Big Data (AES 690) Project Proposal
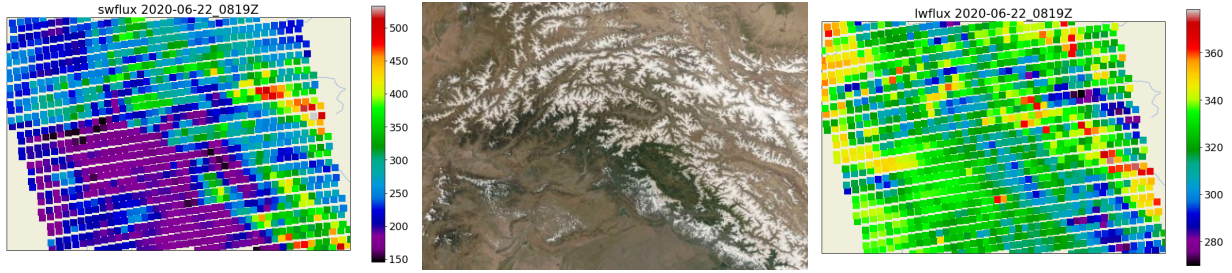
Mitchell Dodson

March 8, 2024

Figure 1: Comparison between the CERES shortwave flux (left), coincident MODIS image (middle), and CERES longwave flux (right). Captured by Aqua over the Hindu Kush Himalayan region on June 22, 2020 at 0819Z. Note the contrast between cold and bright mountaintop snow, warm dark vegetation, and warm bright desert surfaces.

Abstract

Our modern understanding of Earth's energy budget relies heavily on the direct observation of outgoing radiance by polar-orbiting satellites. Space-based platforms like the Earth Observation System (EOS) Terra and Aqua satellites are indispensable for characterizing the governing processes and magnitudes of fluxes leaving Earth's surface and atmosphere since they have frequent global coverage and offer direct measurement of the outgoing energy at the top of a full atmospheric column. Nonetheless, broad-band bolometers like EOS' Clouds and Earth's Radiant Energy System (CERES) are limited in that they generally have coarse spatial resolutions, and the conversion from the observed directional radiance to an estimate of full-sky flux depends on Angular Distribution Models (ADMs), which themselves require simultaneous data from finer-resolution narrow-band radiometers like EOS' Moderate Resolution Imaging Spectroradiometer (MODIS). In order to address both of these shortcomings, I propose a novel machine learning methodology with the potential to (1) emulate the ADMs in order to predict the coarse longwave and shortwave fluxes from CERES given only MODIS pixels and viewing geometry, and (2) sharpen the fluxes to the MODIS resolution.

Background

The utility of satellites for building a holistic picture of Earth's outgoing radiation has been understood since the dawn of the space age. The earliest satellite-derived estimates of Earth's overall albedo (30%) and emission temperature (253 K) were acquired from the TIROS-generation satellites in the mid-1960s, and later improved by the ESSA III platforms into the 1970s [1][2]. One major challenge became quickly apparent: the quantity of interest is the full-sky *flux* ($W\,m^{-2}$) exiting the hemisphere above a surface, but the bolometers can only ever observe directional *radiance* ($W\,m^{-2}\,sr^{-1}$) from a surface into the aperture. Many surfaces are highly anisotropic, so assuming they uniformly radiate the observed amount of energy in every direction is erroneous. The next generation of broad-band radiometers were launched aboard Nimbus-7, NOAA-10, and NOAA-11 in the 1970s as part of the Earth's Radiation Budget Experiment (ERBE) program. These satellite platforms also host Advanced Very High Resolution Radometer (AVHRR) narrow-band sensors which provide enough spectral information to group surfaces into categories with similar anisotropy models [3][4].

The EOS satellites continued with this approach by carrying CERES and MODIS sensors, which have a data record of about 25 years. CERES has a sub-satellite instantaneous field of view (hereafter *footprint*) of about 16×32 km, and MODIS has a 1 km pixel resolution, so many MODIS pixels can be directly co-located within each CERES footprint. During the L2 preprocessing of CERES SSF data, the MODIS data are used to identify land surfaces, mask water, and derive cloud and aerosol properties. Once a category is determined for each pixel in a footprint, the corresponding ADMs are selected and evaluated for the viewing geometry, and their returned anisotropy factors are combined by weighted sum according to the CERES point spread function (PSF). The subsequent aggregated anisotropy factor is the value used to convert the footprint's

