

Response to Reviewer #2:

Dear reviewer #2,

We are very grateful to your very helpful and insightful suggestions which have significantly improved the manuscript.

The modifications made during this revision can be seen in the **appended 'track change' version of the revised manuscript**. Our responses (in black) to the comments/suggestions (in blue) please find below. We also included the major changes in the manuscript to address the comments (in the form of tables).

Sincerely,

Johannes Gensheimer on behalf of all co-authors

Reviewer #2 comments:

This manuscript presents a downscaled SIF product 'SIFnet' (x10 improvement to 500m) based on combining TROPOMI native low resolution gridded SIF, climate products, and remotely sensed gridded products using a convolutional neural network approach. The method development and product verification are well laid out and justified, with a comparison to a recent downscaled SIF product nicely incorporated. I am not a ML expert, so cannot comment on the specifics of the algorithm chosen, apart from that a substantial and commendable effort was made by the authors to: check model sensitivity to the input parameters (including developing a feature importance metric), overfitting, multicollinearity, validation, and intercomparison to other products.

We appreciate the reviewer's comments and feedback.

My main suggestions are to: i) incorporate more information on data filtering and quality assurance, including data coverage and missing values and their impact on final product values, ii) incorporating a more physiological perspective on what SIF is, and outline the implications of how choices in the methodology affect the interpretation of SIFnet SIF, iii) include a more detailed explanation for the apparent low correlation of TROPOMI SIF with OCO SIF and the implications for those using SIF products for GPP proxies, and iv) detailing why solar radiation at surface level was not included as an input as one of the main drivers of SIF and APAR (i.e. taking into account clouds). These general comments and more specific minor comments are found below.

General comments

Data filtering: the information provided is severely lacking. First, be specific about what data filtering methods and thresholds were applied for each dataset, and then specify the amount of data that was filtered (or available) after filtering. This is important on many fronts, especially as filtering will be non-uniformly distributed across space and time (e.g. cloudiness in the tropics, snow affected regions in winter etc.). Also, there will be a strong bias towards clear sky conditions as the reflectance indices won't produce valid data with clouds, yet SIF may still be retrievable through some clouds.

We have added the following part to the manuscript (L.163-171):

During training, all missing values in the data are set to zero. That mainly affects water regions, as the share of missing values in the used SIF data is caused by 91.2% by water. In case there is a missing value in the SIF training sample, all feature values of this pixel are also set to zero to ensure the network does not learn false relationships between the predictors and the target variable. For the MODIS bands we applied the quality index value 0 (best quality only). This filtering also removes pixels that include clouds. To ensure a high coverage we interpolated in time for MODIS. Further, training and test folds are selected based on coverage, i.e. the regions near the equator (between $\pm 22.5^\circ$) are not included in the data cubes as MODIS reflectance is sensitive to clouds which appear frequently in these regions (compare Supplemental Figure S2 and S4). All static variables have full coverage on land. ERA5 data has full spatial and temporal coverage. We did not apply any further quality filtering on SMAP soil moisture data. The data is provided as a Level 3 product on Google Earth Engine.

Physiology perspective: Missing a more detailed explanation of what SIF is. Also, a physiology perspective (on the implications of method choices) beyond the intro is lacking seeing the emphasis of the downscaled SIF product is to monitor veg dynamics (e.g. GPP).

We have added the following part to the manuscript (L.35-39):

The electromagnetic signal SIF is emitted by chlorophylls during photosynthesis. SIF is emitted in the red-far-red wavelengths of 650-850 nm (Magney et al. 2020). It is a way, beside photochemistry and nonphotochemical quenching, for de-excitation of the chlorophylls (Köhler et al. 2018, Turner et al. 2020). Even though the link between chlorophyll fluorescence and photosynthesis is nonlinear at leaf and canopy scale (Magney et al. 2020), it does not hold for satellite scales, where a linear relationship of SIF to GPP is frequently reported (e.g., Turner et al. 2021, Magney et al. 2020, Frankenberg et al. 2011, Joiner et al. 2011).

TROPOMO vs OCO SIF: Can you explain why the r^2 values are reasonably low (0.56-0.61) for TROPOMI vs OCO SIF? (Fig. S19 in S7)? This is important because they are both a measurement of the same metric and has implications for interpreting SIF. Is this purely time of day mismatch and spatial heterogeneity? This is especially important for putting into context the r^2 values achieved in this study. Is the difference due to anything physiologically related, i.e. is OCO SIF more sensitive to GPP than TROPOMI (I cannot see why, but maybe the retrieval algorithms play a role)? A detailed comparison will be beyond the scope of this

