Remote Sensing of Environment 285 (2023) 113376

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimates of broadband upwelling irradiance from GOES-16 ABI

Kevin McCloskey [a],[*], Sixing Chen [a], Vincent R. Meijer [b], Joe Yue-Hei Ng [a], Geoff Davis [a], Carl Elkin [a], Christopher Van Arsdale [a], Scott Geraedts [a]

[a] Google, Inc, United States of America

[b] Laboratory for Aviation and the Environment, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, 02139, MA, United States of America

ARTICLE INFO

Edited by Menghua Wang

Dataset link: https://github.com/google-research/google-research/tree/master/collocated_irradiance_network/

Keywords:
Artificial neural networks
Spatial resolution
Temporal resolution
CERES
GOES-16 ABI
Top of atmosphere flux
OLR
RSR

ABSTRACT

Satellite-derived estimates of the Earth's radiation budget are crucial for understanding and predicting the weather and climate. However, existing satellite products measuring broadband outgoing longwave radiation (OLR) and reflected shortwave radiation (RSR) have spatio-temporal resolutions that are too coarse to evaluate important radiative forcers like aircraft condensation trails. We present a neural network which estimates OLR and RSR based on narrowband radiances, using collocated Cloud and Earth's Radiant Energy System (CERES) and GOES-16 Advanced Baseline Imager (ABI) data. The resulting estimates feature strong agreement with the CERES data products (R^2 = 0.977 for OLR and 0.974 for RSR on CERES Level 2 footprints), and we provide open access to the collocated satellite data and model outputs on all available GOES-16 ABI data for the 4 years from 2018–2021.

1. Introduction

Direct measurements of top-of-atmosphere (TOA) outgoing longwave radiation (OLR) and reflected shortwave radiation (RSR) flux are essential to our understanding of the Earth system. While the CERES sensor (Wielicki et al., 1996) aboard the polar-orbiting Terra and Aqua satellites provides this capability, the product's spatio-temporal resolution is too coarse to study the radiative impact of short-lived and/or individually small features. Such phenomena can be studied using geostationary satellite imagery, as has been done for aircraft condensation trails (contrails), ship tracks, and atmospheric convection (Meijer et al., 2022; Schreier et al., 2010; Cintineo et al., 2020). However, since this imagery consists of narrow band radiance measurements it cannot directly be used to quantify the radiative effects these phenomena have on the earth system. This paper develops a method that uses the narrowband radiance measurements by geostationary satellite imagers to estimate the OLR and RSR broadband quantities at a spatio-temporal resolution that is higher than currently possible using CERES measurements alone. These high resolution estimates enable the quantification of the radiative impact of the aforementioned climate forcers, which will inform mitigation efforts.

Several methods have been introduced that perform a regression using geostationary narrowband radiance measurements as input and produce single-kilometer-scale resolution estimates of OLR and RSR as output. The majority of these methods (reviewed in detail in Section 2) have used radiative transfer simulations to generate datasets for developing the OLR and RSR regression models.

The approach developed here differs fundamentally from those in that we developed a neural network regressor that estimates OLR and RSR using only collocated data from GOES-16 ABI and CERES aboard Terra and Aqua satellites. Because of this, we call the method **CO**llocated **I**rradiance **N**etwork, or COIN. The primary strength of this approach is that we avoid potential discrepancies between modeled and measured broadband fluxes which can lead to decreased performance of regression models that are developed using such simulation outputs; this phenomenon is generally known as "covariate shift" (Shimodaira, 2000; Huang et al., 2006). Our design choice to rely solely on collocated geostationary/CERES data is associated with several other advantages and disadvantages, which are detailed further in Section 5.

In addition to reporting aggregated evaluation of the approach we developed here, we report model error characteristics on subsets of the

[*] Corresponding author.
E-mail address: mccloskey@google.com (K. McCloskey).

https://doi.org/10.1016/j.rse.2022.113376
Received 22 August 2022; Received in revised form 16 November 2022; Accepted 18 November 2022

data and sliced along different dimensions to identify opportunities for future modeling improvements. We also make all the collocation data used to develop the model and the estimates from applying COIN to all available GOES-16 ABI data in the four years from 2018 to 2021 available for public use: see Data Availability section.

2. Related work

There is a many-decades long history of estimating top-of-atmosphere flux from satellite sensors; for a broad review, consider Liang et al. (2019). In this section we largely restrict our overview to the techniques most similar to COIN, which is designed to estimate top-of-atmosphere flux using geostationary satellite radiances as input (in this work specifically GOES-16 ABI Schmit et al., 2017).

As noted, multiple previous works have developed OLR and RSR regressors based on radiative transfer simulation datasets. The most similar to COIN are those which have evaluated their TOA flux estimates not just on their respective simulation datasets but also on CERES data from satellites. The following four works fall into this category, though they vary considerably in the amount of CERES data used in validation. For example, while Vázquez-Navarro et al. (2013) simulated SEVIRI radiances aboard the Meteosat Second Generation satellite using the radiative transfer model libRadTran (Mayer and Kylling, 2005; Pinker et al., 2022) simulated measurements by the Advanced Baseline Imager (ABI) aboard GOES-16 and GOES-17 using MODTRAN (Berk et al., 1987), both of these works compared their flux estimates to less than 10 h worth of CERES data which did not span the yearly cycle of seasons, preventing assessment of seasonal error characteristics. The OLR estimates by Kim and Lee (2019) (which simulated the Advanced Himawari Imager (AHI) aboard Himawari using SBDART) were evaluated on CERES data from two days each month in 2017, but because each day was only evaluated on less than 2 h worth of CERES data, assessing flux error across the diurnal cycle was not possible. Lee et al. (2018) which also simulated the Advanced Himawari Imager (AHI) aboard Himawari using SBDART (Ricchiazzi et al., 1998) reports the most extensive evaluation we are aware of against CERES data prior to this work, by reporting error characteristics for its RSR estimates compared to most of the collocated daytime CERES data from Terra, Aqua, and Suomi National Polar-orbiting Partnership (S-NPP) on the 15th day of each month for 20 consecutive months — omitting however any solar/viewing zenith angles greater than 80° or determined to be affected by sunglint (Kay et al., 2009).

In contrast, here we report evaluations comparing to more extensive CERES data: our validation dataset is comprised of 20% of the hours spanning 3 years, and we make publicly available the underlying COIN-estimated full-disk OLR and RSR for every 10–15 min for each day across 4 years. Additionally, this is the first report including a chronological holdout set: our test set is the entire year 2021 which was held out from all model development, as a way to assess COIN being applied to future GOES-16 ABI data without any further adjustments.

We note also that all four of the OLR and RSR regressors reviewed above make comparisons to CERES data that has been spatio-temporally averaged/regridded, and consequently it is possible they under-report their models' error. In this work comparisons are made directly to individual CERES L2 Single Scanner Footprints, by aggregating COIN's 2 km-nominal flux estimations as weighted by each footprint's point spread function (PSF) (Green and Wielicki, 1997). The CERES PSF is an ovaloid shape with exponential decay towards its edges, and we contribute as part of this work an open source efficient calculation of the CERES PSF for future works to use to directly compare to CERES L2 SSF flux labels without the need for regridding or spatio-temporal averaging.

Most of the regressors reviewed above contain explicitly separate steps for estimating the broadband radiance from observed narrowband radiances, followed by the flux inversion to estimate the irradiance for all outgoing angles. The modeling approach taken in this work however, is most similar to the RSR regressor introduced by Vázquez-Navarro et al. (2013) where a neural network learns to approximate the sequential composition of those two functions as a single operation; while (Vázquez-Navarro et al., 2013) trained on a radiative transfer simulation dataset, here we train directly on CERES flux labels by incorporating the CERES PSF as a layer in the neural network, and back-propagating errors directly through it.

The literature also contains many works of satellite-driven flux estimates which are designed for slightly different purposes than COIN's goal of top-of-atmosphere flux estimates at geostationary spatio-temporal resolution. For example, Gupta et al. (2016) used CERES data collocated with measurements from the Ozone Monitoring Instrument on the low-earth-orbiting Aura satellite. The resulting regression model was used to extend the record of RSR data back to as early as 1979, by using historical ozone measurements. Another example are techniques that make estimates of TOA albedo as an intermediate value to facilitate estimating earth surface fluxes, which can then be used in estimating snowmelt, flood forecasts, soil moisture, and assimilation by numerical weather models (Huang et al., 2019). However, there are substantial challenges in comparing TOA flux estimates to surface station measurements, due primarily to land cover heterogeneity (Li et al., 1995; Cescatti et al., 2012) and attenuation by the intervening atmosphere. Given that the same physical factors driving TOA irradiance also play a large role in surface irradiance (primarily scattering/absorption by clouds, aerosols and water vapor) we do expect COIN can be extended to the purpose of surface solar irradiance estimation, but leave this effort to future work.

3. Methods

3.1. Regression dataset

We train and validate our model using 3 data sources: CERES Level 2 single scanner footprint (SSF) (Loeb et al., 2016; Minnis et al., 2008, 2011b,a; Su et al., 2015a,b) from the Terra and Aqua satellites; CERES Level 3 SYN1deg product (Doelling et al., 2013, 2016); and collocated GOES-16 ABI data (Schmit et al., 2017).

The CERES instrument (Wielicki et al., 1996) is a broadband scanning bolometer with a roughly ovaloid spatial response function (the "footprint") that is nominally 20 km diameter at nadir but at high viewing angles can stretch to more than 100 km. The OLR and RSR measured by CERES can thus be viewed as a spatial, weighted average of the fluxes inside the footprint. The relative contribution of the locations within the footprint are quantified using the point spread function (PSF), which is a function of the orbital geometry and viewing angle (Green and Wielicki, 1997). The CERES L2 SSF data product provides estimates of the top of atmosphere irradiance at a single location directly viewed by a CERES instrument. The L3 SYN1deg product combines data from multiple satellite sensors (including CERES on Terra and Aqua and ABI on GOES-16) to produce hourly estimates of top-of-atmosphere flux on a 1 × 1 degree grid of the earth.

We train and validate on a mix of CERES L2 and L3 data because collectively they provide a more accurate set of flux labels across the viewing extent/times of the GOES-16 ABI. The CERES L2 SSF data contain broadband radiance measurements from the CERES instrument, converted to flux using anisotropic factors determined by empirical angular distribution models (ADMs, in this work we use Edition 4 A) (Su et al., 2015a). This provides an unprecedented empirical measurement record of the earth's radiation budget, however, operational CERES instruments are only present on sun-synchronous polar-orbiting satellites; the L3 SYN1deg product provides flux estimates throughout the day and night at all locations (e.g., including mid-latitude sunrise/sunset) as well as ensuring sun glint (Kay et al., 2009) is not under-represented.

The GOES-16 ABI (Schmit et al., 2017) is a 16-band imaging radiometer, with spectral bands covering the visible, near-infrared, and infrared. In this work we use the full disk L1b radiance product, which

Table 1
Searched ranges and selected values of model hyperparameters.

Parameter	Range searched	Selected value
Learning rate	1e-5 to 1e−1	0.00067
Learning rate drop patience	10 to 500	115
Learning rate drop factor	0.1 to 1.0	0.72
Dropout	0.0 to 0.5	0.0
First layer size	100 to 1000	525
Layer size scale factor	0.1 to 0.5	0.34
Number of layers	3, 4, or 5	4
Activation	sigmoid, relu, swish or leaky_relu	leaky_relu
Loss function	MSE or MAE	MAE
Example weight clip	1 to 1e5	90

Table 2
Prediction bias and error of broadband flux estimates from narrowband imagers aboard geostationary satellites, which were validated against the CERES L2 SSF product. All values are W/m^2.

Satellite	OLR bias	RMSE	RSR bias	RMSE	Validation set
GOES-16[a]	0.08	6.64	−0.88	24.64	2018, 2019, 2020
GOES-16/17[b]	–	–	19.14	85.15	7 h in 2019
Himawari-8[c]	2.28	11.03	–	–	2017
Himawari-8[d]	–	–	−2.34	52.12	1 day/mo, 2015–2017
Meteosat-9[e]	−0.945	9.065	−17.79	37.42	8 swaths in 2004

[a] COIN (this work).
[b] Pinker et al. (2022).
[c] Kim and Lee (2019).
[d] Lee et al. (2018).
[e] Vázquez-Navarro et al. (2013).

has a refresh rate of 10 min (15 min before April 2, 2019), with a nadir pixel size of 2 km for its infrared bands.

COIN is developed using a fraction of the collocation dataset ("the training data"). We monitor the progress of COIN as we optimize its parameters using the training data by setting aside a random 20% of the hours in the three years 2018–2020 to use as validation data. The performance of COIN on this validation dataset is used as a proxy for the model's performance on unseen data. Finally, all data for the year 2021 is used as the test dataset: COIN's performance on this data is the best estimate of its ability to be used in future years without retraining it. CERES L2 footprints and L3 1 × 1 degree gridboxes are randomly sampled from within each hour box keeping 25% and 6% respectively to generate roughly equal numbers of L2 and L3 training examples. The footprints/gridboxes were then filtered by three criteria. First, to limit limb-darkening and parallax effects (Joyce et al., 2001), the location of an L2 footprint/L3 gridbox centroid must be within 7258 km from the GOES-16 sub-satellite point at −75.2° longitude, and the CERES viewing zenith angle for an L2 footprint must be less than or equal to 60 degrees. Second, no GOES-16 ABI pixels inside the footprint/gridbox can have an invalid value, within any of the 16 bands. Third, to limit how much the atmospheric state can evolve between the collocated measurements, the L2 footprint acquisition time must be less than 120 s from the GOES-16 ABI scan time at that location. To implement this last filter, we created a pixel-wise map of ABI scan times for modes 3 and 6 A following (Kalluri et al., 2018). This procedure yielded train, validation and test set sizes of 48,759,836; 12,196,071; and 22,982,710 CERES footprints/gridboxes.

The GOES-16 ABI full-disk refresh rate changed from 15 min to 10 min on Apr 2, 2019 at 00:00 UTC; All training data are split into before and after that time, and two (otherwise-identical) models are trained — one for before that time and one for after. The figures, tables and metrics reported in this paper are from these two models applied to the validation data in their respective date ranges.

3.2. Model architecture

The neural network model takes as input the radiances measured in all 16 bands of the GOES-16 ABI for a single pixel as well as 5 auxiliary inputs: the latitude, the longitude, the solar zenith angle, the solar azimuth angle, and the day of the year. Higher resolution ABI radiances were downscaled (by averaging) to the 2 km nadir resolution of the ABI infrared channels. The neural network processes the inputs through a series of fully-connected layers with non-linear activation functions, and outputs both $\widehat{OLR}_n$ and $\widehat{RSR}_n$ for each GOES-16 ABI pixel n within a CERES footprint/gridbox having a total of N ABI pixels in it. Because the fully-connected layers are applied to each GOES-16 ABI pixel separately, it is equally valid to call these layers 1 × 1 convolutional layers over a 2D image of GOES-16 ABI pixels. During training, all of the $\widehat{RSR}_n$ and $\widehat{OLR}_n$ are aggregated in a weighted sum before being compared with the CERES fluxes. The L2 footprints use the CERES PSF weights, while L3 gridboxes use normalized uniform weights, both denoted as P_n

$$\widehat{RSR} = \sum_{n=1}^{N} P_n \cdot \widehat{RSR}_n \qquad (1)$$

$$\widehat{OLR} = \sum_{n=1}^{N} P_n \cdot \widehat{OLR}_n$$

The loss function $\mathcal{L}$ then penalizes errors compared to the CERES data product as follows

$$\mathcal{L} = \left|\widehat{RSR} - RSR_{\text{CERES}}\right| + \left|\widehat{OLR} - OLR_{\text{CERES}}\right| \qquad (2)$$

where RSR_{CERES} and OLR_{CERES} are the irradiance values reported by the CERES L2 SSF or L3 SYN1deg data product. The network weights are iteratively adjusted to minimize this loss function by using backpropagation (Rumelhart et al., 1985). After training is complete, the weighted aggregation is no longer applied to the network output, and the ABI pixel-by-pixel predictions $\widehat{OLR}_n$ and $\widehat{RSR}_n$ are used directly as the model outputs.

