOS Key comparison and estimate the uncertainties associated with each of them

Representatives from 10 countries and 13 organizations took part in either one or both years of the campaigns. Measurements were made in the visible and near-infrared (VNIR), and shortwave infrared (SWIR) and simulated the calibration of sensors with spatial resolution varying from tens of meters to a kilometer. The data sets collected also permit evaluation of the repeatability and accuracy of vicarious calibration that are a critical part of the calibration of earth imagers sufficient to create a set of long-term, absolutely-calibrated set of observations for the study of global change [3]. Such efforts require multiple sensors on multiple platforms with each sensor typically having its own calibration team. Most of the sensor teams include some form of vicarious calibration in their calibration plans. Thus, it is essential to ensure that different groups performing these calibrations obtain consistent results to prevent biases between different sensors while at the same time producing accurate results with SI-traceability.

The vicarious calibration of land imagers using the reflectance-based method [1] requires the measurement of the reflectance factor ( $\rho$ ) of the site surface at the time of the satellite overpass. The terminology for reflectance quantities follows the definitions given in optical remote sensing [4]. The  $\rho$  of the test site surface is calculated from radiance measurements performed with a portable spectroradiometer typically made in comparative mode against a reference panel of known reflectance.

In the interest of brevity, specific details on the test site, its location, instrumentation, and methodologies are omitted. The aim of this paper is to describe the key lessons learned and make recommendations for best practices for the

retrieval of surface reflectance. Recommendations are highlighted throughout the text. The discussion of the lessons learned concentrates on those from the 2010 campaign because of the larger number of participants and days of clear-sky data. The lessons learned are specifically related to understanding the equipment and measurement protocols, processing schemes, and traceability to SI quantities. A discussion of how the 2009 campaign influenced the work during the 2010 campaign is also included.

## 2. 2009 CAMPAIGN

The 2009 campaign to Tuz Golu was a smaller campaign than the 2010 campaign but still included groups from six different countries and seven organizations and was intended as a pilot for the main 2010 comparison. The campaign included:

- 1) Comparison of field radiometers in the laboratory
- 2) Comparison of field radiometers in the field against a common reference panel
- 3) Characterization of participants' reference panels in the laboratory
- 4) Site surface characterization
- 5) Site BRDF characterization
- 6) Atmospheric characterization.

The results from the 2009 effort gave insight into the accuracy and precision of in situ measurements for vicarious calibration. The major lesson learned from the 2009 campaign related to the 2010 campaign was that the timeframe of 4 days was too short in 2009 for all activities when imperfect weather is taken into account. The 2010 campaign was scheduled for two weeks of which nine days were in the field and the other days in the laboratory before and after the campaign. The other lesson learned that influenced how the 2010 campaign was conducted was that the individual groups' data formats and content delayed the data analyses and templates were provided before the comparison in 2010.

## 3. MEASUREMENT PROTOCOLS

The participants in 2009 and 2010 used multiple measurement approaches to determine surface reflectance. All groups, however, used the same basic methodology of comparing surface-leaving radiance measurements to measurements from a white reference to obtain reflectance. Major differences in how the white reference is used included the size of the reference, the laboratory calibration of the reference to spectral reflectance, use of hemispheric-directional reference calibration versus bi-directional, and placement of the reference relative to instrumentation.

The surface-leaving radiance is spatially sampled and individual samples are averaged in all cases to provide a surface reflectance representing a large area (either 1 km<sup>2</sup> or

.03 km<sup>2</sup> in the 2010 campaign). The sampling approaches from the groups included continuous sampling as the user walked the site with periodic white reference data based on specified time intervals or after specified distances, a stop and stare approach with white reference data at each stopping point, or a stop and stare approach with white reference collections at specified time intervals. The methodology selected by each group is based on the number of personnel available and length of time to collect. The fastest method is continuous sampling with white reference at specified distance intervals while the longest is the stop and stare with white reference at each point. The methods for carrying the equipment and panel also varied between groups and this plays a role in possible interference between the user and the retrieved surface reflectance.

One lesson learned is that even with all of the differences in collection between groups, the differences in reflectance are currently dominated by instrumental artifacts including knowledge of the white reference calibration, length of time between white reference data, and instrument maintenance. Instrument performance can vary due to factors such as ambient temperature, length of instrument operation, and instrument warm up time. Methods to understand the field radiometer's performance at the time of reflectance collection are vital to improving the accuracy of the retrievals.

**Recommendation:** *Use of an invariant standard before and after site characterizations is needed to evaluate instrument performance.*

The goal of the measurement protocols is to improve the methods to a point where differences begin to be dominated by sampling strategy and atmospheric influences (hemispheric versus bi-directional reflectance).