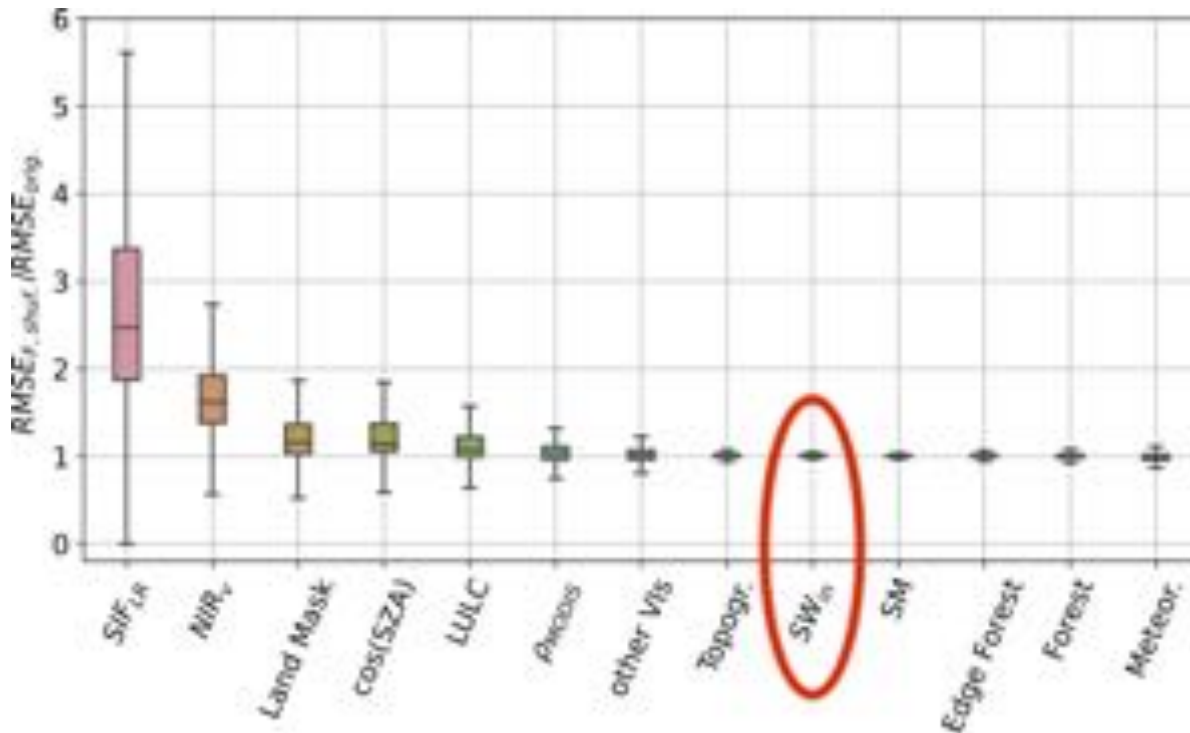
paper, but a figure and some stats and more explanation are warranted seeing OCO SIF is used as verification.

We have added a figure to compare the daily corrected SIF of TROPOMI and OCO-2/3 (Supplemental Figure S19). Here, 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 Koehler 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, Koehler et al. 2018 found a r^2 of 0.67, only additional aggregation lead 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 ($\sim 0.4 \text{ mW} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{nm}^{-1}$). In our analysis, from the 16 day product from TROPOMI SIF for April 2018 until March 2021 at 0.05 degree, we observe an average error in the TROPOMI SIF of $0.43 \text{ mW} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{nm}^{-1}$ for the CONUS. That error is close to the RMSE between instantaneous TROPOMI SIF and instantaneous OCO-2 SIF ($0.37 \text{ mW} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{nm}^{-1}$). 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 degree). As we aggregate multiple OCO-2 or OCO-3 footprints to match one TROPOMI grid cell at 0.05 degree the certainty of the OCO measurements increases, and therefore the RMSE between TROPOMI and OCO measurements decreases.

The explanation above is also included as a paragraph in the supplemental material (section S7)

Inputs: why not use incoming solar radiation at surface level from a gridded climate product rather than a clear sky flux equivalent? This would factor in cloudiness and is closer to the first driver of SIF magnitude in veg (PAR or APAR).

Thank you for this great suggestion! Indeed, PAR does strongly influence SIF and therefore we did a test to include the two variables "surface_net_solar_radiation" and "surface_solar_radiation_downwards" from ERA5 land (https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_HOURLY) as features into our model. The model does not rely on it as shown in following feature importance figure (feature cluster SWin, marked by red circle):



A value of 1 means that the error of the model output with shuffled features is the same as with unshuffled features. That means that the model does not rely on the features, as the value does not influence the model output.

We think that this is due to the following reasons. The spatial resolution of available shortwave incoming radiation products is coarser than our prediction resolution during the training process (solar radiation variables of ERA5 land are at 11132 meters and our model output during training at 0.05 degree). Therefore several pixels at the resolution of 0.05 degree have the same value in shortwave radiation, but different values in the target variable SIF. Further, shortwave incoming radiation has (compared to other variables) a weak spatial gradient. Low resolution SIF that is ten times coarser than the model output is part of the feature space. As PAR strongly influences SIF and due to the weak spatial gradients of PAR the information of it is included in the SIF features.

