Supplement of Biogeosciences, 19, 1777–1793, 2022
https://doi.org/10.5194/bg-19-1777-2022-supplement

Biogeosciences Open Access EGU

Supplement of

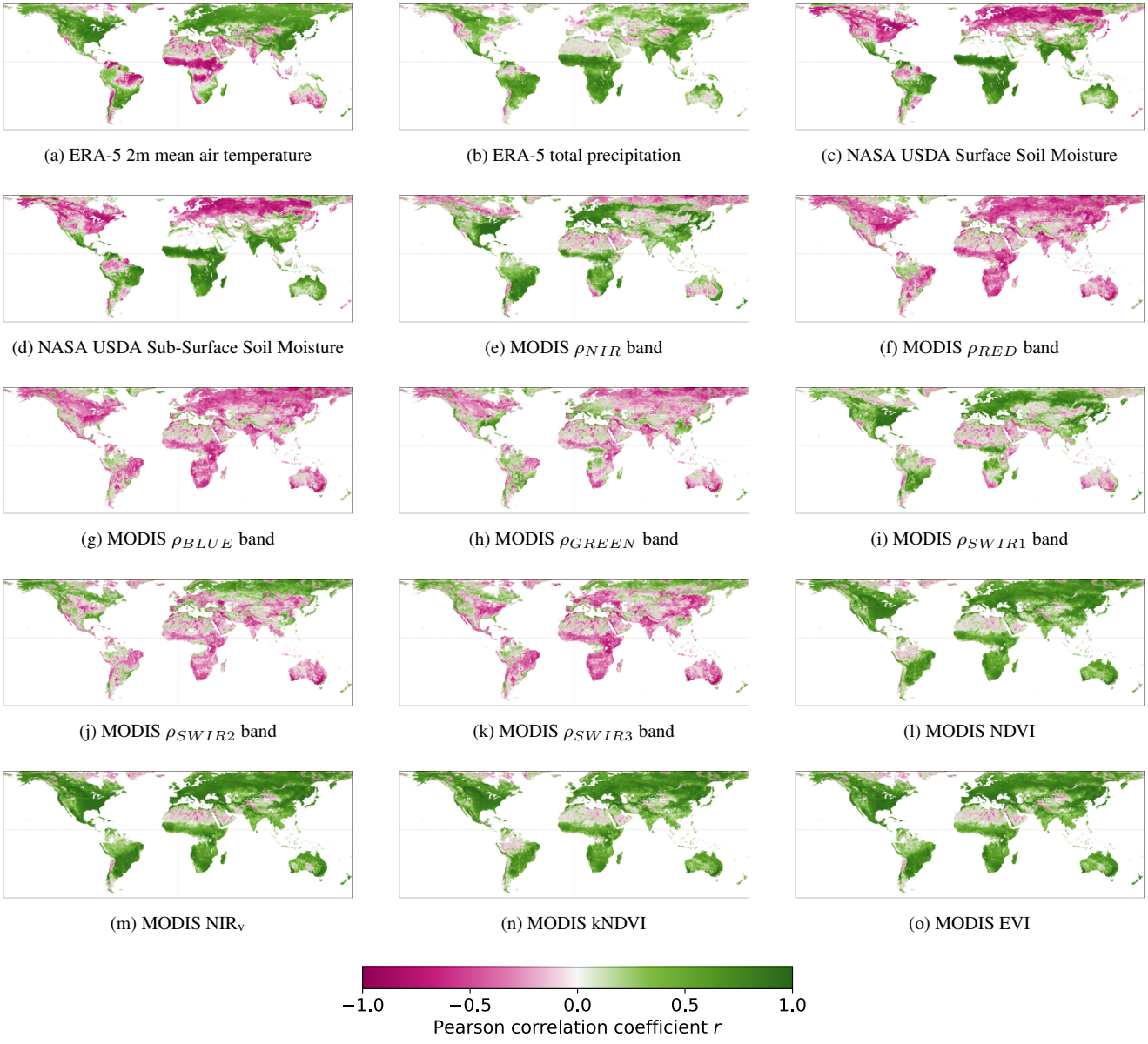
A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet)

Johannes Gensheimer et al.

Correspondence to: Johannes Gensheimer (johannes.gensheimer@bgc-jena.mpg.de), Jia Chen (jia.chen@tum.de), and Alexander J. Turner (turneraj@uw.edu)

S1 Data Correlations of Used Datasets

In Figure S1 the Pearson correlation coefficient on a global map for each grid cell is shown. It stands out that for all products there is only a weak correlation in northern Africa. The ERA-5 air temperature data is in parts very well positive but also negative correlated to SIF. Positive correlations are especially in the northern hemisphere, whereas strong negative correlations appear in southern parts of the world. Soil moisture (Figure S1c and d) is well correlated with SIF in the southern hemisphere (except for central Australia and southern part of South America). In the northern hemisphere, however, the correlation is notably weaker. All MODIS bands show diverse correlation patterns. Combining some of the bands to compute vegetation indices leads to a much stronger correlation to SIF. All vegetation indices have fundamental difficulties to capture the SIF signal in northern Africa or northern North America. In contrast to NIR_v the other vegetation indices show also weak correlations in the amazon region. Correlations show a dependency on latitude and satellite coverage in Figure S2. Especially in regions near the equator the coverage of the MODIS instrument is only around 20-30%. TROPOMI shows also a weaker coverage but is still over 90%. Interestingly, at latitudes around +-22.5° the correlations of SIF to the vegetation indices is weaker but this does not correlate to the reduced satellite coverage. At even higher latitudes the correlations and satellite coverage get weaker.



(a) ERA-5 2m mean air temperature

(b) ERA-5 total precipitation

(c) NASA USDA Surface Soil Moisture

(d) NASA USDA Sub-Surface Soil Moisture

(e) MODIS ρ_{NIR} band

(f) MODIS ρ_{RED} band

(g) MODIS ρ_{BLUE} band

(h) MODIS ρ_{GREEN} band

(i) MODIS ρ_{SWIR1} band

(j) MODIS ρ_{SWIR2} band

(k) MODIS ρ_{SWIR3} band

(l) MODIS NDVI

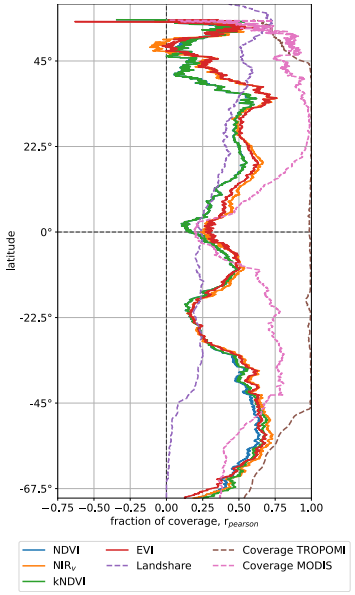
(m) MODIS NIR_v

(n) MODIS kNDVI

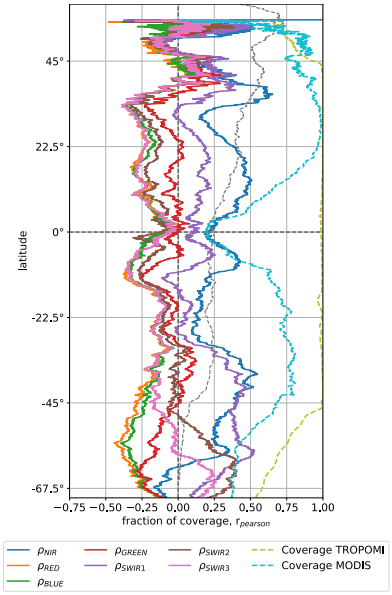
(o) MODIS EVI

(p)

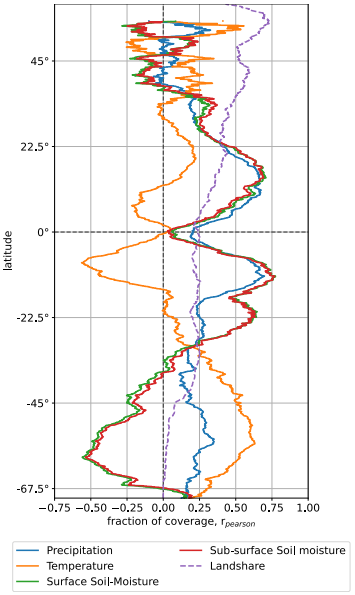
Figure S1. Pearson correlation coefficient of auxilary data to TROPOMI SIF. Data is compared in 16-day resolution starting in April 2018 until March 2021. Comparison in lowest resolution of the two products: For ERA-5 at 0.25°, for USDA at 0.1° and for all MODIS data at 0.05°.



(a) Latitude dependent $r_{pearson}$ of MODIS Vegetation Indices and TROPOMI SIF at 0.05° resolution



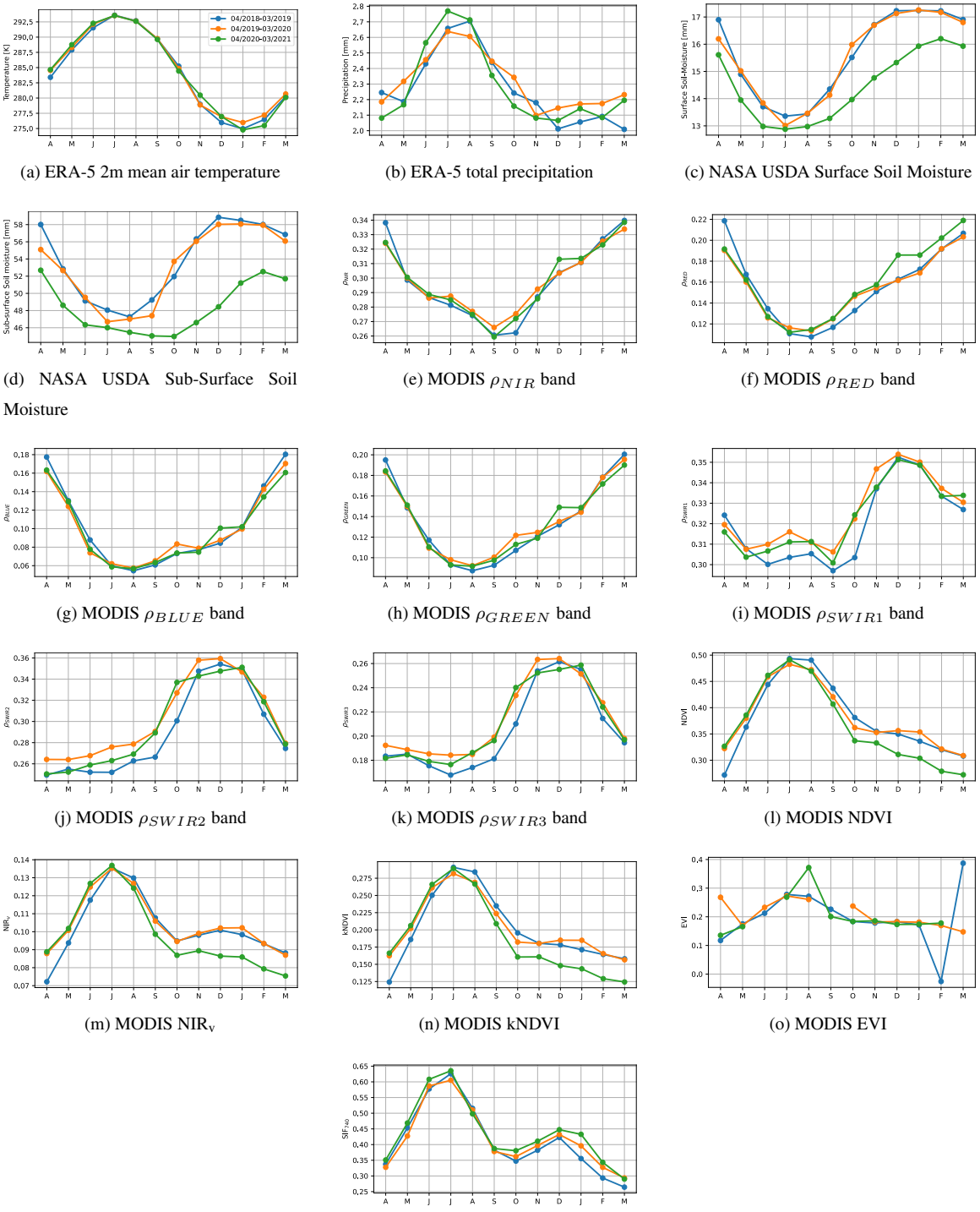
(b) Latitude dependent $r_{pearson}$ of MODIS bands and TROPOMI SIF at 0.05° resolution