While the COIN model published here predicts both OLR and RSR simultaneously from the same network ("dual-task"), single-task models were also trained which predicted just OLR and just RSR. However, the improvement in average single-task RMSE or bias was marginal compared to dual-task models (not shown). Therefore we selected the dual-task mode for operational convenience.

3.3. Data preprocessing

The GOES-16 radiances, CERES fluxes, and 5 auxiliary inputs detailed above are all preprocessed to standard ranges before being passed to the model, as follows. During training data generation, the GOES-16 radiances, CERES fluxes, and latitude and longitude were sampled and their mean and standard deviation recorded, so that they can be clipped at 5.5 standard deviations and normalized to the range [0.0, 1.0]. The normalization is done by subtracting the mean, dividing by 11 times the standard deviation, adding 0.5, and clipping the result to the range [0, 1]. The solar zenith angle is passed through a cosine function before being presented to the network. Similarly, the day of the year is multiplied by $2\pi/365$ and also passed through the cosine function. The solar azimuth angle is calculated as radians/2π and presented to the model with the convention that it is in the range [0, 0.5] when the hour angle is negative and in [0.5, 1.0] when the hour angle is positive. The model outputs are mapped to units of irradiance (W/m^2) by inverting the scaling procedure used to normalize the CERES flux labels (i.e., subtracting 0.5, multiplying by 11 times the flux standard deviation, and adding back the flux mean). This predicted flux is then clipped to a minimum of 0.0 W/m^2. For any GOES-16 ABI pixel with an input solar zenith angle larger than 90 degrees, the model was forced to return 0.0 W/m^2 RSR.

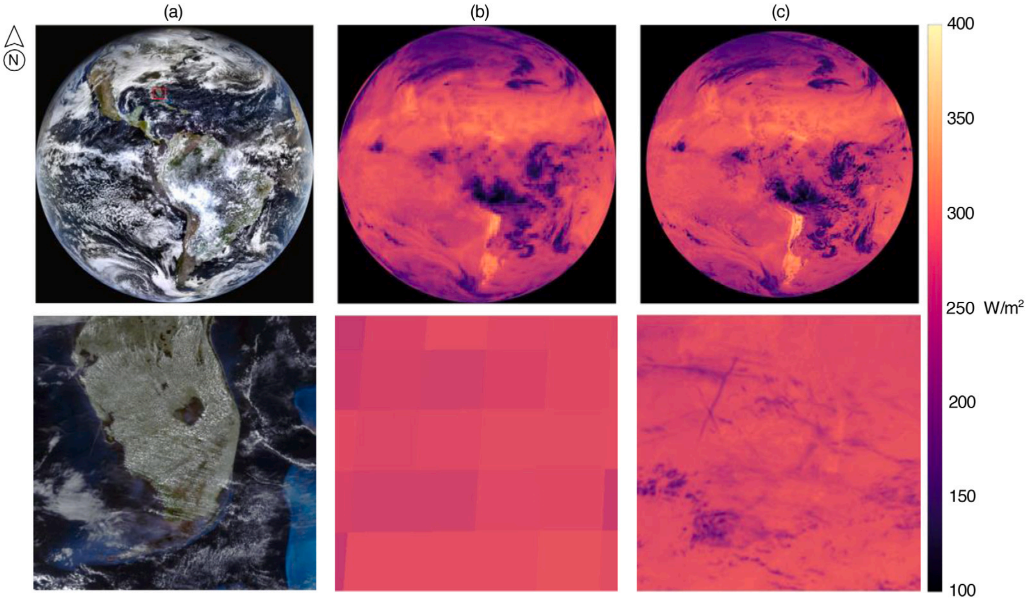

Fig. 1. (a) GOES-16 true color product for 2021-02-10 17:00:00 UTC, with a region including southern Florida outlined in red shown below. **(b)** The CERES Level 3 1 × 1 degree OLR product for the same scene. **(c)** COIN-predicted OLR for the same scene. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

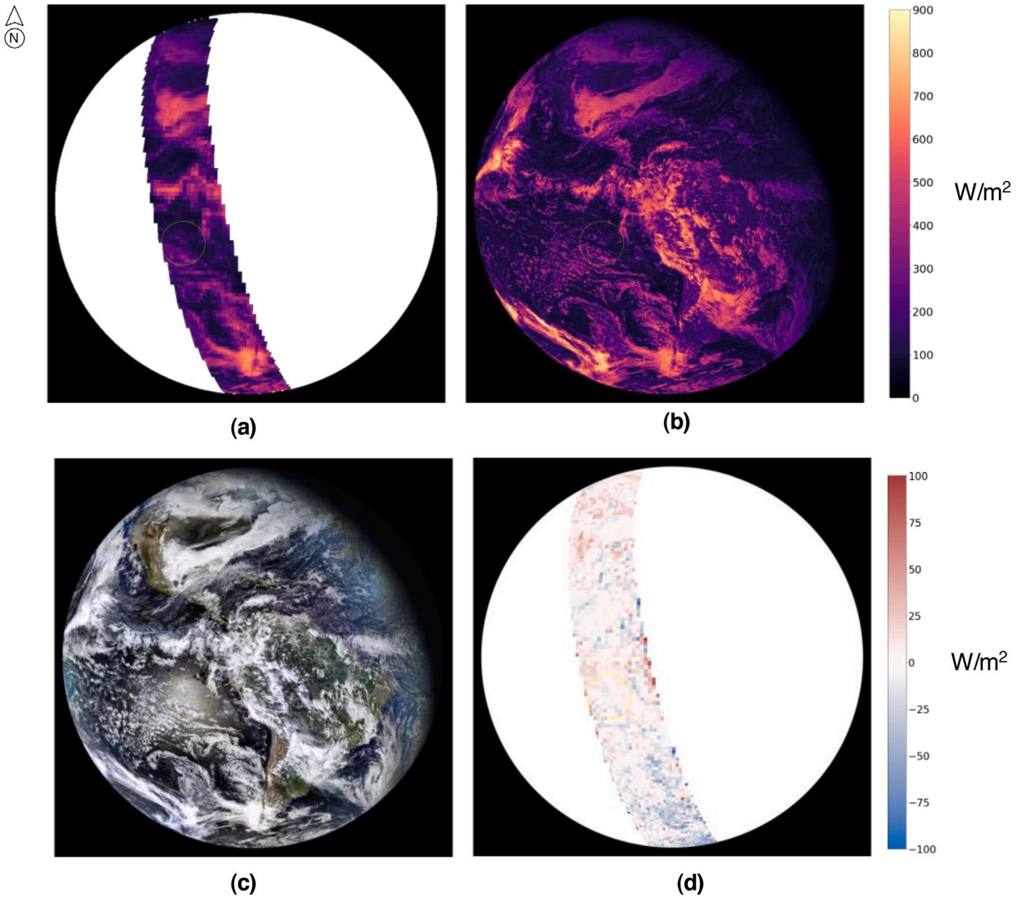

Fig. 2. A scene in GOES-16 perspective for the hourbox centered at 2018-01-01 19:30:00 UTC, with obvious sunglint in the Pacific ocean, which we have highlighted by adding a thin yellow circle in each panel. **(a)** The CERES Level 2 RSR swath observed from Aqua in the hourbox, averaged on a 1 × 1 degree grid. **(b)** COIN-estimated RSR from the GOES-16 19:30:00 UTC scan. **(c)** The GOES-16 true color product (hourbox-averaged) where sunglint is clearly visible. **(d)** Bias resulting from averaging the COIN estimates to a 1 × 1degree grid and subtracting the CERES Level 2 RSR product. GOES-16 true color and CERES product have been trimmed to show only points within 7258 km of the GOES-16 sub-satellite point to match the COIN estimation extent. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

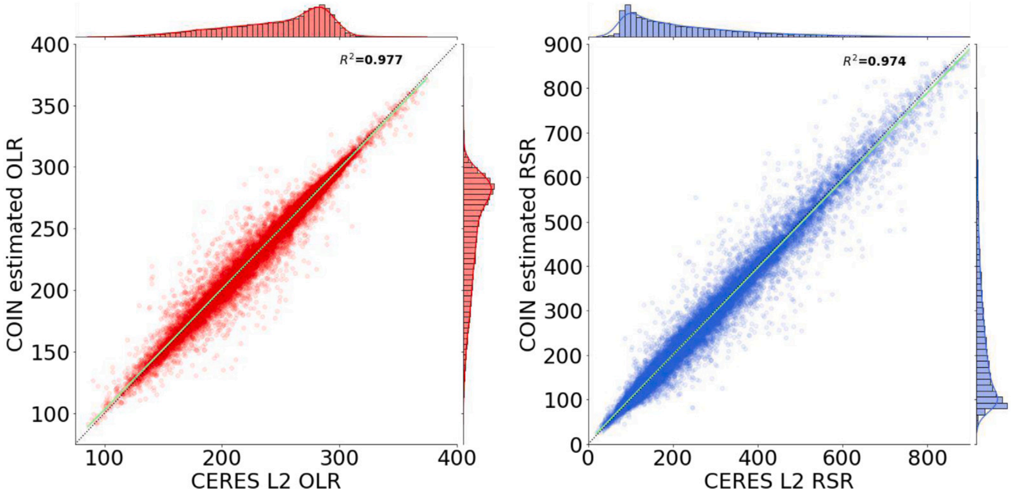

Fig. 3. Comparison between COIN and CERES L2. The pale green line is a linear fit to the data. The black dotted line corresponds to the 1:1 line. Units are W/m^2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

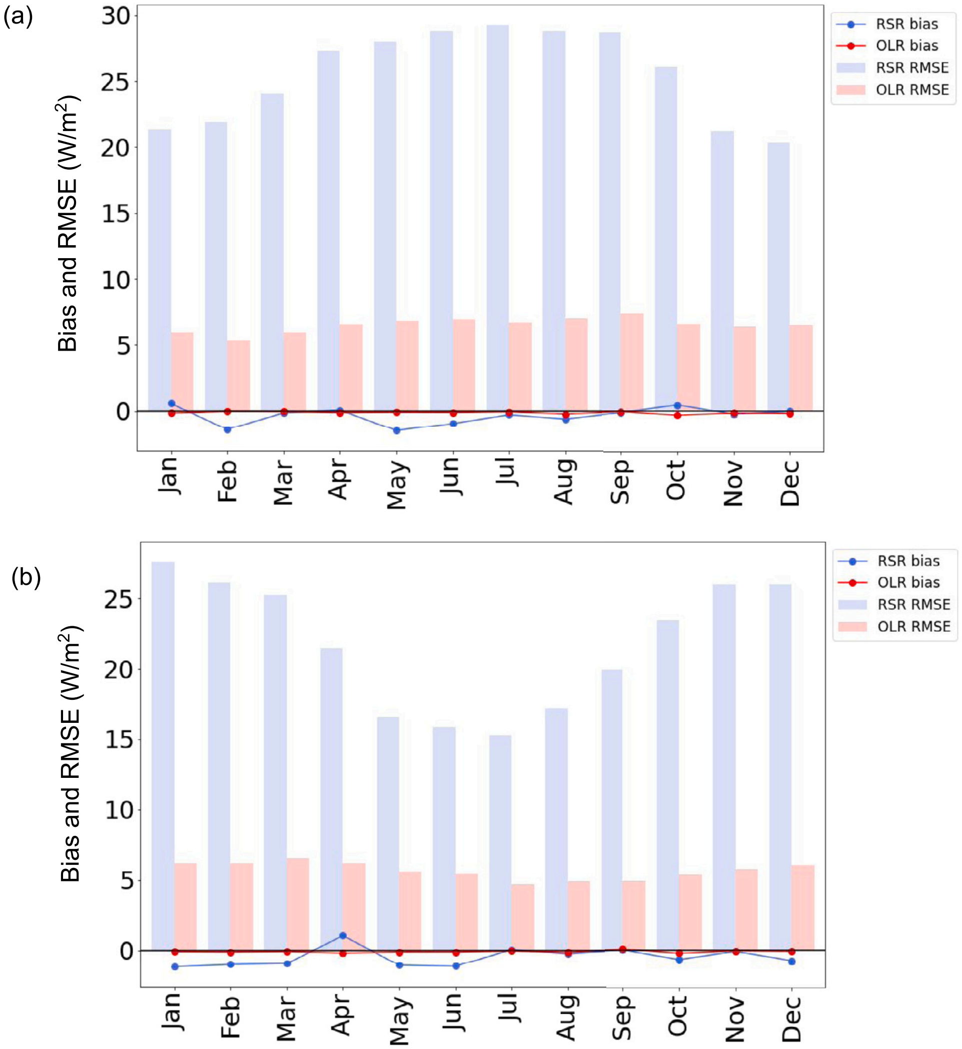

Fig. 4. Comparison of COIN to CERES L2 broken out by month for the **(a)** Northern hemisphere and **(b)** Southern hemisphere.

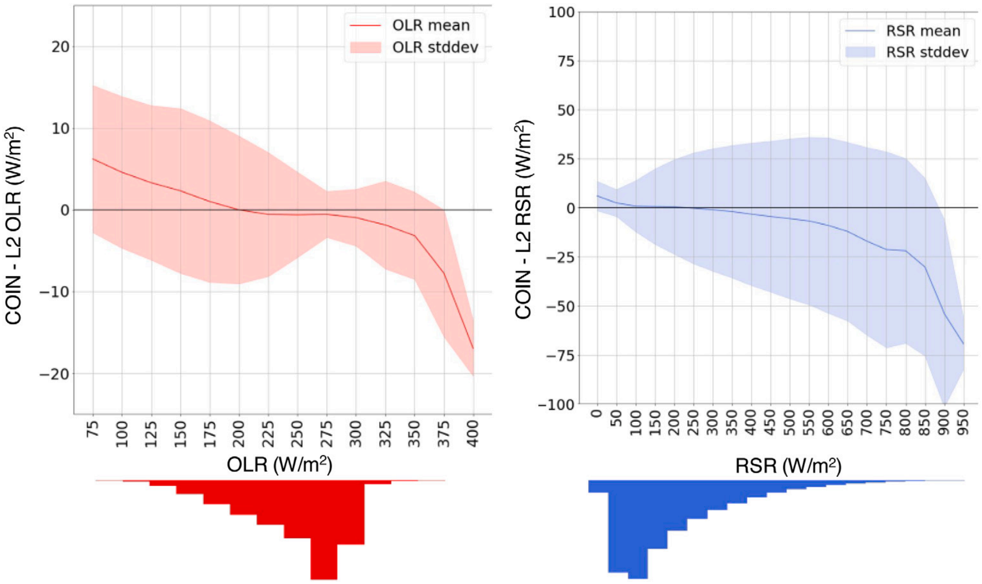

Fig. 5. Comparison between COIN and CERES L2 for bins of OLR and RSR magnitude. The validation set frequency for OLR and RSR magnitudes are shown as upside-down histograms at the bottom. Footprints with solar zenith angle > 90 degrees are not included in the RSR data. The validation set frequency by magnitude is shown as an upside-down histogram at the bottom.