Due to the weak influence of the two variables “surface_net_solar_radiation” and “surface_solar_radiation_downwards” from ERA5 land in on the model output which might be due to reasons listed in the previous paragraph, we propose to not include SW_{in} in the feature space of our analysis.

Minor

Thank you for all the comments! We have worked on all of them and marked them with the symbol ✓ after resolving. In some cases we have added further explanations or mention the changes in the manuscript.

✓ Figure 1: what function was fit for each subplot to get the goodness of fit? Add info to figure caption.

We have added the following sentence to the caption of Figure 1.

To quantify the goodness of fit we compute the Pearson correlation coefficient (r) for each subplot (Benesty et al., 2009).

✓ L 120 remove 'In' before Figure 2.

✓ L 123: any explanations offered as to why TROPIMI SIF and GOME-2 SIF show different behaviors of correlation with NIRv to Camps-Valls et al. (2021)?

Several aspects may contribute to the differences: 1) GOME-2 has a substantially lower spatial resolution, 2) the number of soundings of TROPOMI is much higher (factor of ~100), 3) differences in overpass time (morning vs midday), 4) differences in viewing-illumination geometry, 5) higher potential of cloud contamination of GOME-2 soundings due to the increased footprint size.

✓ L130: expand 'NN' (first usage)

✓ L136: check grammar of sentence beginning 'As such'

✓ L161: remove 'are' after 'we'

✓ L 166: perhaps explain the implication of setting missing values and to zero, especially over vegetated regions.

(same answer as to your comment on data filtering)

We have added the following part to the manuscript (L.163-171):

During training, all missing values in the data are set to zero. That mainly affects water regions, as the share of missing values in the used SIF data is caused by 91.2% by water. In case there is a missing value in the SIF training sample, all feature values of this pixel are also set to zero to ensure the network does not learn false relationships between the predictors and the target variable (that also applies to vegetated regions). For the MODIS bands we applied the quality index value 0 (best quality only). To ensure a high coverage we interpolated in time for MODIS. Further, training and test folds are selected based on coverage, i.e. the regions near the equator (between $\pm 22.5^\circ$) are not included in the data cubes as MODIS reflectance is sensitive to clouds which appear frequently in these regions (compare Supplemental Figure S2 and S4). All static variables have full coverage on land. ERA5 data has full spatial and temporal coverage. We did not apply any further quality filtering on SMAP soil moisture data. The data is provided as a Level 3 product on Google Earth Engine.

- ✓ Fig. 4: add SIF units to subplots
- ✓ L 199: remove 'did' before 'yielded'. Also a double full stop evident.
- ✓ L225: add space after 'notably,'
- ✓ L223: remove 'do' after 'approach'
- ✓ L276: both high res SIF products performing poorly in urban areas suggest they should be treated separately in a model and/or point to the requirement for even higher spatial res products (e.g. 5 or 10m).

Thank you for pointing this out! Yes, indeed there might be a need to treat urban areas differently. Potentially, we need to scale to higher resolutions (like you suggested) or focus the model training on urban areas. Optimizing the model on urban areas is an active research going on right now at our institute and out of scope of the current study.

- ✓ L317: check grammar 'us to spatial'

Thank you for pointing this out, we modified the sentence.

The interpretable machine learning approach also allowed us to spatial regions of importance for the different parameters.	The interpretable machine learning approach also allowed us to identify spatial regions of importance for the different parameters.
--	---

- ✓ L322: repeated point of less training data – suggest removing or consolidating with the previous mention in the conclusion.

Supplementary Minor

- ✓ L29: remove 'is' after 'Australia'
- ✓ L51: see question mark
- ✓ Fig 6S: can remove one upper or lower diagonal as the plot is mirrored. Should subplots c & d have 1SD error bars seeing they represent a feature group?

We would like to keep the mirrored axis in the subplots a and b. In supplemental Figure S6 c and d, we show the explained variance of each principal component (PC). They do not represent feature groups but there is the same number of PCs as features. For the time variant case there are 19 features and therefore 19 PCs that each represents orthogonal axes and explain a certain share of the variance.

✓ L80: 'similar' to 'similarly' and be quantitative (provide a range).

✓ Fig S11c: can you comments on the mean value of rho_NIR compared to the equivalent plot in the main body for mean NIRv? Is the second feature (rho_NIR) more important than the second feature NIRv for the different input parameterisations? This point is worthy of another sentence or two in the main text.

We have added the following part to the manuscript (L.224-225):

This may result from the collinearity between input features.	This may result from the collinearity between input features or suggests that another combination of rho_NIR and RED is better correlated to SIF than e.g. NIRv or kNDVI.
---	---

References

Benesty, J., Chen, J., Huang, Y., and Cohen, I.: Pearson correlation coefficient, in: Noise reduction in speech processing, pp. 37–40, Springer, 2009.

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.

Joiner, J., Yoshida, Y., Vasilkov, A., Middleton, E., et al.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, *Biogeosciences*, 8, 637–651, 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.

Magney, T. S., Barnes, M. L., and Yang, X.: On the covariation of chlorophyll fluorescence and photosynthesis across scales, *Geophysical Research Letters*, 47, e2020GL091 098, 2020.

