Improving Near-Real-Time CERES Surface Solar Radiative Flux Estimates with Machine Learning

Kuan-Man Xu[1] , P. K. Sawaengphokhai[2], Hunter F. Winecoff[3], Jay Garg[4], Walter F. Miller[2], and Paul W. Stackhouse Jr.[1]

[1]NASA Langley Research Center, Hampton, Virginia
[2]ADNET Systems, Inc., Hampton, Virginia
[3]Analytical Services and Materials, Inc., Hampton, Virginia
[4]Allen Control Systems, Austin, Texas



Motivations, Objective and Datasets

Motivations

1) Standard CERES products have a latency of several months and are not suitable for near-real-time applications
2) The CERES FLASHFlux product addresses this need by using empirical algorithms to produce global hourly and daily surface radiative flux, typically available within a few days
3) These algorithms rely on simplified representations of atmospheric radiative transfer processes and limited meteorological, cloud and aerosol inputs
4) Previous efforts to improve the existing parameterizations had very limited success

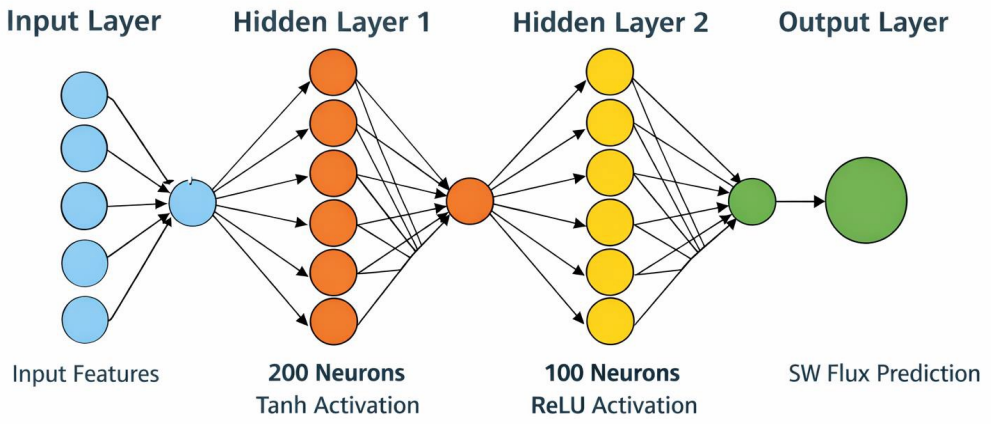
Objective

- Evaluate the use of artificial neural networks (ANNs) to replace the Langley Surface Parameterization models for estimating solar radiation in the FLASHFlux production

Datasets

- Two years of the CERES-Terra Cloud Radiative Swath (CRS) dataset (footprint data, ~20 km at nadir), with surface radiative fluxes computed using the Fu-Liou radiative transfer model
  - 2019 for training and 2020 for testing
- BSRN surface measurements for independent evaluation

Artificial Neural Network (ANN) Configuration & Training Approaches



Input variables ("features")

- Radiation (TOA insolation, Reflected SW, Albedo)
- Cloud (Cloud fraction, Cloud optical depth)
- Meteorology (Precipitable water)
- Aerosol (Optical depth, SSA, Asymmetry factor)

Output: Surface downward SW flux

Activation functions

- Tanh: Captures smooth, continuous physical relationships
- ReLu: Represent nonlinear effects to approximate regime transitions

Early stopping (ES): Avoid overfitting

Training approaches:

- Train monthly/seasonal/annual models
- Change ANN configuration: ES & learning rate
- Use input sequence lengths of 4, 6 or 8 days
- Change/Add input cloud features
- Weights from training will be used in future FLASHFlux (FF) production

Key Input Variables ("features")

Radiation:

1) TOA solar insolation flux
2) TOA reflected solar radiative flux
3) Solar zenith angle (removed after initial testing because it is redundant with insolation flux)

Cloud: three sets of cloud features; total, two-layer, and two-layer with effective radius

1) Cloud fraction (CF): Combined two-layer fraction → separate cloud-layer fractions
2) Cloud optical depth (COD): Combined cloud-area mean → footprint mean [CF × ln(COD)] for each cloud layer
3) Effective radius: Phase-dependent liquid or ice effective radius added as an additional input feature

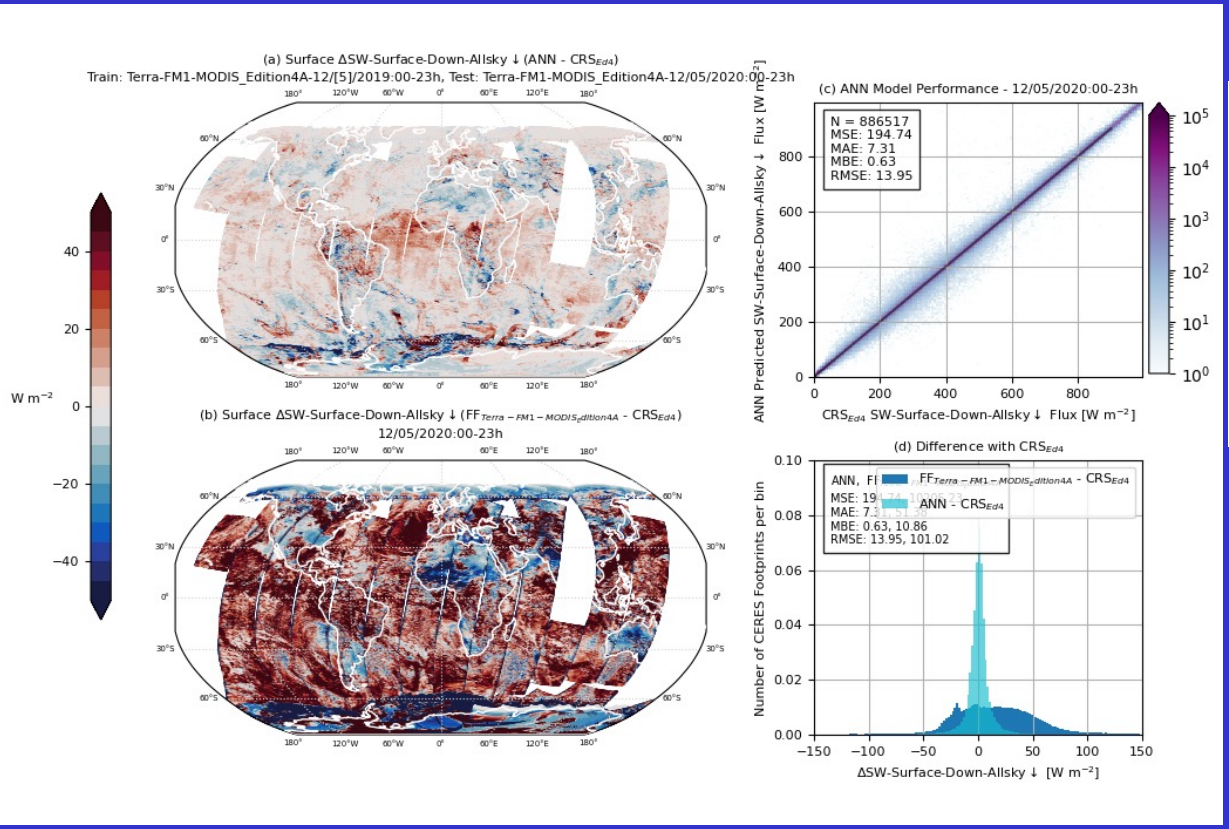
Meteorology: Precipitable water

Aerosol:

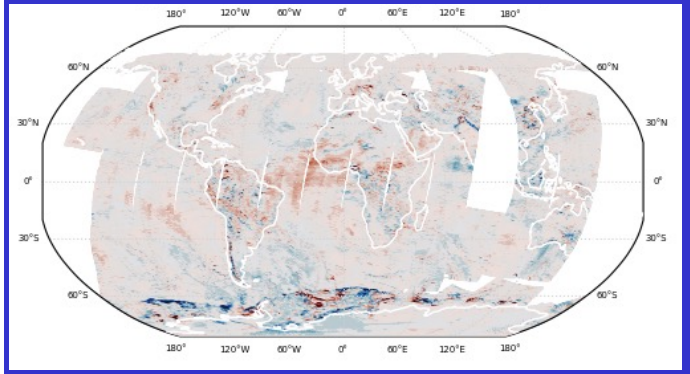
1) Aerosol optical depth
2) Monthly-mean single-scattering albedo and asymmetry factor