**Recommendation:** *A standardised radiometer should be developed that can act as transfer standard to link test-sites traceability.*

Such a radiometer would have limited bands, field of view, or portability that would limit its use in characterizing the test site. The radiometer would, however, provide a means to ensure the calibration of white reference panels and field radiometer behavior across multiple groups. Ideally, each group would have its own individual radiometer, but the advantage to a detector-based approach is that the radiometer can be used as a travelling standard, allowing a few groups to shoulder the costs of developing and operating the radiometer.

## 4. PROCESSING METHODOLOGY

Processing differences between the groups include accounting for sun angle changes during the collections and how the white reference calibration is implemented within the reflectance retrieval. The processing methodology currently plays a limited role in the surface reflectance differences between the groups. Processing methodology

plays a much larger role when the surface reflectance is included with atmospheric characterization to provide a prediction of the radiance at the top of the atmosphere.

The dominant lesson learned related to processing methodology is that a clear bias is created when groups rely on a hemispheric-directional calibration for the white reference. Recommendations related to this topic are described in the next section as part of the discussion related to error budgets. A further lesson learned from both the 2009 and 2010 campaigns is the difficulty comparing results from the separate groups caused by differences in data formats.

**Recommendation:** *A standardised format should be established for reflectance-based calibration measurements to enable data from such site characterisations to be easily compared.*

An important part of the standardised format is the inclusion of appropriate documentation of errors and uncertainties needed to determine the Type A and Type B uncertainties.

## 5. TRACEABILITY AND ERROR BUDGETS

As described, the basic equipment used by each group included a field spectroradiometer and a white reference. Absolute calibration of the field spectroradiometers were supplied by the instrument manufacturer for all but one participant. The manufacturer-supplied calibrations are based on knowledge of the output of a spherical lamp illuminated integrating source traced to a third-party source of spectral irradiance. The field instrument's radiometric calibration is periodically updated by the manufacturer when requested by the individual group. The lone group not relying on a manufacturer-supplied calibration had its field instrument calibrated in a National Metrology Institute (NMI) directly to SI standards.

White reference calibration followed a similar approach. Most groups relied on the calibration supplied by the manufacturer, one group characterized their own reference in their own laboratory, and one group relied on a third party to characterize their reference. All manufacturer-based calibrations were in terms of a hemispheric-directional characterization while the other two groups obtained bi-directional characterizations of their references.

The primary lesson learned from this information is that none of the groups had performed an error budget or traceability study sufficient to assess the precision and accuracy of their results. The reasons for this ranged from the limited experience of the group to reliance on error budgets previously published by other groups. The fact that scientists operating in the field have backgrounds far different from metrologists operating in the laboratory means that an education process is required to ensure that a full understanding of traceability, error budgets, and estimation of sources of uncertainties is needed.

Such an understanding is not crucial at the 5-10% absolute uncertainty in vicarious calibration that most groups are currently achieving. Improvement to the state of the art level of 2-3% absolute uncertainty demonstrated by a select number of groups requires a well developed error budget. Achieving vicarious calibration uncertainties <2% will require dramatic shifts in the philosophy of the field groups towards a more rigorous collection approach that is similar in philosophy to laboratory practices.

For example, the radiometer performance or so-called "noise" can be calculated as a Type A standard uncertainty and includes the repeatability and reproducibility of radiance measurements against a known source (Type A is the uncertainty resulting from the statistical analysis of the data). This value was calculated using three independent runs, where each run had ten measurements. The Type A values were reported by the participants and a value was also calculated by NPL using the Guide to the Expression of Uncertainty in Measurement (GUM). As expected, most instruments show a similar Type A standard uncertainty as calculated by NPL but these values did not necessarily agree with those computed by the individual groups.

For each site, the laboratories provided the associated Type A evaluated and Type B evaluated standard uncertainties for the estimated reflectance factor for the site and at a set of prescribed wavelengths. The Type A evaluated standard uncertainty, which is calculated from a statistical analysis of measured values obtained by repeating the measurement at the same location and at different locations across the site, describes effects such as measurement repeatability and the variability of the site. No group collected data that would be traditionally viewed as a suitable data set for evaluating repeatability for calculating Type A errors.

**Recommendation:** *Perform "repeatability measurement" before and during site characterisation based on a ratio of repeated panel views to repeated views of a single surface location*

Specifics of such a measurement are still under discussion to determine the optimal number of data points needed to provide sufficient statistical significance.

