Reviewer #2 Responses

This study presents an interesting data-driven approach, combining satellite observations, neural networks, and Monte Carlo uncertainty estimation, to derive Arctic absorbing aerosol direct radiative forcing (ADRF) trends over a 15-year period. The topic is highly relevant and timely, given the sensitivity of the Arctic climate. The manuscript is generally clear, but several areas require improvement to strengthen its scientific rigor and clarity.

Response: We thank the reviewer for constructive comments

1. *Data quality is critical for this analysis. Although the authors utilize several satellite products, many of these have primarily been validated over low- to mid-latitudes. Thorough validation over the Arctic region is necessary. More importantly, uncertainties associated with cloud, aerosol, and surface classification must be quantified. How reliable are the cloud-free and aerosol-free conditions as defined? Similar validation is needed for other aerosol and cloud products.*

   RESPONSE: We thank the reviewer for the comments. Regarding the quality of the surface classification dataset, we have added the following paragraph to Section 2.4 describing the accuracy of the SSMIS sea ice concentration dataset:

   "The SSMIS SIC dataset used in this study is one of two key SIC datasets provided by the NSIDC and has been used extensively in the scientific community to study Arctic sea ice trends. The algorithm used in the dataset, developed by NASA (Cavalieri et al., 1984), has been included in several SIC validation studies (Cavalieri et al., 1992; Ivanova et al., 2015; Kern et al., 2019, 2020; Meier, 2005; Steffen and Schweiger, 1991). Overall, and as reported in the NSIDC dataset user guide (https://nsidc.org/sites/default/files/documents/user-guide/nsidc-0051-v002-userguide.pdf), errors in the SIC dataset are less than 5% in the wintertime but can be as large as 15% in the summertime (Cavalieri et al., 1992). Some recent studies have reported that the SIC dataset may underestimate SIC by up to 10% (Kern et al., 2019, 2020), with the underestimation being partly caused by surface melt ponds in the summer months (Steffen and Schweiger, 1991). Additionally, microwave-based sea ice concentrations have been found to be sensitive to areas of thin ice (Ivanova et al., 2015). Nevertheless, despite some limitations, the algorithm and associated SIC dataset are widely used to represent Arctic SIC."

As for the MODIS cloud dataset used in the study, we note that additional checks are already included in MODIS cloud masking over polar regions, including the use of the observed radiance difference between the 6.7 and 11 μm channels. Still, there are known issues in cloud masking over the polar regions, especially associated with misclassification of snow and ice surfaces or smoke plumes as clouds (e.g. Fig. 2). It is for this reason that additional checks, such as the use of OMI AI and MODIS observations at 2.1 μm were included to mediate the issues (e.g. as shown in Fig. 2).

Regarding the quality of the aerosol information, we describe the methods by which we mitigate uncertainty in the OMI UVAI dataset in Section 2.1. Sorenson et al. (2023) developed a "perturbing method" to remove systematic biases and uncertainties in Arctic OMI UVAI data for use in quantitative Arctic aerosol studies. This method removes substantial viewing geometry and surface condition-related biases and uncertainties in the Arctic OMI UVAI data, and in addition to removing row anomaly-related uncertainty in the data following Sorenson et al. (2023), the OMI UVAI dataset is prepped for analysis in this study.

Regarding validation of the CERES data over the Arctic region, we have added the following paragraph to the end of Section 2.2 to describe the validation of CERES data in the Arctic region:

"CERES data have been used extensively for investigating changes in Arctic radiative energy budgets for both TOA (Duncan et al., 2020; Kay and L'Ecuyer, 2013; Riihelä et al., 2013) as well as the surface (Boeke and Taylor, 2016; Hegyi and Taylor, 2017). Previous studies have also worked to validate Arctic CERES surface radiative fluxes (Di Biagio et al., 2021; Riihelä et al., 2017) and TOA fluxes (Taylor et al., 2022), with the latter seeking to validate CERES TOA radiative fluxes against aircraft-based upwelling radiative flux observations. While Taylor et al. (2022) noted some error in the Arctic CERES Level-2 SSF TOA upwelling SWF resulting largely from errors in the imager-based sea ice concentrations used in the scene classification, the CERES observations compared well overall with the aircraft observations (differences between the CERES and aircraft observations were within 2σ uncertainty). The authors concluded that CERES TOA radiative flux data are suitable for polar climate studies (Taylor et al., 2022)."