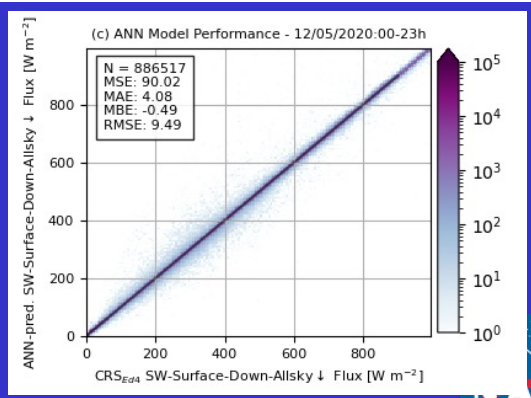


Performance improvement with the ANN model—An example



MSE: 195 → 90 W^{2}m^{-4}





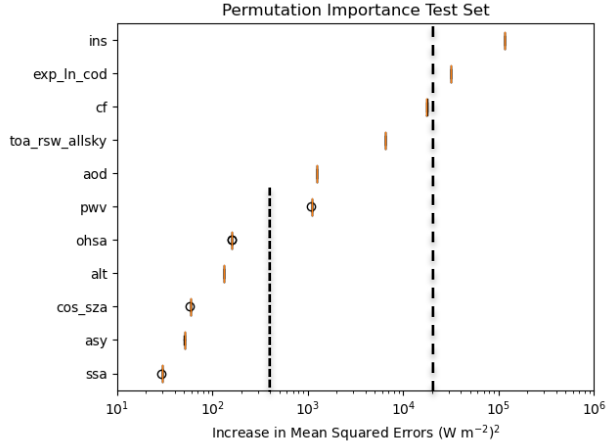
Comparison of prior and current ML models against SOFA parameterization in FF

Which Input Variable is More Important? -- Permutation Test

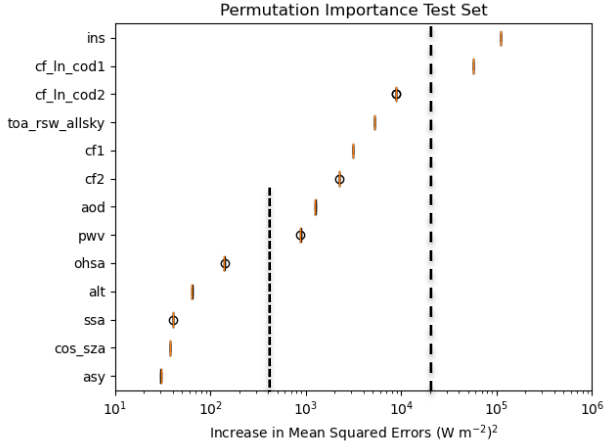
Permutation importance: For a trained model, randomly permute the values of one input feature while leaving all other inputs unchanged, and measure the increase in mean squared error (MSE). A larger increase indicates greater dependence on that feature.

This test breaks the relationship between that feature and the target (SW flux). It answers this question: How much does the trained ANN rely on this variable when making predictions?

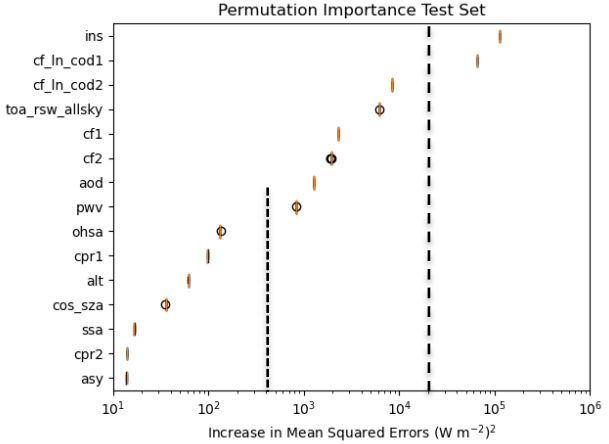
(a) Total CF and COD



(b) Two-layer CF and COD



(c) Two-layer & effective radius

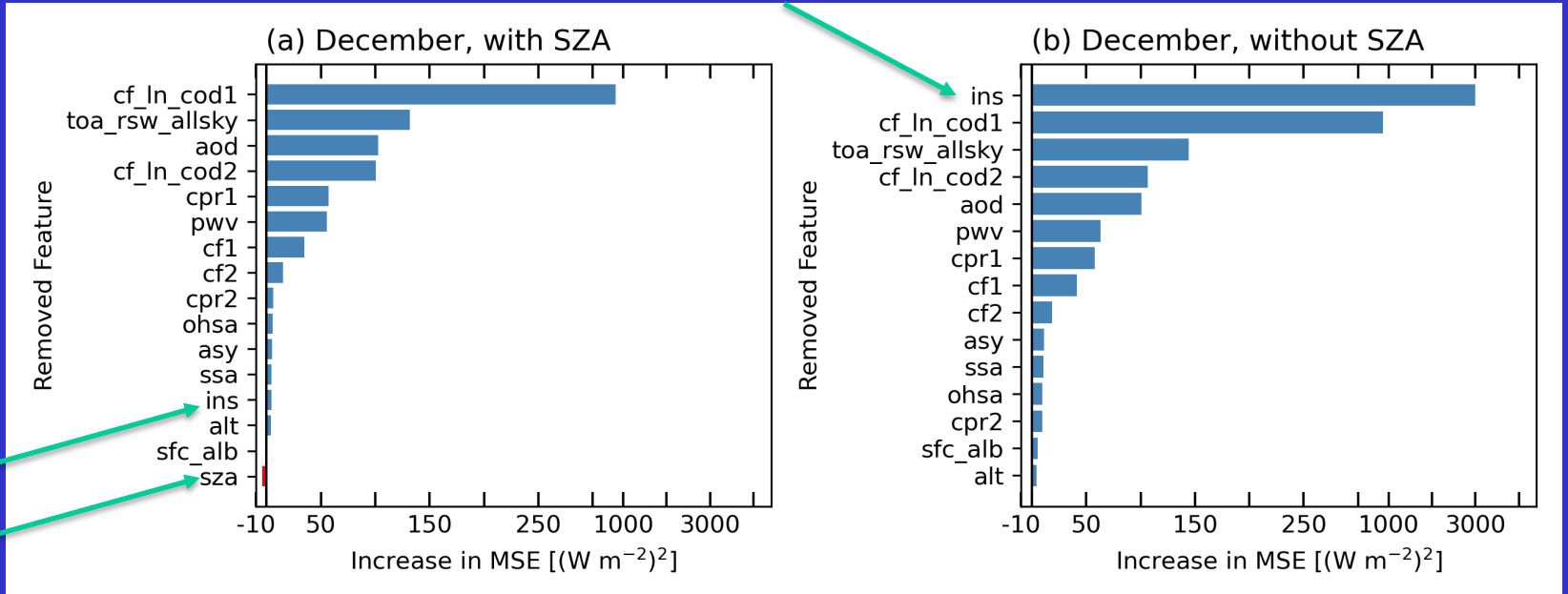


COS_SZA is surface albedo

Which Input Variable is More Important? -- LOFO Test

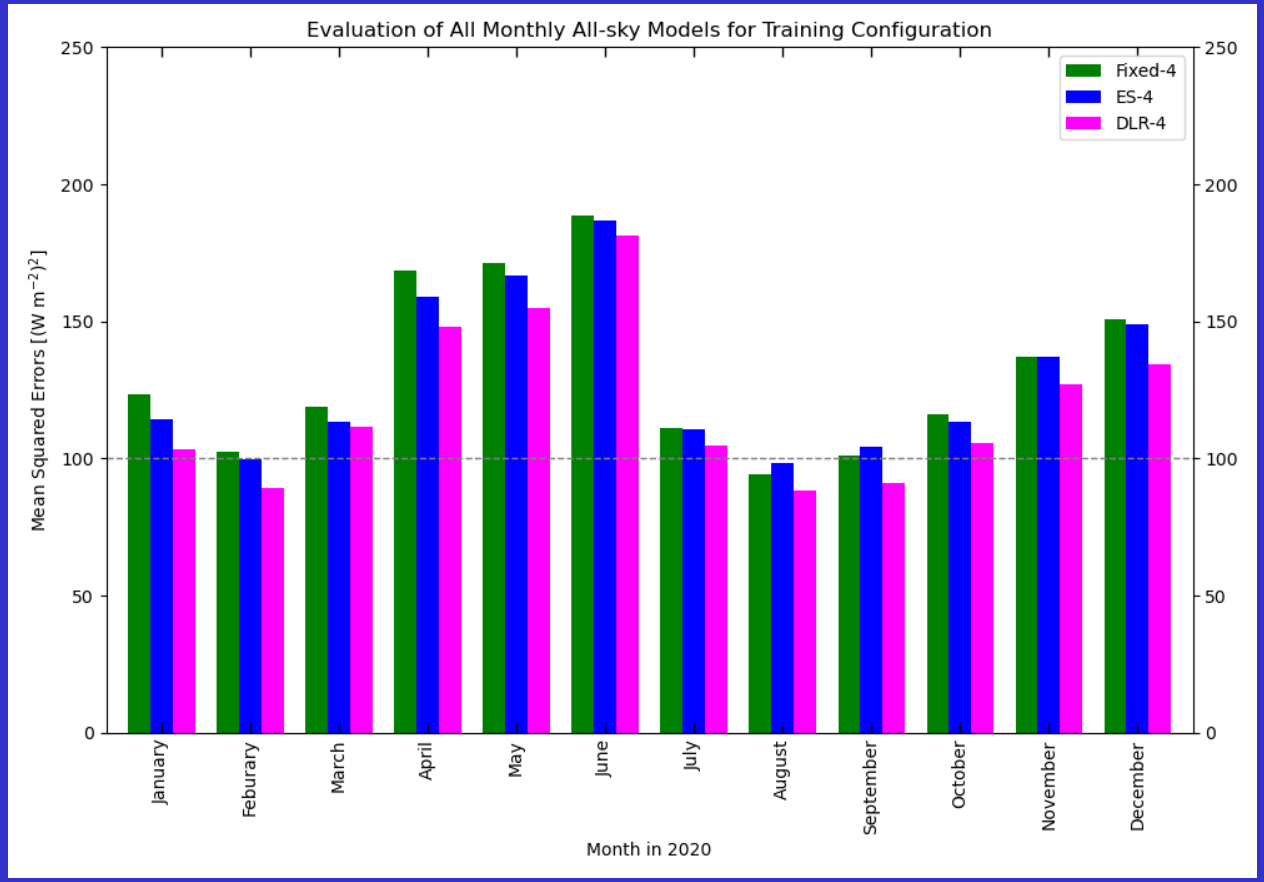
LOFO (Leave-One-Feature-Out): Retrain the model after removing one input feature, then compare its mean squared error (MSE) with that of the full model trained using all input features. A larger increase in MSE indicates that the removed feature contributes more to model performance (see ** below).