(c) Latitude dependent $r_{pearson}$ of ERA-5 total precipitation and temperature and NASA USDA Soil Moisture data and TROPOMI SIF at 0.1° resolution

Figure S2. Latitudinal pearson correlation coefficient of auxiliary data to TROPOMI SIF. Data is compared in 16-day resolution starting in April 2018 until March 2020 and March 2021 for ERA-5 data and all others, respectively. Comparison in lowest resolution of the two products: For ERA-5 and USDA soil moisture at 0.1° and for all MODIS data at 0.05°.

Figure S3 shows the time series of the time variant data sets. Differences between the years are visible for several data sets. Outstanding is the reduction in soil moisture in 2021 (Fig. S3c and d). The MODIS bands show some variance between the years that also influence the vegetation indices. In the three vegetation indices $NDVI$, NIR_v, and $kNDVI$ (Fig. S3l,m,n) the period from September 2020 until March 2021 show the lowest measured values of the three year period. In contrast to that the TROPOMI SIF measurements are highest for this period, indicating a fundamental different measured event. The reduction in soil moisture for this period might also reduce the photosynthetic activity, which is not confirmed by the SIF measurements. The vegetation indices therefore show a very contrary behavior than the SIF measurements.



(a) ERA-5 2m mean air temperature

(b) ERA-5 total precipitation

(c) NASA USDA Surface Soil Moisture

(d) NASA USDA Sub-Surface Soil Moisture

(e) MODIS ρ_{NIR} band

(f) MODIS ρ_{RED} band

(g) MODIS ρ_{BLUE} band

(h) MODIS ρ_{GREEN} band

(i) MODIS ρ_{SWIR1} band

(j) MODIS ρ_{SWIR2} band

(k) MODIS ρ_{SWIR3} band

(l) MODIS NDVI

(m) MODIS NIR_v

(n) MODIS kNDVI

(o) MODIS EVI

(p) TROPOMI SIF

Figure S3. Time Series of the used data sets. Data is shown for three years in monthly resolution and global means.

S2 Study Regions

Figure S4 shows the six study regions on a global land cover map from Buchhorn et al. (2020). There are five folds (Europe and northern Africa, central Asia, east Asia, southern Africa, and South America) that are used for training the CNN. The set over North America is used as the validation set.

Table S1 summarizes the exact longitude and latitude borders of the six regions. In Figure S5 the land cover shares are shown. East Asia and the two folds in the southern hemisphere show significantly higher shares of the land cover type "others" which refers to all non-vegetated lands. This also includes water bodies which take up a big share in these regions (compare Figure S4). Using the same region sizes was a necessary step for efficiently training the model. Regions were selected due to linear correlations (Figure S1) and satellite coverage (Figure S2). Australia due to its big share of shrubs and very noisy signals not considered in this work. Further, the amazon rainforest and the region between fold 1 and fold 4 show very weak satellite coverage of MODIS (Figure S2). Except for these regions the data set covers nearly the whole land share of the earth.

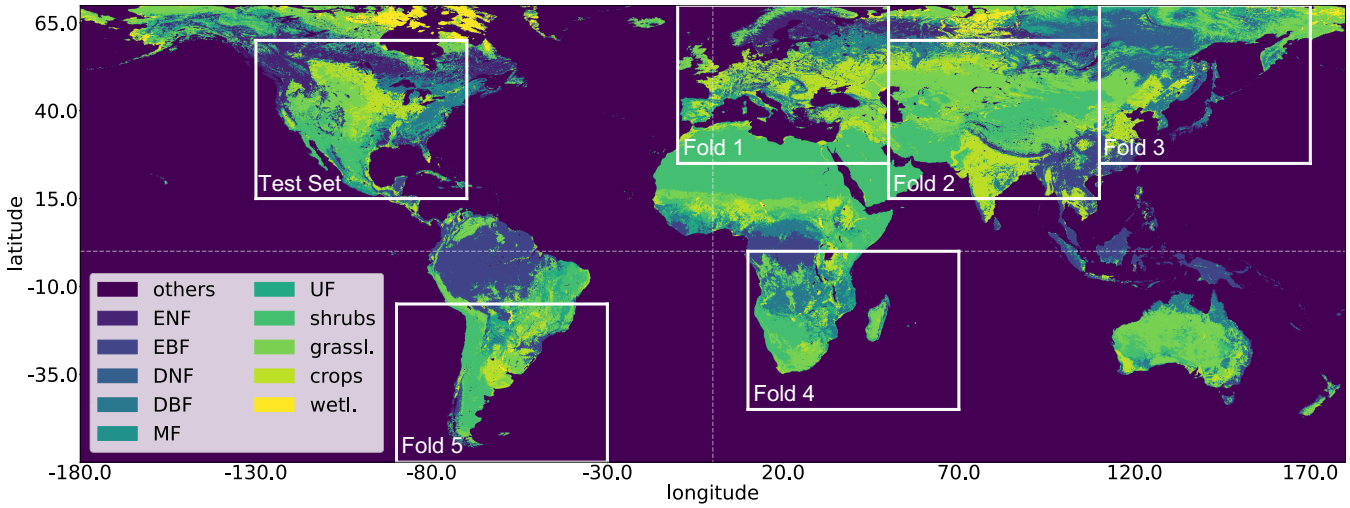


Figure S4. Regions of 5-fold cross validation and test set over dominant land cover type. The land cover type is at a resolution of 0.05° from Corine LULC.

Table S1. Longitude and Latitude borders of cross validation folds

Fold	Longitude borders	Latitude borders	Approximate covered area
1	-10° to 50°	25° to 70°	Europe and North Africa
2	50° to 110°	15° to 60°	Central Asia
3	110° to 170°	25° to 70°	East Asia
4	10° to 70°	-45° to 0°	South Africa
5	-90° to -30°	-60° to -15°	South America
Test set	-126° to -66°	15° to 60°	North America

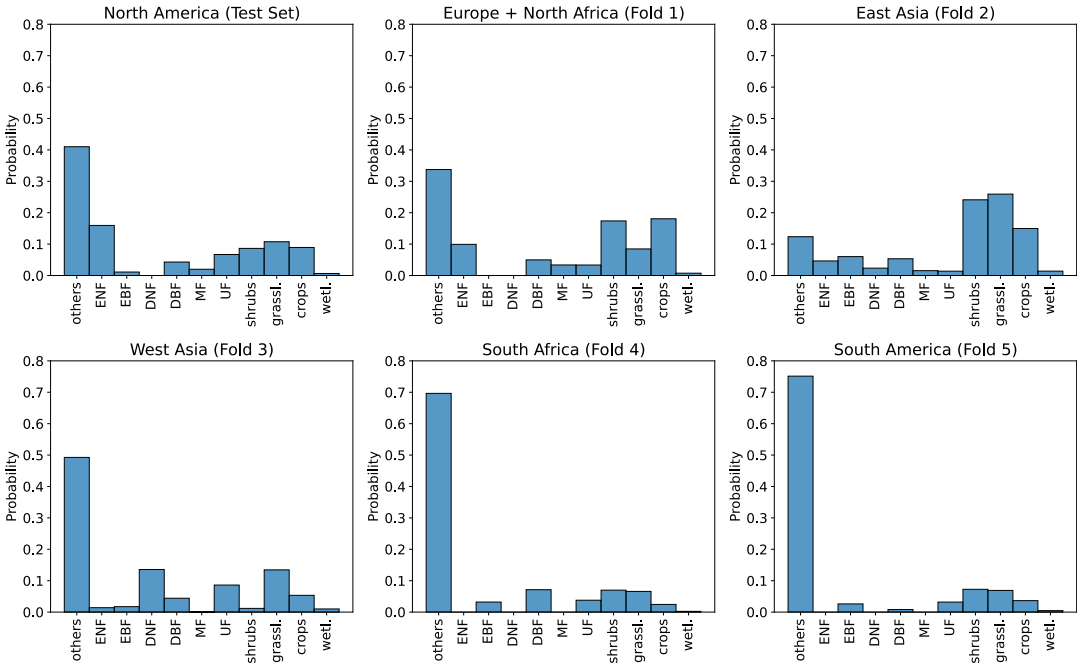


Figure S5. Land cover share in cross validation folds and test fold. Will probably be moved to supplemental.

S3 Input Feature Investigations

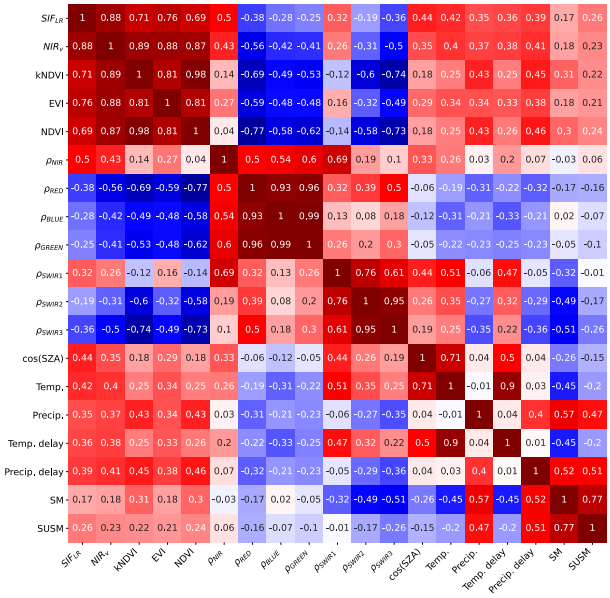
In Figure 1 (main text) high correlations of all vegetation indices to SIF are observed. Vegetation indices are all computed by the same MODIS bands (ρ_{NIR}, ρ_{RED}, and for EVI also ρ_{BLUE}). High common linear correlations of the features to SIF might also indicate a high collinearity between the features, which is investigated in the following section.