2. *The authors should clarify whether all retrieved data were used or if any quality control measures (e.g., quality flags) were applied.*

RESPONSE: We thank the reviewer for the comment and suggestion. The quality control methods applied to the data are summarized in Table 1 in the paper. Following the Arctic OMI quality control methods described by Sorenson et al. (2023), we use the OMI row anomaly quality control flag to remove L2 OMI pixels from rows with known row anomaly issues. Then, we follow Sorenson et al. (2023) to remove additional row anomaly-affected rows that exhibit contamination over the Arctic. These quality control methods have cascading impacts on the rest of the datasets used in this study, as the L2 OMI UVAI data served as the basis for the co-location of the other satellite products. The surface type flag in the SSMIS sea ice concentration dataset was applied to remove pixels classified as "coastline" or as being too near the North Pole (i.e. in the "pole hole").

We have modified paragraph 2 of section 2.1 with the following text (new text is in italics here): "…with about 50% of the OMI rows currently being contaminated, so we *apply the row anomaly quality control flag in the OMI dataset to* exclude all flagged, row anomaly-affected rows from our analysis."

We also added the following text to paragraph 2 of Section 3 to describe the application of the quality control flags during the colocation process:

"As described in Section 2.1, we used the L2 OMI quality control flags and the methods described by Sorenson et al. (2023) to exclude pixels with flagged or unflagged OMI row anomaly contamination. from the colocated dataset. "

"We excluded pixels from the colocated dataset with the SSMIS surface type flag denoting coastline pixels or pixels too close to the North Pole (i.e. in the "pole hole")."

3. *The dataset used for training the neural network appears limited, which could introduce biases, particularly given the uncertain data quality. This needs careful discussion.*

RESPONSE: We thank the reviewer for the comments and suggestions. When counting the training data in terms of the number of swaths/granules, the training dataset seems limited. However, each swath/granule contains sufficient data points with different observing conditions and viewing geometries. The training dataset consisted of co-located L2 pixels from over 1100 OMI swaths over the Arctic across the boreal sunlit months of 2005 - 2020, equating to about **4 million pixels** for training and **400,000 pixels** for testing the neural network during the training process. The validation dataset used in the study consisted of about 50 independent swaths with a total size of about **200,000 pixels**.

To test for systematic biases in the neural network-based estimates of ADRF, we binned the validation dataset first by the different surface types, and then by the different COD ranges to determine if systematic biases were associated with either variable. The NN error distributions binned by the SSMIS surface type and by the MODIS COD are shown in the figure below. We found that the mean SWF errors for all of these swaths are largely small, with magnitudes primarily less than 3 Wm^{-2}. The peaks of nearly all the error distributions for the different surface types and CODs are around 0, suggesting little systematic bias in the system overall associated with the different surface types and CODs. The mean error for the land distribution is slightly larger at -5.5 Wm^{-2}, suggesting a slight negative bias over land. We suspect that this is related to the lack of information about the land-based surface type in the system. If the NN is primarily trained on dark land surfaces, but it is applied to brighter-than-normal land surfaces (e.g. snow- and ice-covered land), the NN will predict lower upwelling SWF than is actually seen by CERES. When excluding data from April and May from this analysis, the mean error for the over-land data is much smaller, supporting our hypothesis that the slight negative shift in the land-based error distribution is related to the land surface brightness that is unaccounted for in this system. Given that the majority of the smoke events analyzed in the study occurred in the summer months (June – August), we do not expect this potential low bias of the NN over bright land surfaces to significantly impact the results of our study.

To further test for biases in the system, we expanded the validation dataset with an additional 25 randomly-selected aerosol-free OMI swaths from the 2005 – 2020 boreal summer data set, bringing the total number of validation pixels up to **300,000 pixels**. When including these additional swaths, the results of the comparison were nearly identical to the results when using the original 50 validation swaths, further suggesting that the system does not contain significant systematic bias.