LOFO test retrains the model without the feature, allowing the remaining features to compensate for its absence. It answers the question: How much predictive skill would be lost if this variable were unavailable?
**However, if two features (SZA, insolation) are highly correlated, removing either one does not change the MSE.



Effect of Training Configuration on Monthly Model MSE

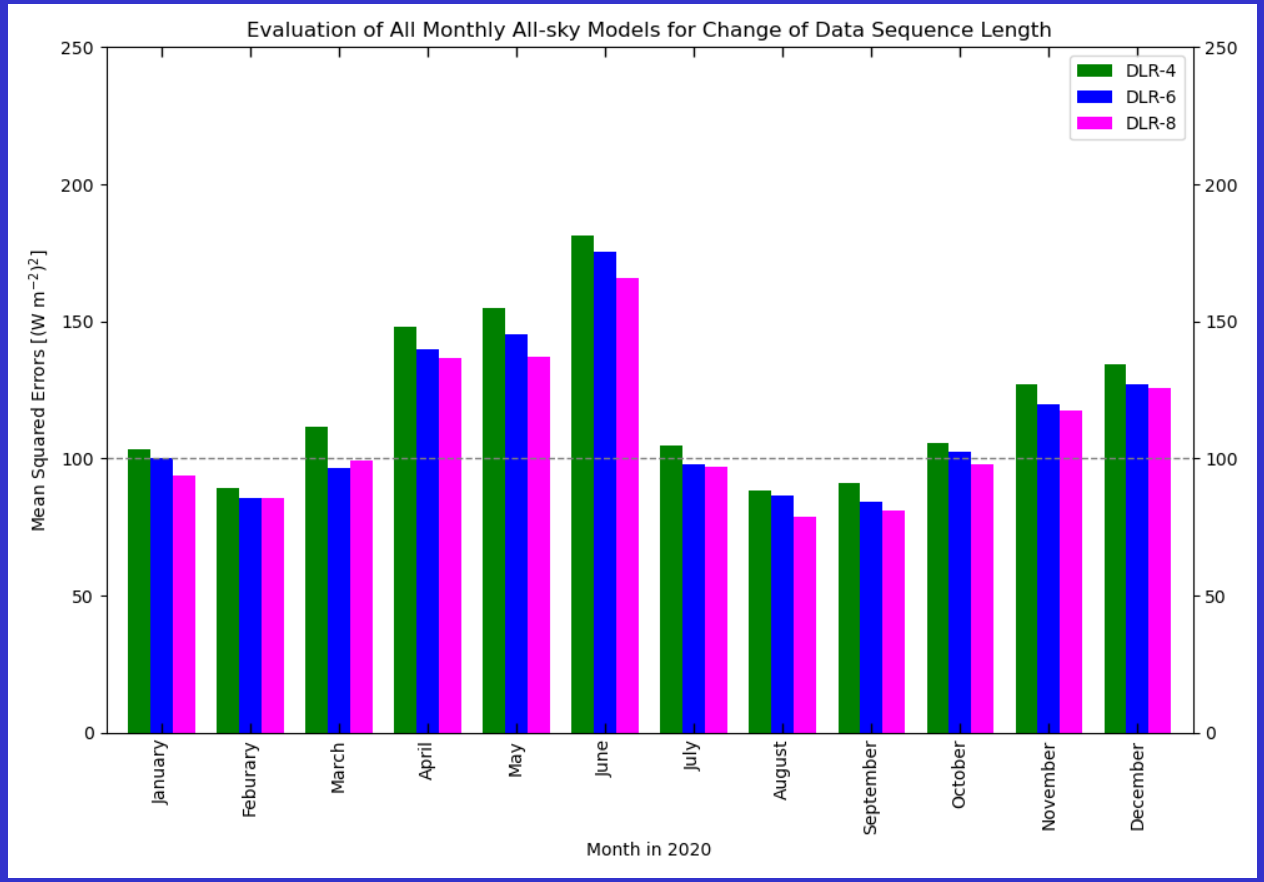
Fixed 40-epochs (green), Early Stopping (blue), and ES & Lower Learning Rate (magenta) Trained with a 2-layer cloud feature set w/o effective radius and 4-day data sequence



- Model performance exhibits strong seasonal variability
- The best performance occurs during the transitional months, while the largest errors occur in **June,** followed by **May, April, December,** and **November**
- The impact of training configuration varies by month (up to 17 **W² m⁻⁴**)
- Combining early stopping with a lower learning rate (**25 x too high**) consistently reduces the MSE by > **10 W² m⁻⁴** in 7/12 months (compare green with magenta)

Effect of Input Data Sequence Length on Monthly Model MSE

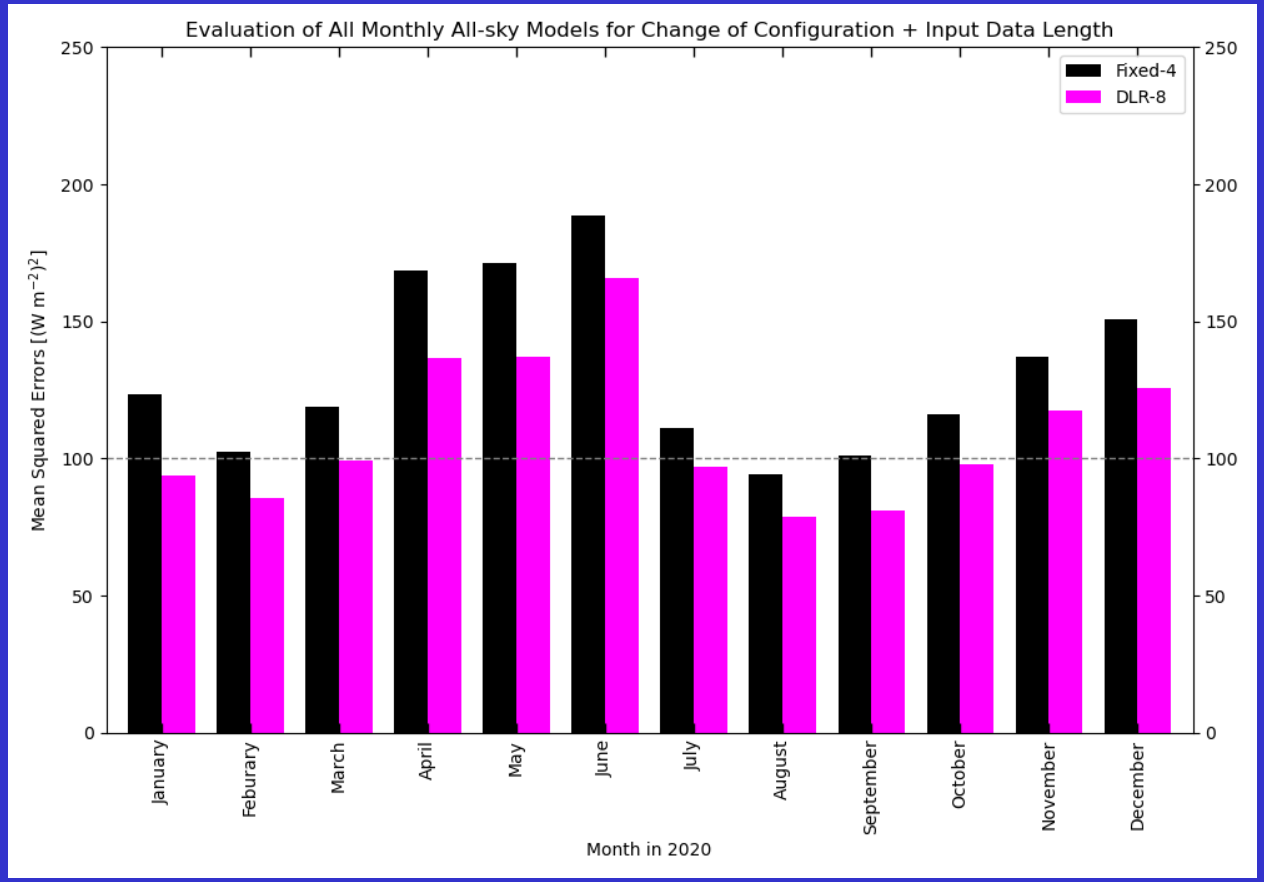
4-day (green), 6-day (blue) and 8-day (magenta) input sequences
Trained with a 2-layer cloud feature set w/o effective radius