**Recommendation:** *Individual site "point measurements" should consist of a statistically significant number.*

This recommendation aims to ensure that a sufficient level of sampling occurs to give adequate understanding of the Type A uncertainties caused by the uncertainty of the field radiometers.

The Type B evaluated standard uncertainty, which is quantified by means other than a statistical analysis of data, describes effects including the calibration of the laboratories' reference panels. An important lesson learned during the 2010 campaign is that the largest Type B error is attributed to the use of the hemispheric-directional reflectance for the reference characterization. A bi-directional characterization is likewise not exact but creates

far lower Type B errors at longer wavelengths and offers the opportunity for a correction of diffuse-light effects at shorter wavelengths.

**Recommendation:** *Assignment of reflectance factor to a white reference panel and subsequently the test site should be based on a bi-directional (Gonio) characterisation at appropriate angle(s)*

In some cases such as solar zenith angles near overhead, a hemispherical-based calibration may be adequate. The bi-directional characterization permits the assessment of cases when a more simplistic processing scheme is feasible. Characterizing bi-directional reflectance is not a trivial approach requiring goniometric facilities that may not be readily available to all laboratories. Evaluations of the white references during this work implied that most references behave in a similar fashion.

**Recommendation:** *A look-up table of panel BRF for a range of incident angles will be published on an accessible cal/val portal as a first order correction.*

## 6. CONCLUSIONS

The 2009 and 2010 CEOS-led campaigns to Tuz Golu, Turkey provided a unique opportunity to evaluate the state of vicarious calibration. The large number of terrestrial imagers being operated by multiple countries each with their own reference standards and sometimes independent routes of “traceability” creates challenges for the earth sciences community to develop a coherent data set suitable for climate-level studies. While the Type B errors of each group and the differences between groups are important, the more important point is that few groups have fully determined their Type B errors and no groups collect data sufficient to robustly determine Type A uncertainties.

Such conclusions are not meant to disparage the groups involved in the Tuz Golu effort but more to illustrate the differences between the measurement approaches of laboratory-based metrology versus that of field-based measurements. A more rigorous approach to the field measurements is clearly needed to develop multi-sensor, multi-national data sets. Thus, the most important recommendation from this work is that a collaborative effort between NMIs and vicarious calibration laboratories is essential. Related to this recommendation is that the vicarious calibration data sets must include suitable measurements to determine Type A uncertainties and development of SI-traceable error budgets. Such error budgets would have led to a clear assessment that the use of hemispheric-directional calibration for a field reference is inadequate for the reflectance-based method. This specific conclusion is currently being formulated into a specific recommendation by CEOS to those making such measurements in the future to avoid biases which can be of the order of several percent.

The major lessons learned from both the 2009 and 2010 field campaigns are:

- 1) A basic set of measurement protocols, or best practices, should be developed to ensure SI traceability and the development of proper error budgets. Note that the best practices are not intended to provide consistency between groups. Consistency without error budgets and traceability only provides improved precision but not improved accuracy.
- 2) Future comparisons must seek to include a greater diversity of field instrumentation to evaluate whether systematic biases are present due to instrumentation
- 3) A clearer understanding of each group’s error budget is needed especially knowledge of the difference between systematic and random biases/errors.
- 4) More rigorous attention to sampling methodology and its impact on instrument behavior is needed

The results of the 2009 and 2010 joint campaigns clearly demonstrate both the need and utility of such campaigns. A key recommendation from this effort is that such comparisons must continue in the future to ensure a coherent set of data that can span multiple sensor types and multiple decades. It is also essential that the results of such comparisons are also visible to the community so that all lessons can be learnt. The full results of this comparison are available on the GEO/CEOS Cal/Val portal [5]

## 6. REFERENCES

- [1] Slater, P.N., S.F. Biggar, R.G. Holm, R.D. Jackson, Y. Mao, M.S. Moran, J.M. Palmer and B. Yuan, “Reflectance- and Radiance-Based Methods for the In-Flight Absolute Calibration of Multispectral Sensors,” *Remote Sensing Environment*, 22, pp. 11-37, 1987.
- [2] <http://QA4EO.org>
- [3] Slater, P. N., Biggar, S. F., “Vicarious radiometric calibrations of EOS sensors,” *J. of Atmos. and Oceanic Tech.*, 13, 376-382, 1996.
- [4] Schaepman-Strub, G., M. E. Schaepman, T. H. Painter, S. Dangel, and J. V. Martonchik. “Reflectance Quantities in Optical Remote Sensing - Definitions and Case Studies,” *Remote Sensing of Environment*, 103 (1), pp. 27-42, 2006.
- [5] <http://calvalportal.ceos.org>