We have added the following text and figure to the paper:

3.2 "Validation of the NN against CERES

Once trained, the NN was first applied to the 50 reserved aerosol-free validation swaths (independent from the 131 aerosol swaths) to validate the NN output against CERES observations. The 50 validation swaths contained about 200,000 pixels to use for validation; we note that similar validation results were obtained when increasing the size of the validation dataset to about 300,000 pixels by adding 25 additional aerosol-free OMI swaths (and co-located MODIS, SSMIS, and CERES data) randomly chosen from the 2005 – 2020 boreal summer study period. Errors were calculated between the NN-estimated aerosol-free SWF and the associated CERES TOA SWF observations, and the distribution of the errors from the 50 validation swaths is shown in Fig. 6a. The error distribution peaks at about 0 Wm^{-2}, suggesting little overall bias in the NN-estimated aerosol-free SWF values. To further test for systematic biases in the NN-estimated aerosol-free SWF, we binned the validation dataset first by the SSMIS SIC and surface type, and then by MODIS COD. The NN error distributions binned by the SSMIS surface type and the MODIS COD are shown in Fig. 6b and Fig. 6c, respectively. We found that the mean SWF errors for the error distributions binned by SSMIS SIC and MODIS COD are largely small, with magnitudes primarily less than 3 Wm^{-2}. The peaks of nearly all the error distributions for the different surface types and CODs are around 0 Wm^{-2}, suggesting little systematic bias in the system associated with the different surface types and CODs. The mean error for the land distribution (Fig. 6b, brown) is slightly larger at -5.5 Wm^{-2}, suggesting a slight negative bias over land. We suspect that this is related to the lack of information about the land-based surface type in the system. If the NN is primarily trained on dark land surfaces, but it is applied to brighter-than-normal land surfaces (e.g. snow- and ice-covered land), the NN will predict lower upwelling SWF than is observed by CERES. When excluding data from April and May from this analysis, the mean error for the over-land data is much smaller, supporting our hypothesis that the slight negative shift in the land-based error distribution is related to the land surface brightness that is unaccounted for in this system. Given that the majority of the smoke events analyzed in the study occurred in the summer months (June – August), we do not expect this potential low bias of the NN over bright land surfaces to significantly impact the results of our study."

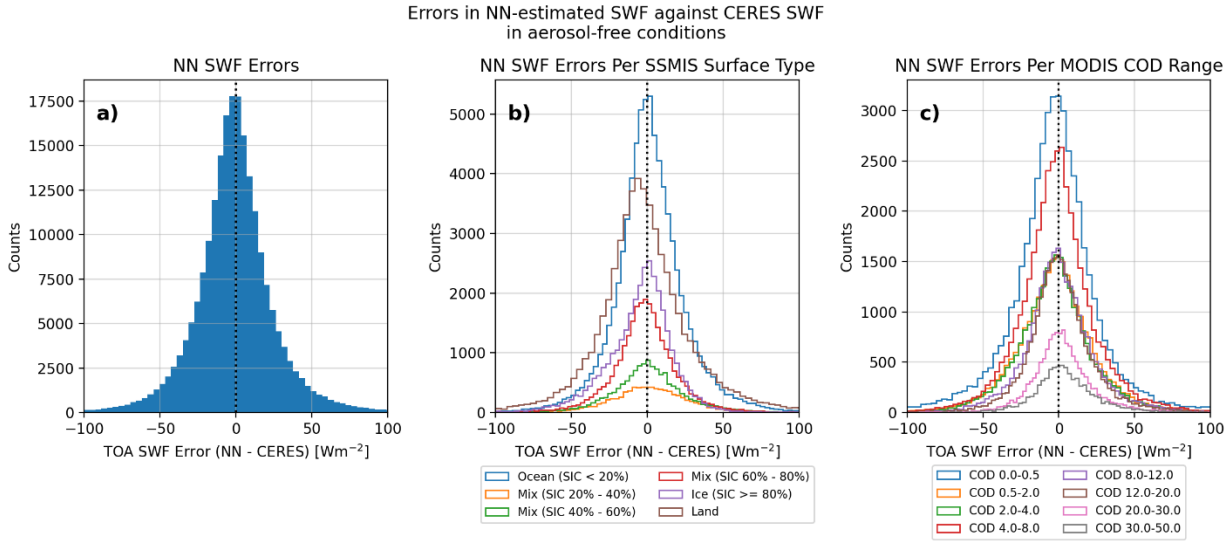

Figure 6. a) Distribution of errors in the neural network (NN)-estimated aerosol-free shortwave flux (SWF) relative to CERES TOA upwelling SWF observations for the 50 validation swaths reserved from the NN training dataset. b) As in (a), but with the errors binned by the SSMIS sea ice concentration (SIC) and surface type. c) As in (a), but with the errors binned by MODIS cloud optical depth (COD).