Figure S6a and b show the Pearson correlation coefficient of all input features with each other. Notably, is first upscaled to 0.5° and then resampled to 0.05° (same as in training process, compare section 3 of main text). The analysis is divided by time variant and invariant data. Here, the correlations are computed for North America from April 2018 until March 2021 in 16 day

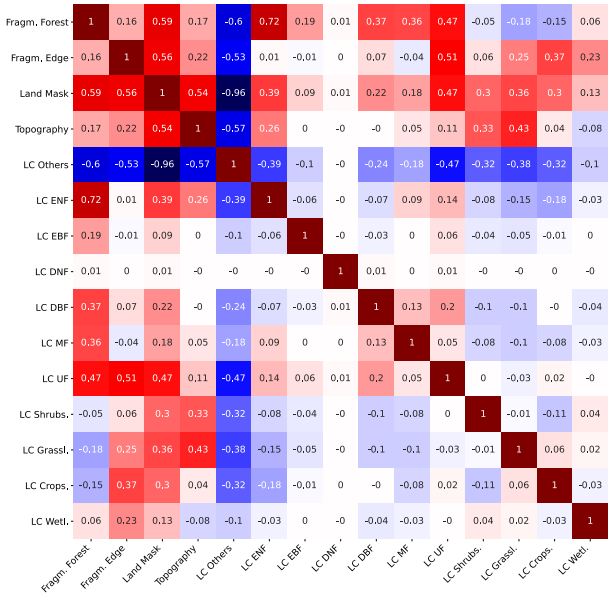
resolutions.

In the time variant features (Figure S6a) a high collinearity between MODIS features (MODIS vegetation indices and individual bands) stands out. Further, moderate correlations between meteorology to MODIS and soil moisture data is detected. Similarly, the cosine of Solar Zenith Angle ($cos(SZA)$) shows moderate correlations to most other features. The time invariant maps (Figure S6b) show a notably weaker collinearity to each other. Forest shares of the fragmentation data show moderate correlations to forest land cover data. The Land Mask is moderate correlated to the fragmentation data and highly correlated to the LC class Others (non-vegetated). Land Mask and the LC class Others are computed from the same data with the difference that Land Mask distinguishes land and water and Others gives the share of non-vegetated in the pixel. Therefore, only in urban areas, which contribute with a rather small area to our global study, these two measure strongly deviate from each other.

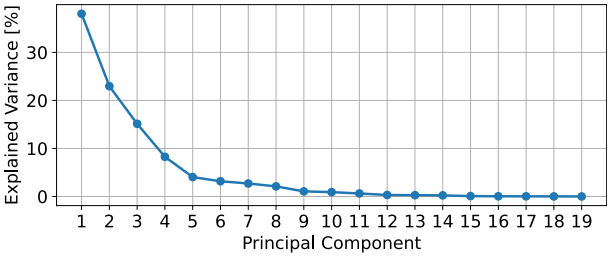
To further investigate the collinearity and co-variance of the feature data a PCA is conducted in Figure S6c and d for the North America Abdi and Williams (2010). PCA is a dimension reduction method which can be applied to investigate how much variance there is in the data set. The PCA applies a singular vector decomposition and computes principle components (PC). Each PC explains a share of variance in the data set (Bishop, 2014). In Figure S6c and d the explained variance of each PC is shown for the time variant and time invariant data, respectively. In the time variant data the first nine PCs carry 99% of the variance in the data (of a total of 19 PCs). Therefore more than half of the data does not carry additional information than the others. In the time invariant data a different pattern is observed. The 99% variance is carried by 13 of the 15 PCs. Therefore the time invariant data is more diverse. Unfortunately, PCA is not interpretable and NNs can handle collinearity, in contrast to e.g. linear regression methods De Veaux and Ungar (1994). Therefore the PCA is not applied as a data pre-processing step.



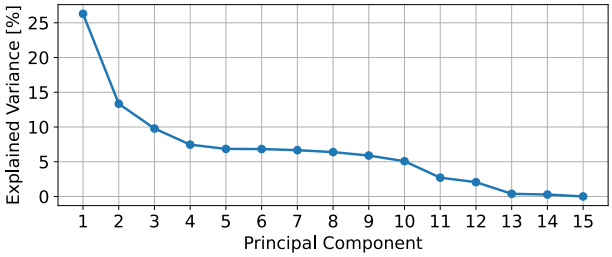
(a) CNN input feature correlation of time variant features



(b) CNN input feature correlation of time invariant features



(c) Explained Variance of PCs for time variant features



(d) Explained Variance of PCs for time invariant features

Figure S6. Input feature Pearson correlation and explained variance by the principal components for North America. (a) shows the feature correlation of time variant variables. (b) shows the feature correlation of time invariant variables (c) shows the explained variance of the PCs of timely changing variables. (d) shows the explained variance of the PCs of static variables. All data is in 0.05° resolution. SIF was first upscaled to 0.5° and then resampled to 0.05° (same as input). Data is investigated from April 2018 until March 2021 in 16 day time steps.

S4 Model Development

S4.1 Hyperparameter Tuning

For hyperparameter tuning the Python library Optuna is used to optimize the parameters learning rate, weight decay, and epoch of SIFnet (Akiba et al., 2019). Here, a TPE-Sampler suggest the parameters of the next trial which is based on a Gaussian Mixture Model. Table S2 summarizes the optimized model parameters.

Figure S7 shows the losses for all investigated hyperparameters that were suggested by Optuna. For visualization a 10-value rolling mean is applied and a threshold of 0.4 is set for the losses as some trials exceeded 2500.

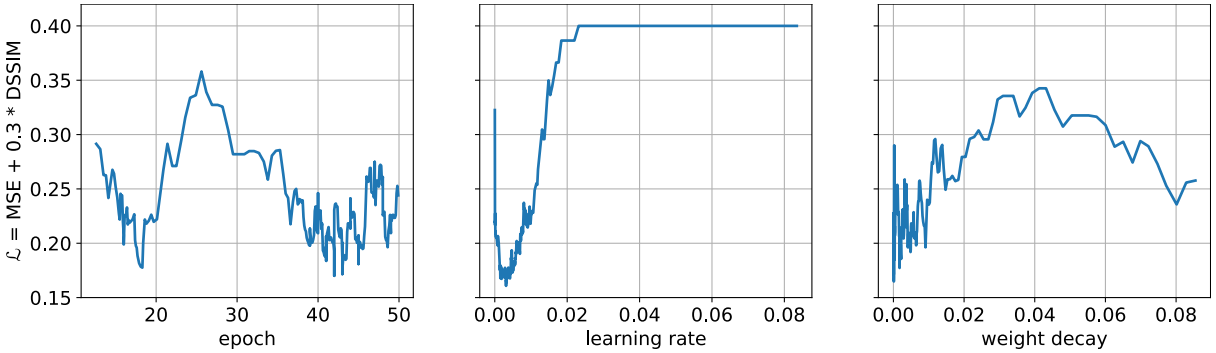


Figure S7. Loss dependent on hyperparameters. y axis shows the losses, x axis the investigated values for the hyperparameters. For visualization a 10-value rolling mean is applied to the data.

Table S2. Model parameters

Parameter	Optimized value
Batchsize	8
Optimizer	Adam
Learning rate	0.0018
Weight decay	0.000057
Epoch	42

S4.2 Cross Validation and Losses of Training Process

In Figure S8a to Figure S8e the results from a cross validation on the five training areas (shown in Figure S4) are presented. All the training runs were conducted with the model parameters shown in Table S2). The region in the title refers to the validation fold. In Figure S8f the losses of the training data (all five folds) and the test data (North America) for all epochs are shown. In contrast to the actual model training the data augmentation technique is also applied to the validation and test set. This was a necessary step as the regions have different shares of ocean in their footprint (compare Figure S5) which results in not comparable losses. Water bodies and missing values are set to zero in the training process, as CNNs can not compute NaNs. Without applying the data augmentation technique (which samples random crops with a maximum share of missing values of 20%) the losses can not be compared between regions. Water bodies are estimated with a high accuracy, as the network only needs to set them to zero.

According to Figure S8a to Figure S8e no sign of overfitting is detected. While the training loss in both Asian regions is lower than the validation loss, for South America and Europe/Northern America the validation loss is lower. In Southern Africa both

losses are in the same order of magnitude. After the cross validation the network is trained with all training folds and validated on the test set over North America (Figure S8f). Also here, the test loss follows the training loss.

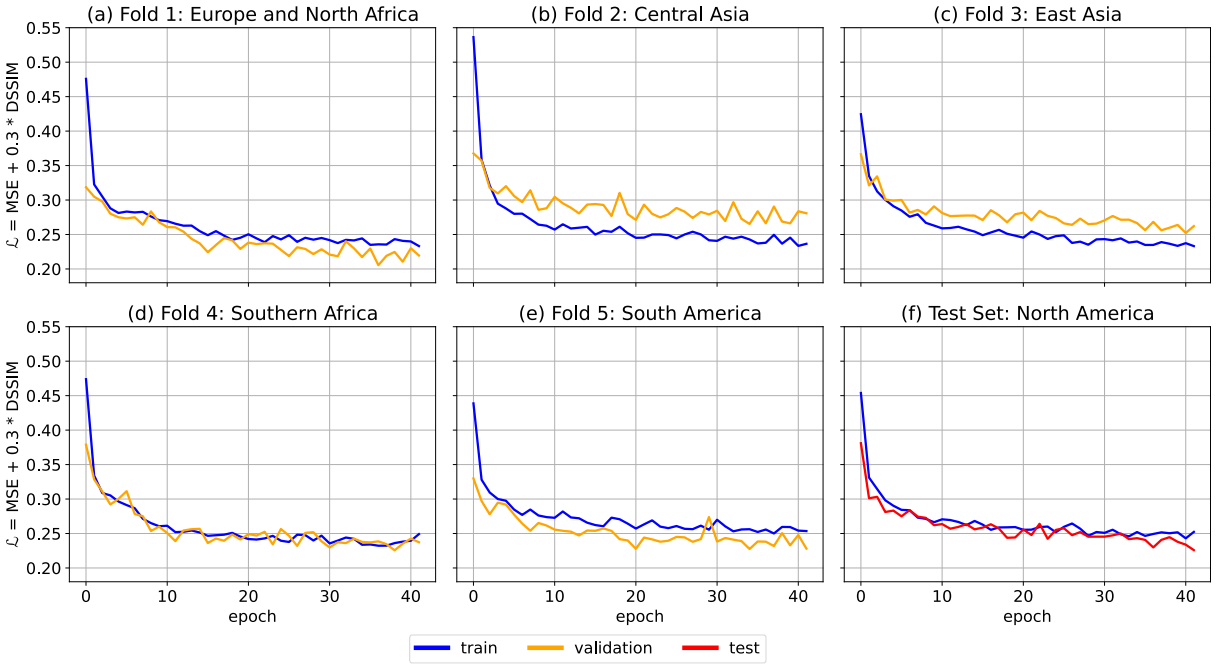
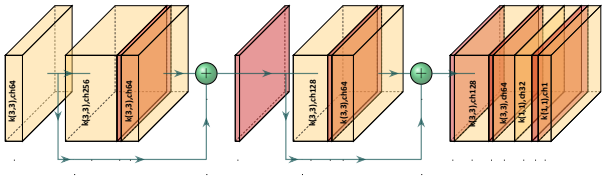