- The effect of input sequence length is generally modest (~10 W^2 m^{-4} for most months), with larger effects observed in **March-June**
- Model performance generally improves with increasing sequence length, with the 4-day input sequence consistently yielding the poorest performance
- However, extending the input sequence from 6 to 8 days yields little improvement in 7/12 months
- Overall, these results suggest that an 8-day input sequence provides the optimal balance of sequence length and model performance

Effect of Input Length+Training Configuration on Monthly Model MSE

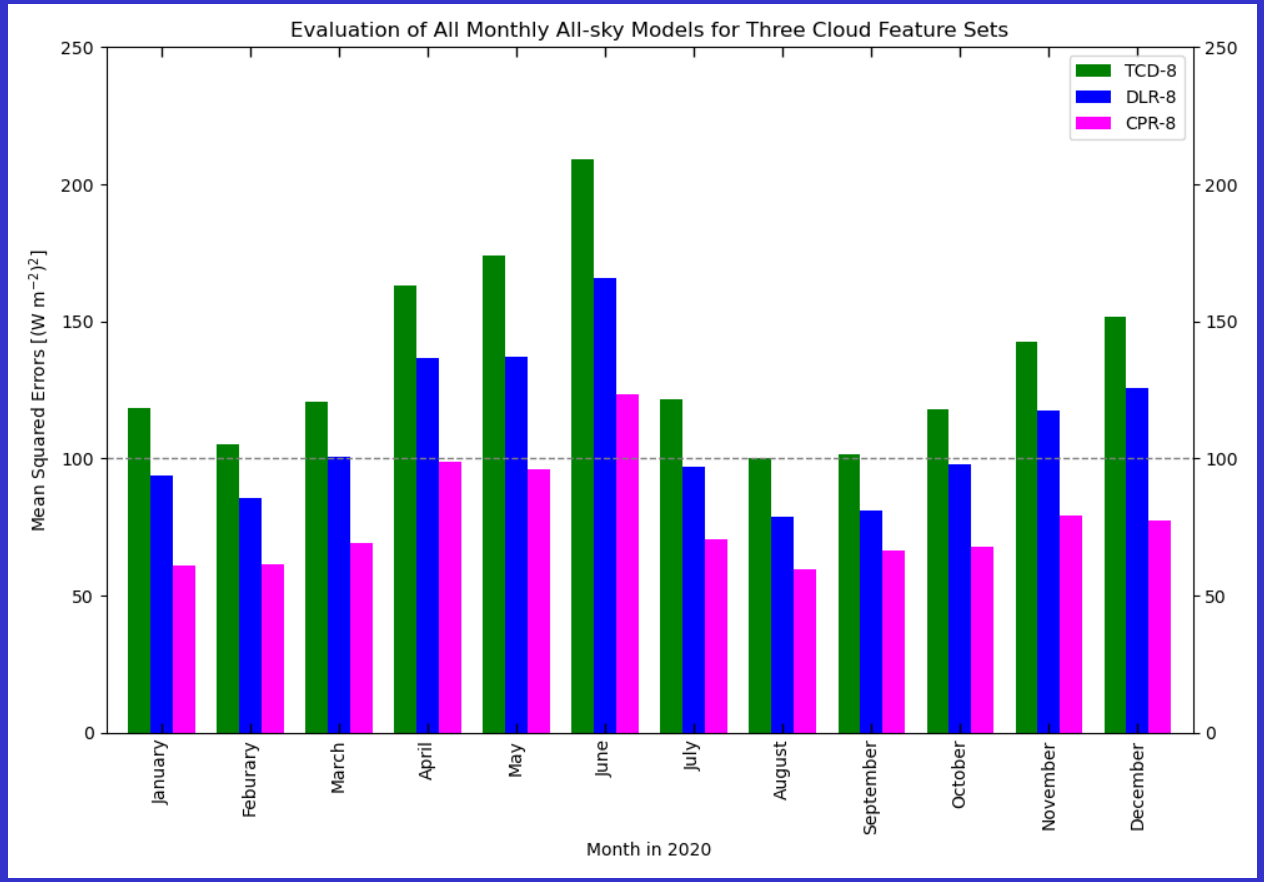
Trained with a 2-layer cloud feature set w/o effective radius



- The combined effect of modifying the training configuration (early stopping and a lower learning rate) and increasing input sequence length (from 4 to 8 days) is comparable in magnitude to the impact of changes in cloud feature selection, discussed next
- MSE is reduced by ~20 $\mathbf{W^2\ m^{-4}}$ in most months, with reductions ranging from 14 to 34 $\mathbf{W^2\ m^{-4}}$

Effect of Cloud Input Features on Monthly Model MSE

Total cloud (green), two-layer cloud (blue), and two-layer cloud with effective radius (magenta) cloud feature sets



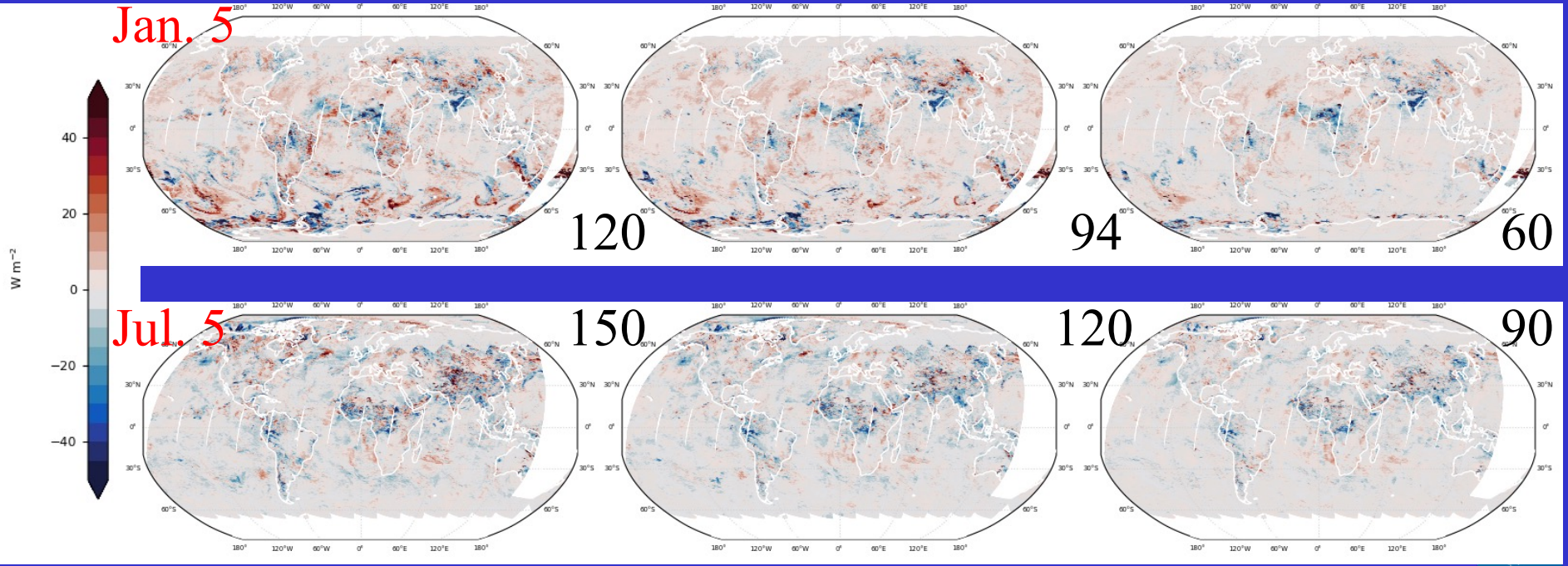
- Increasing cloud detail consistently improves model performance
- **Using two-layer cloud feature set:**
  - **7/12** monthly models achieve MSE < 100 W^2 m^{-4}
  - MSE is reduced by 15-44 **W^2 m^{-4}** relative to the total-cloud feature set
  - The largest improvement (> 30 **W^2 m^{-4}**) occur in **May-June**
- **Adding cloud effective radius to the two-layer cloud feature set:**
  - **11/12** (except **June**) monthly models achieve MSE < 100 W^2 m^{-4}
  - MSE further reduced by 15-48 **W^2m^{-4}**
  - The largest improvement (> 30 **W^2 m^{-4}**) occurs in **March-June** and **November-January**