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.

Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: Extreme events driving year-to-year differences in gross primary productivity across the US, *Biogeosciences Discussions*, pp. 1–13, 2021.

A Convolutional Neural Network for Spatial Downscaling of Satellite-Based Solar-Induced Chlorophyll Fluorescence (SIFnet)

Johannes Gensheimer^{1,2}, Alexander J. Turner³, Philipp Köhler⁴, Christian Frankenberg⁴, and Jia Chen¹

¹Environmental Sensing and Modeling, Technical University of Munich (TUM), Munich, Germany

²Max Plank Institute for Biogeochemistry, 07745 Jena, Germany

³Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA

⁴Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

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

Abstract. Gross primary productivity (GPP) is the sum of leaf photosynthesis and represents a crucial component of the global carbon cycle. Space-borne estimates of GPP typically rely on observable quantities that co-vary with GPP such as vegetation indices using reflectance measurements (e.g., NDVI, NIR_v , and $kNDVI$). Recent work has also utilized measurements of solar-induced chlorophyll fluorescence (SIF) as a proxy for GPP. However, these SIF measurements are typically coarse resolution while many processes influencing GPP occur at fine spatial scales. Here, we develop a Convolutional Neural Network (CNN), named SIFnet, that increases the resolution of SIF from the TROPospheric Monitoring Instrument (TROPOMI) on board of the satellite Sentinel-5P by a factor of 10 to a spatial resolution of 500 m. SIFnet utilizes coarse SIF observations together with high resolution auxiliary data. The auxiliary data used here may carry information related to GPP and SIF. We use training data from non-US regions between April 2018 until March 2021 and evaluate our CNN over the conterminous United States (CONUS). We show that SIFnet is able to increase the resolution of TROPOMI SIF by a factor of 10 with a r^2 and $RMSE$ metrics of 0.92 and 0.17 $mW m^{-2} sr^{-1} nm^{-1}$, respectively. We further compare SIFnet against a recently developed downscaling approach and evaluate both methods against independent SIF measurements from Orbiting Carbon Observatory 2 and 3 (OCO-2/3). SIFnet performs systematically better than the downscaling approach ($r = 0.78$ for SIFnet, $r = 0.72$ for downscaling), indicating that it is picking up on key features related to SIF and GPP. Examination of the feature importance in the neural network indicates a few key parameters and the spatial regions these parameters matter. Namely, the CNN finds low resolution SIF data to be the most significant parameter with the NIR_v vegetation index as the second most important parameter. NIR_v consistently outperforms the recently proposed $kNDVI$ vegetation index. Advantages and limitations of SIFnet are investigated and presented through a series of case studies across the United States. SIFnet represents a robust method to infer continuous, high spatial resolution SIF data.

1 Introduction

Photosynthesis represents the single largest CO₂ flux between the atmosphere and the biosphere. At the canopy level, the sum of all leaf photosynthesis is termed Gross Primary Productivity (GPP) and accurate characterization of GPP represents a major uncertainty in the carbon cycle (Friedlingstein et al., 2019). Directly measuring GPP from remote sensing systems (e.g.,
25 satellites) is not presently possible. Instead, previous work has utilized stationary measurements of net ecosystem exchange (NEE) from flux towers that can be decomposed into GPP and respiration (e.g., Reichstein et al., 2005). Observable quantities from satellites (e.g., vegetation indices computed from reflectance data) are then related to GPP inferred from flux towers (e.g., Huete et al., 2006; Jung et al., 2019; Sims et al., 2006; Zeng et al., 2020) in light use efficiency (Mahadevan et al., 2008) or machine learning models Jung et al. (2019) to derive global estimates of GPP.

30 Vegetation indices such as NDVI and NIR_v combine two (or more) spectral bands with different absorption characteristics (Hanes, 2013) to infer quantities related to plant physiology and canopy structure. The MODIS instrument was launched on the Terra and Aqua satellites in 1999 and ~~2002.~~2002, respectively. This instrument has proved particularly useful due, in part, to the long operational lifetime. More recently launched satellites, like Sentinel-5P, carry instruments with the necessary signal to noise ratio and spectral resolution to retrieve solar-induced chlorophyll fluorescence (SIF),~~which is a measure of~~
35 ~~re-emitted photons by the chlorophyll during photosynthesis (Köhler et al., 2018).~~The electromagnetic signal SIF is emitted by chlorophylls during photosynthesis. SIF is emitted in the red-far-red wavelengths of 650-850 nm (Magney et al., 2020). It is a way, beside photochemistry and nonphotochemical quenching, for de-excitation of the chlorophylls (Turner et al., 2020; Köhler et al., 20
. Even though the link between chlorophyll fluorescence and photosynthesis is nonlinear at leave and canopy scale that does not hold for satellite scales, where a linear relationship of SIF to GPP is frequently reported (e.g., Turner et al., 2021; Magney et al., 2020; Fran
40 ~

Vegetation indices (also termed greenness) can be regarded as a measure of photosynthetic capacity (Sellers, 1985) whereas SIF indicates photosynthetic activity(Sellers, 1985). SIF has been shown to be a powerful proxy for estimating GPP (Magney et al., 2019; Turner et al., 2021), to capture the impact of drought on photosynthetic activities across different vegetation types (Shekhar et al., 2020a; Castro et al., 2020), and to assess the regional source of carbon emissions (Shekhar et al., 2020b).