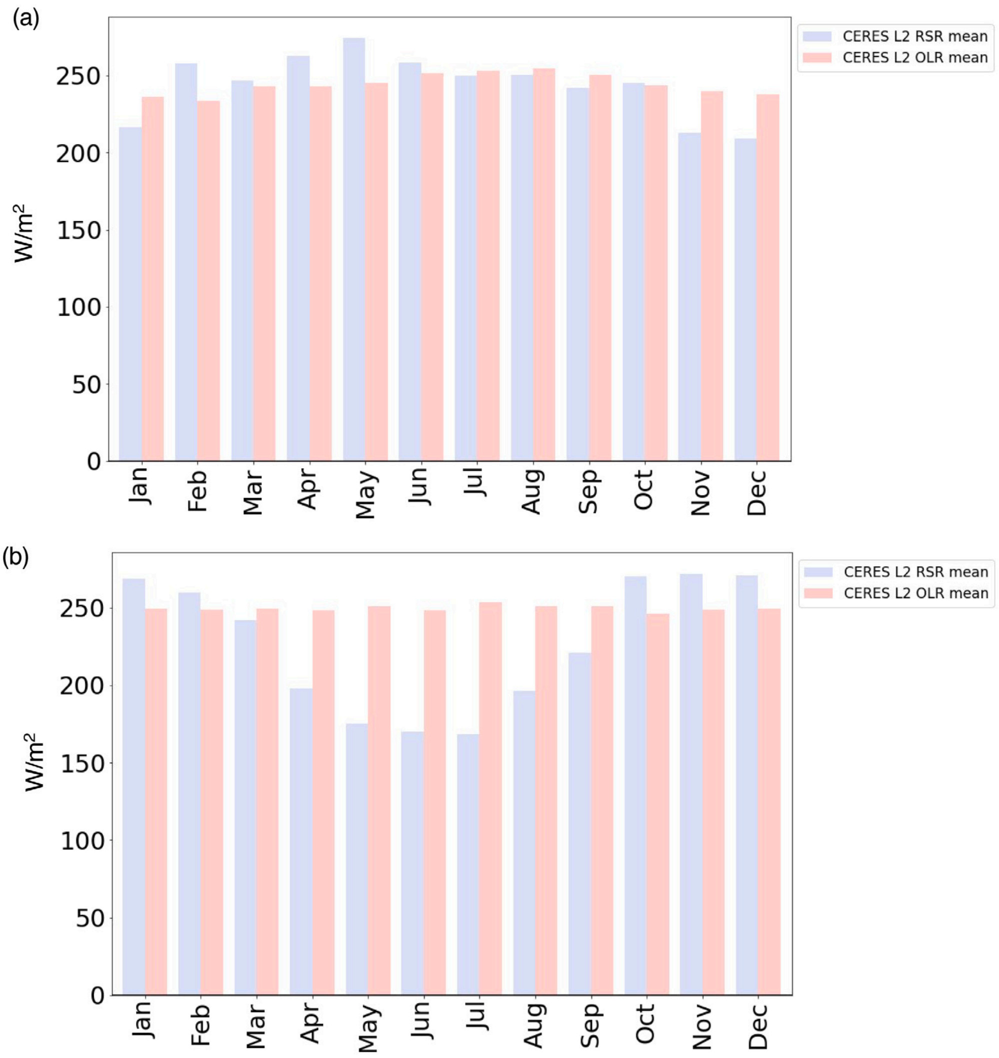

Fig. 6. CERES L2 TOA flux broken out by month for the **(a)** Northern hemisphere and **(b)** Southern hemisphere.

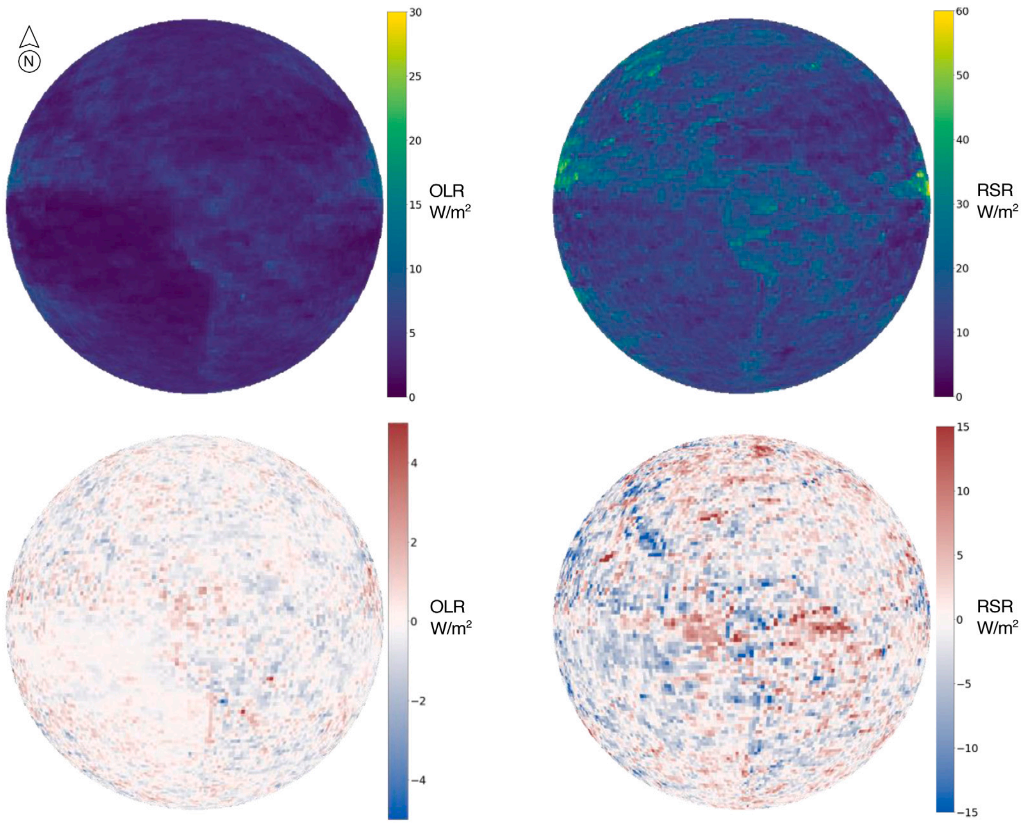

Fig. 7. Top row: mean absolute deviation between spatio-temporally averaged COIN outputs and CERES L2 data product for each 1 × 1 latitude and longitude gridbox, across 3 years. Bottom row: mean bias of same.

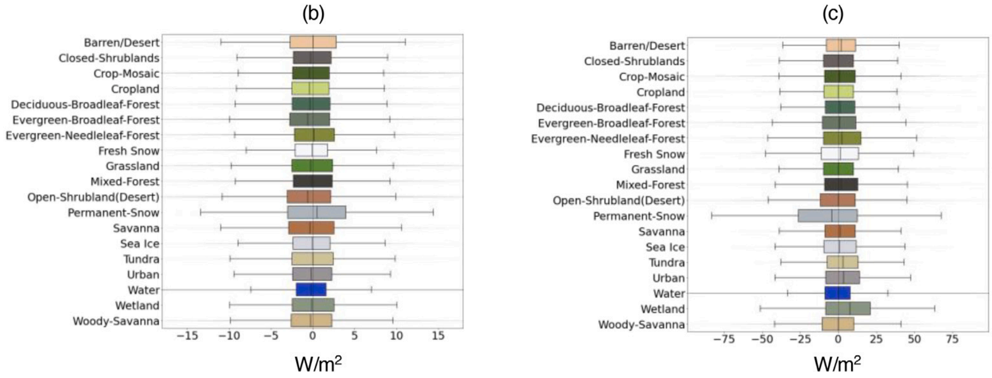

Fig. 8. Comparison between COIN and CERES L2 SSF for the most common surface type of the SSF. **(a)** A plot of CERES earth surface types, color-coded to match the error plots showing **(b)** OLR differences and **(c)** RSR differences.

3.4. Weighting scheme

We applied a loss weighting scheme during training, where each training example (footprint/gridbox) has its loss up- or down-weighted according to its inverse frequency in the training set. The solar zenith angle, OLR magnitude, and RSR magnitude for examples were each discretized into buckets of 10 (degrees), 25 (W/m^2), and 50 (W/m^2) respectively to determine the example's frequency.

Each training example has one OLR flux label and one RSR flux label from CERES L2 or L3 and up to 1700 collocated GOES-16 ABI input pixels. A sizeable portion of the generated dataset's CERES L3 training examples are in the Meteosat −11 longitude domain (≥ −37.5 degrees) and so the CERES L3 SYN1deg product uses (CERES-normalized) Meteosat −11 SEVIRI radiances to determine the flux in hours without CERES measurements from low earth orbit (Doelling et al., 2013). In this domain SEVIRI radiances can differ substantially from ABI radiances due to scan time differences alone: SEVIRI scans south to north (Aminou et al., 1997) while ABI scans north to south (Kalluri et al., 2018), so the mean scan time of the 4–6 scans within the hour box can be 10–15 min different for a given location. We mitigated this issue by applying a time-weighted random sampling of the ABI input pixels so that the mean ABI scan time matches the mean SEVIRI scan time within the hour box.

3.5. Hyperparameters

We applied the black box optimization method known as Batched Gaussian Process Bandits (Golovin et al., 2017) to tune the hyperparameters of the neural network (such as number of layers and which non-linear activation function to use). The full list of parameters,

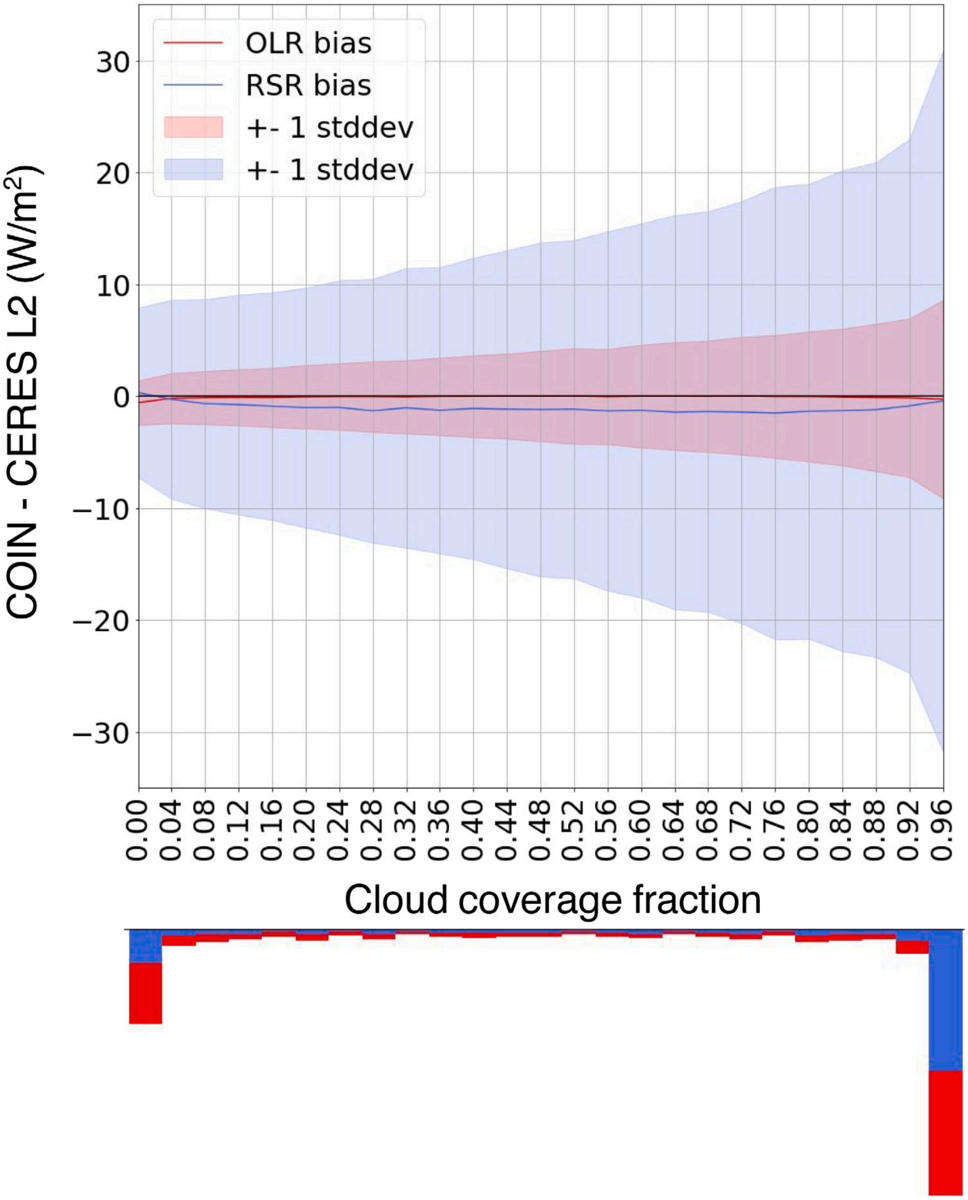

Fig. 9. Comparison between COIN and CERES L2 SSF for different levels of (GOES-16 ABI Level 2 Binary Cloud Mask) cloud cover within the CERES footprint. OLR is in red, RSR is in blue; the shaded areas contain one standard deviation of prediction error on both sides of the (solid line) mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ranges searched, and selected values are in Table 1. The mini-batch size (number of footprints/gridboxes per training step) is 64. After every 100 training steps, the model's mean absolute error (MAE) is evaluated on 100 mini-batches from the validation set. The learning rate is decreased automatically as training progresses by multiplying it with the "learning rate drop factor" whenever the performance on the validation sample does not improve for a set number of training steps (the "learning rate drop patience"). The size of hidden layers of the neural network are set by two hyperparameters: the "first layer size" and the "layer size scale factor". For example, a 4-layer network with first layer size equal to 525 and layer size scale factor of 0.34, has latent layer sizes of 525, 178, 60, and 20.

Dropout (Srivastava et al., 2014) was considered but not found to be used in any of the best performing models, in alignment with the model not showing any evidence of overfitting. Several activation functions were considered (Ramachandran et al., 2017) as well as two possible loss functions: the mean squared error (MSE) and mean absolute error (MAE). The "example weight clip" parameter is the maximum value allowed to be used as an example weight (i.e., multiplied against the training loss for a given example prior to back-propagation). Each example's (unclipped) weight is based on its inverse frequency of occurrence in the training set as described in Section 3.4.

Training COIN in the selected configuration takes about 9 h on a single Nvidia P100 GPU.

4. Results

4.1. Qualitative

Fig. 1 provides a comparison between a CERES Level 3 SYN1deg data product and the corresponding output of COIN. Qualitative agreement in OLR magnitude between CERES Level 3 and COIN can be seen across cloud structures and various surface types, while relatively small-scale features such as contrails and coastlines are only resolved in the output of COIN.