Improved performance due to changes of cloud input features

Total CF, COD

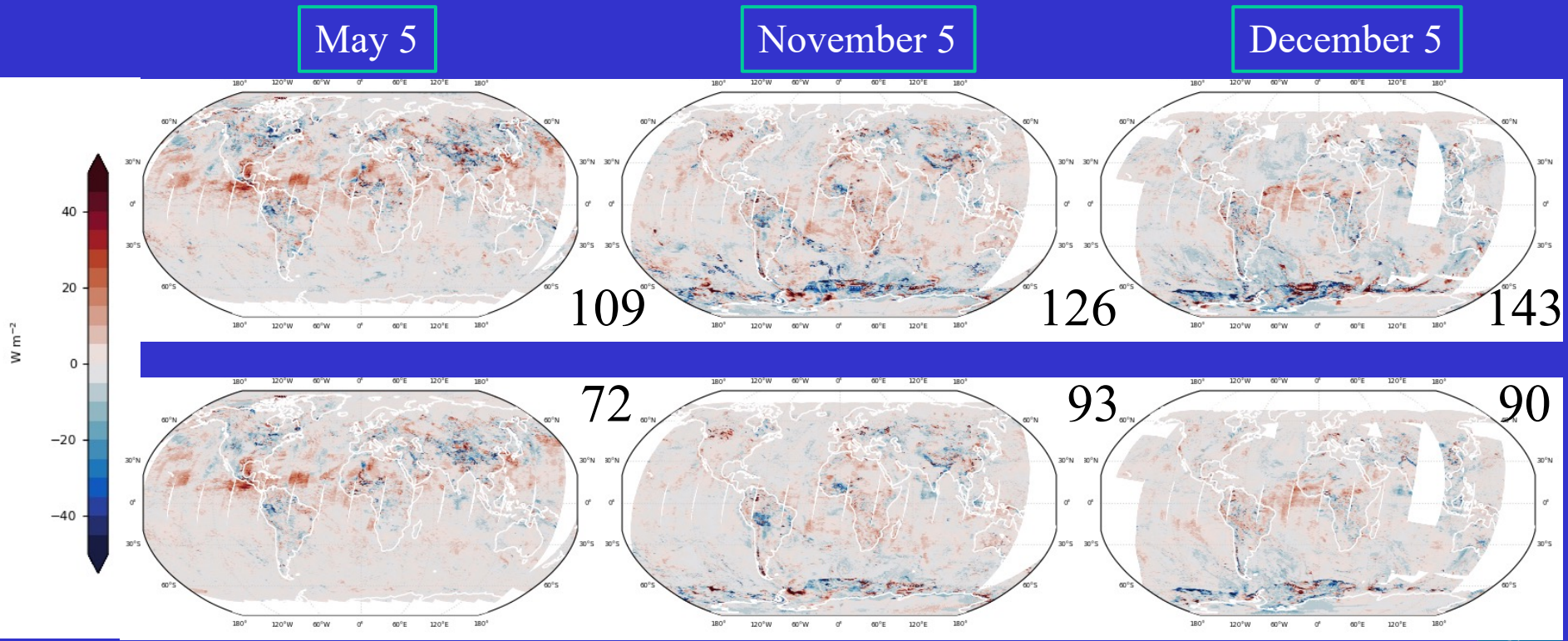
2-layer CF & COD

Add cloud effective radius



Differences in SW flux between ANN test and CERES-CRS data for January 5 (top row) and July 5, 2020 (bottom row). ML training uses 2019 data. Black number is MSE in unit of $W^2\ m^{-4}$.

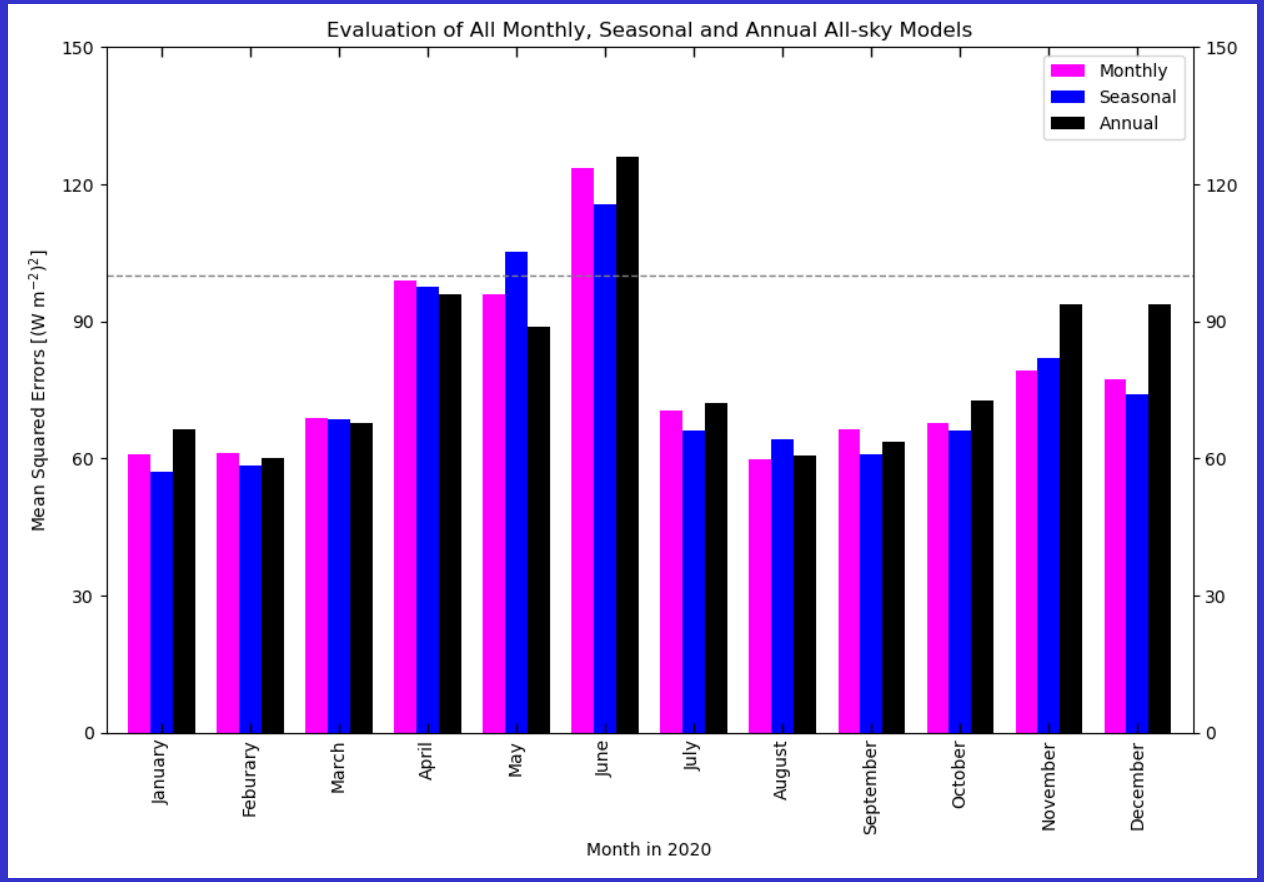
Improved performance due to adding cloud effective radius as an input feature: More examples



Differences in surface downward SW flux between ANN test and CERES-CRS data for May 5, Nov. 5, and Dec. 5, 2020. Cloud effective radius is added to training (bottom row). Black number is MSE in unit of W^2 m^{-4}

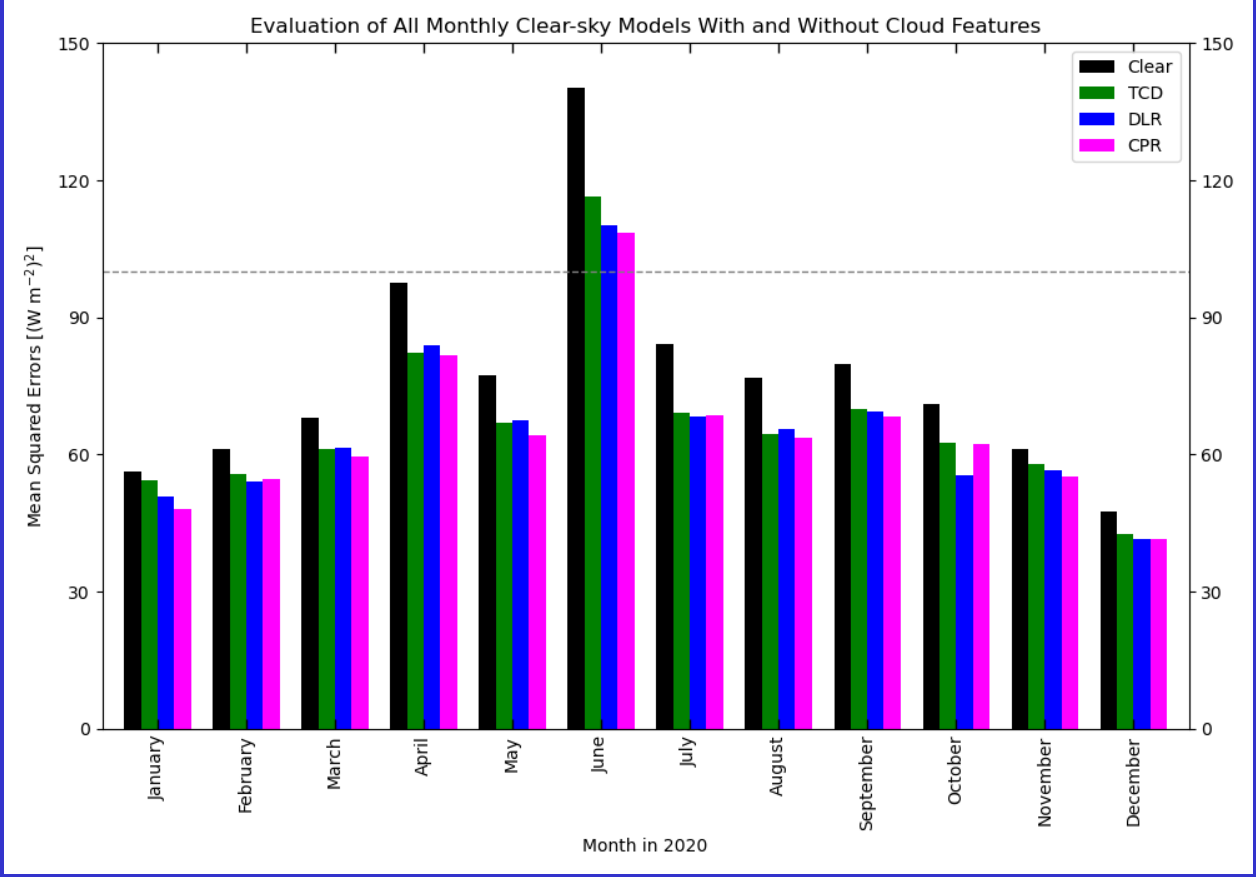
Comparison of Monthly, Seasonal, and Annual models