45 Köhler et al. (2018) described the first retrievals of SIF from TROPOMI, the sole instrument on the Sentinel-5P satellite. The TROPOMI instrument has an equatorial crossing time of 13:30 local solar time (~~LST~~) and a 16 day orbit cycle. TROPOMI has a wide swath (2600 km across track) that allows for near-daily temporal resolution and a spatial resolution of 5.5x3.5 km. This was a substantial ~~improvent~~improvement to previous satellite instruments measuring SIF that were limited to 40x40 km² spatial resolution (Joiner et al., 2013). Despite the ~~improves~~higher spatial resolution of TROPOMI, there have been efforts to
50 estimate SIF at finer spatial scales (e.g., Turner et al., 2020). This is motivated by the importance of fine-scale phenomena in the carbon cycle such as ecosystem fragmentation (e.g., Haddad et al., 2015).

Globally, 20% and 70% of the remaining forests are within a distance of 100 m and 1 km, respectively, from the forest edges, meaning that most of the forests are fragmented (Haddad et al., 2015). Reinmann and Hutrya (2017) show that the carbon uptake and storage of trees near the forest edge increase up to 13±3% and 10±1%, respectively. On the other hand most of

55 our understanding about forest carbon fluxes comes from intact ecosystems, resulting in a mismatch between the ecosystems we are trying to quantify and the data we are using to do so (Smith et al., 2018). Higher resolution estimates of photosynthetic activity might enable to include fragmentation effects of ecosystems to global carbon cycle estimates or biosphere models like, Jung et al. (2019); Wu et al. (2021); Turner et al. (2021). Additionally, recent work has shown the importance of fine-scale variations in the urban biosphere on the overall carbon flux for a city (e.g., Miller et al., 2020).

60 There has been some recent work with the goal of increasing the resolution of existing global SIF estimates through downscaling methods (i.e., physics based methods). For example, Turner et al. (2020, 2021) used NIR_v to partition SIF within a particular TROPOMI scene and oversampled it using a 16 day window afterwards, resulting in a daily 500m SIF estimate over CONUS. Duveiller et al. (2020) downscaled GOME-2 satellite SIF from 0.5° to 0.05° using a parameterization with a term for the fraction of absorbed photosynthetically active radiation (fAPAR), one for water stress, and one for heat stress based on
65 MODIS data. Siegmann et al. (2021) used airborne data to downscale far-red SIF from canopy to leaf level.

Machine learning has also been used to create global, high resolution SIF data sets. Li and Xiao (2019), Yu et al. (2019), and Zhang et al. (2018) used spectral bands from MODIS as input to neural networks that were trained with OCO-2 SIF data to build global continuous SIF products at 0.05° resolution. OCO-2 has a narrow swath and, therefore, the networks are trained only in the regions where OCO-2 SIF is available by using MODIS data as input. After training, the global MODIS data is
70 used as input to estimate SIF on a global scale. Gentine and Alemohammad (2018) uses MODIS reflectance data as input and predicts GOME-2 normalized by clear-sky irradiance. Multiplying that with a MODIS derived photosynthetic active radiation product results in a MODIS only estimated SIF, termed RSIF. Zhang et al. (2021) ~~trains a CNN~~ trained a Convolutional Neural Network (CNN) with MODIS data on the artificial GOSIF data set (Li and Xiao, 2019) at a resolution of 0.05° and ~~uses~~ used
75 the trained network and MODIS data at a resolution of 0.008° to estimate SIF at 0.008° . The physics based downscaling approach from Turner et al. (2020) can only consider one variable for weighting the SIF signal while the machine learning based approaches in the literature can consider more than one variable – but many do not use SIF data as an input to their model, meaning that they estimate SIF based on reflectance data.

Here we build a convolutional neural network to obtain high-resolution SIF, named SIFnet. SIFnet increases the spatial resolution of TROPOMI SIF by considering coarse resolution SIF together with high resolution auxiliary data as input. This
80 auxiliary data consists of either proxies of SIF or photosynthetic drivers. SIFnet is trained using data with near-global coverage. Different model parameters (structure, input features, and scaling factors) are compared and evaluated. After training the model, the resolution of TROPOMI SIF is refined by a factor of ten to a spatial resolution of 0.005° . This product is then compared against a recent downscaling method from literature (Turner et al., 2020). Both high resolution estimates are validated over CONUS against the independent SIF measurements of the OCO-2 and 3 instruments (OCO-2 and OCO-3, respectively).
85 (OCO-2 Science Team/Michael Gunson, 2020; OCO-3 Science Team/Michael Gunson, 2020).