4. *Figure 5 is not very informative for understanding the neural network architecture. A clearer schematic illustrating the network structure and flow is recommended.*

RESPONSE: We thank the reviewer for the recommendation. We have remade Fig. 5 to more clearly illustrate the neural network structure, as shown below, and have inserted this figure into the paper in place of the old version.

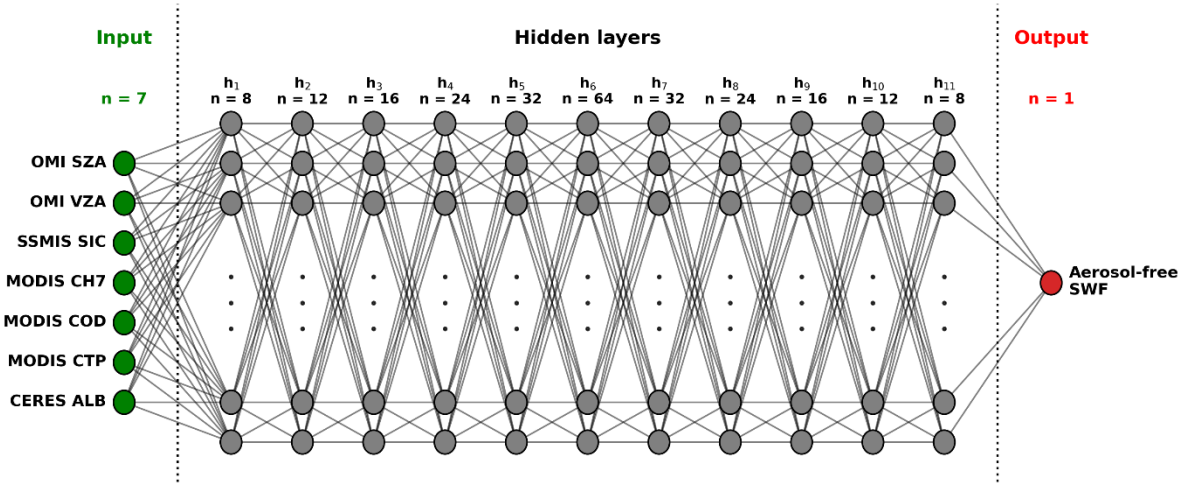

Figure 5. Architecture of the neural network for estimating L2 aerosol-free SWF from L2 input values of solar zenith angle (SZA), viewing zenith angle (VZA), sea ice concentration (SIC), 2.1 μm reflectance (CH7), cloud optical depth (COD), cloud top pressure (CTP), and surface albedo (ALB). Green circles represent nodes in the input layer, gray circles represent nodes in the hidden layers, and the red circle represents the node in the output layer. All nodes in the neural network are fully connected to the nodes in the next layer, as illustrated by the lines connecting the circles.

5. *The method used for trend estimation should be described in detail. Was data uncertainty incorporated into the trend analysis? In Section 4.3, the error analysis is not pixel-based—how representative is this approach?*

RESPONSE: We thank the reviewer for the comments and questions. Our methods for trend estimation, in addition to the methods by which we quantify trend error, are discussed in Section 4.4. Daily ADRF estimates were derived from the daily averaged OMI UVAI value and the ADRF look-up table. Then, a randomly-generated error value matching the error distribution derived in Section 4.3 was added to the derived daily ADRF estimate. Then, all daily ADRF estimates were averaged into monthly values at each grid point. To estimate the trend at each grid point, linear regression was applied to the time series of monthly averaged ADRF estimates, and the difference between the end and beginning points of the fitted trend line represent the trend. This process was repeated 600 times, through a stochastic approach, so that at each grid point there were 600 independent estimates of the monthly ADRF trend from 2005 – 2020.