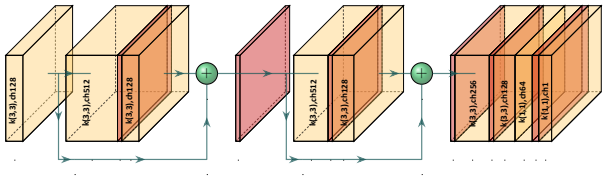
Figure S8. Loss of cross validation. (a) to (e) shows the losses of the 5-fold cross validation. The region in the title of the subfigure refers to the validation fold. All other four folds are used for training. (f) shows the losses of training on all five training folds and testing on the test set.

S4.3 Comparison of Model Structures

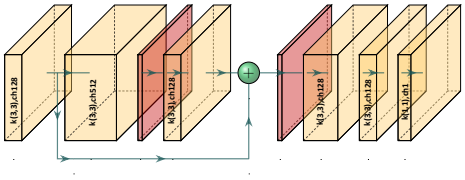
Figure S9 visualizes eight investigated CNN model structures. Model complexity is decreasing from a to h. All models perform similarly well (Table S3). Model structure 5 (e) is chosen as the best trade off between complexity and performance. Further, the input feature collinearity and PCA shown in Section S3 shows that some input features show high correlations with each other which suggests a feature reduction model structure with decreasing the number of channels in the first convolutional layers (in model 5, from 34 to 16 in the first layer).



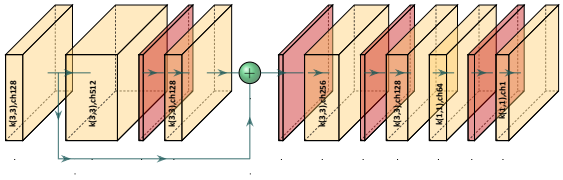
(a) Model Structure 1



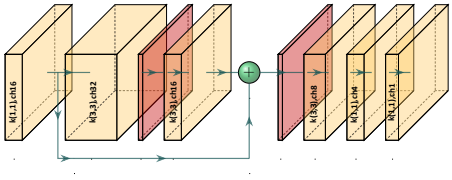
(b) Model Structure 2



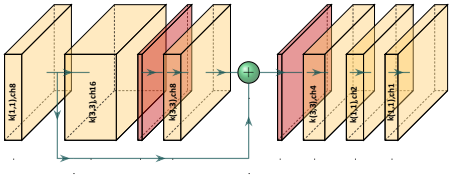
(c) Model Structure 3



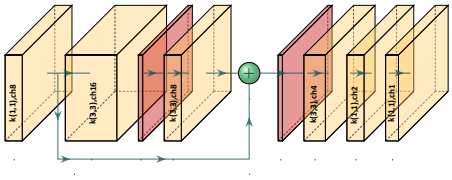
(d) Model Structure 4



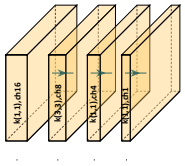
(e) Model Structure 5 (chosen for SIFnet)



(f) Model Structure 6



(g) Model Structure 7



(h) Model Structure 8

Figure S9. Investigated Model Structures. (a) to (h) show compared model structures with decreasing complexity. Results after optimization are shown in Table S3.

Table S3. Results of Model Structures after Optimization for the test set over North America.

Structure	RMSE	SSIM	r^2
1	0.16	0.88	0.93
2	0.16	0.89	0.93
3	0.17	0.87	0.93
4	0.16	0.89	0.93
5 (chosen for SIFnet)	0.17	0.87	0.92
6	0.18	0.88	0.92
7	0.17	0.86	0.92
8	0.21	0.77	0.89

S4.4 Individual loss function

In Figure S10 the mean high resolution estimate of SIF for the San Francisco Bay Area is shown. Further, two modified versions of that image are appended. One with random noise and the other with a constant value added. The RMSE and SSIM metrics are computed for the modified images to the original image. Both modified images show the same RMSE but they deviate strongly in their SSIM. The SSIM metrics is notably lower for the random noise image than for the one with the constant added. This shows that there are multiple versions of the same RMSE deviation. In the loss function used in this work, both metrics are considered and the model is therefore optimized on the overall deviation and on structural similarity from the model output to the ground truth.

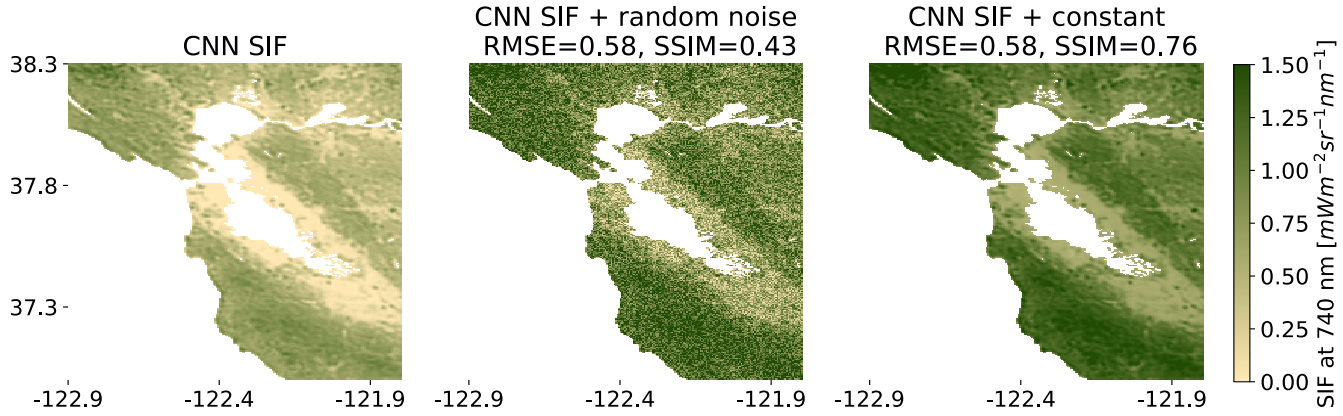
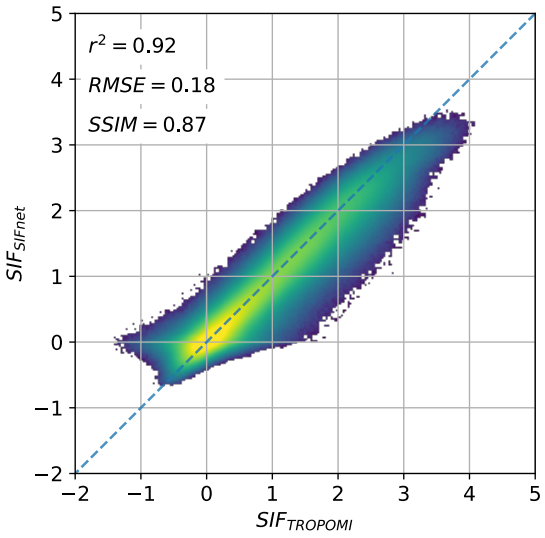


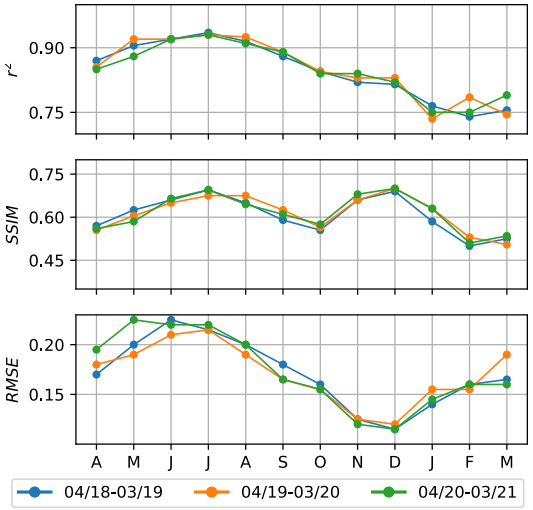
Figure S10. Advantage of own loss function

S4.5 Varying Input Data Combinations

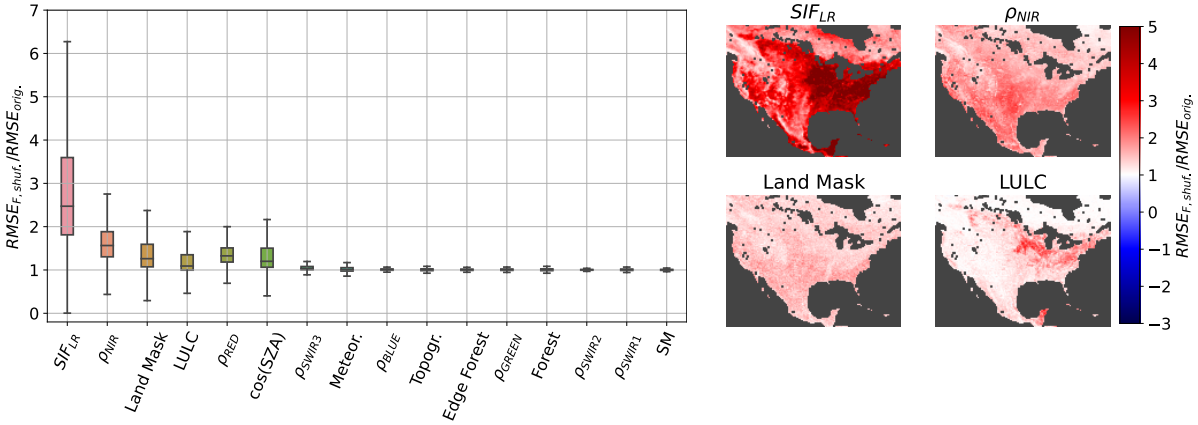
Three further input feature combinations are evaluated in Figures S11, S12, and S14. In the results from Figure S11 the MODIS vegetation indices (NIR_v, $NDVI$, EVI, and $kNDVI$) are not used as features. Therefore, the network needs to get the vegetation information directly from the MODIS bands. Only a minor reduction in performance is observed in the $RMSE$ metrics from 0.17 (with vegetation indices) to 0.18 (without vegetation indices). The metrics $SSIM$ and r^2 are unchanged for North America. Feature importance still suggests the low resolution SIF feature as the most important one, followed by the MODIS NIR band and the land cover information from Buchhorn et al. (2020).



