Deep Learning Methods for Estimating 1km Broad-band TOA Flux with CERES and MODIS Imagery

Mitchell Dodson

1 Introduction

Satellite observation of broad-band outgoing longwave and shortwave radiance is an indespensible resource for understanding Earth's energy budget thanks to the unparalleled spatial coverage of in-orbit radiometers, and their ability to directly measure the directional solar reflectance and thermal emissions from an entire atmospheric path. These properties make them a vital basis for evaluating atmospheric and land surface reanalysis datasets like ERA5 [1], reducing error in surface-based bulk flux approximations [2], investigating the extent and theoretical fidelity of climate feedbacks like clouds and aerosols [3][4], and characterizing the radiative effects of large-scale phenomena like the El Niño Southern Oscillation processes and poleward atmospheric energy transport [5][6].

Nonetheless, satellite-based instruments measuring wide-band radiance generally have coarse spatial resolution on the order of tens of kilometers in order to maintain sufficient precision over the full range of their spectral response. This is in spite of the fact that many features that have a considerable impact on Earth's energy budget like clouds, urban development, and variations in vegetation type and coverage have characteristic scales which are often much smaller. Modern satellite projects like the NASA Earth Observation System (EOS), Suomi NPP, and JPSS which support wide-band instruments also typically provide data in narrow visible and infrared bands at a much finer resolution. While previous studies have used statistical regression techniques to merge multi-instrument observations of a scene into data products estimating TOA flux at finer resolutions [7][8], there is general a lack of investigation of deep learning methods that may be suited to the task (with one exception, addressed later [9]). As such, in this document I make the argument that unsupervised methods for machine learning are appropriate for extracting spatial and spectral characteristics of imager observations that are relevant for estimating TOA flux, and propose a novel approach for training several deep learning models with the end-goal of developing a 1km TOA shortwave and longwave flux product based on MODIS data.

2 Satellites and Earth's Radiation Budget

The value of satellite observations of outgoing radiation was recognized early in the devopment of space technology, with the first estimates of global shortwave and longwave fluxes being derived from the TIROS and ESSA III platforms in the late 1960s [10][11]. These efforts were formalized with the launch of broad-band Earth's Radiation Budget (ERB) instruments alongside Advanced Very High Resolution Radiospectrometer (AVHRR) imagers aboard a constellation of polar-orbiting satellites including Nimbus 7, NOAA-9, and NOAA-10 [12]. The ERB instruments offered a high signal-to-noise ratio across a wide spectrum in the shortwave $(.2 - 5\,\mu\text{m})$ and longwave $(5 - 50\,\mu\text{m})$ regions, and the AVHRR instruments provided 4-6 bands with spatial resolution at $4\,\text{km}$, or $1\,\text{km}$ on a limited domain [13].

Following the success of the ERB experiment, the Clouds and Earth's Radiant Energy System (CERES) instrument aboard the Terra and Aqua satellites has provided a near-continuous data record of top-of-atmosphere (TOA) broad-band fluxes for over two decades. CERES has three channels in the shortwave $(.3 - 5.0\,\mu\text{m})$ window $(8.0 - 12.0\,\mu\text{m})$ and total-spectrum $(.3 - 200\,\mu\text{m})$ ranges, which allow for measurement of solar reflectance
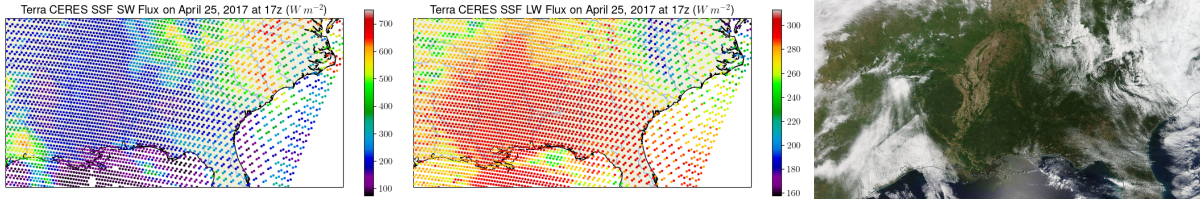
Figure 1: Comparison between shortwave and longwave flux from CERES Single Scan Footprint (left and middle), and cotemporal MODIS truecolor imagery (right) over the field of regard for this project. The bulk influence of clouds and land-surface features is apparent in the CERES footprint data, for example the increase in outgoing shortwave and depression in outgoing longwave radiation associated with the cloud deck over Virginia, however kilometer-scale features like the cumulus clouds over Florida are only distinguishable in the MODIS imagery.

and terrestrial emission with and an overall precision up to three times that of the ERB instruments within an elliptical receptive area (footprint) varying between about $20\,\mathrm{km}^2$ and $40\,\mathrm{km}^2$ depending on the panoramic distortion at the viewing angle [14]. The Terra and Aqua satellites also include Moderate Resolution Imaging Spectroradiometer (MODIS) instruments that contribute 36 narrow spectral bands distributed throughout the $0.41 - 14.1\,\mu\mathrm{m}$ range with nadir resolutions of $1\,\mathrm{km}^2$ or finer.