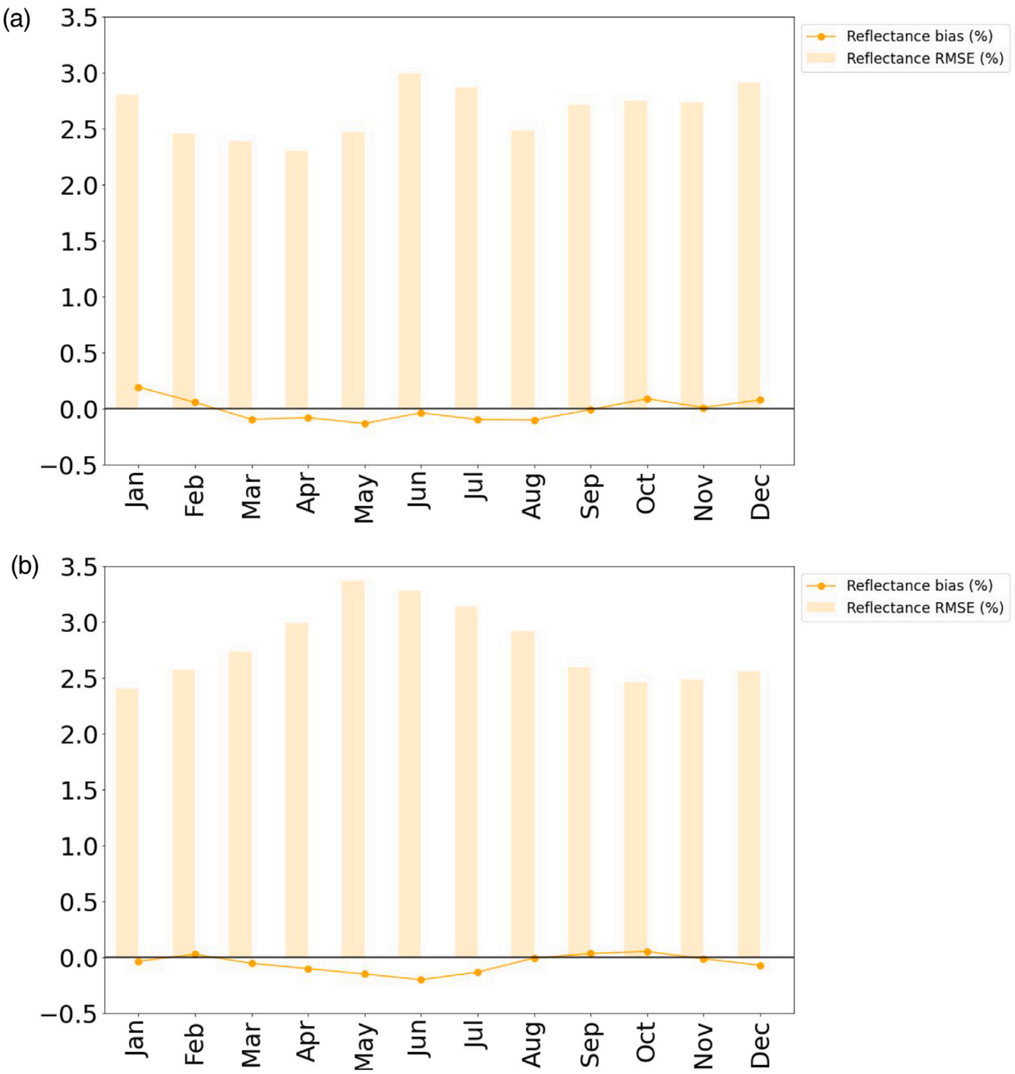

Fig. 10. Comparison of COIN to CERES L2 Reflectance percentage (outgoing flux divided by TOA incoming solar flux), broken out by month for the **(a)** Northern hemisphere and **(b)** Southern hemisphere.

RSR flux estimates which do not observe multiple viewing angles of the same location near the same time must generally account for sun glint, where the sun reflects strongly off of water at certain angles (Kay et al., 2009). COIN does not explicitly model sun glint, instead learning to correct for it based on the CERES flux labels provided when it occurs (Su et al., 2015a; Doelling et al., 2013). Without sufficient training examples having sun glint and correct flux labels, one might expect the model to over-estimate the RSR flux (because the apparent radiances are large for that viewing angle but not isotropically for all angles), but we see no evidence of that error in Fig. 2.

4.2. Comparison to literature models

To provide a more quantitative comparison, we begin with the prediction error mean (bias) and root mean square error (RMSE) on the CERES Level 2 SSF data product for the validation set. Bias in this work is always reported as COIN minus CERES. The results are shown, amongst the existing works in the literature, in Table 2.

While COIN reports the least error, it is worth noting that in general the numbers are not necessarily directly comparable across geostationary satellites because they observe different parts of the Earth, which have different difficulty (e.g. different proportions of land, water and clouds). However, we have compared the COIN-estimated RSR in our dataset to CERES Level 2 footprints in the same Continental United States (CONUS) region over the same 7 hourboxes which were reported on by Pinker et al. (2022), and find the bias and RMSE to be 2.99 and 22.46 W/m^2, which is a substantial and directly comparable improvement over the (Pinker et al., 2022) bias and RMSE of 19.14 and 85.15 W/m^2.

4.3. Seasonality

For assessing radiative effects of climate-relevant phenomena with seasonal covariance, it is important to characterize how COIN's performance varies across the seasons. While the general correlation between COIN and CERES is quite good as seen in Fig. 3, and Fig. 4 shows no strong evidence of COIN bias being covariant with season, we do see a pattern of COIN RSR RMSE being clearly higher in the summer months of each hemisphere.

The seasonality in the RMSE is understandable by inspecting the error characteristics of COIN as a function of OLR and RSR magnitude. Fig. 5 plots this comparison, and shows the clear trend in COIN RSR prediction error standard deviation being higher for larger RSR magnitudes. As can be seen in Fig. 6, the summer months in each hemisphere have mean RSR magnitudes approximately 60–100 W/m^2 higher than the winter months. For example, in the Southern hemisphere in July the mean CERES L2 RSR is 168 W/m^2 but in November it is 271 (an increase of 62%). Returning to Fig. 5, we can see for the 150–200

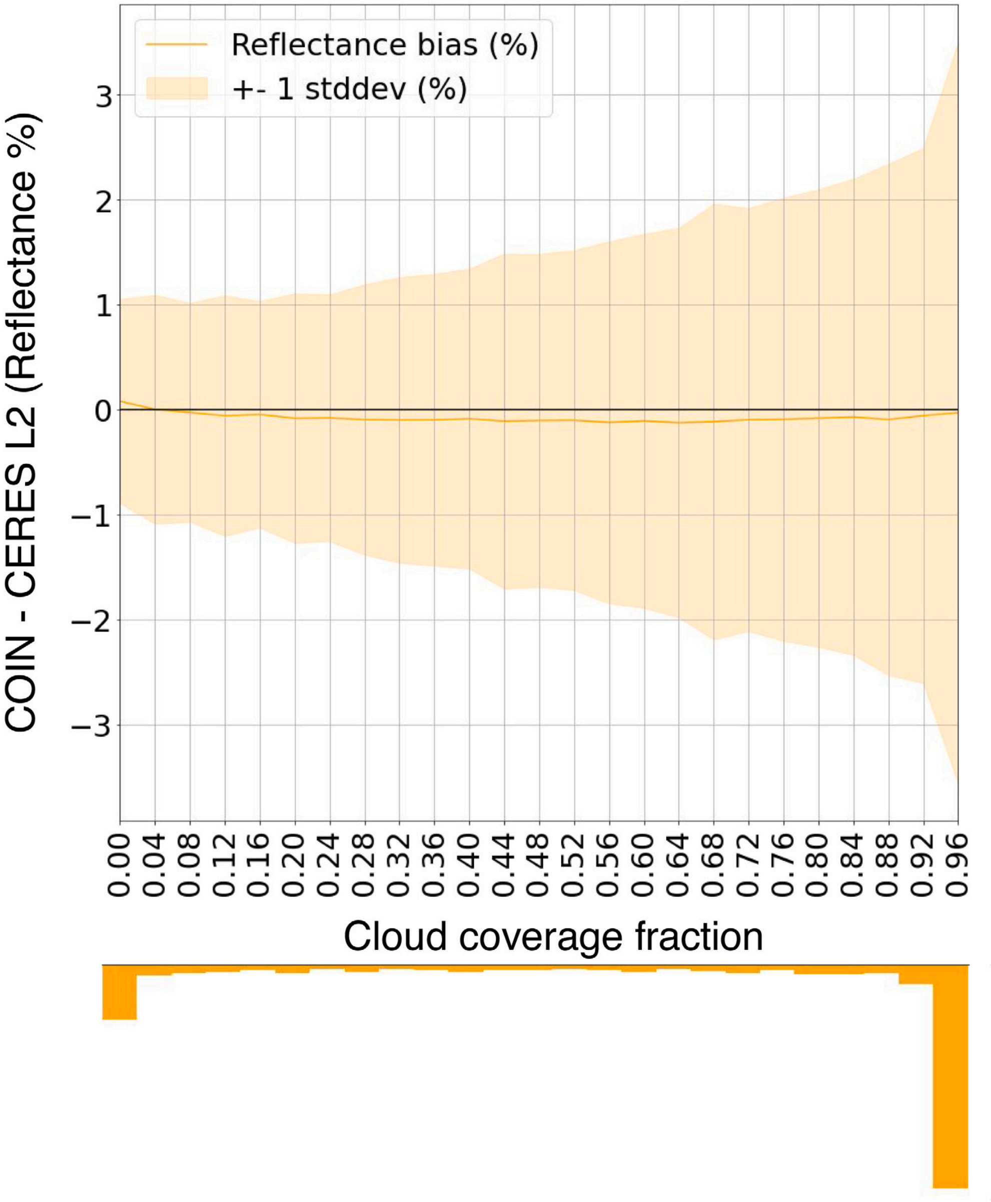

Fig. 11. Comparison between COIN and CERES L2 SSF Reflectance percentage (outgoing flux divided by TOA incoming solar flux), for different levels of cloud cover (GOES-16 ABI Level 2 Binary Cloud Mask) within the CERES footprint. the shaded areas contain one standard deviation of prediction error percentage on both sides of the (solid line) mean.

W/m^2 bucket that COIN's prediction error standard deviation is about 19 W/m^2 while it is about 28 W/m^2 in the 250–300 W/m^2 bucket (an increase of 67%). The trend (in these RSR magnitude ranges) of proportionate increase in COIN RSR uncertainty per increase in RSR magnitude lead us to believe that the seasonal RSR RMSE swings can be addressed by improving COIN certainty on higher-magnitude RSR scenes.

Fig. 5 also shows an almost monotonic trend in prediction bias of COIN compared to CERES L2 flux, for both OLR and RSR. That is, when the CERES measured OLR or RSR is below its respective average in the validation set, COIN over-estimates these fluxes; but when the CERES OLR or RSR is above average, COIN under-estimates these fluxes. While the trend is clear, only the most infrequently-occurring magnitudes are strongly affected: 97.4% of the validation set has an OLR magnitude with less than 2.25 W/m^2 of absolute bias, and 94.1% of the validation set has an RSR magnitude with less than 5.0 W/m^2 of absolute bias.

4.4. Scene types

Our investigation into the performance of COIN on different scene types begins with Fig. 7, where the mean absolute deviations and biases between COIN outputs and CERES L2 footprints are plotted in the latitude/longitude gridbox in which they were observed. We observe similar patterns for both OLR and RSR in this analysis. The lowest deviations and biases for both OLR and RSR occur above ocean surfaces, in areas that less frequently experience clouds (King et al., 2013).

There is a noticeable 'striping' artifact visible in the RSR deviations (upper right panel of Fig. 7). These higher-error regions correlate with areas where our validation dataset contains relatively few closely-raymatched CERES L2 footprints. This phenomena is analyzed in more detail in Figs. 14 and 16 and discussed in Section 5.

Land surfaces generally have higher deviations than water, with continental coastlines clearly visible. This characterization is reinforced

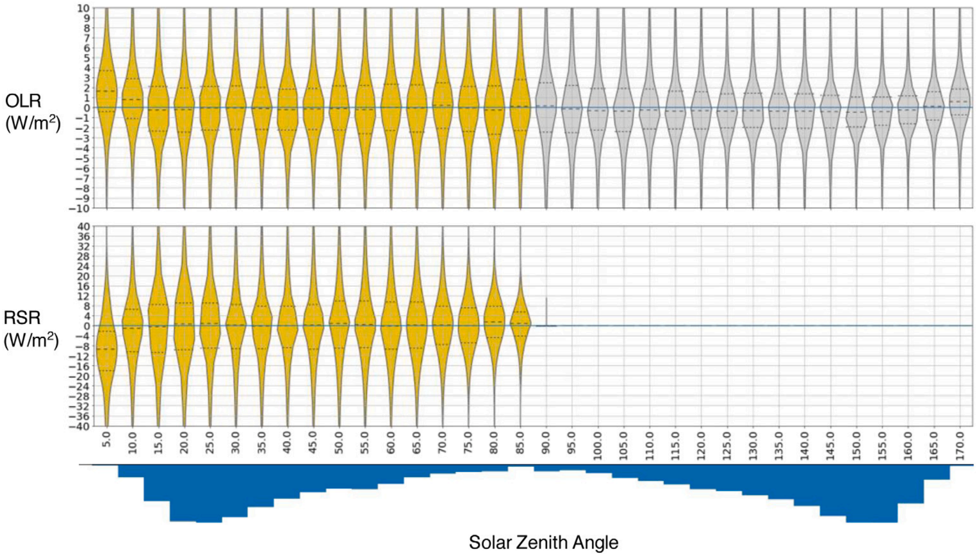

Fig. 12. Comparison between COIN and CERES L2 for discretized bins of solar zenith angle. Daytime violin plots are in yellow, nighttime violin plots are in gray, with the median marked as a dashed line and first and third quartiles as dotted lines. The validation set frequency for solar zenith angles is shown as an upside-down histogram at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

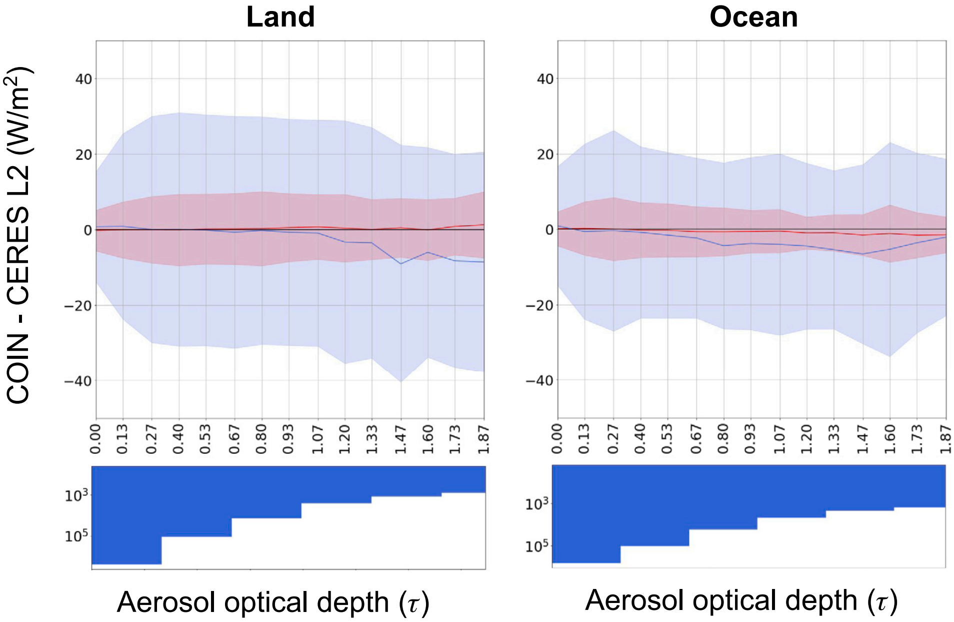

Fig. 13. Comparison between COIN and CERES L2 for discretized bins of aerosol optical depth, on the test set. OLR is in red, RSR is in blue; the shaded areas contain one standard deviation of prediction error on both sides of the (solid line) mean. The test set frequency (in log-scale) for the aerosol optical depths is shown in an upside-down histogram at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

by Fig. 8 which shows the breakdown of COIN compared to CERES L2 by earth surface type. This can again be interpreted as a flux magnitude-driven modeling issue as described in Section 4.3 because land surfaces generally have higher OLR and RSR flux than water. Consistent with Fig. 5 this also affects the bias in COIN; for example CERES L2 footprints with a primarily desert surface type (Barren or Open-Shrubland) are under-predicted by COIN by an average of 0.3055 W/m^2.

We should note that we do not provide COIN with an explicit surface type as a model input, a decision that was made for the operational convenience of not needing to incorporate a separate surface-type dataset to run model inference on future GOES-16 ABI inputs. However, this choice likely has contributed to infrequent surface types (such as Wetlands and Permanent snow) having high absolute biases and RMSE.