We account for data uncertainty in this trend approach by perturbing the daily forcing values by an error that matches the distribution from Section 4.3, and through a Monte Carlo method. In that section, we quantified how errors in the input components (i.e. uncertainty in the NN output, uncertainty in applying the LUT to estimate ADRF, uncertainty in the SSMIS SIC / surface type, and uncertainty in the MODIS COD) affect the final daily L3 ADRF estimate. Thus, by conducting 600 independent trend estimates while adding errors that match the combined distribution of the errors from Section 4.3, the spread of the 600 trends at each grid point represents the impacts of data uncertainty on the final trend. We chose to conduct a bulk error analysis rather than a pixel-based error analysis to simplify the error analysis and include more data in each error distribution.

6. *CERES data are used as the reference for shortwave flux (SWF) validation. However, the authors must first assess and validate the accuracy of CERES data specifically over the Arctic.*

RESPONSE: We thank the reviewer for the suggestion. We have modified the end of Section 2.2 to include the following paragraph that describes the validation of CERES data in the Arctic region:

"CERES data have been used extensively for investigating changes in Arctic radiative energy budgets for both TOA (Duncan et al., 2020; Kay and L'Ecuyer, 2013; Riihelä et al., 2013) as well as the surface (Boeke and Taylor, 2016; Hegyi and Taylor, 2017). Previous studies have also worked to validate Arctic CERES surface radiative fluxes (Di Biagio et al., 2021; Riihelä et al., 2017) and TOA fluxes (Taylor et al., 2022), with the latter seeking to validate CERES TOA radiative fluxes against aircraft-based upwelling radiative flux observations. While Taylor et al. (2022) noted some error in the Arctic CERES Level-2 SSF TOA upwelling SWF resulting largely from errors in the imager-based sea ice concentrations used in the scene classification, the CERES observations compared well overall with the aircraft observations (differences between the CERES and aircraft observations were within 2σ uncertainty). The authors concluded that CERES TOA radiative flux data are suitable for polar climate studies (Taylor et al., 2022)."

7. *On line 564, the authors assume that daily Level 3 ADRF errors are normally distributed. This assumption should be justified with supporting analysis.*

RESPONSE: We thank the reviewer for the comments and suggestions. Our assumption that the daily Level 3 ADRF errors are normally distributed is based on our component-based error analysis we conducted in Section 4.3. A normal distribution provides a good fit for the first two error components, which are the errors in the NN-based L2 ADRF estimates and errors in the application of the ADRF LUT. Though we calculate mean error and error standard deviations for the error distributions arising from the impacts of SSMIS SIC errors and MODIS COD errors on the estimated L3 ADRF values, as shown in Fig. 10 (now Fig. 11), those error distributions are not normally distributed. The vast majority of the errors are equal to 0 because of the size of the SIC and COD bins, so a fitted curve from a normal distribution would likely exhibit a wider spread than is actually present in the error

distribution. However, a normal fit likely provides an overestimate of the spread of the errors for those components. Thus, this overestimate leads to the L3 ADRF errors likely being an overestimate of the true L3 ADRF error. We chose to accept this overestimation of the error to have higher confidence in the trends found to be significant in the study.

References

Boeke, R. C. and Taylor, P. C.: Evaluation of the Arctic surface radiation budget in CMIP5 models, J. Geophys. Res. Atmospheres, 121, 8525–8548, https://doi.org/10.1002/2016JD025099, 2016.

Cavalieri, D. J., Gloersen, P., and Campbell, W. J.: Determination of sea ice parameters with the NIMBUS 7 SMMR, J. Geophys. Res. Atmospheres, 89, 5355–5369, https://doi.org/10.1029/JD089iD04p05355, 1984.

Cavalieri, D. J., Crawford, J. P., Drinkwater, M. R., Emery, W. J., Eppler, D. T., Farmer, L. D., Fowler, C. W., Goodberlet, M., Jentz, R. R., and Milman, A.: NASA Sea Ice Validation Program for the Defense Meteorological Satellite Program Special Sensor Microwave Imager, 1992.

Di Biagio, C., Pelon, J., Blanchard, Y., Loyer, L., Hudson, S. R., Walden, V. P., Raut, J.-C., Kato, S., Mariage, V., and Granskog, M. A.: Toward a Better Surface Radiation Budget Analysis Over Sea Ice in the High Arctic Ocean: A Comparative Study Between Satellite, Reanalysis, and local-scale Observations, J. Geophys. Res. Atmospheres, 126, e2020JD032555, https://doi.org/10.1029/2020JD032555, 2021.