Trained using a two-layer cloud feature set with cloud effective radius and 8-day data sequences from each month



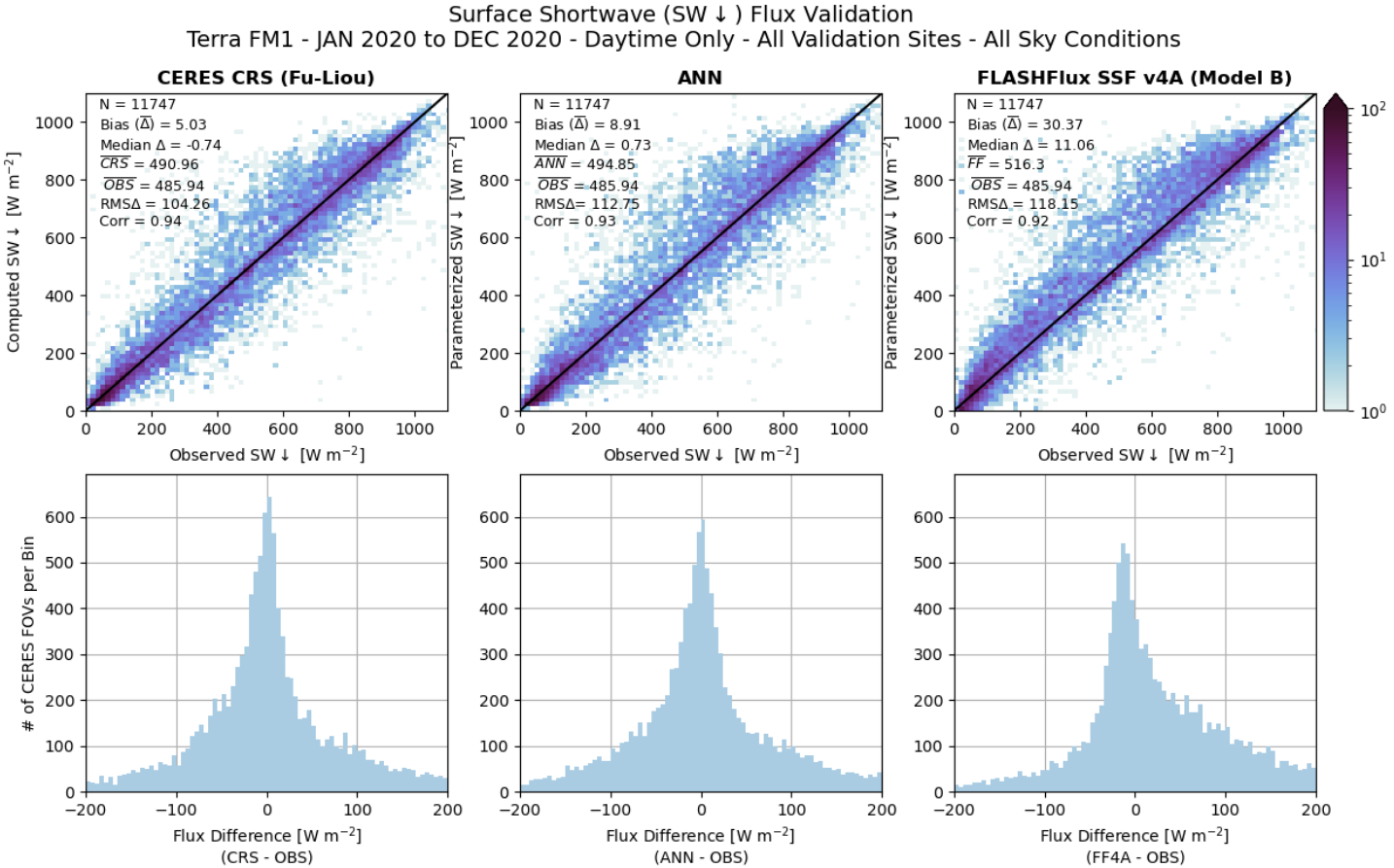
- A total of 17 models:
  - 12 monthly models
  - 4 seasonal models (DJF, MAM, JJA, and SON)
  - 1 annual model
  - All models evaluated against individual months
- Seasonal models show modestly improved performance in most months, except May, August and November, which are the final month of their respective seasons
- Annual model performs worst in six months (Jan., June, July, Oct.-Dec.) but comparable in other months (slightly better in May)

Comparison of Monthly Clear-sky Flux Models with/without Cloud Features as Input



- Trained clear-sky flux models with (green, blue, and magenta) and without cloud features (black)
  - Ideally, clear-sky flux models should exclude cloud features as input
  - However, daily aerosol information is unavailable, except for aerosol optical depth, so monthly aerosol information is used instead
- Results indicate that including cloud features consistently improve model performance
- Overall, model performance shows modest/no sensitivity to the included cloud feature set, as opposed to all-sky flux training

Independent Evaluation against BSRN Surface Measurements, all sites



Summary

- ANN-based models successfully emulate surface SW flux using the CERES-CRS dataset, with substantial improvements achieved through optimization of the training configuration and cloud input features.
  - Combining early stopping with a lower learning rate reduces MSE by > 10 $W^2\ m^{-4}$ for monthly models. Extending the input sequence yield somewhat smaller improvement. The combined impact is comparable to that of changing the cloud feature selection.
  - Representing cloud fraction and cloud optical depth in two separate cloud layers further reduces MSE by 15–44 $W^2\ m^{-4}$, with the largest improvement (> 30 $W^2\ m^{-4}$) occurring from **May-July**.
  - Adding phase-dependent cloud particle effective radius provides the greatest improvement, reducing MSE by 15–48 $W^2\ m^{-4}$. The largest reductions (> 30 $W^2\ m^{-4}$) occur in **March-June** and **November-January**, indicating that the models better capture cloud scattering effects, particularly over midlatitude regions in summer.

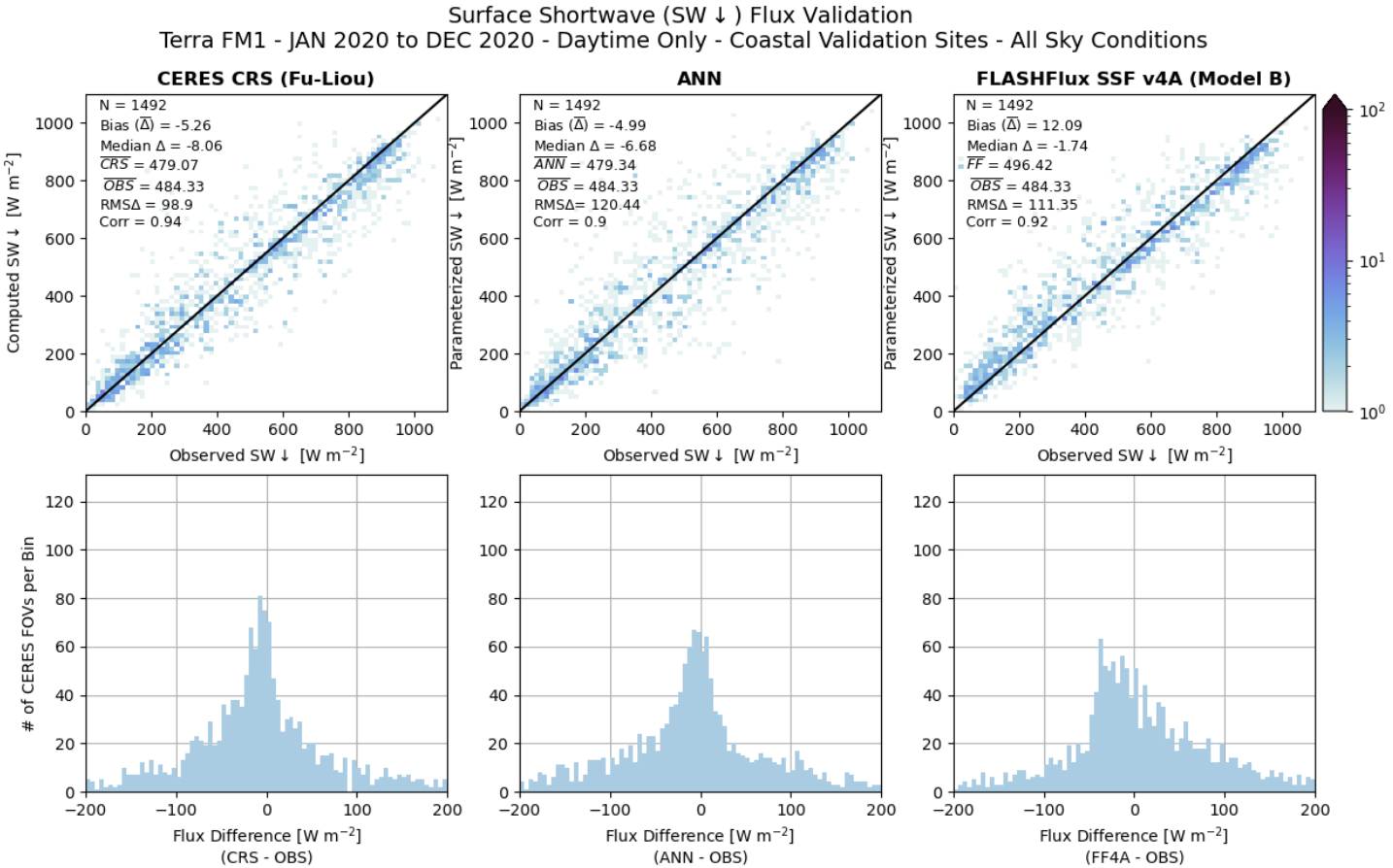
Summary (cont.)