The juxtaposition of imager data having fine spatial/spectral resolution with radiometers providing broad-spectrum shortwave/longwave radiance proved a critical innovation that facilitated the development of Angular Dependence Models (ADMs), which considerably improve the reliability of full-sky TOA flux estimates. In particular, ADMs are lookup tables that use averages derived from large ensembles imager data to characterize the distribution of radiance reflected and emitted by a certain surface type given a specific sun/pixel/satellite geometry [15][16]. This is necessary for an accurate understanding of the overall energy budget because many surfaces are strongly anisotropic, so the assumption that the radiance observed for a feature is linearly proportional to the total hemispherical exitance doesn't often hold.

In practice, the surface type of an imager pixel is discretely determined by comparing its spectral properties in multiple bands to clusters of similar pixels, then the table associated with the selected surface type is inverted using the scene's angular geometry to select an anisotropic factor. This coefficient is finally used to scale the radiance to an estimate of the radiant flux integrated over all viewing angles, which is the quantity most directly related to Earth's energy budget. Since the imager pixels used to determine incident surface types have a much finer spatial resolution than the broadband footprints from which flux measurements are derived, the footprints often contain a mixture of multiple surface types. Therefore, the anisotropic factors used to determine flux values associated with a broadband observation are generally calculated as an average of the imager properties, which is spatially weighted by the Point Spread Function (PSF) of imager pixels in the receptive area with respect the footprint's centroid.

3   Regression Strategies in Machine Learning

Machine learning describes a broad category of powerful constructs for iteratively developing a representation of the intrinsic properties and relationships between complex datasets. At a fundamental level, most machine learning algorithms function by accepting an arbitrary-dimensional numerical vector as an input, then composing a series of learned

affine matrix transformations (hidden layers), each followed by element-wise operations that are differentiable but nonlinear (activation functions). Ultimately, the model produces an arbitrary-dimensional vector as an output. During the training process, a large volume of input samples are individually transformed by the model's hidden layers until the vector of output values (prediction) is arrived at. The prediction is subsequently compared to the ideal output corresponding to the input sample by a loss function which characterizes the skill of the model for evaluating that sample.

Since the affine transformation and element-wise activation function are necessarily differentiable at each layer of the model, the chain rule can be used to determine the output vector's gradient of sensitivity with respect to each element of the affine transformation matrix at every layer starting with the closest one to the output. Provided this distribution of sensitivity values throughout the model, an optimization algorithm is used to systematically adjust the matrix element values – which constitute the learned weights and biases of the model – such that the total error over collections of input samples is progressively reduced [17][18]. When multiple hidden layers are used, the model gains the capacity to learn increasingly sophisticated and nonlinear abstractions, which is a now-ubiquitous approach known as deep learning [19].

In contemporary times, much of the progress in data science applications for deep learning models involves developing novel high-level architectures built on top of the same basic principles outlined above. By modulating the depth and breadth of connections between layers and by introducing regular sub-structures, and flow of information and error through the model can be optimized for a particular domain of problems. For example, Convolutional Neural Networks (CNNs) are a category of model architectures that is currently very popular for problems involving data which is gridded on two or more orthogonal axes.

Rather than learning a single large vector of weights which are applied to the entire input sample simultaneously, the CNN strategy entails learning weights that constitute a series of kernel grids that are generally much smaller than the input layer. The kernels are typically learned independently to each other by convolving their weights over the input grid to generate the next layer of activations. This approach is useful because imbues the model with translation-invariance to input data and encodes a default spatial scale over which the kernels are receptive [20].

