

# Determination of CERES TOA Fluxes Using Machine-Learning Algorithms

Bijoy Vengasseril Thampi<sup>1</sup>, Takmeng Wong<sup>2</sup>, Constantine Lukashin<sup>2</sup>,  
Norman G Loeb<sup>2</sup>

<sup>1</sup>Science Systems and Applications, Inc., Hampton, VA

<sup>2</sup>NASA Langley Research Center, Hampton, VA

The Clouds and the Earth's Radiant Energy System (CERES) instrument onboard NASA Terra and Aqua satellites provide continuous monitoring of earth's top of the atmosphere (TOA) radiation, which is critical to our understanding of the earth's climate and its variability with time. The CERES mission produces a number of different data products with various level of complexity, ranging from the simple CERES ERBE-like data product to the most advanced CERES EBAF data set (Loeb et al. 2009). The CERES ERBE-like data product use the ERBE algorithm, including ERBE Scene ID (Wielicki and Green 1989), ERBE Angular Directional Models (ADM) and ERBE time and space averaging method (Brook et al. 1986), to produce a data set that is compatible with historical ERBE mission (Barkstrom, 1984). The CERES ERBE-like data product does not use MODIS scene information and is based purely on standalone CERES broadband data. There are many advantages and disadvantages of using the CERES ERBE-like data product. The most obvious advantage is the simplicity and standalone nature of this product, which allow user a quick access to the CERES TOA radiance and flux data. CERES instrument-working group (IWG) is currently using ERBE like products as part of the calibration/validation effort to determine instrument drift artifacts. Another advantage of ERBE-like data is that it provides a consistent long-term backup dataset in case of imager instrument pre-mature failure since the advanced CERES data product can no longer be produced without imager information. The most noticeable disadvantage of CERES ERBE-like data is that it is based on a 30-year old ERBE algorithm. The ERBE TOA fluxes are known to have larger uncertainty than the CERES TOA fluxes due to Scene ID and ADM errors. In order to improve the standalone CERES TOA fluxes, these two deficiencies must be corrected. This abstract describes a new CERES algorithm for

improving the standalone CERES TOA fluxes without the use of coincident MODIS data. This new CERES algorithm is based on a subset of modern artificial intelligence (AI) paradigm called Machine Learning (ML) algorithms.

Development of these new CERES algorithms can be explained in two steps. The first step includes the validation of the ML algorithm called Random Forests (RF), which is used to classify the CERES broadband footprint measurement into clear and cloudy scenes. The second step involves the conversion of CERES TOA clear-sky and all-sky directional radiance to TOA fluxes using a ML algorithm called artificial neural networks (ANN). Random forests or random decision forests are an ensemble learning method for classification and use decision tree classifiers as the base learner (Breiman 2001). In the first part of the study, supervised RF classification of CERES radiances (TOA LW and SW) into clear and cloudy classes is carried out. For this purposes we have used CERES Single Scanner Footprint (SSF) data set (day and night time) for all the 12 months from 2003-2014. Input variables are split in to two groups; CERES variables and ancillary variables. CERES variables used in the analysis are solar zenith angle (SZA), viewing zenith angle (VZA), relative azimuth angle (RAZ), CERES TOA LW and SW broadband radiances and surface type information. Ancillary variables used in the analysis are LW surface emissivity, broadband surface albedo, surface skin temperature, surface wind speed and atmospheric precipitable water. Our primary goal was to train and test the efficiency of RF algorithm in classifying the CERES radiances in to clear and cloudy classes. For this purpose, a training and test dataset is developed using the multi-year SSF data. A typical monthly CERES SSF dataset contains millions of CERES footprints spread all over the globe and are very difficult to process simultaneously. In order to create more compact training dataset, the SSF dataset is stratified in the variable of interest (SZA, VZA, RAZ, TOA radiances) and corresponding mean values are used in the analysis. Using the Random forest algorithm and training dataset, a decision tree forest is built and saved. Using this saved forest and RF algorithm, the test dataset is then classified in to clear and cloudy classes. RF classification of CERES data in to clear and cloudy classes show very good results with average classification error being  $< 5\%$  for most of the surface types. This study shows that using Random forest method, it is possible to successfully classify the CERES TOA radiances into clear and cloudy scenes

without using any imager information.

Once the TOA radiances are classified into ‘clear-sky’ and ‘cloud-sky’ scenes, next step in the analysis involve conversion of CERES TOA radiances to flux. For this purpose, a feed-forward error back propagation Artificial Neural Network (ANN) algorithm is used to produce CERES ADMs that can be used for the conversion of CERES TOA radiances to flux. A clear-sky and all-sky ANN based ADMs are developed for TOA SW and LW flux retrieval. Results from the ANN based analysis is tested by comparing TOA fluxes estimated using ANN and that from CERES SSF products. Compared to the all-sky only ANN approach used by Loukachine and Loeb (2003), the new combined clear-sky and all-sky ANN approach allow much better determination of clear-sky flux, all-sky flux and the cloud radiative effect, which is critical in understanding the radiative effect of clouds in our climate system. Based on the ANN analysis using ~1.5 million clear sky SSF test data points (over Ocean), ANN-clear sky approach for SW TOA flux produced lower bias values for ~60-70% of test cases compared to the all-sky ANN method. For TOA LW cases, using ANN-clear sky method produced relatively lower bias values for ~85-90% for test cases.