(a) Scatter result and metrics for the test set over North America.



(b) Metrics depend on time.

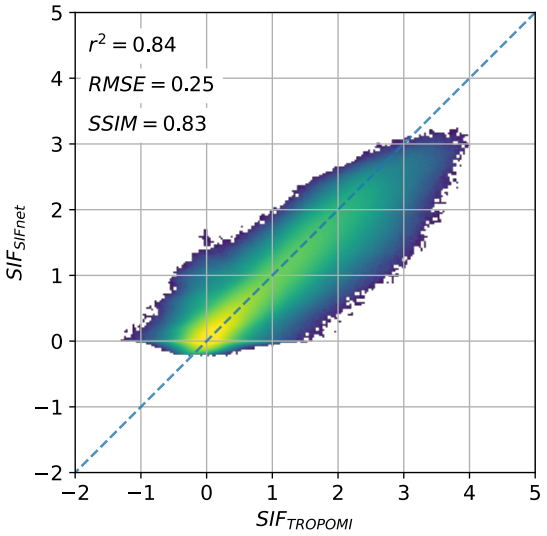


(c) Feature Importance

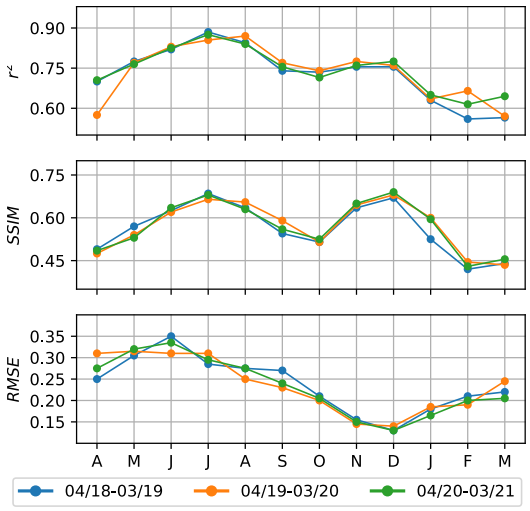
Figure S11. Results of Data Option without any MODIS vegetation indices. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (b) shows the metrics r^2, $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards. (c) shows the feature importance of the optimized SIFnet. Some input variables are clustered and all variables of that class are permuted at the same time. LULC: all eleven land cover classes; SM: surface soil moisture and sub-surface soil moisture; Mereor.: temperature, precipitation, temperature with 16 day delay, precipitation with 16 days delay.

Figure S12 shows the results for North America, if the low resolution SIF feature is not considered as a input variable. Interestingly, in contrast to the findings from Figure 5 (main text) the vegetation index NIR_v is not considered by the model anymore. Rather it relies strongly on the cosine of the solar zenith angle, the MODIS NIR band and the two vegetation indices $kNDVI$ and $NDVI$. NIR_v is strongly linear correlated to the other vegetation indices (Figure S6).

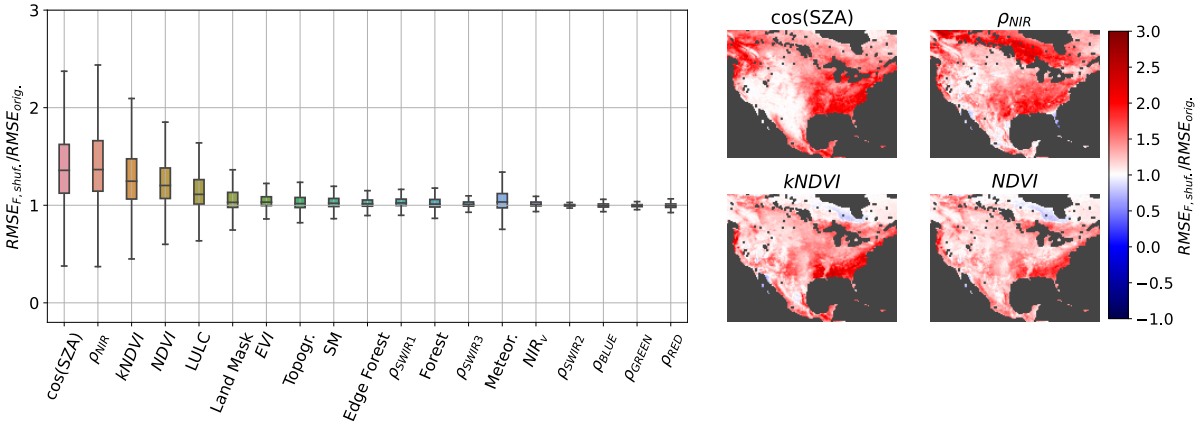
These findings suggests that there are several solutions to the optimization problem that depend on the input feature combination. Reasons for this can be the high collinearity between the features (Section S3). Further, CNNs can process non-linear relationships and can therefore compute the vegetation indices or similar indices by themselves. But also optimizing another CNN model structure (structure 2 in Figure S9) results in a similar feature importance (shown in Figure S13). The vegetation index $kNDVI$ stays more important than NIR_v. But the model relies most on $cos(SZA)$.



(a) Scatter result and metrics for the test set over North America.

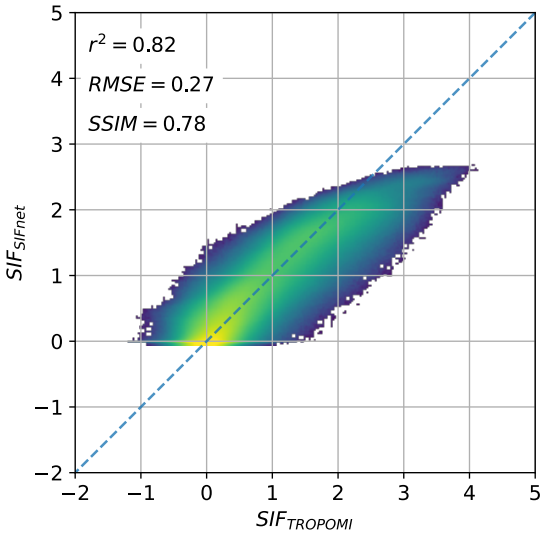


(b) Metrics depend on time.

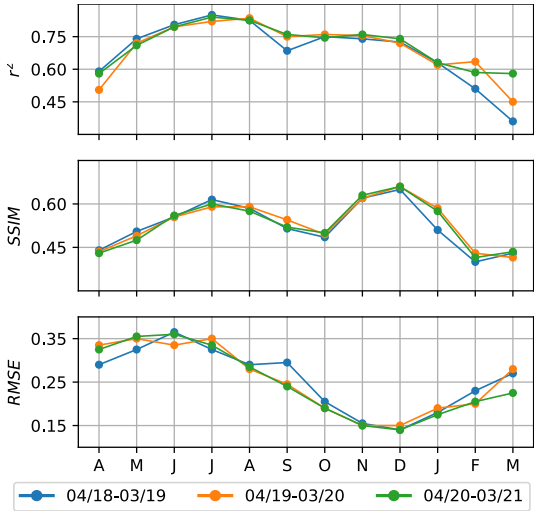


(c) Feature Importance

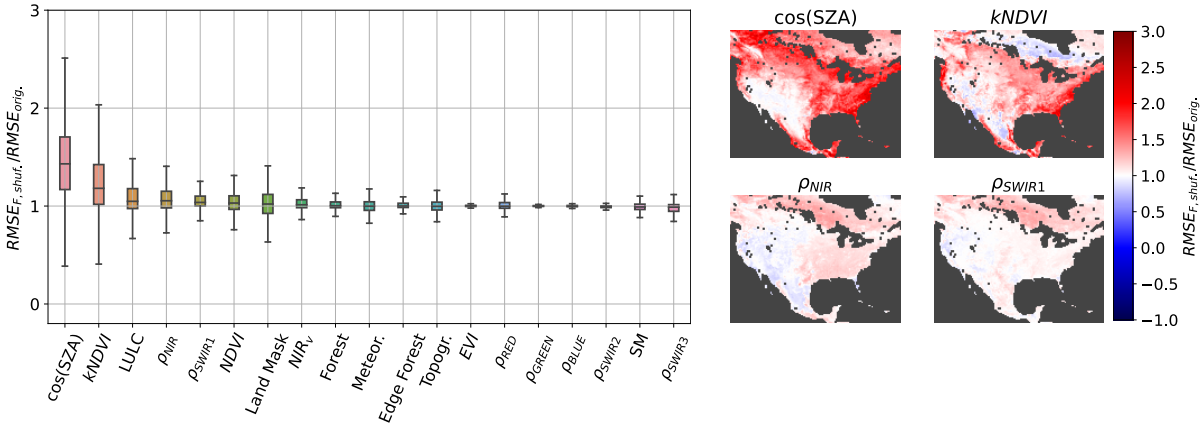
Figure S12. Results of Data Option without low resolution SIF as feature. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (b) shows the metrics r^2, $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards. (c) shows the feature importance of the optimized SIFnet. Some input variables are clustered and all variables of that class are permuted at the same time. LULC: all eleven land cover classes; SM: surface soil moisture and sub-surface soil moisture; Mereor.: temperature, precipitation, temperature with 16 day delay, precipitation with 16 days delay.



(a) Scatter result and metrics for the test set over North America.



(b) Metrics depend on time.

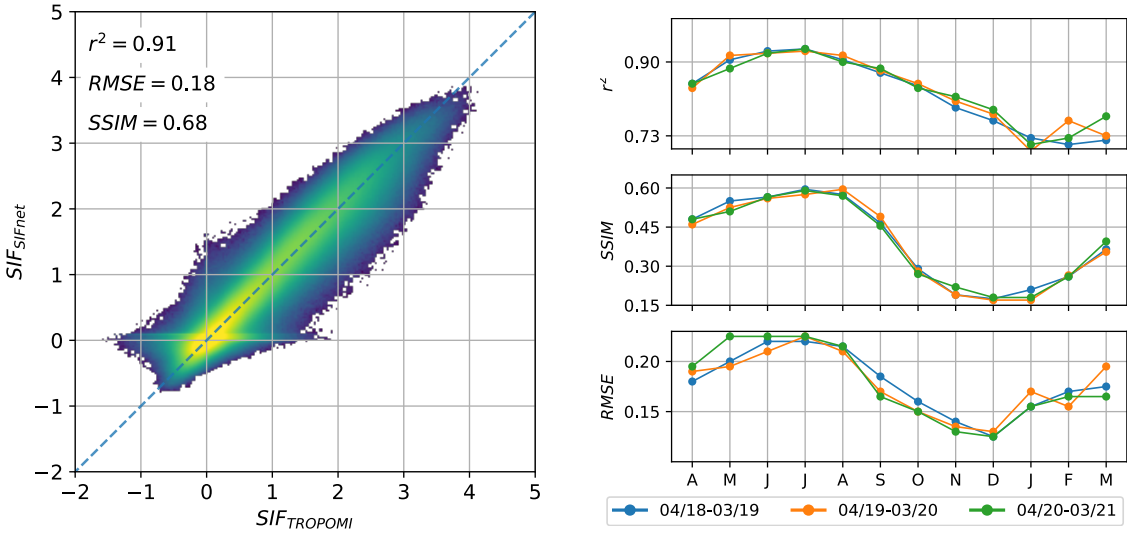


(c) Feature Importance

Figure S13. Results of Data Option without low resolution SIF as feature with model structure 2 from Figure S9 (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (b) shows the metrics r^2, $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards. (c) shows the feature importance of the optimized SIFnet. Some input variables are clustered and all variables of that class are permuted at the same time. LULC: all eleven land cover classes; SM: surface soil moisture and sub-surface soil moisture; Mereor.: temperature, precipitation, temperature with 16 day delay, precipitation with 16 days delay.