All of the prior literature I found that applies deep learning techniques to MODIS data with the goal of producing broad-band fluxes utilized architectures derived from CNNs. Li et al. implemented CNN-based Deep Belief Networks and ResNet models with multiple kernel sizes, and used ERA5 reanalysis data as a truth basis. While their models were successful in producing consistent fine-resolution flux estimates based on ERA5 reanalysis data, the outputs are limited to daily estimates of net radiative flux rather than instantaneous observations [9]. Zhan et al generated a 21-year dataset of TOA 1km longwave fluxes based on MODIS data with respect to CERES observations that have been resampled onto a regular grid using bilinear interpolation. Their technique employed a gradient-boosted regression network similar to a CNN architecture which showed impressive skill in validation [21]. Nonetheless, there are several shortcomings that arise due to the method's reliance on a gridded dataset. First, CNN-like architectures require that the input conforms to a fixed grid shape. This means that the MODIS data used to initialize the model as well as the flux data used as a truth basis need to be resampled onto large and regular image grids in order to train and execute the model. This is appropriate for some applications, however it limits the utility of the architecture as

an analysis tool for investigating the direct correspondence between MODIS imagery and outgoing flux. Furthermore, the resampling process may contaminate the comparison between CERES fluxes and MODIS pixels since the radiative effects of sub-footprint features such as cloud edges are extrapolated over a wider area than the corresponding MODIS pixels would actually represent.

In addition to the gridded problem domains most commonly accommodated by CNNs, machine learning is also widely adapted to a sequence-to-sequence (seq2seq) problem type, which entails a model architecture that accepts and produces a number of fixed-length vectors. Common applications for seq2seq problems include natural language processing and time series regression, both of which benefit in many cases from the flexibility to accept variable-length inputs, and from the capacity to reference the full context of the input sequence while generating the output. The specific architectures applied to seq2seq problems vary widely, however two rather different strategies of interest for this project which boast both of these abilities are Long Short-Term Memory networks (LSTMs) and Transformer models.

Since CERES single-scan footprints don't reliably conform to a fixed grid, but have receptive areas that can be directly associated with near-simultaneous MODIS observations, I'm proposing to reformulate the problem as a seq2seq-style regression, with each input and output sequences consisting of a collection of MODIS pixel vectors clustered around a single CERES footprint's centroid. This has several advantages, namely that (1) since neither the CERES flux footprints nor the MODIS pixels have been re-gridded, there is a more direct correspondence between observations by the two instruments, with fewer artifacts from interpolation, (2) relieving the dependence on a fixed grid would make the model
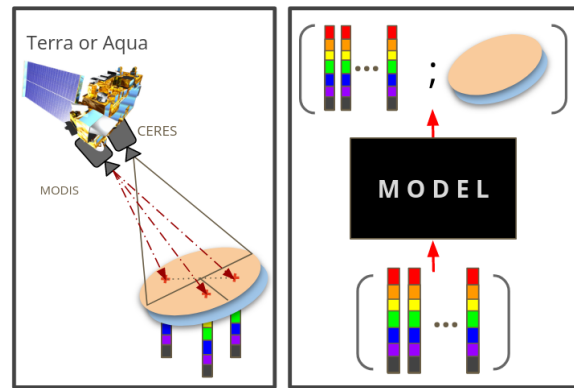

Figure 2: Schematic diagram demonstrating how estimation of outgoing flux from MODIS pixels co-located with CERES data can be posed as a sequence-to-sequence regression problem using an autoencoder-like deep learning model.

more useful as an analysis tool since small subsets of MODIS pixels may be processed independently from an entire granule, and (3) while spatial features on the 20km native scale of a CERES footprint will be initially learned inputs at the default MODIS resolution, the freedom to train on variable-length inputs lends itself to fine-tuning the model on sparse data, or even different datasets with entirely different resolutions.

LSTMs are a style of Recurrent Neural Network (RNN), which implies that they apply a consistent sub-structure to each element of the input or output sequence using the same learned weights. For each additional element in the sequence, the activation values from the previous element are passed to the recurrent component, allowing for the propagation of a so-called context vector containing information from the previous inputs. Since the same weights are used for each element, it's generally straightforward and efficient to train models to be agnostic to the length of the sequence and to accept skipped data values using basic masking techniques. LSTMs advance this concept by maintaining a context vector that is only operated on indirectly and element-wise by the input sequence via the activations from 3 learned information gates. This improves

the stability of the network over long sequences since the persisting information isn't transformed by a matrix operation after every element. [22][23]

Transformers are a more recent approach to seq2seq problems which have achieved impressive new states of the art in many use cases. Unlike RNN-based seq2seq architectures, transformers accept the entire input sequence at once, and apply an innovative technique called multi-headed attention (MHA) in order to make predictions by dynamically focusing on different parts of the input [24]. Although transformers are especially well-known for their aptitude in solving NLP problems, they have been adapted to computer vision tasks with great success and have been shown to be skillful in learning spatial properties of input data [25][26].