Locations which more frequently experience optically thick clouds (including the Atlantic region most strongly influenced by Saharan dust/aerosol outbreaks (Li et al., 2004)) are associated with higher RMSE, as is typical in the other models from the literature which estimate broadband flux from narrowband geostationary measurements. Fig. 9 confirms this unambiguously: COIN has mean RMSE for OLR and RSR (respectively) of only 2.02 and 7.46 W/m^2 in clear-sky L2 footprints but 8.41 and 30.62 W/m^2 in fully overcast footprints. We also see here the same magnitude-driven trend that was visible in Fig. 5, because clear-sky RSR is generally lower than fully overcast scenes: in clear-sky scenes COIN over-estimates CERES L2 RSR by 0.33 W/m^2 and for most partially cloudy footprints COIN under-estimates CERES L2 by about 0.8 W/m^2. However, in fully overcast footprints (despite having higher RSR flux) COIN bias has decreased to −0.12

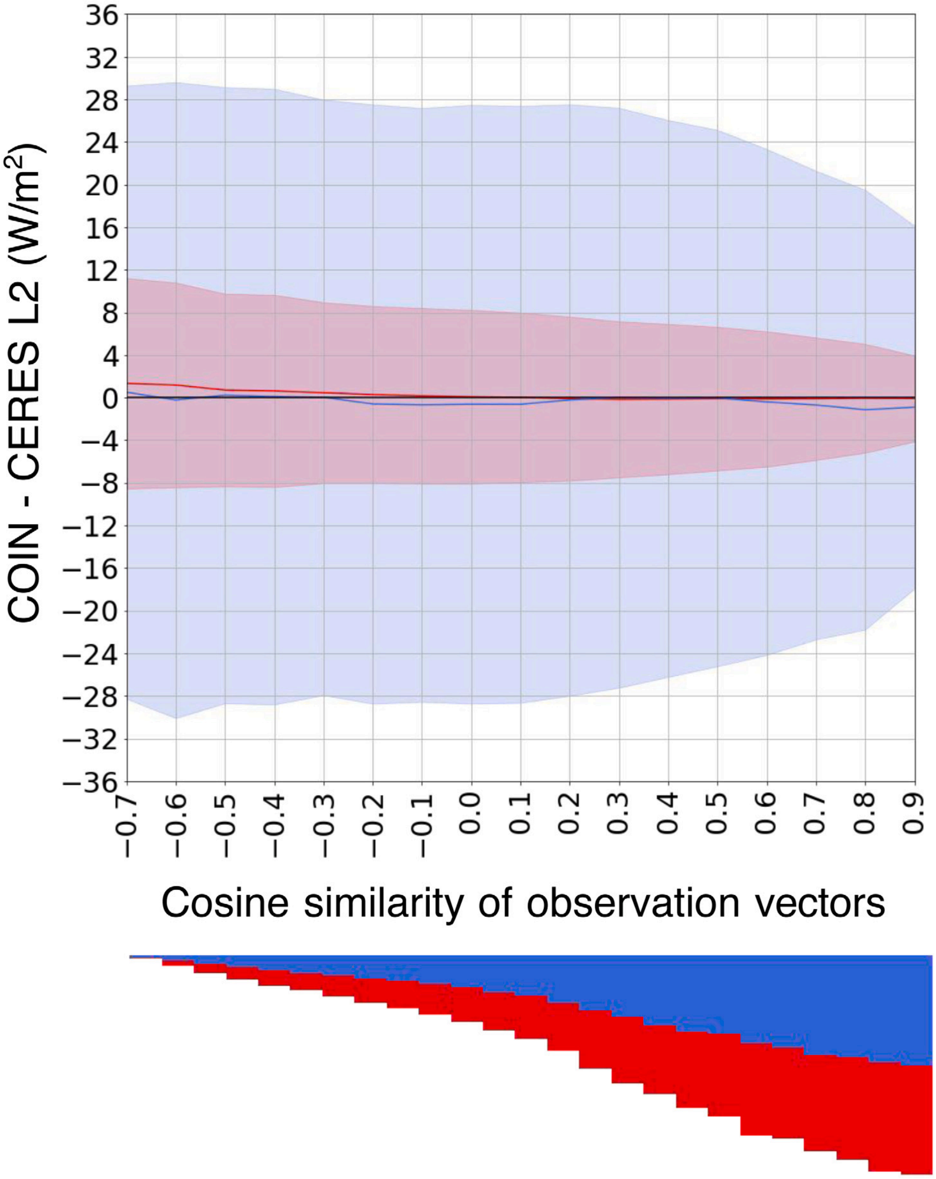

Fig. 14. Comparison between COIN and CERES L2 for discretized bins of cosine similarity of viewing angle between GOES-16 ABI and CERES aboard Terra/Aqua, on the test set. Cosine similarity of 1.0 indicates identical vectors, 0.0 are orthogonal vectors, and −1.0 are opposite vectors. OLR is in red, RSR is in blue; the shaded areas contain one standard deviation of prediction error on both sides of the (solid line) mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

W/m^2, perhaps because it is mitigated by the relative abundance of fully-overcast footprints available in the training data.

There are a number of factors contributing to the challenge of estimating flux in scenes containing clouds, including surface radiance coming through semi-transparent clouds, movement of clouds in between the imager times of collocated imagery, and parallax issues due to non-raymatched viewing angles. See Section 5 for a more extensive discussion of the modeling trade-offs made in this work and worthy of future investigation.

Given we have now seen higher COIN RSR RMSE in summer months as well as overcast footprints, we use Figs. 10 and 11 to clarify whether this increased RMSE is driven by larger solar insolation or by reflective clouds. In these figures we have normalized the COIN-estimated and CERES L2 RSR by dividing by the CERES-provided TOA incoming solar flux, to yield unitless reflectance we report as a percentage. We see that the seasonal trends of Fig. 4 are disrupted, but the higher RMSE in overcast footprints remains, and conclude the RMSE is primarily cloud-driven. Further, this analysis argues that COIN may benefit from insolation being provided as a model input; we had hoped the neural network could implicitly estimate insolation from the provided day-of-year and solar angles if it was beneficial, but perhaps not.

COIN's performance as a function of solar zenith angle (SZA) in Fig. 12 shows no substantial trends except for biases in the most infrequently occurring SZAs in the CERES L2 data due to its sun-synchronous orbit, those less than 10 degrees. Adding in the CERES L3 SYN1deg data to the analysis, in Fig. B.1 we can see that COIN in fact learns these "sun directly overhead" scenes reasonably well. It should be noted that the CERES L3 SYN1deg flux labels are based in part on geostationary radiance inputs including GOES-16 ABI, so it is expected COIN should perform better in general in comparisons to CERES L3 data.

We also note that some CERES footprints having SZA between 90 and 95 degrees do not have zero variance, despite the COIN output being forced to return 0.0 RSR for any solar zenith angle greater than 90 degrees. The reason for this is because Fig. 12 groups footprints/gridboxes by the SZA of their centroid latitude/longitude. The

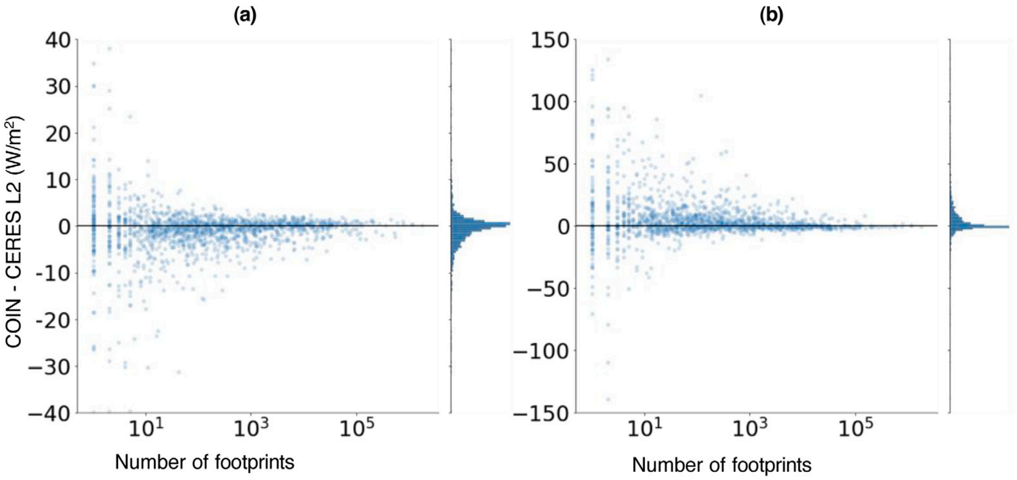

Fig. 15. Mean COIN bias with respect to CERES L2 SSF, as a function of number of footprints with the same CERES "Cloud Classification" code (which also incorporates surface type), in the test set for **(a)** OLR and **(b)** RSR. Note the *x*-axis is log-scale.

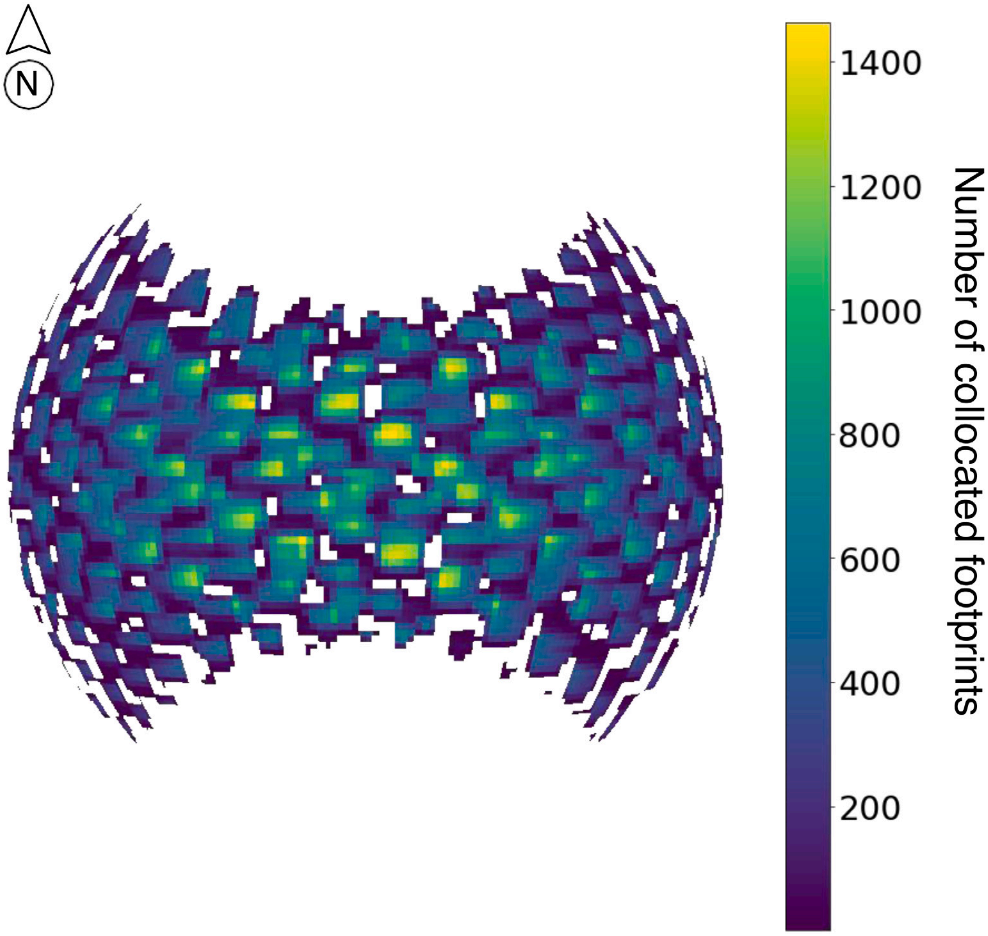

Fig. 16. Spatial distribution (in GOES-16 perspective) of CERES L2 footprints which are ray-matched with > 0.9 cosine similarity to GOES-16 ABI.

COIN model operates on GOES-16 ABI pixels, and a CERES footprint/gridbox contains many ABI pixels: at sunrise or sunset, some may have SZA > 90 while the centroid has SZA ≤ 90.

A comparison between COIN and CERES L2 as a function of aerosol optical depth (CERES PSF-weighted MODIS retrieval at 550 nm) is given in Fig. 13. We see here again the pattern that infrequently-occurring scene types (in this case higher aerosol optical depths) can lead to bias in COIN. Similar to land-surface type not being explicitly passed as a model input, we had hoped that COIN could implicitly learn the effect of aerosols on flux directly from the radiance data without an explicit aerosol retrieval being supplied as an input. We see here that hope is only partially manifest, as footprints over land with aerosol optical depths $\tau > 1$ and over ocean $\tau > 0.5$ are consistently under-estimated by COIN.

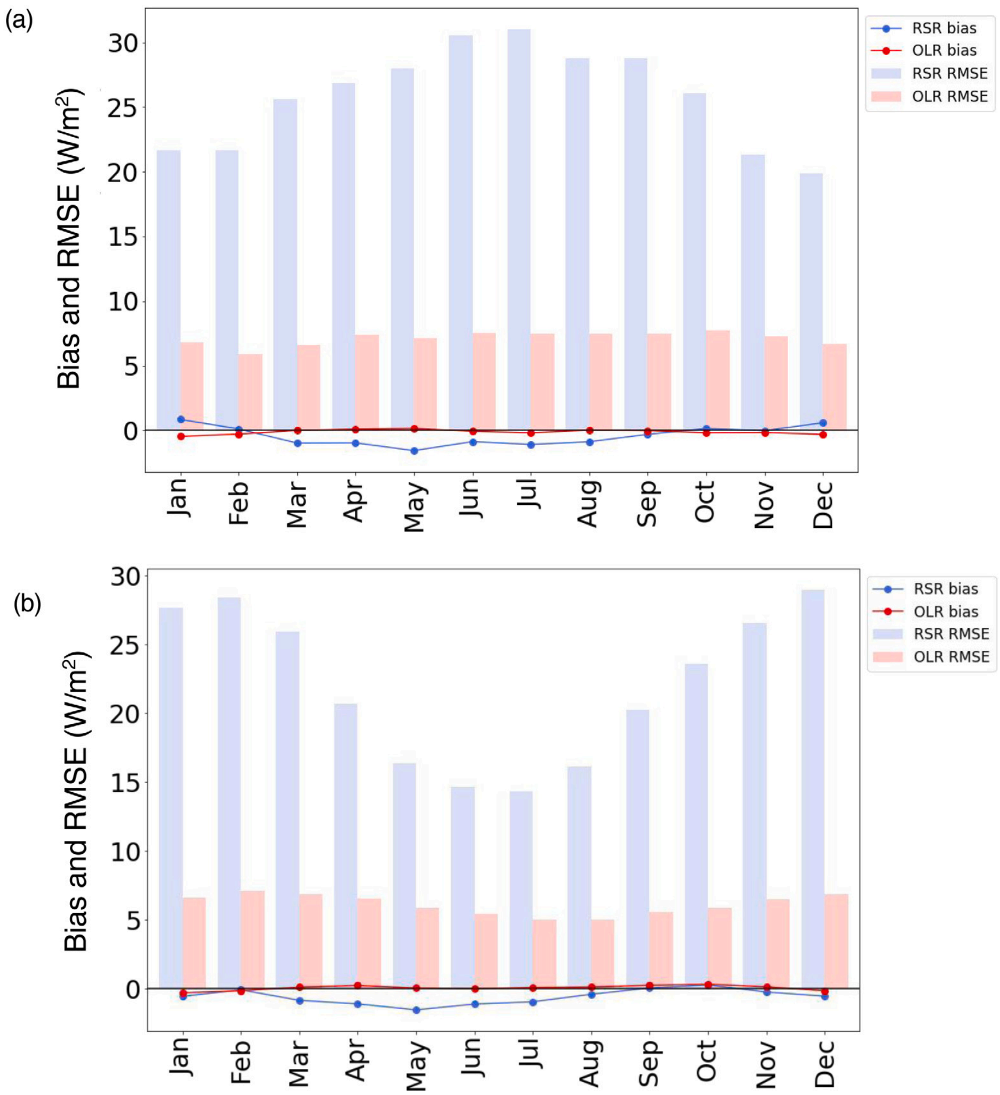

Fig. A.1. Comparison of COIN to CERES L2 on the test set broken out by month for the **(a)** Northern hemisphere and **(b)** Southern hemisphere.