Figure S14 shows the results for North America, if only the low resolution SIF (SIF_{LR}) and the vegetation index NIR_v are considered as input to the model. These have been shown to be the two most important predictors (Figure 5 main text). Interestingly, the metrics r^2 and $RMSE$ only worsen by 0.01 each. In contrast to that the $SSIM$ reduces from 0.87 to 0.68

in average (here the average SIF signal of the three investigated years are compared). The importance of $SSIM$ is explained in Figure S10. The overall deviation of the model output to the validation data is similar to the model output with all input features, but the structural patterns deviate strongly. This finding suggests that the auxiliary data beyond NIR_v especially contribute to estimating structural similarity (e.g., $cos(SZA)$ for very high latitudes).



(a) Scatter result and metrics for the test set over North America.

(b) Metrics depend on time.

Figure S14. Results of Data Option with only SIF_{LR} and NIR_v as features. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (b) shows the shows of each investigated month the metrics r_2, SSIM, and RMSE. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards.

S4.6 Investigating Different Scaling Factors during Training the CNN

Higher scaling factors between low and high resolution SIF mean that in the optimization process a model is trained to resolve to a higher spatial resolution from the initial situation of 0.05°. Using a scaling factor of 20 or 50 estimation resolution of 0.0025° and 0.001° can be achieved, respectively.

Using a upscaled 1° SIF product as input, which trains the model to resolve with a scaling factor of 20, the r^2 reduces to 0.9 (Figure S15). Increasing the scaling factor further to 50, where the input SIF is at a resolution of 2.5° the r^2 metrics decreases to 0.87. A potential estimated high resolution SIF product would be at 0.001° (appr. 100 m) resolution. Interestingly, leaving the coarse SIF feature out of the model results in an r^2 of 0.84 (Figure S16) which is lower than using this very coarse 2.5° SIF product. The 2.5° SIF still contributes to the result as the second most important feature following NIR_v. The SIF product might contribute by providing information about radiation, seasonal cycle or the order of magnitude of SIF in the area.

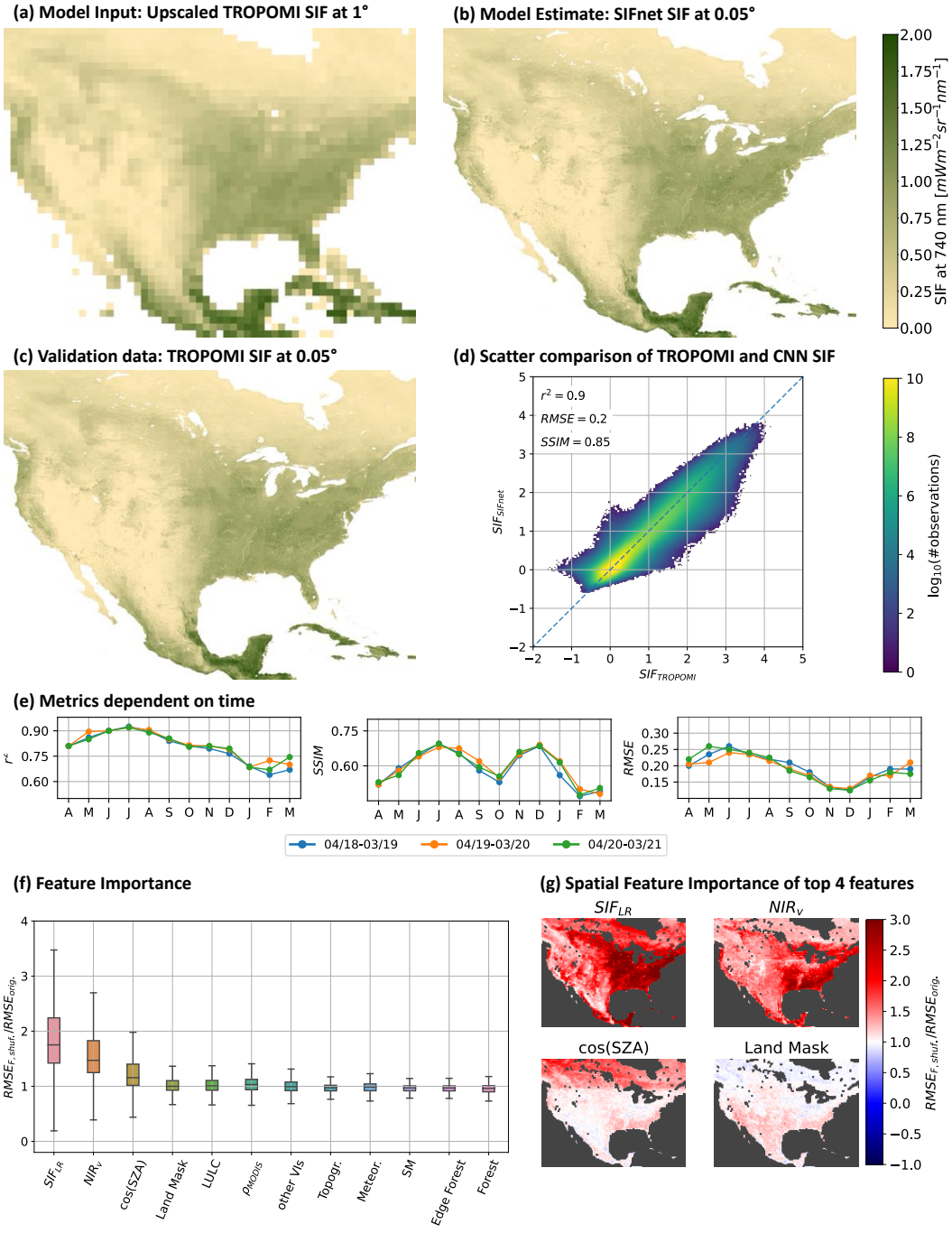


Figure S15. Test set results of SIFnet training at 0.05° with a scaling factor of 20. (a) shows low resolution SIF that is used as model input. (b) shows the estimated SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (e) shows the shows of each investigated month the metrics r^2, $SSIM$, and $RMSE$. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards. (f) shows the feature importance.

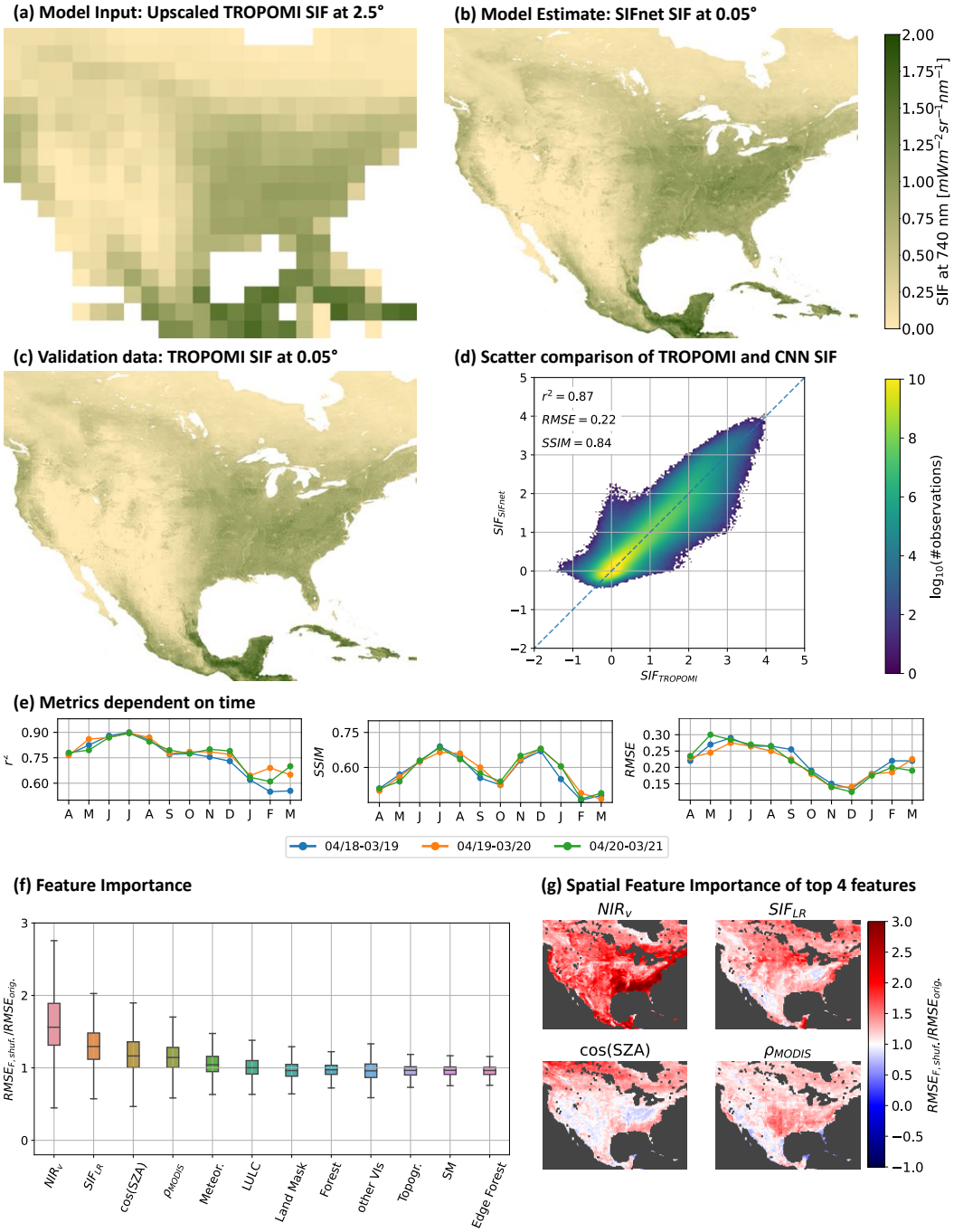


Figure S16. Test set results of CNN training at 0.05° with a scaling factor of 50. (a) shows low resolution SIF that is used as model input. (b) shows the estimated SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05°. (e) shows the shows of each investigated month the metrics r^2, $SSIM$, and $RMSE$. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards. (f) shows the feature importance.