As a final piece of context, all of the above machine learning model varieties can be pretrained in an unsupervised fashion as autoencoders. An autoencoder is an architecture which initially learns to reduce the input data to a vector representation that is much smaller than its original size, then re-produces the input with the highest possible fidelity. If the model is reliably skillful in generating the original data after having its degrees of freedom substantially reduced, then the intermediate vector, known as the latent space representation, must have properties that correspond to abstract and expressive qualities of the original dataset. [27][28][29] Once the autoencoder is trained, the portion of the model that decodes the intermediate representation is discarded and replaced by a new and untrained decoder, which subsequently learns to accept the information-rich latent vector from the encoder as an input, and translate it into an output that has properties correlated with the characteristics extracted and generalized from the original data.

4   Methodology

During the process of this project I aim to develop and compare a series of autoencoder networks based on the LSTM and transformer architectures which will initially learn to reproduce the spatial and spectral features from MODIS pixels embedded in a circular domain surrounding the center of CERES footprints.

Band	Wavelength (μm)	Role
8	.41	Near UV
3	.46	Blue
4	.55	Green
1	.64	Red
2	.86	Near-Infrared
18	.935	NIR Water Vapor
26	1.38	Cirrus
6	1.64	Cloud Phase
7	2.11	Particle Size
20	3.7	Shortwave IR
27	6.5	Upper Water Vapor
30	9.7	Ozone
31	10.9	Clear IR Window
33	13.3	CO_2

Figure 3: MODIS bands used as input features for model training.

1. Download MODIS L1b imager data and CERES SSF longwave and shortwave fluxes for 2015, 2017, 2019, and 2021 within a region covering the South East US.

2. For each of N CERES footprints in a satellite overpass (swath), collect the nearest K MODIS imager pixels. Encode their position as the great circle distance and azimuth to the CERES centroid.

3. Store each swath as a (N,1,C) array of C CERES features, and a (N,K,M) array of M MODIS features, each along with a list of strings labeling each feature axis.

4. Quality-check the data by making sure there are no NaN values or data outside of the acceptable range. Remove any swaths that have no valid data from the dataset.

5. Gauss-normalize all in-range data and collect it into memory-mapped arrays for training and validation data.

6. Train an LSTM based autoencoder architecture to reproduce a sequence of up to K of the nearest imager pixels, optimizing for spatial interpolation of masked values within the 400px domain.

7. Train an LSTM based autoencoder architecture optimizing for spectral interpolation of masked values.

8. Develop a decoder head that converts the latent-space output of both above models into an estimate of the longwave and shortwave flux contribution of each MODIS pixel, using a custom loss function that compares the spatial average to the bulk CERES footprint values for flux.

9.

References

[1] N. G. Loeb, M. Mayer, S. Kato, J. T. Fasullo, H. Zuo, R. Senan, J. M. Lyman, G. C. Johnson, and M. Balmaseda, "Evaluating Twenty-Year Trends in Earth's Energy Flows From Observations and Reanalyses," *Journal of Geophysical Research: Atmospheres*, vol. 127, no. 12, p. e2022JD036686, 2022. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022JD036686.

[2] L. Yu, "Global Air–Sea Fluxes of Heat, Fresh Water, and Momentum: Energy Budget Closure and Unanswered Questions," *Annual Review of Marine Science*, vol. 11, no. 1, pp. 227–248, 2019. _eprint: https://doi.org/10.1146/annurev-marine-010816-060704.

[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] J. Zhang and S. A. Christopher, "Longwave radiative forcing of Saharan dust aerosols estimated from MODIS, MISR, and CERES observations on Terra," *Geophysical Research Letters*, vol. 30, no. 23, 2003. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2003GL018479.

[5] N. G. Loeb, D. A. Rutan, S. Kato, and W. Wang, "Observing Interannual Variations in Hadley Circulation Atmospheric Diabatic Heating and Circulation Strength," *Journal of Climate*, vol. 27, pp. 4139–4158, June 2014. Publisher: American Meteorological Society Section: Journal of Climate.

[6] M. Mayer and L. Haimberger, "Poleward Atmospheric Energy Transports and Their Variability as Evaluated from ECMWF Reanalysis Data," *Journal of Climate*, vol. 25, pp. 734–752, Jan. 2012. Publisher: American Meteorological Society Section: Journal of Climate.