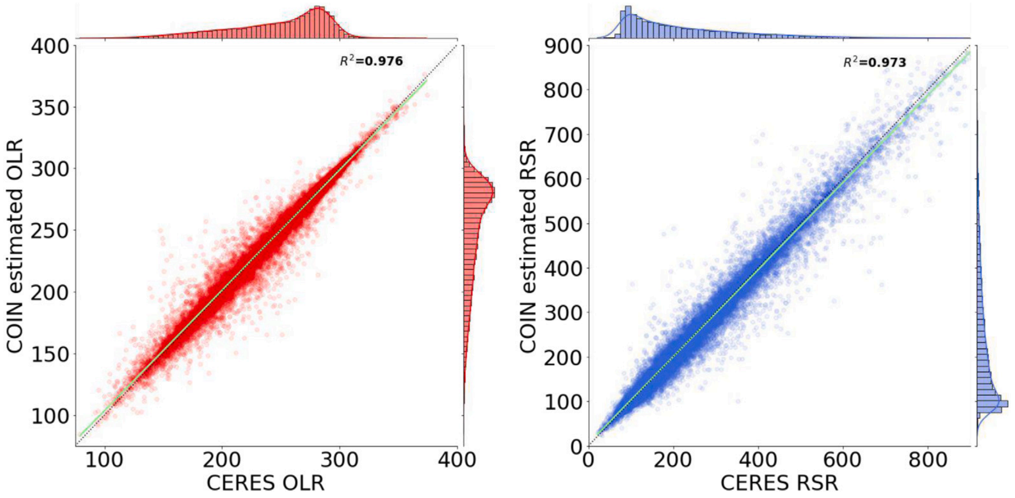

Fig. A.2. Test set comparison between COIN and CERES L2 and L3. The pale green line is a linear fit to the data. The black dotted line corresponds to the 1:1 line. Units are W/m^2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Analysis of COIN estimates as a function of viewing geometry are available in Fig. 14. To collapse the varying viewing geometries to a single dimension, we calculate the cosine similarity between the vector from each satellite (GOES-16 and Terra or GOES-16 and Aqua) to the

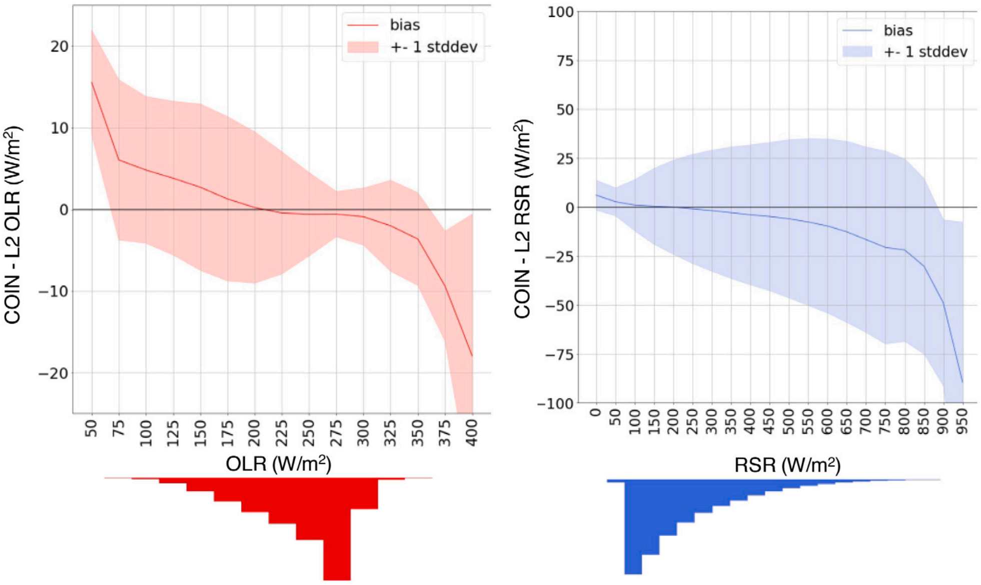

Fig. A.3. Test set comparison between COIN and CERES L2 for bins of OLR and RSR magnitude. The validation set frequency for OLR and RSR magnitudes are shown as upside-down histograms at the bottom. Footprints with solar zenith angle > 90 degrees are not included in the RSR data. The validation set frequency by magnitude is shown as an upside-down histogram at the bottom. In the 2021 test set there were 14 total L2 SSF footprints with OLR in the 50–75 W/m^2 bin, while there were 0 such footprints in the validation set hour boxes across 2018, 2019 and 2020. The validation set frequency by magnitude is shown as an upside-down histogram at the bottom.

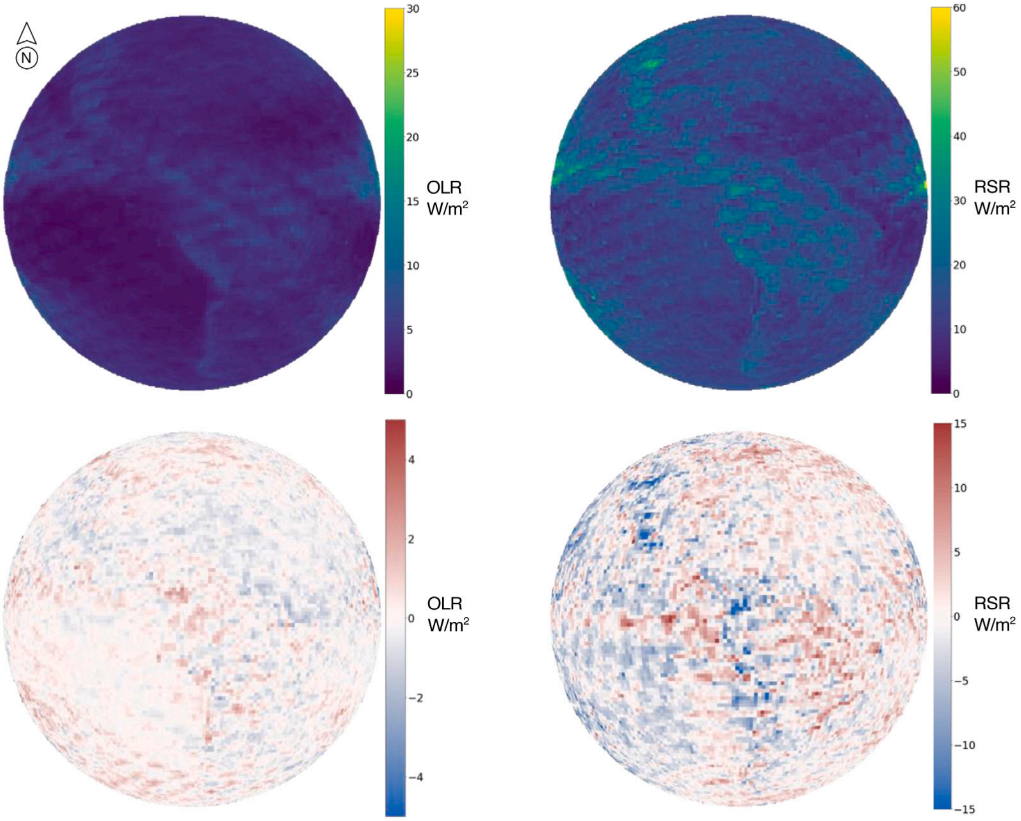

Fig. A.4. Top row: mean absolute deviation between spatio-temporally averaged COIN outputs and CERES L2 data product on the test set for each 1 × 1 latitude and longitude gridbox, across 3 years. Bottom row: mean bias of same.

CERES footprint centroid on the earth's surface that both satellites' sensors are observing. Footprints where the GOES-16 ABI data are well ray-matched (cosine similarity of the observation vectors > 0.9) with the CERES observations report approximately half the RMSE for both OLR and RSR compared to the most poorly ray-matched footprints. This could be caused by a few factors. First, collocated footprints that have clouds would not suffer from parallax effect if they are ray-matched. Second, even in clearsky footprints not subject to parallax effect, when viewed from a different angle than CERES, COIN's estimate of the flux will be subject to compounded uncertainty in cases where the CERES ADM itself has uncertainty, which can be on the order of a few percent (Su et al., 2015a). Finally, the trend is also aligned with pattern

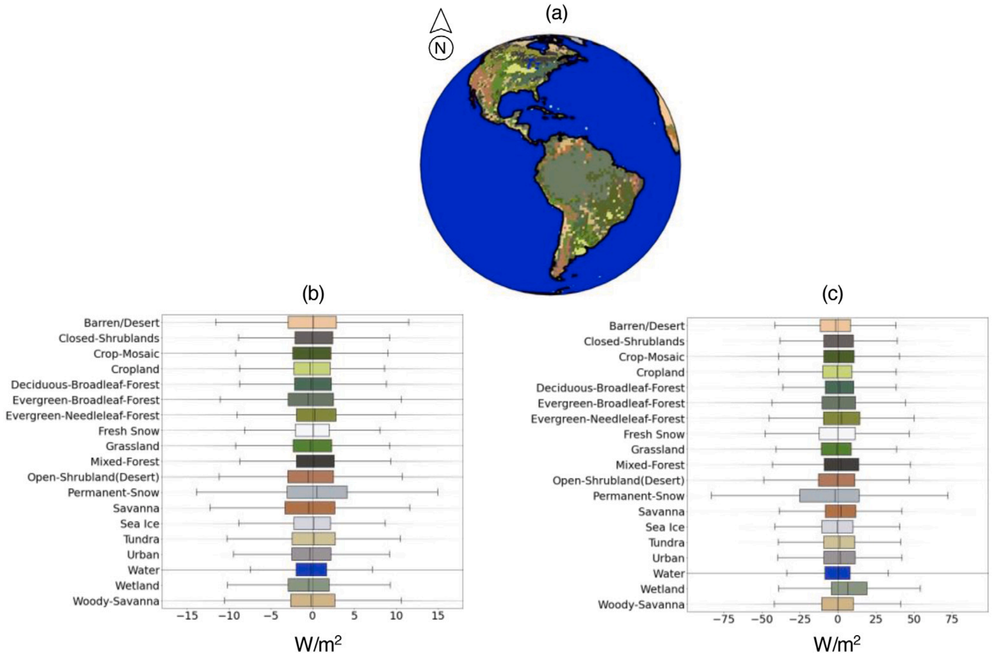

Fig. A.5. Test set comparison between COIN and CERES L2 SSF for the most common surface type of the SSF. **(a)** A plot of CERES earth surface types, color-coded to match the error plots showing **(b)** OLR differences and **(c)** RSR differences.

seen elsewhere that the most frequently-occurring data has the lowest RMSE.

Finally, to further confirm the effect of scene-type frequency on COIN's performance, we render in Fig. 15 COIN's mean bias for each CERES footprint type — according to the CERES "Cloud Classification" code. The Cloud Classification code succinctly represents MODIS cloud retrieval data and also incorporates 10 earth-surface types. It encodes up to two cloud layers of varying coverage fraction, optical thickness and overlap at up to 3 altitude ranges, as described in Ham et al. (2014). The trend in the figure is clear: under-represented scene types are more likely to be subject to bias in COIN flux estimates.

4.5. Test set evaluation

See Appendix A for copies of the tables and figures above applied to the held out year 2021, as a means of assessing the accuracy of COIN when applied to future years of GOES-16 ABI data without being retrained. Based on these tables and figures we do not observe any substantial differences in the performance of COIN on the validation set (years 2018, 2019, and 2020) compared with the test set (2021).

5. Discussion

We have introduced a collocation dataset between GOES-16 ABI and CERES SSF/SYN1deg data products, which we believe to be more accurate than previously reported geostationary/CERES collocation datasets because it is restricted to 120 s of time difference between the satellite sensors (previous works have allowed 300 s) and we utilize a precise calculation of the CERES point spread function: previous works have downsampled and regridded their validation data (Vázquez-Navarro et al., 2013; Lee et al., 2018; Kim and Lee, 2019; Pinker et al., 2022). It is also more complete because we report our analysis on several million collocated footprints that span all seasons, include the full diurnal cycle, and do not omit sunglint.

While comparing COIN to previous works in the literature comes with the caveat that they are mostly evaluated on different fields of view of the Earth, it is promising to see COIN report aggregate RMSE values drop by factors of 1.37 to 1.66 for OLR and 1.52 to 3.32 for RSR (See Table 2). Compared to the sole previous work where we can provide a direct comparison on the same satellite sensors and times (Pinker et al., 2022), COIN estimates of RSR have bias decreased by a factor of 6.4 and RMSE decreased by a factor of 3.8. We believe the improvement shown by COIN is due primarily to it being trained on real satellite observations, so it is less affected by covariate shift and sample selection bias (Shimodaira, 2000; Huang et al., 2006). This design choice is a trade-off we take in exchange for losing some ability to interpret the model and run controlled sensitivity studies on it, which are easier with models trained directly on radiative transfer simulation datasets.

COIN does in fact sometimes learn from radiative transfer simulations indirectly through the CERES data product labels, which depend on angular distribution models (ADMs) that utilize simulations in at least two instances in Edition4 A (Su et al., 2015a): (1) non-glint clear-sky ocean scenes use aerosol optical properties from Hess et al. (1998) to fill un-observed angular bins, and (2) cloudy land scenes have their top of atmosphere RSR distinguished from surface contributions using an analytical ADM with cloud albedo and transmittance coefficients calculated by Fu and Liou (1993). Additionally, researchers studying TOA outgoing irradiance in scenes which may have significant amounts of high-altitude horizontal-path transmission through the atmosphere should be aware that COIN may pass along the < 0.35 W/m^2 flux error due to the fixed global 20 km TOA reference level adopted by CERES rather than scene-specific reference levels (Loeb et al., 2002). These nuances highlight an additional strength of COIN's design: by learning from the CERES data products, we expect to take advantage of improvements in future editions of the CERES data products by simply retraining COIN with the labels from the new edition.