radiance observed by CERES to the full-sky flux estimates included in the CERES SSF product [5]. In the past 40 years, considerable attention has been directed towards identifying clusters in surface types' angular dependence [6], and developing associated ADMs that are carefully-validated with respect to ground station data and specialized satellite sensors like EOS' Multi-angle Imaging SpectroRadiometer (MISR) [7] [8].
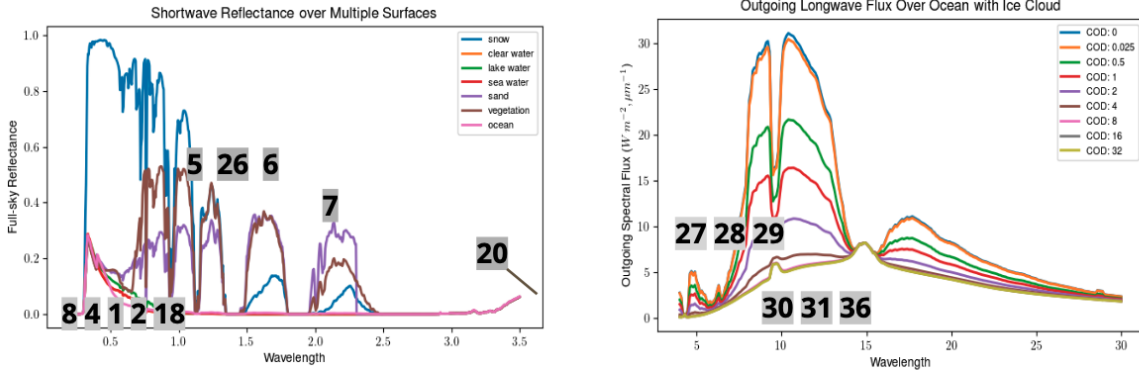

Figure 2: Comparison of top-of-atmosphere outgoing spectral flux of multiple surfaces in the shortwave and longwave ranges, with several MODIS bands' center wavelegnths approximately labeled. CERES effectively observes the area under each spectral curve, and MODIS pixels may be considered point data at the specific wavelengths. Data were generated with SBDART [9].

Hypothesis

The full-sky fluxes inverted from broad-band CERES radiances by modern ADMs have been well-validated, but flux estimates still intermediately rely on sophisticated cloud and aerosol property retrievals and annualized discrete surface classes to pick an anisotropy model. Since the cloud properties and surface classes informing ADMs are derived from MODIS radiance data, the information needed to determine surface characteristics should be available in the pixel's spectral properties.

In fact, since MODIS pixels include narrow bands that are intentionally centered at salient wavelengths throughout most of the solar and terrestrial emission spectra, the spectral response of a collection of MODIS pixels within a CERES footprint may be able to directly predict the well-validated full-sky longwave and shortwave fluxes of the footprint. In particular, reflected solar irradiance in the $.45 - 3.6 \, \mu m$ range is measured by 21 shortwave bands including visible bands that help characterize color and albedo, near infrared bands that distinguish vegetation, surface water, hydrometeor phase, and particle size, and shortwave infrared bands for fires and emissivity. Furthermore, terrestrial emissions in the $3.6 - 14.4 \, \mu m$ range are captured by 16 longwave infrared bands, which help characterize atmospheric O_3 and CO_2 content, surface temperatures, and the distribution of water vapor in the atmosphere.

As such, my primary hypothesis is that a neural network trained on MODIS pixels can learn to predict the full-sky flux values of CERES SSF footprints over multiple surfaces and viewing geometry configurations. If the above model is performant, my secondary hypothesis is that an additional network module can be trained to estimate longwave and shortwave flux at the MODIS 1km resolution.

Methodology

The model starts with an encoder producing a grid of latent vector embeddings by applying the same weights independently to each MODIS pixel. This may be thought of as a dimensionality reduction technique encouraging the model to find a condensed latent representation of information about each pixel's surface and radiative properties. Next, the latent outputs are aggregated by a weighted sum according to the CERES point spread function, which adjusts the contribution of each latent vector in the footprint to the spatial sensitivity of the CERES instrument. The aggregated latent vector is then concatenated with the footprint viewing geometry and passed to a

decoder. The decoder component of the model serves to emulate the ADM, using the distilled and aggregated information on radiative and surface properties from MODIS in combination with the whole scene's viewing geometry (solar zenith, viewing zenith, and relative azimuth) to estimate the CERES-observed shortwave and longwave full-sky fluxes.

If this approach is successful, I will extend the model to predict the flux contribution of each MODIS pixel. With the viewing geometry and the full latent grid as inputs, the extended model will impose the aggregate flux estimates from the previous model as a constraint on the PSF-weighted sum of its predictions in the loss function. It will be interesting to investigate the pixel-wise information encoded in each vector of the latent grid, and to experiment with the consequences of modulating the viewing geometry while holding the vector constant. It's also worth noting that since the encoder applies the same weights independently to each MODIS pixel, and since the latent pixels are aggregated by a simple weighted sum, the trained network will be agnostic toward the number or spatial properties of the input pixels.
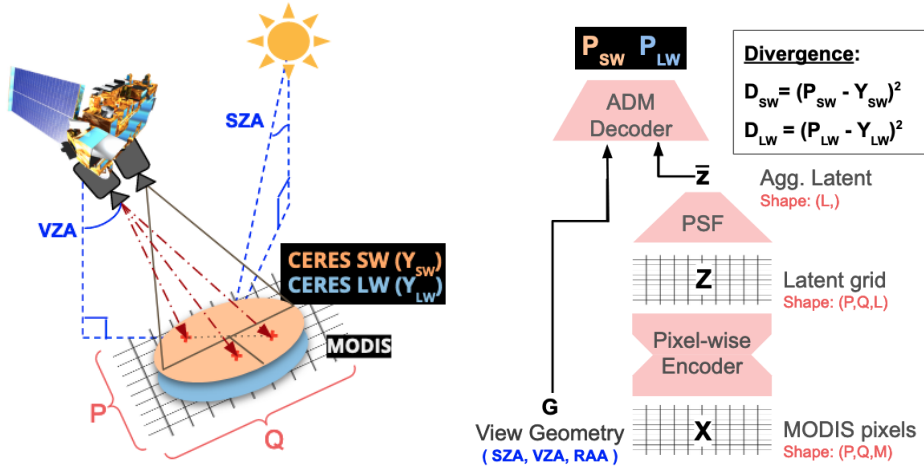