- June remains the most challenging month to model accurately. Seasonal models perform slightly better in some months, whereas the annual model generally perform worst in most months.
- Clear-sky flux models trained with and without cloud features show that including cloud features provide a modest improvement in clear-sky flux prediction.
- Independent validation against surface radiation measurements is encouraging, with performance comparable to the CERES-CRS reference data and substantially better in mean and median biases than FLASHFlux, except for polar and coastal sites.

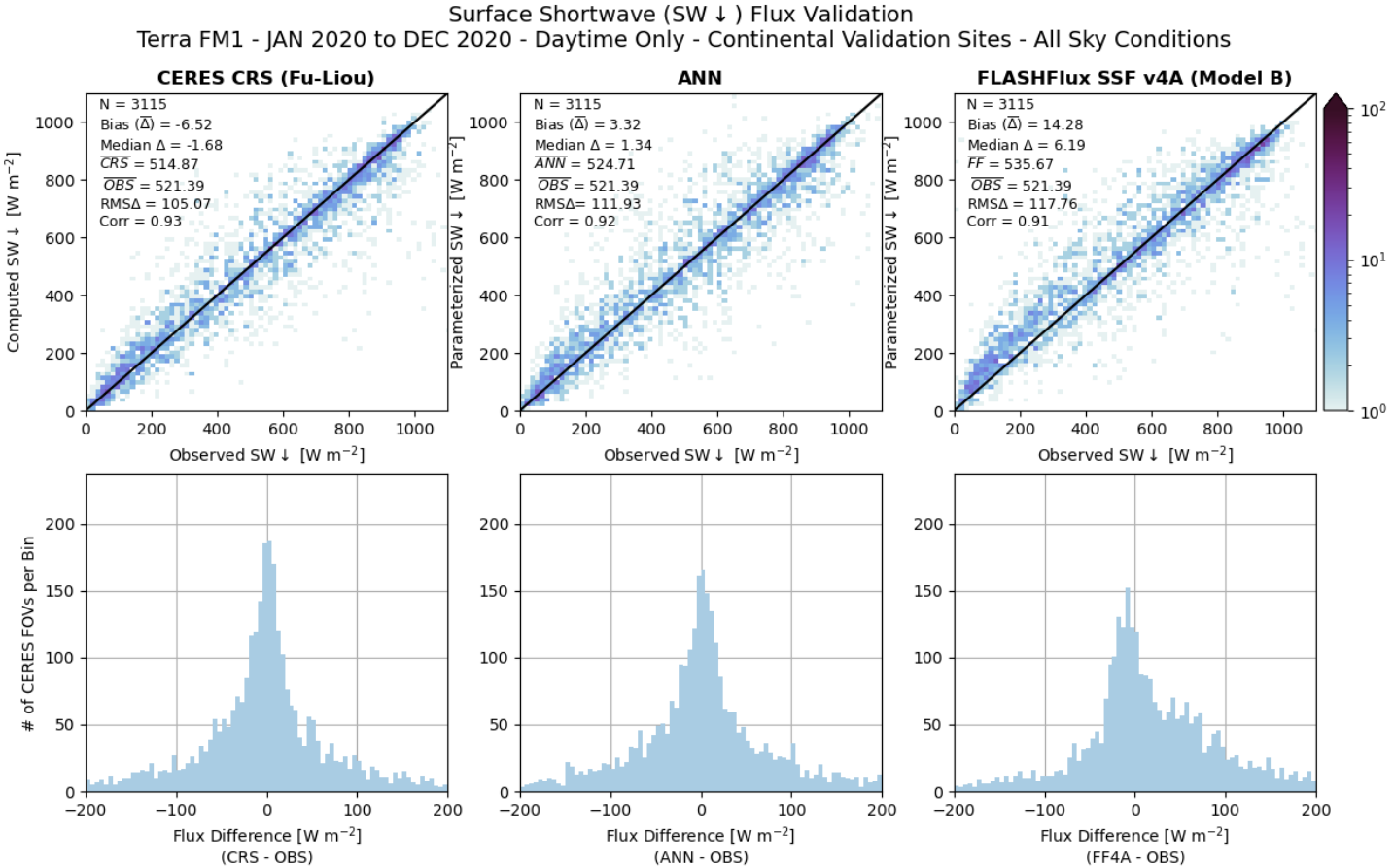
Surface validation site type	Number of samples	Mean Biases			Median Biases		
		CRS	NN	FF	CRS	NN	FF
Global	11748	**-5.03**	8.91	30.37	-0.74	**0.73**	11.06
Continental	3115	-6.52	**3.32**	14.28	-1.68	**1.34**	6.19
Coastal	1492	-5.26	**-4.99**	12.09	-8.06	-6.68	**-1.74**
Desert	594	-17.16	**-15.98**	19.48	-12.22	**-10.35**	-17.43
Island	614	**25.57**	37.27	44.4	**1.02**	4.33	16.03
Polar	3293	-22.81	-20.98	**12.06**	-11.0	-10.1	**-5.06**
Buoy	2640	**59.39**	59.85	90.47	**46.65**	48.63	77.54

Backup slides

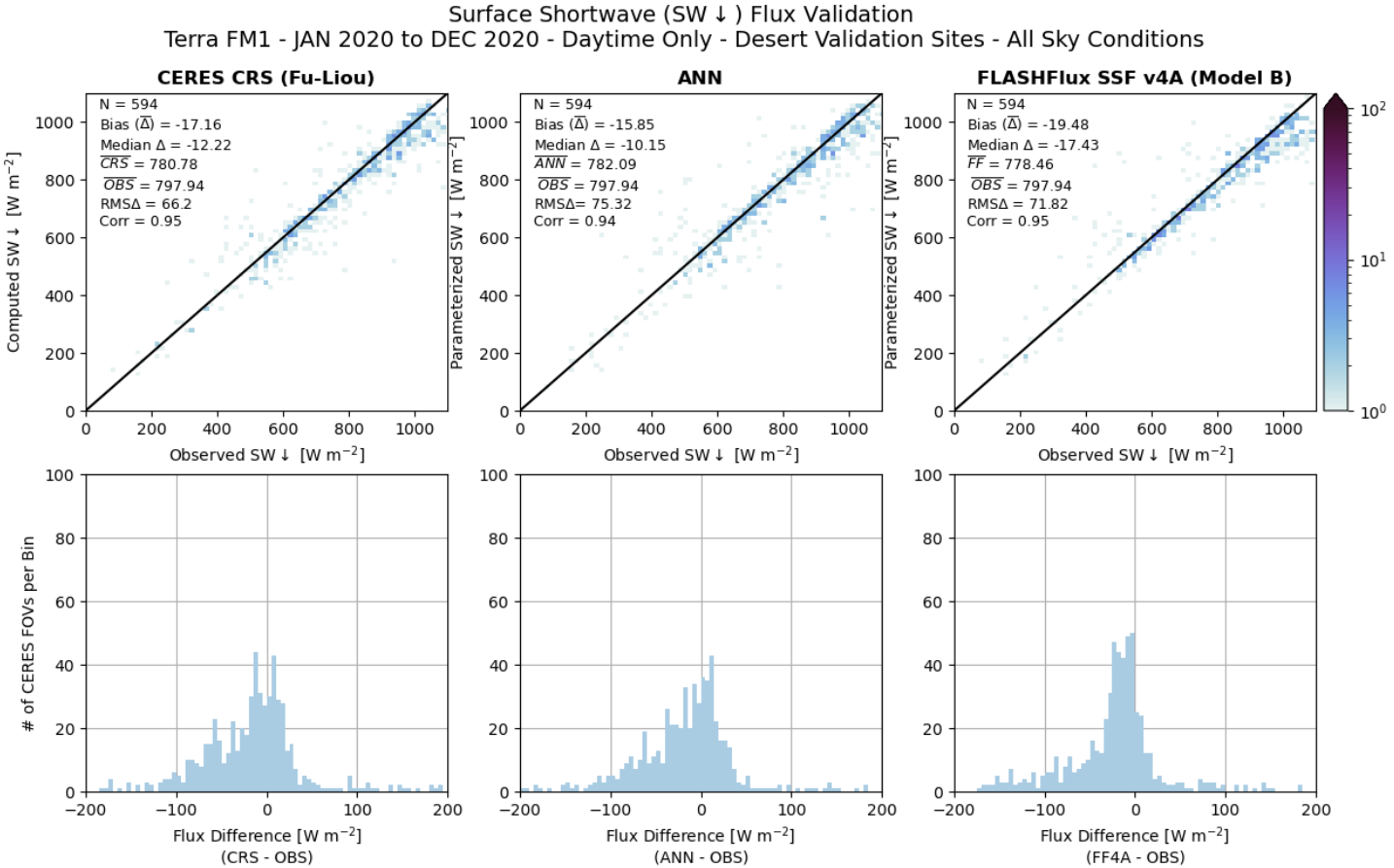
Independent Evaluation against BSRN Surface Measurements, coastal sites