Besides the improvement that CERES editions may bring, there are many avenues to making improvements to COIN that are suggested by the results in Section 4. The highest RMSE is repeatedly seen in the least frequently-occurring atmosphere/surface scene types and OLR/RSR flux magnitudes: this is a classic symptom of machine learning on "imbalanced data". Potential remedies may be found in recent training techniques from the literature such as "label and feature distribution smoothing" (Yang et al., 2021), and post-training calibration methods (Guo et al., 2017). If under-represented scene types still report higher error, one could consider augmenting our collocation dataset

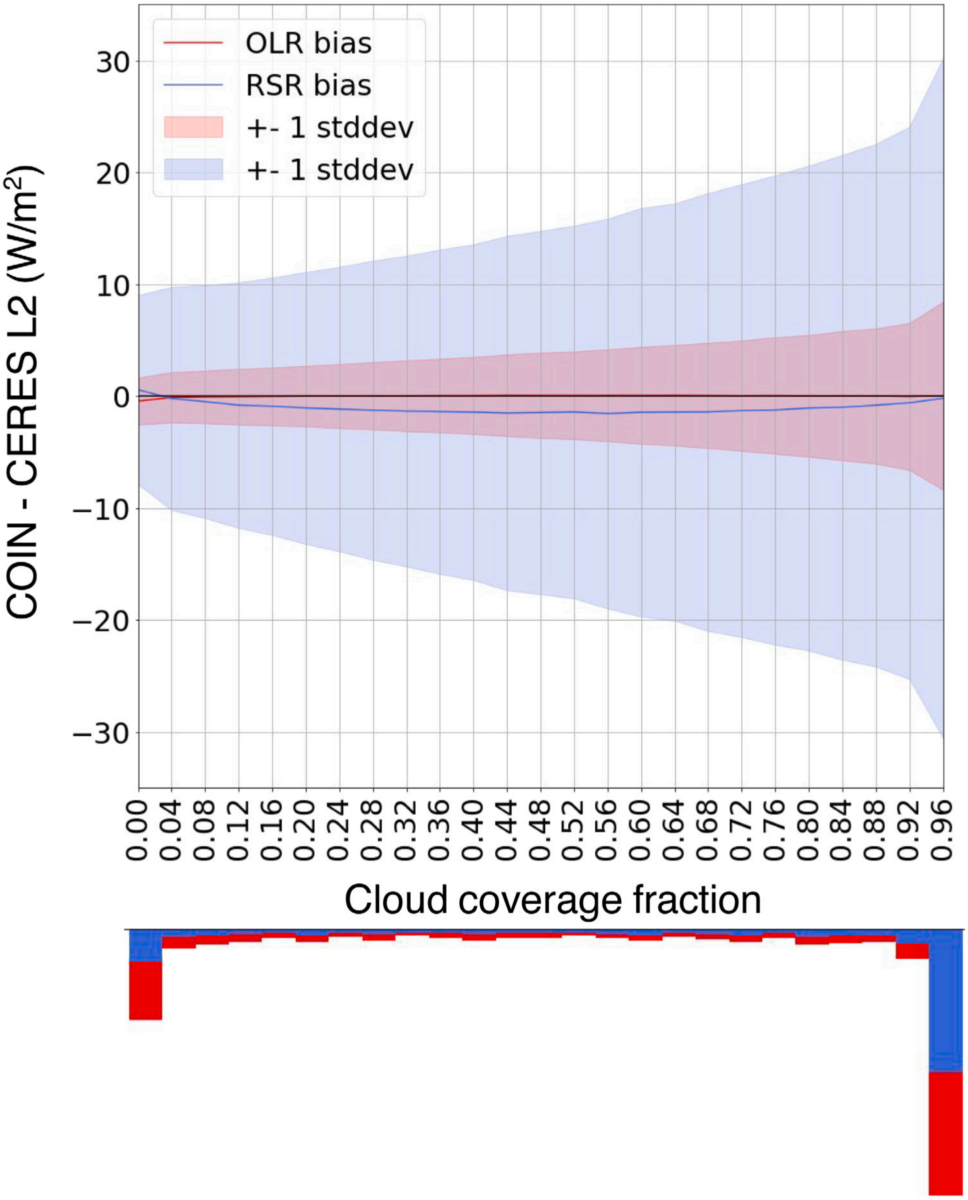

Fig. A.6. Test set comparison between COIN and CERES L2 SSF for different levels of (GOES-16 ABI Level 2 Binary Cloud Mask) cloud cover within the CERES footprint. OLR is in red, RSR is in blue; the shaded areas contain one standard deviation of prediction error on both sides of the (solid line) mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with radiative transfer simulations, while still carefully monitoring our extensive validation set for signs of covariate shift.

There are a number of cases which seem likely to be improved by the addition of auxiliary loss terms while training COIN. Given that the CERES L2 SSF flux labels are calculated as observed broadband radiance divided by a scene-specific anisotropic factor (Equation 2 from Su et al. (2015a)), we suspect COIN can also be improved by regularizing it with an auxiliary loss term that would penalize it for being unable to identify the footprint's ADM scene type from its geostationary narrowband radiances; an auxiliary loss term for broadband radiance observed by CERES (for ray-matched footprints) also seems likely to improve the model RMSE. For improved performance on higher aerosol optical depths and for infrequently observed earth-surface types, we are optimistic that an auxiliary loss term teaching COIN to predict the MODIS-retrieved aerosol optical depth as well as the earth-surface type can benefit under-represented scene types.

Another class of possible improvements are suggested by Fig. 14 which shows COIN has substantially lower RMSE on footprints where the GOES-16 ABI and Terra/Aqua CERES viewing rays are similar to each other. We conjecture this is due at least in part to poorly ray-matched footprints having higher effective CERES flux "label noise" due to parallax effect: for example in partially cloudy conditions GOES-16's viewing ray to the footprint's surface location could observe ground-reflected radiances while CERES' different viewing ray observes cloud-reflected radiances, making it challenging to estimate the CERES flux. It is tempting to consider restricting the dataset to only ray-matched footprints, however this idea is in tension with the dataset's spatial coverage and would likely incur covariate shift — for example Fig. 16 shows the heatmap of valid collocated footprints if they are required to have > 0.9 cosine similarity of viewing angles and still less than 120 s of collocation time difference. We are more optimistic about two other possible ways to address the issue. First, parallax correction could be a mitigation by using a cloud height estimate such as the ABI cloud height product (Heidinger et al., 2020). Second, given we still allow up to 120 s of collocation time difference during which clouds can enter/leave the footprint, even raymatched footprints would suffer

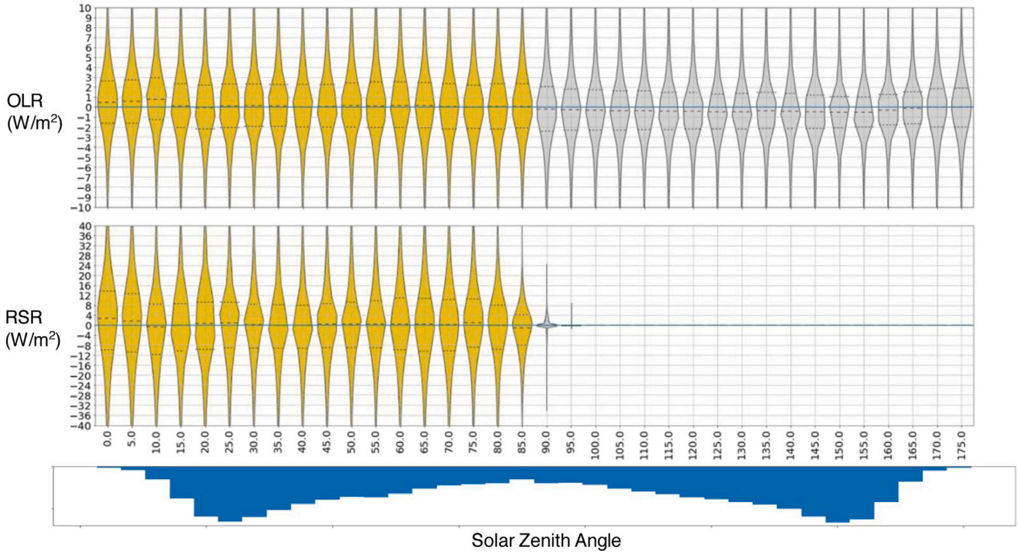

Fig. A.7. Test set comparison between COIN and CERES L2 and L3 for bins of solar zenith angle. Daytime violin plots are in yellow, nighttime violin plots are in gray, with the median marked as a dashed line and first and third quartiles as dotted lines. The validation set frequency for solar zenith angles is shown as an upside-down histogram at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table A.3
Prediction bias and error of broadband flux estimates from narrowband imagers aboard geostationary satellites, which were validated against the CERES L2 SSF product in a separate test set that was not used for training or validation. All values are W/m^2.

Satellite	OLR bias	RMSE	RSR bias	RMSE	Test set
GOES-16[a]	−0.05	6.26	−0.48	23.86	2021

[a] COIN (this work).

label noise, so techniques from the label noise literature such as early stopping (Li et al., 2020) seem prudent to apply.

The 120 s of collocation time difference we allow here was chosen because it is as low as we can go without damaging spatial coverage — there are covariances in the orbit/scanning pattern of CERES on Terra/Aqua and the scan timing of GOES-16 ABI. It may be informative to determine how low RMSE can be pushed for any given neural network architecture in the absence of viewing-ray and temporal mismatch: using MODIS radiances as input to such a neural network and CERES flux as output labels, footprints are trivially collocated with perfect ray-matching and zero collocation time difference given the instruments operate in tandem on Terra/Aqua.

In future works we also look forward to applying COIN to other geostationary satellite sensors such as the Advanced Himawari Imager (AHI) aboard the Himawari 8, and the Spinning Enhanced Visible and InfraRed Imager (SEVIRI) aboard Meteosat Second Generation satellites, because they have similar spectral bands and their data can be collocated with CERES measurements.

6. Conclusion

We have published alongside this paper a collocation dataset between GOES-16 ABI and CERES SSF/SYN1deg data products, which can be used for broadband flux estimation from narrowband radiances. We believe it is the most accurate and complete that has been reported to date because it has the shortest collocation time difference, spans all seasons, includes the full diurnal cycle, and does not omit sunglint regions.

On this dataset, we have developed a neural network that directly back-propagates error gradients through the CERES point spread function. This avoids the need for a bespoke radiative transfer simulation dataset and mitigates covariate shift, achieving substantially lower bias and RMSE than previous broadband flux estimates made from geostationary narrowband radiances.

Through extensive analysis presented here, we have identified some remaining weaknesses in our modeling approach, and have highlighted several techniques to further improve the flux estimates.

With the model inference outputs we provide for the years 2018, 2019, 2020 and 2021, climate researchers can analyze important climate forcers visible to GOES-16 with new levels of precision and accuracy.

CRediT authorship contribution statement

Kevin McCloskey: Conceptualization, Software, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Sixing Chen:** Software, Writing – review & editing, Visualization. **Vincent R. Meijer:** Conceptualization, Software, Writing – review & editing, Visualization. **Joe Yue-Hei Ng:** Software, Writing – review & editing. **Geoff Davis:** Software, Visualization. **Carl Elkin:** Software. **Christopher Van Arsdale:** Software, Supervision. **Scott Geraedts:** Conceptualization, Software, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Some authors are employees of Google Inc as noted in their author affiliations. Google is a technology company that sells machine learning services as part of its business.

Data availability

The data presented in this study are openly available in Google Cloud Storage at gs://upwelling_irradiance/.

The code to generate CERES SSF point spread function weights, and examples of reading the training data and model outputs, are available at https://github.com/google-research/google-research/tree/master/collocated_irradiance_network/.

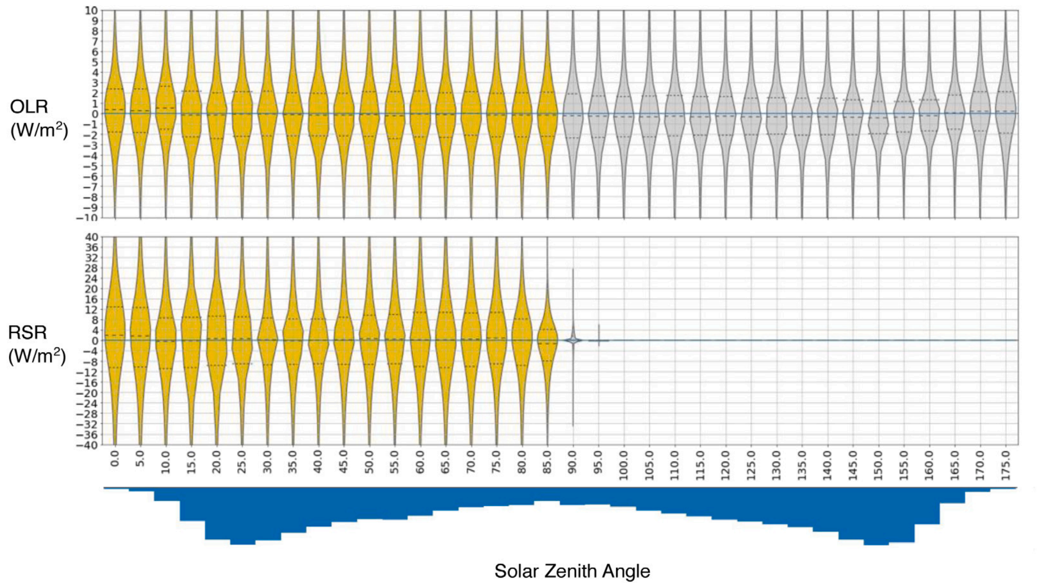

Fig. B.1. Comparison between COIN and CERES L2 and L3 for discretized bins of solar zenith angle. Daytime violin plots are in yellow, nighttime violin plots are in gray, with the median marked as a dashed line and first and third quartiles as dotted lines. The validation set frequency for solar zenith angles is shown as an upside-down histogram at the bottom. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

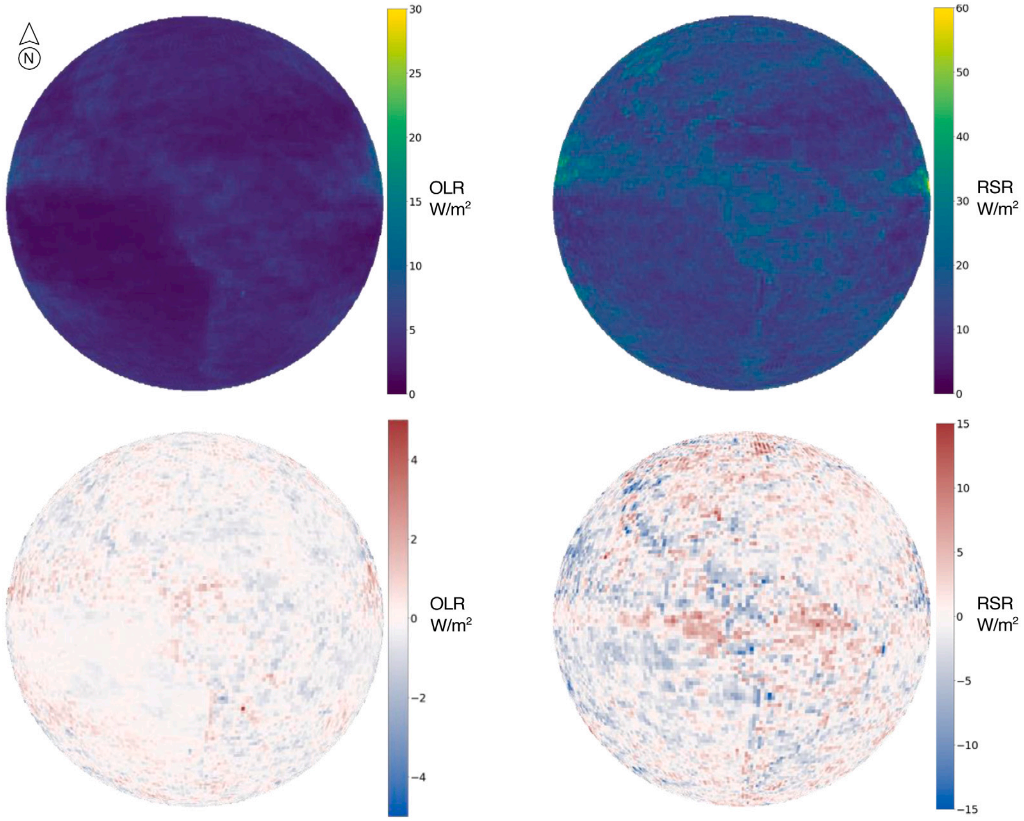

Fig. B.2. Top row: mean absolute deviation between spatio-temporally averaged COIN outputs and CERES L2 and L3 data product for each 1 × 1 latitude and longitude gridbox, across 3 years. Bottom row: mean bias of same.

Acknowledgments

Contributions from V.R.M. were partially supported by NASA, USA research grant 80NSSC19K0943. CERES data were obtained from the NASA Langley Research Center Atmospheric Science Data Center.

Appendix A. Test set evaluation, the year 2021

See Table A.3 and Figs. A.1–A.7.

Appendix B. Selected CERES L3 comparisons

See Figs. B.1–B.3.