S5 Estimation of SIF at 0.005° Resolution and its Comparison to Downscaled SIF

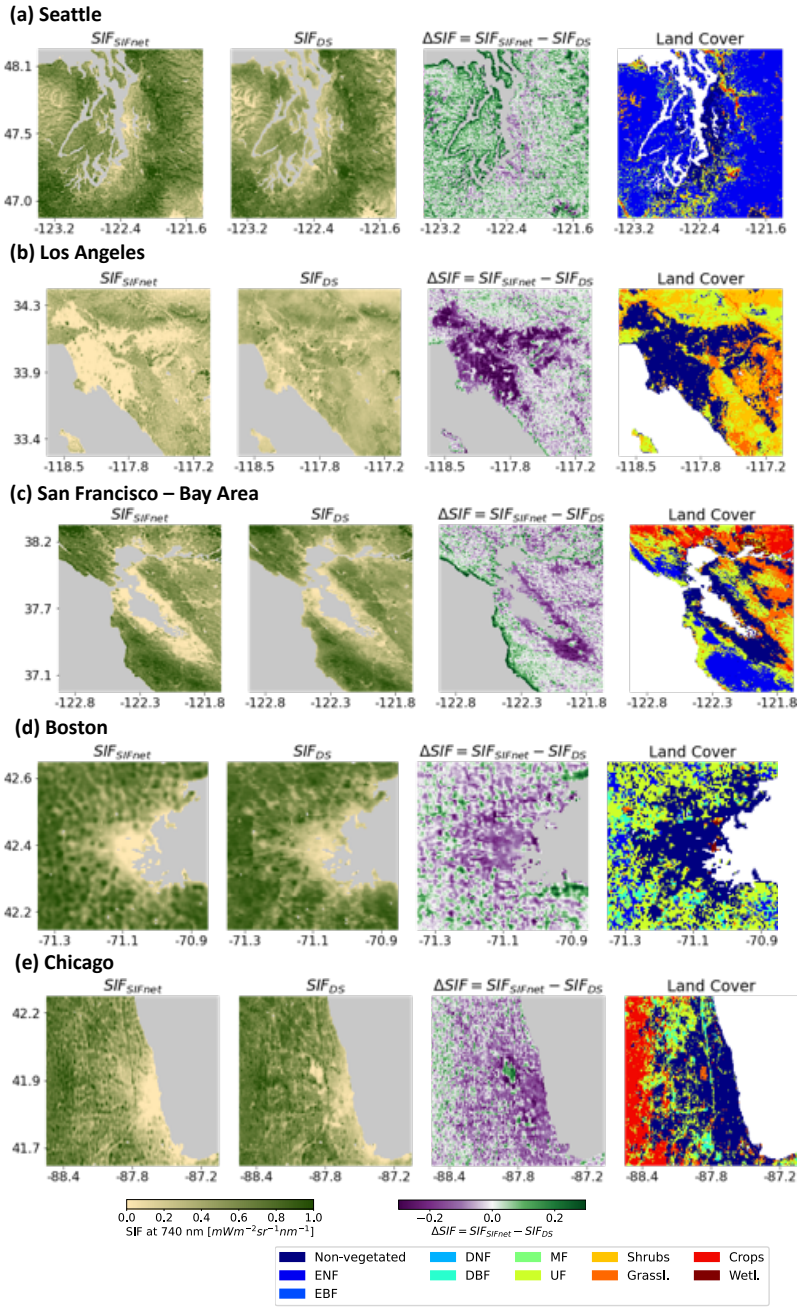


Figure S17. SIFnet and downscaled SIF, the difference between these, and the dominant land cover type for five urban regions across CONUS. First column shows the SIFnet estimate, second the downscaled SIF from Turner et al. (2020), third the difference between SIFnet and downscaled SIF and the last shows the dominant land cover type from Buchhorn et al. (2020).

S6 Training SIFnet on daily integrated SIF estimates

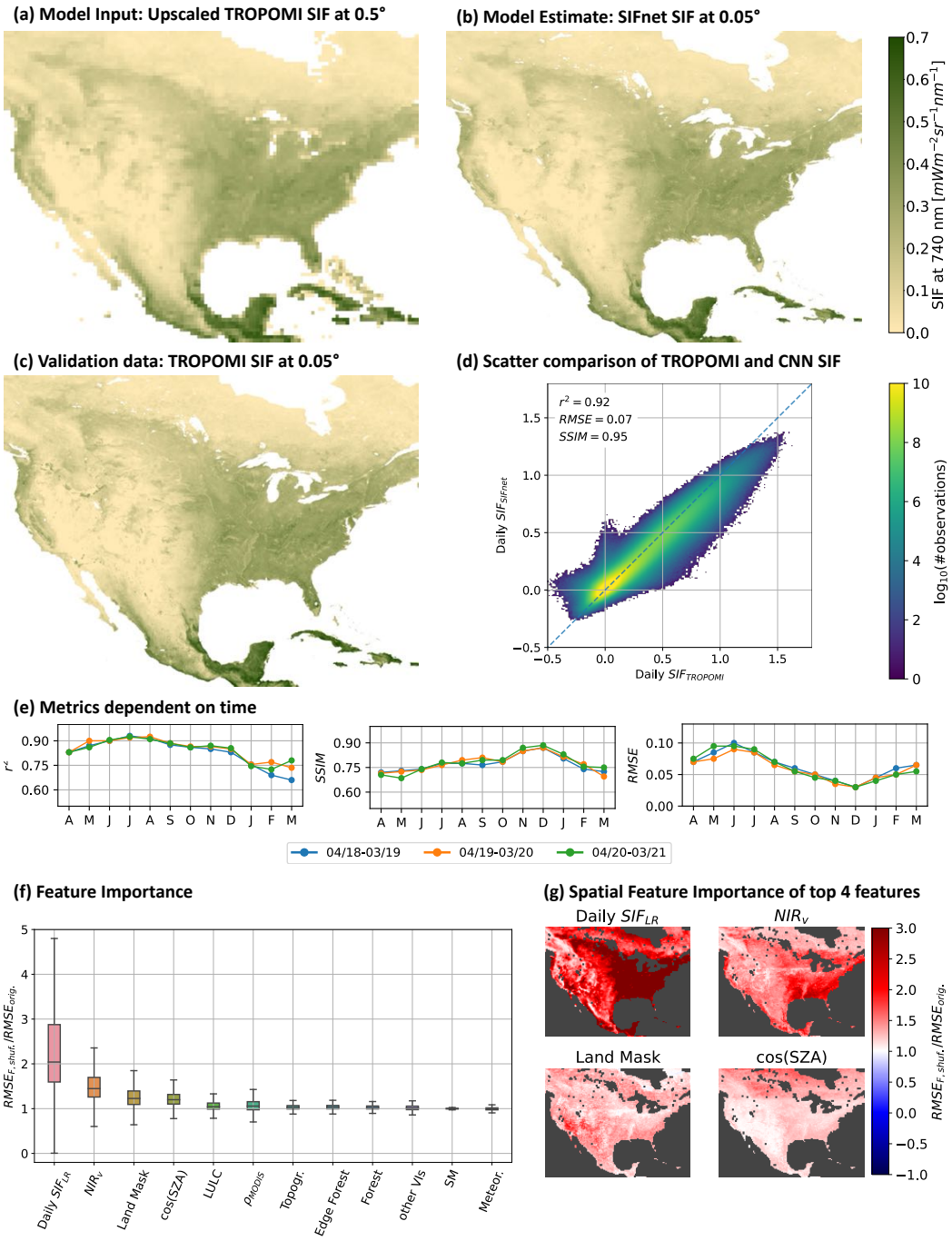
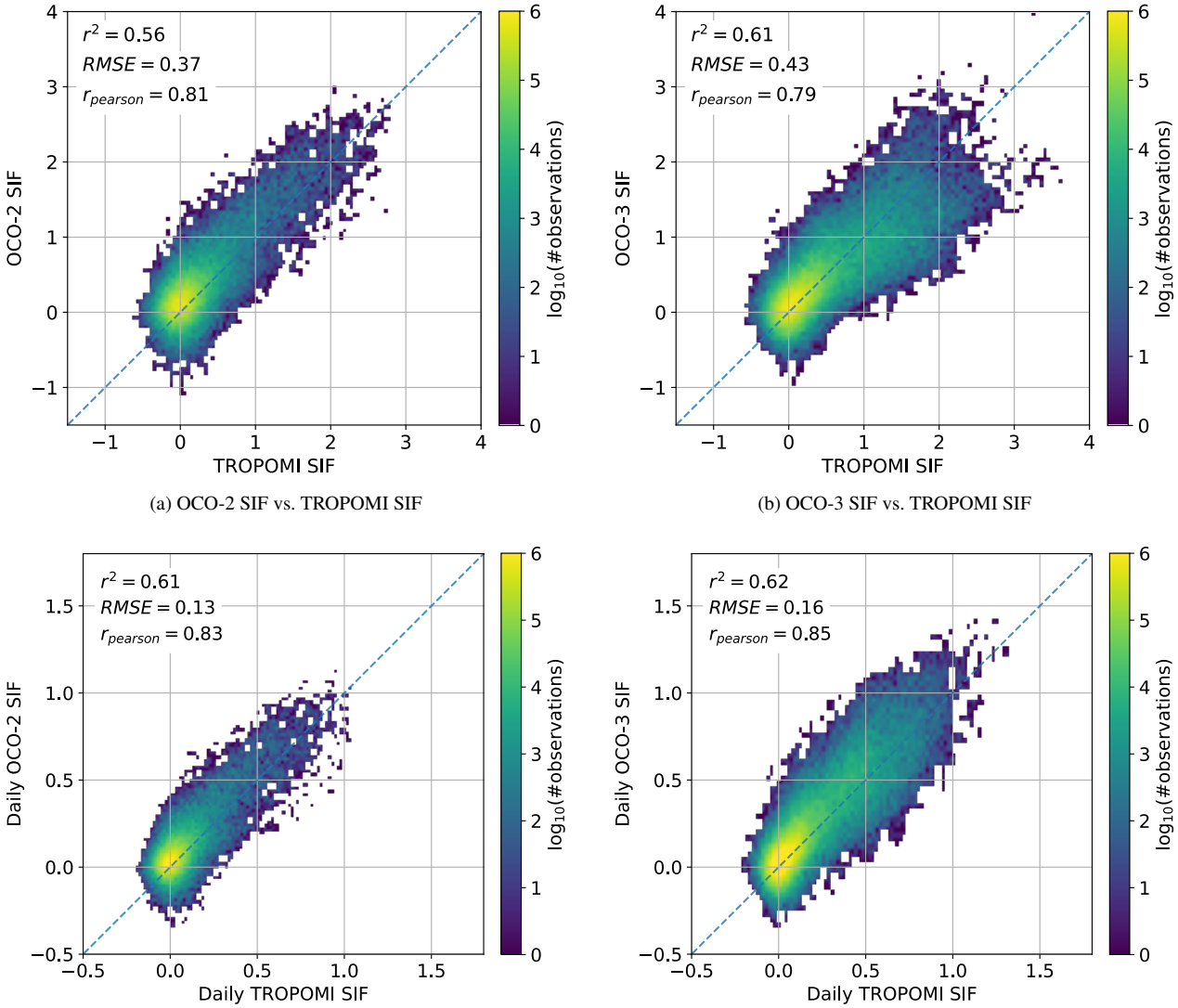


Figure S18. Test set results of SIFnet training at 0.05° using daily SIF estimates (a) shows low resolution Daily SIF that is used as model input. (b) shows the estimated Daily SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI Daily SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between TROPOMI Daily SIF and the SIFnet estimate at 0.05°.