Figure 3: Schematic diagrams of the sun/pixel/sensors system and aggregate flux predictor architecture. The model applies the same weights to each size M pixel on the (P,Q) shaped grid, and generates a size L latent vector for each. The subsequent (P,Q,L) shaped grid is aggregated to a single size L latent vector by the PSF, then concatenated with the 3-element viewing geometry. Finally, a feedforward decoder takes the size L+3 combined vector as input, and tries to predict the CERES-observed longwave and shortwave flux.

Development Agenda

1. Acquire multiple years of EOS satellite overpasses within several small regions selected to have a wide variety of surface types and overpass times. Calculate the satellite-relative viewing angles for both datasets, and save the MODIS subgrid, CERES footprints, and geometric information for each overpass as an HDF5 file identified by its region and capture time.

2. Create a TensorFlow Dataset generator that selects CERES footprints, extracts their MODIS subgrids, and calculates the point spread function over each subgrid. The generator must open multiple overpass HDF5 files in separate processes and interleave their data in order to support training generalization.

3. Train the above architecture with a variety of hyperparameters such as network width, layer count, activation and, crucially, latent vector size. Attempt latent variational distribution.

4. Test the robustness of the model's flux predictions by modulating the viewing geometry, inputs, and latent values. Verify that the predictions remain physically realistic.

References

[1] T. H. V. Haar and V. E. Suomi, "Measurements of the Earth's Radiation Budget from Satellites During a Five-Year Period. Part I: Extended Time and Space Means," *Journal of the Atmospheric Sciences*, vol. 28, pp. 305–314, Apr. 1971. Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences.

[2] V. E. Suomi, K. J. Hanson, and T. V. Haar, "The theoretical basis for low-resolution radiometer measurements from a satellite," *Annual Report, Grant WBG-27, Dept. of Meteorology, University of Wisconsin*, pp. 79–100, 1967.

[3] V. Ramanathan, R. D. Cess, E. F. Harrison, P. Minnis, B. R. Barkstrom, E. Ahmad, and D. Hartmann, "Cloud-Radiative Forcing and Climate: Results from the Earth Radiation Budget Experiment," *Science*, vol. 243, pp. 57–63, Jan. 1989. Publisher: American Association for the Advancement of Science.

[4] G. L. Smith, R. N. Green, E. Raschke, L. M. Avis, J. T. Suttles, B. A. Wielicki, and R. Davies, "Inversion methods for satellite studies of the Earth's Radiation Budget: Development of algorithms for the ERBE Mission," *Reviews of Geophysics*, vol. 24, no. 2, pp. 407–421, 1986. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/RG024i002p00407.

[5] N. Loeb, W. Su, D. Doelling, T. Wong, P. Minnis, S. Thomas, and W. Miller, "Earth's Top-of-Atmosphere Radiation Budget," in *Comprehensive Remote Sensing*, pp. 67–84, Elsevier, 2018.

[6] J. T. Suttles, R. N. Green, P. Minnis, G. L. Smith, W. F. Staylor, B. A. Wielicki, I. J. Walker, D. F. Young, V. R. Taylor, and L. L. Stowe, "Angular radiation models for Earth-atmosphere system. Volume 1: Shortwave radiation," July 1988. NTRS Author Affiliations: A & B Design Engineering Co., Inc., Planning Research Corp., National Oceanographic Data Center, NASA Langley Research Center, National Oceanic and Atmospheric Administration NTRS Report/Patent Number: L-16414 NTRS Document ID: 19880018293 NTRS Research Center: Legacy CDMS (CDMS).

[7] N. G. Loeb, S. Kato, and B. A. Wielicki, "Defining Top-of-the-Atmosphere Flux Reference Level for Earth Radiation Budget Studies," *Journal of Climate*, vol. 15, pp. 3301–3309, Nov. 2002. Publisher: American Meteorological Society Section: Journal of Climate.

[8] W. Su, J. Corbett, Z. Eitzen, and L. Liang, "Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: methodology," *Atmospheric Measurement Techniques*, vol. 8, pp. 611–632, Feb. 2015. Publisher: Copernicus GmbH.

[9] P. Ricchiazzi, S. Yang, C. Gautier, and D. Sowle, "SBDART: A Research and Teaching Software Tool for Plane-Parallel Radiative Transfer in the Earth's Atmosphere," *Bulletin of the American Meteorological Society*, vol. 79, pp. 2101–2114, Oct. 1998. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.