2 Datasets

2.1 Data sources

The input data to the neural network is listed in Table 1. These diverse global data products are expected to capture a broad range of photosynthetic drivers. The table divides the data into time-varying or time-invariant and training or validation. The native spatial resolution is shown in the last column of Table 1. In a first step, all data sets are aggregated to 0.05° spatial resolution and 16 day time steps. In case of a higher native spatial resolution, the data is regridded by computing the mean value that falls into the coarse resolution grid cell. In case of coarser resolutions than 0.05° it is resampled to the common grid. Quality control flags and cloud filtering are applied when necessary.

MODIS measures the reflected radiance from the earth surface in seven different spectral bands covering the visible and infrared spectral region. Vegetation indices are computed by combining the near-infrared (where chlorophyll is non-absorbing) and the red band (where chlorophyll is highly absorbing) (Hanes, 2013). Specifically, the normalized difference vegetation index (NDVI) (Tucker, 1979), near-infrared vegetation index (NIR_v) (Badgley et al., 2017), the kernel NDVI (kNDVI) (Camps-Valls et al., 2021), and the enhanced vegetation index (EVI) (Huete et al., 2002) are computed as:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}} \quad (1)$$

$$NIR_v = \rho_{NIR} \cdot NDVI \quad (2)$$

$$kNDVI = \tanh(NDVI^2) \quad (3)$$

$$EVI = G \cdot \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + C_1 \cdot \rho_{RED} - C_2 \cdot \rho_{BLUE} + L} \quad (4)$$

EVI coefficients for MODIS are: $L=1$, $C_1=5$, $C_2=7.5$, $G=2.5$ (Huete et al., 2002). ρ_{NIR} is the near infrared band, ρ_{RED} is the red band, ρ_{BLUE} is the blue band from the MODIS satellites.

Temperature and precipitation is taken from ERA5-Land data at the time step of interest and with a delay of one time step. Soil moisture has been shown to be a strong driver of global photosynthesis due, in part, to its impact on vapor-pressure deficit (Humphrey et al., 2021). Here we use the coarse resolution NASA USDA SMAP soil moisture (Entekhabi et al., 2010) as a model input and explore its correlation with TROPOMI SIF. The cosine of the SZA is a proxy for [photosynthetically active radiation \(PAR\)](#) under cloud-free conditions (Chen et al., 2020; Turner et al., 2021).

Time invariant data sets consist of elevation data, fractional land cover classification, and forest fragmentation data. The land cover classification (Buchhorn et al., 2020) is resampled to 11 fractional classes. The forest fragmentation data consists of two [channel-bands](#) and has a native resolution of 30 m. One [channel-band](#) describes the share of forest within the grid cell (forest share) and the other how much of that forest is edge forest (defined as a maximum distance to an edge or other land cover type of 30 m). OCO-2 and OCO-3 have high spatial resolution (2.25 km x 1.29 km) but small swaths (10 km) and a 16-day revisit time.

Table 1. Data sets used in this work.

Data		time invariant	training	validation	spatial resolution
Sentinel-5P TROPOMI ¹		SIF@ 740 nm		X	0.05°
MODIS MCD43A4.v006 (v06) ²	MODIS bands	NIR, RED, BLUE, GREEN, SWIR1, SWIR2, SWIR3		X	500m
	Vegetation Indices	NIRv, kNDVI, NDVI, EVI		X	500m
ERA5-Land Hourly - ECMWF Climate Reanalysis ³	temperature	mean air temperature, mean air temperature with 16 days delay		X	0.1°
	precipitation	total precipitation, total precipitation with 16 days delay		X	0.1°
NASA USDA Enhanced SMAP Soil Moisture ⁴		surface soil moisture, sub-surface soil moisture		X	10 km
Solar zenith angle ⁵		cosine of the solar zenith angle		X	computed
USDA GMTED2010: Global Multi-resolution Terrain Elevation Data 2010 ⁶		elevation	X	X	7.5 arc seconds
Copernicus Corine global land cover classification (CLC2018) ⁷		Non-vegetated, ENF, EBF, DNF, DBF, MF, UF, shrubs, grassl., crops,wetland	X	X	100m
Forest fragmentation ⁸	forest share		X	X	30m
	edge share		X	X	30m
OCO-2 ⁹		SIF @ 740 nm		X	2.25 x 1.29 km
OCO-3 ¹⁰		SIF @ 740 nm		X	2.25 x 1.29 km

References: ¹ Köhler et al. (2018); ² Schaaf, C., Wang, Z. (2015); ³ ERA (2017); ⁴ Entekhabi et al. (2010); ⁵ PySolar (2021); ⁶ Danielson and Gesch (2011); ⁷ Buchhorn et al. (2020); ⁸ Morreale et al. (2021); ⁹ OCO-2 Science Team/Michael Gunson (2020); ¹⁰ OCO-3 Science Team/Michael Gunson (2020)