Figure S18 shows the result of training on daily SIF estimates. For estimating daily corrected SIF the instantaneous measurement is scaled by the cosine of the solar zenith angle (Equation 7 main text). r^2 is similar to the approach of instantaneous SIF and the $SSIM$ shows a better result. RMSE is lower as the value range of the daily SIF is lower than that of instantaneous SIF.

S7 Validation against OCO-2 and OCO-3 data



(a) OCO-2 SIF vs. TROPOMI SIF

(b) OCO-3 SIF vs. TROPOMI SIF

(c) Daily OCO-2 SIF vs. daily TROPOMI SIF

(d) Daily OCO-3 SIF vs. daily TROPOMI SIF

Figure S19. OCO-2 and OCO-3 SIF vs. TROPOMI SIF at 0.05° spatial resolution over North America. OCO SIF is gridded to a 0.05° grid by the center point of the pixel. Latitudinal borders are from 25° to 50° and longitudinal border are from -126° to -66°. OCO-2 comparison of a three year time period from April 2018 until March 2021 in 16 days averages. OCO-3 comparison from August 2019 until March 2021. Daily correction according to Equation 6 from the main text (Frankenberg et al., 2011)

Figure S19 shows the scatter comparison of TROPOMI SIF and OCO-2 and OCO-3 SIF on a 0.05° grid over North America. Figs. S19a and S19b compare instantaneous SIF and Figs. S19c and S19d daily corrected SIF according to Equation 6 from the

main text (Frankenberg et al., 2011). OCO SIF is gridded to a 0.05° grid by the center point of the pixel. Latitudinal borders are from 25° to 50° and longitudinal border are from -126° to -66.

Applying the daily correction to the SIF data, the r^2 correlation coefficient increased from 0.56 to 0.61 and 0.61 to 0.62 between TROPOMI SIF and OCO-2 and OCO-3 SIF, respectively. Indeed, one might expect better correlations here as both present SIF at 740 nm. However, as pointed out in Köhler et al. (2018), the uncertainty of both TROPOMI and OCO-2 SIF are expected to lead to a certain spread between the data sets. In addition, we do not account for differences in acquisition times and viewing-illumination geometry, which can lead to additional uncertainties in this comparison. For reference, when comparing single footprints of TROPOMI SIF to aggregated OCO-2 SIF for June 2018 globally, Köhler et al. (2018) found a r^2 of 0.67, only additional aggregation leads to a r^2 of 0.88. The mean deviation of TROPOMI SIF to OCO-2 SIF is close to the average standard deviation of TROPOMI SIF (0.4 $mWm^{-2}sr^{-1}nm^{-1}$). In our analysis, from the 16 day product from TROPOMI SIF for April 2018 until March 2021 at 0.05°, we observe an average error in the TROPOMI SIF of 0.43 $mWm^{-2}sr^{-1}nm^{-1}$ for the CONUS. That error is close to the RMSE between instantaneous TROPOMI SIF and instantaneous OCO-2 SIF (0.37 $mWm^{-2}sr^{-1}nm^{-1}$). To compare TROPOMI and OCO-2/3 SIF we aggregate the OCO-2/3 footprints to the same grid as our TROPOMI data (0.05°). As we aggregate multiple OCO-2 or OCO-3 footprints to match one TROPOMI grid cell at 0.05° the certainty of the OCO measurements increases, and therefore the RMSE between TROPOMI and OCO SIF decreases.

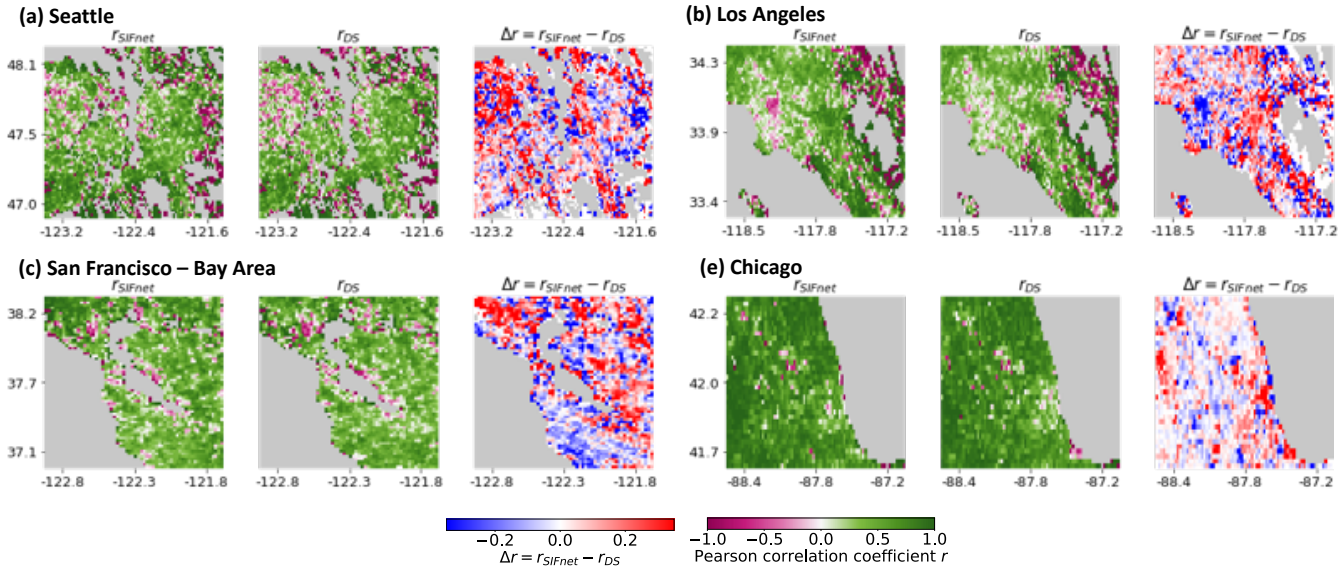


Figure S20. SIFnet and downscaled SIF, the difference between these, and the difference in correlation to OCO-2 and OCO-3 for four urban regions. First column shows the correlation of SIFnet SIF to OCO-X, second the correlation of downscaled SIF from Turner et al. (2020) to OCO-X, third the difference in correlation of the high resolution SIF estimates to OCO-X data on a 0.02° grid and fourth the difference between SIFnet and downscaled SIF.

Table S4. Pearson Correlation Coefficient of SIFnet and Downscaled SIF vs. OCO-2 and OCO-3 SIF dependent on Land Cover Type After Daylength Correction. At least 60% of the grid cell need to be covered with the specific land cover type. There is no data on evergreen broadleaf (EBF) and decidious needle leaf (DNF) available.

	SIFnet			Downscaled			#observations		
	OCO-2	**OCO-3**	**combined**	**OCO-2**	**OCO-3**	**combined**	**OCO-2**	**OCO-3**	**combined**
Non-vegetated	0,77	0,81	0,79	0,71	0,72	0,72	130,309	70,852	201,161
ENF	0,79	0,82	0,80	0,73	0,75	0,74	291,620	162,491	454,111
DBF	0,79	0,83	0,8	0,73	0,76	0,74	233,032	109,318	342,350
MF	0,55	0,60	0,56	0,50	0,52	0,51	11,436	5,520	16,956
UF	0,82	0,84	0,82	0,73	0,76	0,74	292,723	141,232	433,955
Shrubs	0,57	0,63	0,59	0,50	0,56	0,52	943,924	525,789	1,469,713
Grassl.	0,59	0,65	0,61	0,53	0,60	0,55	1,077,786	505,942	1,583,728
Crops	0,75	0,75	0,75	0,70	0,69	0,70	788,982	409,494	1,198,476
Wetl.	0,82	0,86	0,84	0,76	0,71	0,73	7,541	9,800	17,341

Table S5. Pearson Correlation Coefficient of SIFnet and Downscaled SIF vs. OCO-2 and OCO-3 for Different Cities.

	SIFnet			Downscaled			#observations		
	OCO-2	**OCO-3**	**combined**	**OCO-2**	**OCO-3**	**combined**	**OCO-2**	**OCO-3**	**combined**
San Francisco	0,55	0,47	0,49	0,51	0,46	0,47	7669	19906	27575
Los Angeles	0,33	0,35	0,34	0,3	0,33	0,32	15026	24277	39303
New York City	0,72	0,73	0,69	0,67	0,69	0,65	2500	2568	5068
Seattle	0,47	0,43	0,44	0,34	0,44	0,37	10842	15485	26327
Chicago	0,75	0,73	0,73	0,69	0,71	0,69	15519	12913	28432
Boston	0,8	0,77	0,79	0,69	0,73	0,69	1367	1505	2872

References

Abdi, H. and Williams, L. J.: Principal component analysis, Wiley interdisciplinary reviews: computational statistics, 2, 433–459, 2010.

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M.: Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 2623–2631, 2019.

Bishop, C.: Bishop-Pattern Recognition and Machine Learning-Springer 2006, Antimicrob. Agents Chemother, pp. 03 728–14, 2014.

Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: Land Cover 100m: collection 3: epoch 2018: Globe, https://doi.org/10.5281/zenodo.3518038, 2020.

De Veaux, R. D. and Ungar, L. H.: Multicollinearity: A tale of two nonparametric regressions, in: Selecting models from data, pp. 393–402, Springer, 1994.

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., et al.: New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity, Geophysical Research Letters, 38, 2011.

Köhler, P., Frankenberg, C., Magney, T. S., Guanter, L., Joiner, J., and Landgraf, J.: Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: First results and intersensor comparison to OCO-2, Geophysical Research Letters, 45, 10–456, 2018.

Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: A double peak in the seasonality of California's photosynthesis as observed from space, Biogeosciences, 17, 405–422, 2020.