[7] D. Wang and S. Liang, "Estimating high-resolution top of atmosphere albedo from Moderate Resolution Imaging Spectroradiometer data," *Remote Sensing of Environment*, vol. 178, pp. 93–103, June 2016.

[8] N. G. Loeb, W. Sun, W. F. Miller, K. Loukachine, and R. Davies, "Fusion of CERES, MISR, and MODIS measurements for top-of-atmosphere radiative flux validation," *Journal of Geophysical Research: Atmospheres*, vol. 111, no. D18, 2006. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2006JD007146.

[9] S. Li, B. Jiang, S. Liang, J. Peng, H. Liang, J. Han, X. Yin, Y. Yao, X. Zhang, J. Cheng, X. Zhao, Q. Liu, and K. Jia, "Evaluation of nine machine learning methods for estimating daily land surface radiation budget from MODIS satellite data," *International Journal of Digital Earth*, vol. 15, pp. 1784–1816, Dec. 2022. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17538947.2022.2130460.

[10] F. B. House, *The Radiation Balance of the Earth from a Satellite*. PhD thesis, Jan. 1965. Publication Title: Ph.D. Thesis ADS Bibcode: 1965PhDT........47H.

REFERENCES

[11] T. H. Vonder Haar and V. E. Suomi, "Satellite Observations of the Earth's Radiation Budget," *Science*, vol. 163, pp. 667–669, Feb. 1969. Publisher: American Association for the Advancement of Science.

[12] B. R. Barkstrom, "The Earth Radiation Budget Experiment (ERBE)," *Bulletin of the American Meteorological Society*, vol. 65, pp. 1170–1185, Nov. 1984. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[13] W. Shenk, "NOAA-10 Advanced Very High Resolution Radiometer."

[14] B. A. Wielicki, B. R. Barkstrom, E. F. Harrison, R. B. Lee, G. L. Smith, and J. E. Cooper, "Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment," *Bulletin of the American Meteorological Society*, vol. 77, pp. 853–868, May 1996. Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society.

[15] 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).

[16] N. G. Loeb, S. Kato, K. Loukachine, and N. Manalo-Smith, "Angular Distribution Models for Top-of-Atmosphere Radiative Flux Estimation from the Clouds and the Earth's Radiant Energy System Instrument on the Terra Satellite. Part I: Methodology," *Journal of Atmospheric and Oceanic Technology*, vol. 22, pp. 338–351, Apr. 2005. Publisher: American Meteorological Society Section: Journal of Atmospheric and Oceanic Technology.

[17] S. J. Russell and P. Norvig, *Artificial Intelligence: a Modern Approach.* London, 2010.

[18] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway Networks," Nov. 2015. arXiv:1505.00387 [cs].

[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. Number: 7553 Publisher: Nature Publishing Group.

[20] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, pp. 2278–2324, Nov. 1998. Conference Name: Proceedings of the IEEE.

[21] C. Zhan and S. Liang, "Generation of global 1-km daily top-of-atmosphere outgoing longwave radiation product from 2000 to 2021 using machine learning," *International Journal of Digital Earth*, vol. 16, pp. 2002–2012, Dec. 2023. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/17538947.2023.2220611.

[22] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, pp. 1735–1780, Nov. 1997. Conference Name: Neural Computation.

[23] C. B. Vennerød, A. Kjærran, and E. S. Bugge, "Long Short-term Memory RNN," May 2021. arXiv:2105.06756 [cs].

[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," Dec. 2017. arXiv:1706.03762 [cs].

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," June 2021. arXiv:2010.11929 [cs].

[26] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked Autoencoders Are Scalable Vision Learners," Dec. 2021. arXiv:2111.06377 [cs].

[27] M. Quade, M. Abel, J. Nathan Kutz, and S. L. Brunton, "Sparse identification of nonlinear dynamics for rapid model recovery," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, p. 063116, June 2018.

[28] K. Kashinath, M. Mustafa, A. Albert, J.-L. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, A. Manepalli, D. Chirila, R. Yu, R. Walters, B. White, H. Xiao, H. A. Tchelepi, P. Marcus, A. Anandkumar, P. Hassanzadeh, and n. Prabhat, "Physics-informed machine learning: case studies for weather and climate modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, p. 20200093, Feb. 2021. Publisher: Royal Society.

[29] L. M. Gao and J. N. Kutz, "Bayesian autoencoders for data-driven discovery of coordinates, governing equations and fundamental constants," Nov. 2022. arXiv:2211.10575 [cs, stat].