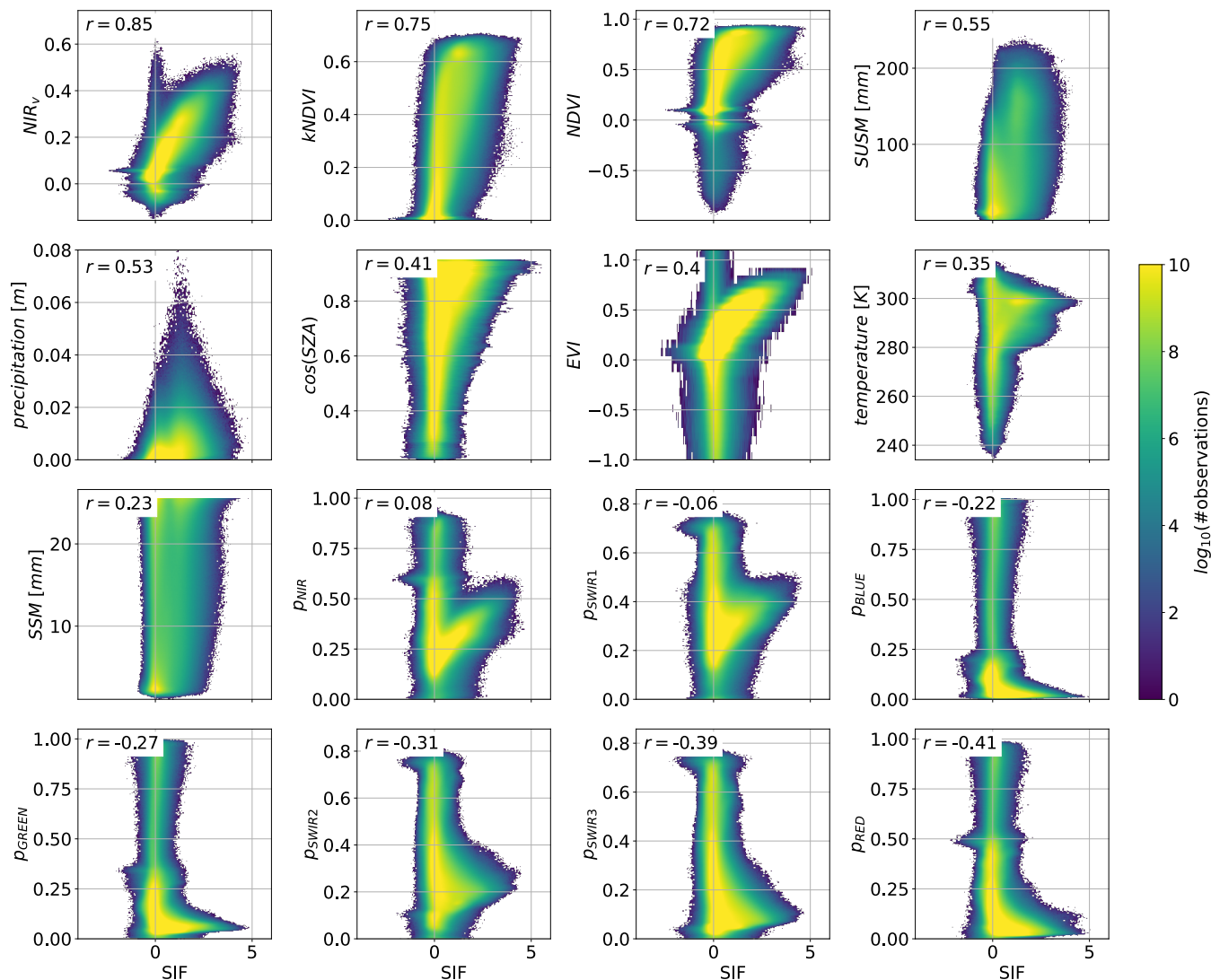


Figure 1. Scatter comparison of SIF to timely changing auxiliary data. The time span of measurements is from April 2018 to March 2021 in 16-day resolution. Longitude and latitude borders are from -180° to 180° and -60° to 70° , respectively. The comparison resolution corresponds to the lowest resolution of the two corresponding products. For all MODIS data the resolution is 0.05° and for precipitation, air temperature, surface soil moisture (ssm), and subsurface soil moisture (susm) 0.1° . [To quantify the goodness of fit we compute the Pearson correlation coefficient \(\$r\$ \) for each subplot \(Benesty et al., 2009\).](#)

We are interested in understanding what these different data sets are telling us about SIF and also how they co-vary with each other. We compare all collected time variant data against TROPOMI SIF in the spatial and temporal domain. As a quantitative

measure, we compute the Pearson correlation coefficient (r) (Benesty et al., 2009). Figure 1 shows a scatter comparison of SIF against the auxiliary data at the lowest resolution of the two corresponding sets. Negative SIF values (on the x-axis in Figure 1) are due to relatively high retrieval errors which scale with radiance levels (Köhler et al., 2018).