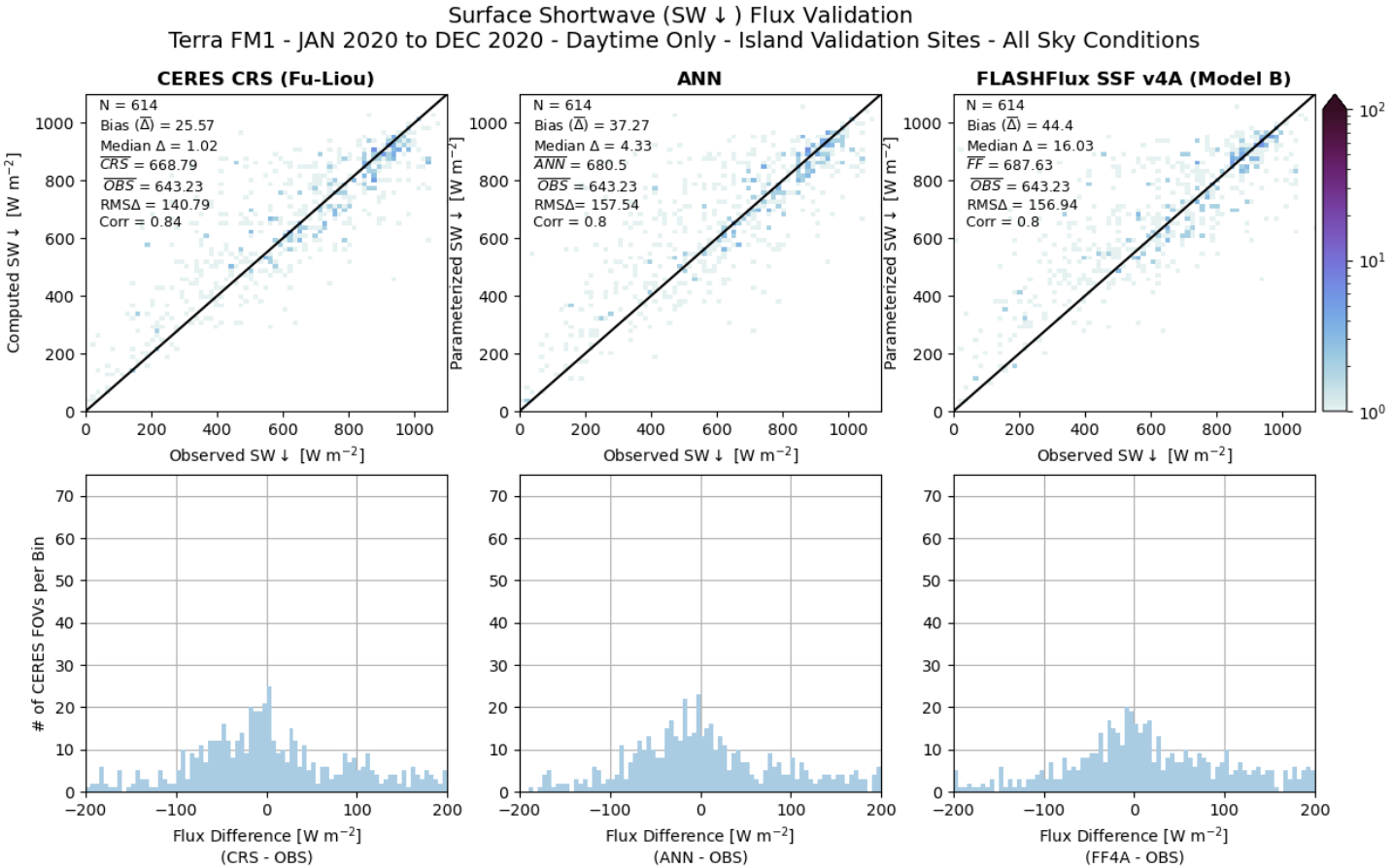
Independent Evaluation against BSRN Surface Measurements, continental sites



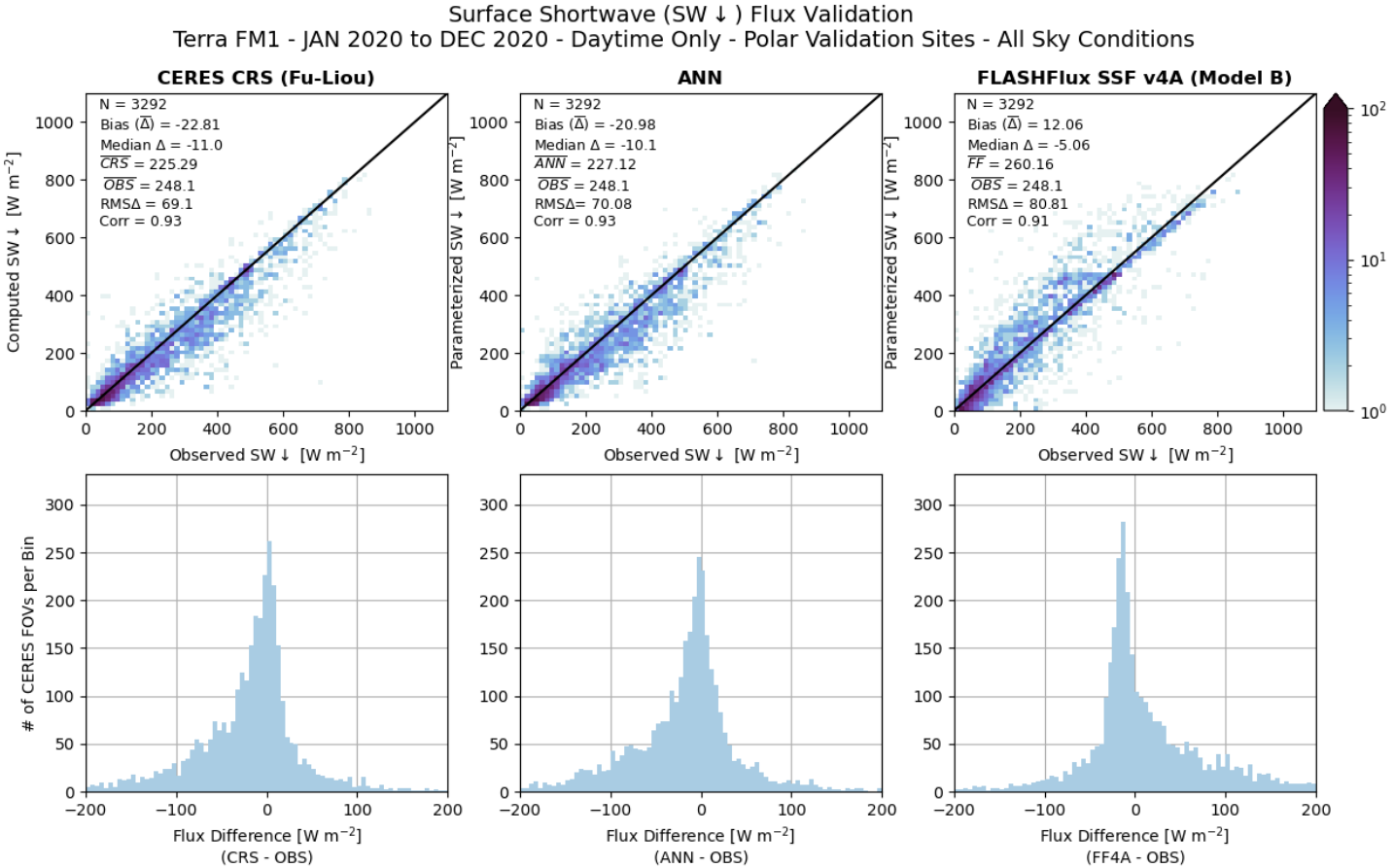
Independent Evaluation against BSRN Surface Measurements, desert sites



Independent Evaluation against BSRN Surface Measurements, island sites



Independent Evaluation against BSRN Surface Measurements, polar sites



Independent Evaluation against BSRN Surface Measurements, buoy sites