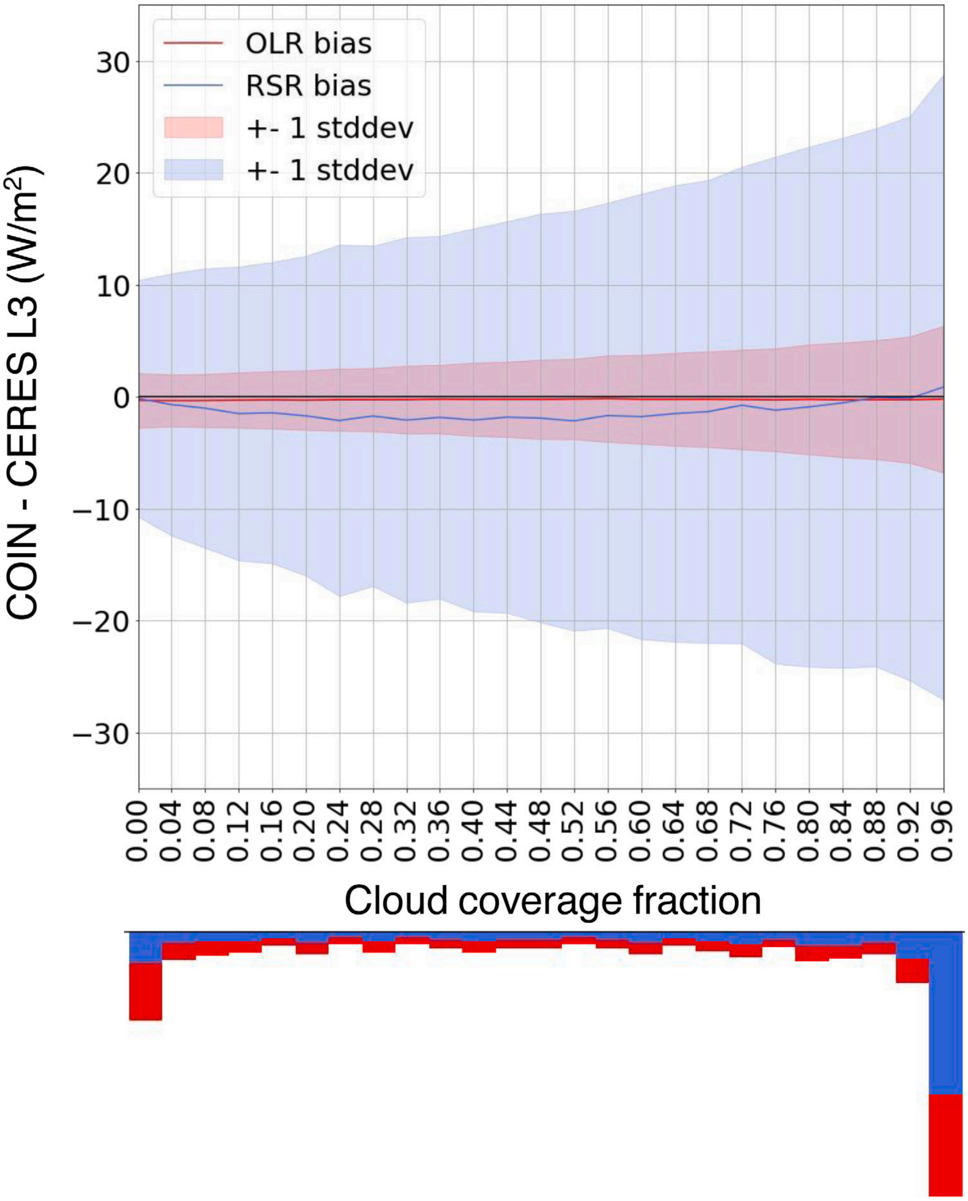

Fig. B.3. Comparison between COIN and CERES L3 SYN1Deg for different levels of (GOES-16 ABI Level 2 Binary Cloud Mask) cloud cover within the CERES gridbox. OLR is in red, RSR is in blue; the shaded areas contain one standard deviation of prediction error on both sides of the (solid line) mean. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Aminou, Donny Maladji A., Jacquet, Bernard, Pasternak, Frederick, 1997. Characteristics of the Meteosat Second Generation (MSG) radiometer/imager: SEVIRI. In: Sensors, Systems, and Next-Generation Satellites. 3221, SPIE, pp. 19–31.

Berk, Alexander, Bernstein, Lawrence S., Robertson, David C., 1987. MODTRAN: A moderate resolution model for LOWTRAN. Technical report, Spectral Sciences Inc Burlington MA.

Cescatti, Alessandro, Marcolla, Barbara, Vannan, Suresh K. Santhana, Pan, Jerry Yun, Román, Miguel O., Yang, Xiaoyuan, Ciais, Philippe, Cook, Robert B., Law, Beverly E., Matteucci, Giorgio, et al., 2012. Intercomparison of MODIS albedo retrievals and in situ measurements across the global FLUXNET network. Remote Sens. Environ. 121, 323–334.

Cintineo, John L., Pavolonis, Michael J., Sieglaff, Justin M., Wimmers, Anthony, Brunner, Jason, Bellon, Willard, 2020. A deep-learning model for automated detection of intense midlatitude convection using geostationary satellite images. Weather Forecast. 35 (6), 2567–2588.

Doelling, D.R., Loeb, N.G., Keyes, D.F., Nordeen, M.L., Morstad, D., Nguyen, C., Wielicki, B., Young, D.F., Sun, M., 2013. Geostationary enhanced temporal interpolation for CERES flux products. J. Atmos. Ocean. Technol. 30 (6), 1072–1090. http://dx.doi.org/10.1175/JTECH-D-12-00136.1.

Doelling, David R., Sun, Moguo, Nguyen, Le Trang, Nordeen, Michele L., Haney, Conor O., Keyes, Dennis F., Mlynczak, Pamela E., 2016. Advances in geostationary-derived longwave fluxes for the CERES synoptic (SYN1deg) product. J. Atmos. Ocean. Technol. 33 (3), 503–521.

Fu, Qiang, Liou, K. No, 1993. Parameterization of the radiative properties of cirrus clouds. J. Atmos. Sci. 50 (13), 2008–2025.

Golovin, Daniel, Solnik, Benjamin, Moitra, Subhodeep, Kochanski, Greg, Karro, John, Sculley, David, 2017. Google vizier: A service for black-box optimization. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1487–1495.

Green, Richard N., Wielicki, Bruce A., 1997. Convolution of imager cloud properties with CERES footprint point spread function (subsystem 4.4). URL http://eospso.nasa.gov/sites/default/files/atbd/atbd-cer-09.pdf.

Guo, Chuan, Pleiss, Geoff, Sun, Yu, Weinberger, Kilian Q., 2017. On calibration of modern neural networks. In: International Conference on Machine Learning. PMLR, pp. 1321–1330.

Gupta, P., Joiner, J., Vasilkov, A., Bhartia, P.K., 2016. Top-of-the-atmosphere shortwave flux estimation from satellite observations: an empirical neural network approach applied with data from the A-train constellation. Atmos. Meas. Tech. 9 (7), 2813–2826. http://dx.doi.org/10.5194/amt-9-2813-2016, URL https://amt.copernicus.org/articles/9/2813/2016/.

Ham, Seung-Hee, Sun-Mack, Sunny, Rose, Fred G., Chen, Yan, Mlynczak, Pamela E., 2014. Variable descriptions of the A-train integrated CALIPSO, CloudSat, CERES, and MODIS merged product (CCCM or C3M).

Heidinger, Andrew K., Pavolonis, Michael J., Calvert, Corey, Hoffman, Jay, Nebuda, Sharon, Straka III, William, Walther, Andi, Wanzong, Steven, 2020. ABI cloud products from the GOES-R series. In: The GOES-R Series. Elsevier, pp. 43–62.

Hess, Michael, Koepke, Peter, Schult, Ingrid, 1998. Optical properties of aerosols and clouds: The software package OPAC. Bull. Am. Meteorol. Soc. 79 (5), 831–844.

Huang, Jiayuan, Gretton, Arthur, Borgwardt, Karsten, Schölkopf, Bernhard, Smola, Alex, 2006. Correcting sample selection bias by unlabeled data. Adv. Neural Inf. Process. Syst. 19.

Huang, Guanghui, Li, Zhanqing, Li, Xin, Liang, Shunlin, Yang, Kun, Wang, Dongdong, Zhang, Yi, 2019. Estimating surface solar irradiance from satellites: Past, present, and future perspectives. Remote Sens. Environ. 233, 111371.

Joyce, Robert, Janowiak, John, Huffman, George, 2001. Latitudinally and seasonally dependent zenith-angle corrections for geostationary satellite IR brightness temperatures. J. Appl. Meteorol. Climatol. 40 (4), 689–703.

Kalluri, Satya, Alcala, Christian, Carr, James, Griffith, Paul, Lebair, William, Lindsey, Dan, Race, Randall, Wu, Xiangqian, Zierk, Spencer, 2018. From photons to pixels: processing data from the advanced baseline imager. Remote Sens. 10 (2), 177.

Kay, Susan, Hedley, John D., Lavender, Samantha, 2009. Sun glint correction of high and low spatial resolution images of aquatic scenes: A review of methods for visible and near-infrared wavelengths. Remote Sens. 1 (4), 697–730.

Kim, Bu-Yo, Lee, Kyu-Tae, 2019. Using the himawari-8 AHI multi-channel to improve the calculation accuracy of outgoing longwave radiation at the top of the atmosphere. Remote Sens. (ISSN: 2072-4292) 11 (5), http://dx.doi.org/10.3390/rs11050589, URL https://www.mdpi.com/2072-4292/11/5/589.

King, Michael D., Platnick, Steven, Menzel, W. Paul, Ackerman, Steven A., Hubanks, Paul A., 2013. Spatial and temporal distribution of clouds observed by MODIS onboard the terra and aqua satellites. IEEE Trans. Geosci. Remote Sens. 51 (7), 3826–3852.

Lee, Sang-Ho, Kim, Bu-Yo, Lee, Kyu-Tae, Zo, Il-Sung, Jung, Hyun-Seok, Rim, Se-Hun, 2018. Retrieval of reflected shortwave radiation at the top of the atmosphere using himawari-8/AHI data. Remote Sens. (ISSN: 2072-4292) 10 (2), http://dx.doi.org/10.3390/rs10020213, URL https://www.mdpi.com/2072-4292/10/2/213.

Li, Mingchen, Soltanolkotabi, Mahdi, Oymak, Samet, 2020. Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks. In: International Conference on Artificial Intelligence and Statistics. PMLR, pp. 4313–4324.

Li, F., Vogelmann, A.M., Ramanathan, V., 2004. Saharan dust aerosol radiative forcing measured from space. J. Clim. 17 (13), 2558–2571.

Li, Zhanqing, Whitlock, Charles H., Charlock, Thomas P., 1995. Assessment of the global monthly mean surface insolation estimated from satellite measurements using global energy balance archive data. J. Clim. 8 (2), 315–328.

Liang, Shunlin, Wang, Dongdong, He, Tao, Yu, Yunyue, 2019. Remote sensing of earth's energy budget: Synthesis and review. Int. J. Digit. Earth 12 (7), 737–780.

Loeb, Norman G., Kato, Seiji, Wielicki, Bruce A., 2002. Defining top-of-the-atmosphere flux reference level for Earth radiation budget studies. J. Clim. 15 (22), 3301–3309.

Loeb, Norman G., Manalo-Smith, Natividad, Su, Wenying, Shankar, Mohan, Thomas, Susan, 2016. CERES top-of-atmosphere Earth radiation budget climate data record: Accounting for in-orbit changes in instrument calibration. Remote Sens. 8 (3), 182.

Mayer, Bernhard, Kylling, Arve, 2005. The libradtran software package for radiative transfer calculations-description and examples of use. Atmos. Chem. Phys. 5 (7), 1855–1877.

Meijer, Vincent R., Kulik, Luke, Eastham, Sebastian D., Allroggen, Florian, Speth, Raymond L., Karaman, Sertac, Barrett, Steven R.H., 2022. Contrail coverage over the United States before and during the COVID-19 pandemic. Environ. Res. Lett. 17 (3), 034039.

Minnis, Patrick, Sun-Mack, Szedung, Chen, Yan, Khaiyer, Mandana M., Yi, Yuhong, Ayers, J. Kirk, Brown, Ricky R., Dong, Xiquan, Gibson, Sharon C., Heck, Patrick W., et al., 2011a. CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part II: Examples of average results and comparisons with other data. IEEE Trans. Geosci. Remote Sens. 49 (11), 4401–4430.

Minnis, Patrick, Sun-Mack, Szedung, Young, David F., Heck, Patrick W., Garber, Donald P., Chen, Yan, Spangenberg, Douglas A., Arduini, Robert F., Trepte, Qing Z., Smith, William L., et al., 2011b. CERES edition-2 cloud property retrievals using TRMM VIRS and Terra and Aqua MODIS data—Part I: Algorithms. IEEE Trans. Geosci. Remote Sens. 49 (11), 4374–4400.

Minnis, Patrick, Trepte, Qing Z., Sun-Mack, Szedung, Chen, Yan, Doelling, David R., Young, David F., Spangenberg, Douglas A., Miller, Walter F., Wielicki, Bruce A., Brown, Ricky R., et al., 2008. Cloud detection in nonpolar regions for CERES using TRMM VIRS and Terra and Aqua MODIS data. IEEE Trans. Geosci. Remote Sens. 46 (11), 3857–3884.

Pinker, Rachel T., Ma, Yingtao, Chen, Wen, Laszlo, Istvan, Liu, Hongqing, Kim, Hye-Yun, Daniels, Jaime, 2022. Top-of-the-atmosphere reflected shortwave radiative fluxes from GOES-r. Atmos. Meas. Tech. 15 (17), 5077–5094.

Ramachandran, Prajit, Zoph, Barret, Le, Quoc V., 2017. Searching for activation functions. arXiv preprint arXiv:1710.05941.

Ricchiazzi, Paul, Yang, Shiren, Gautier, Catherine, Sowle, David, 1998. SBDART: A research and teaching software tool for plane-parallel radiative transfer in the Earth's atmosphere. Bull. Am. Meteorol. Soc. 79 (10), 2101–2114.

Rumelhart, David E., Hinton, Geoffrey E., Williams, Ronald J., 1985. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science.

Schmit, Timothy J., Griffith, Paul, Gunshor, Mathew M., Daniels, Jaime M., Goodman, Steven J., Lebair, William J., 2017. A closer look at the ABI on the GOES-R series. Bull. Am. Meteorol. Soc. 98 (4), 681–698.

Schreier, M., Joxe, L., Eyring, V., Bovensmann, H., Burrows, J.P., 2010. Ship track characteristics derived from geostationary satellite observations on the west coast of southern Africa. Atmos. Res. 95 (1), 32–39.

Shimodaira, Hidetoshi, 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Statist. Plann. Inference 90 (2), 227–244.

Srivastava, Nitish, Hinton, Geoffrey, Krizhevsky, Alex, Sutskever, Ilya, Salakhutdinov, Ruslan, 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Su, W., Corbett, J., Eitzen, Z., Liang, L., 2015a. Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: Methodology. Atmos. Meas. Tech. 8 (2), 611–632.

Su, W., Corbett, J., Eitzen, Z., Liang, L., 2015b. Next-generation angular distribution models for top-of-atmosphere radiative flux calculation from CERES instruments: Validation. Atmos. Meas. Tech. 8 (8), 3297–3313.

Vázquez-Navarro, M., Mayer, B., Mannstein, H., 2013. A fast method for the retrieval of integrated longwave and shortwave top-of-atmosphere upwelling irradiances from MSG/SEVIRI (RRUMS). Atmos. Meas. Tech. 6 (10), 2627–2640. http://dx.doi.org/10.5194/amt-6-2627-2013, URL https://amt.copernicus.org/articles/6/2627/2013/.

Wielicki, Bruce A., Barkstrom, Bruce R., Harrison, Edwin F., Lee III, Robert B., Smith, G. Louis, Cooper, John E., 1996. Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment. Bull. Am. Meteorol. Soc. 77 (5), 853–868.

Yang, Yuzhe, Zha, Kaiwen, Chen, Yingcong, Wang, Hao, Katabi, Dina, 2021. Delving into deep imbalanced regression. In: International Conference on Machine Learning. PMLR, pp. 11842–11851.