Duncan, B. N., Ott, L. E., Abshire, J. B., Brucker, L., Carroll, M. L., Carton, J., Comiso, J. C., Dinnat, E. P., Forbes, B. C., Gonsamo, A., Gregg, W. W., Hall, D. K., Ialongo, I., Jandt, R., Kahn, R. A., Karpechko, A., Kawa, S. R., Kato, S., Kumpula, T., Kyrölä, E., Loboda, T. V., McDonald, K. C., Montesano, P. M., Nassar, R., Neigh, C. S. R., Parkinson, C. L., Poulter, B., Pulliainen, J., Rautiainen, K., Rogers, B. M., Rousseaux, C. S., Soja, A. J., Steiner, N., Tamminen, J., Taylor, P. C., Tzortziou, M. A., Virta, H., Wang, J. S., Watts, J. D., Winker, D. M., and Wu, D. L.: Space-Based Observations for Understanding Changes in the Arctic-Boreal Zone, Rev. Geophys., 58, e2019RG000652, https://doi.org/10.1029/2019RG000652, 2020.

Hegyi, B. M. and Taylor, P. C.: The regional influence of the Arctic Oscillation and Arctic Dipole on the wintertime Arctic surface radiation budget and sea ice growth, Geophys. Res. Lett., 44, 4341–4350, https://doi.org/10.1002/2017GL073281, 2017.

Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L., and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using

passive microwave observations, The Cryosphere, 9, 1797–1817, https://doi.org/10.5194/tc-9-1797-2015, 2015.

Kay, J. E. and L'Ecuyer, T.: Observational constraints on Arctic Ocean clouds and radiative fluxes during the early 21st century, J. Geophys. Res. Atmospheres, 118, 7219–7236, https://doi.org/10.1002/jgrd.50489, 2013.

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., Tonboe, R. T., Saldo, R., and Sørensen, A. M.: Satellite passive microwave sea-ice concentration data set intercomparison: closed ice and ship-based observations, The Cryosphere, 13, 3261–3307, https://doi.org/10.5194/tc-13-3261-2019, 2019.

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., and Tonboe, R.: Satellite passive microwave sea-ice concentration data set inter-comparison for Arctic summer conditions, The Cryosphere, 14, 2469–2493, https://doi.org/10.5194/tc-14-2469-2020, 2020.

Meier, W. N.: Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in arctic peripheral seas, IEEE Trans. Geosci. Remote Sens., 43, 1324–1337, https://doi.org/10.1109/TGRS.2005.846151, 2005.

Riihelä, A., Manninen, T., and Laine, V.: Observed changes in the albedo of the Arctic sea-ice zone for the period 1982–2009, Nat. Clim. Change, 3, 895–898, https://doi.org/10.1038/nclimate1963, 2013.

Riihelä, A., Key, J. R., Meirink, J. F., Kuipers Munneke, P., Palo, T., and Karlsson, K.-G.: An intercomparison and validation of satellite-based surface radiative energy flux estimates over the Arctic, J. Geophys. Res. Atmospheres, 122, 4829–4848, https://doi.org/10.1002/2016JD026443, 2017.

Sorenson, B. T., Zhang, J., Reid, J. S., Xian, P., and Jaker, S. L.: Ozone Monitoring Instrument (OMI) UV aerosol index data analysis over the Arctic region for future data assimilation and climate forcing applications, Atmospheric Chem. Phys., 23, 7161–7175, https://doi.org/10.5194/acp-23-7161-2023, 2023.

Steffen, K. and Schweiger, A.: NASA team algorithm for sea ice concentration retrieval from Defense Meteorological Satellite Program special sensor microwave imager: Comparison with Landsat satellite imagery, J. Geophys. Res. Oceans, 96, 21971–21987, https://doi.org/10.1029/91JC02334, 1991.

Taylor, P. C., Itterly, K. F., Corbett, J., Bucholtz, A., Sejas, S., Su, W., Doelling, D., and Kato, S.: A Comparison of Top-Of-Atmosphere Radiative Fluxes From CERES and ARISE, J. Geophys. Res. Atmospheres, 127, e2022JD037573, https://doi.org/10.1029/2022JD037573, 2022.