125 **In** Figure 2 shows spatial patterns of the Pearson correlation coefficients between both NIR_v and $kNDVI$ to SIF . In both our spatial (Fig. 2) and temporal (Fig. 1) analyses, we find NIR_v is a better predictor for SIF than $kNDVI$, which contradicts the recent findings from Camps-Valls et al. (2021). However, Camps-Valls et al. (2021) used GOME-2 SIF instead of TROPOMI SIF .

130 The vegetation index NIR_v outperforms $kNDVI$ in nearly all vegetated regions. Only central Asia, Sahara, and very high latitudes show a better correlation of $kNDVI$ with SIF . At the same time, these regions generally show a weaker correlation of vegetation indices with SIF .

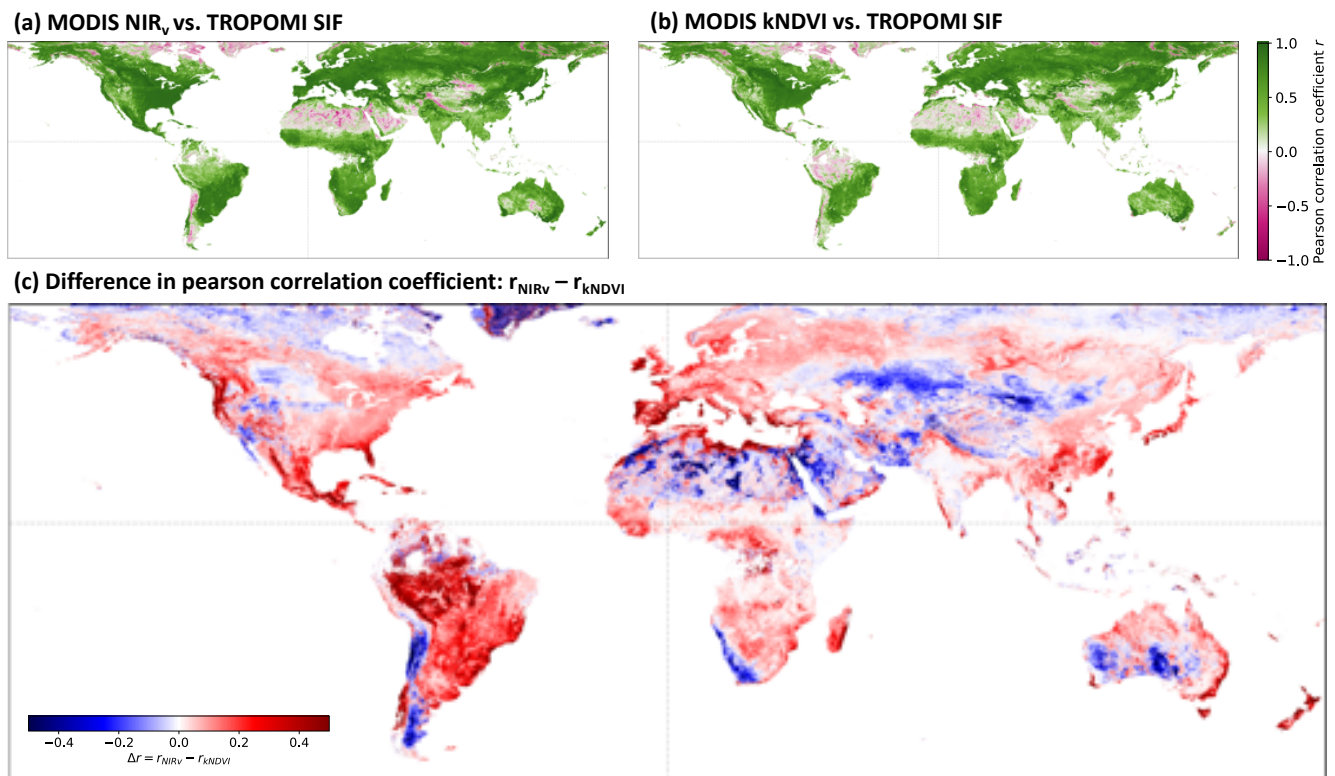


Figure 2. Pearson correlation coefficient of NIR_v and $kNDVI$ to TROPOMI SIF . Data is compared at 0.05° spatial resolution and in 16-day time steps starting in April 2018 until March 2021. The value per grid cell in a) and b) represents the Pearson correlation coefficient of the vegetation index to SIF in time. Panel c) represents the difference in correlation of the vegetation indices to SIF .

135 **Figure 1 shows several features that are highly correlated to SIF (e.g., NIR_v and $kNDVI$), suggesting a collinearity between features. Using principal component analysis (PCA), we find that the first 9 PCs (of in total 19) represent 99% of the variance in the time-varying datasets and that 13 (of in total 15) of the PCs represent 99% of the variance in the time-invariant**

datasets. There are many instances where PCs are clearly interpretable and NNs can handle collinearity, in contrast to e.g. linear regression methods (De Veaux and Ungar, 1994). We therefore keep all input features in our investigation and expect a longer convergence time during training of the CNN.

3 Development and Optimization of SIFnet

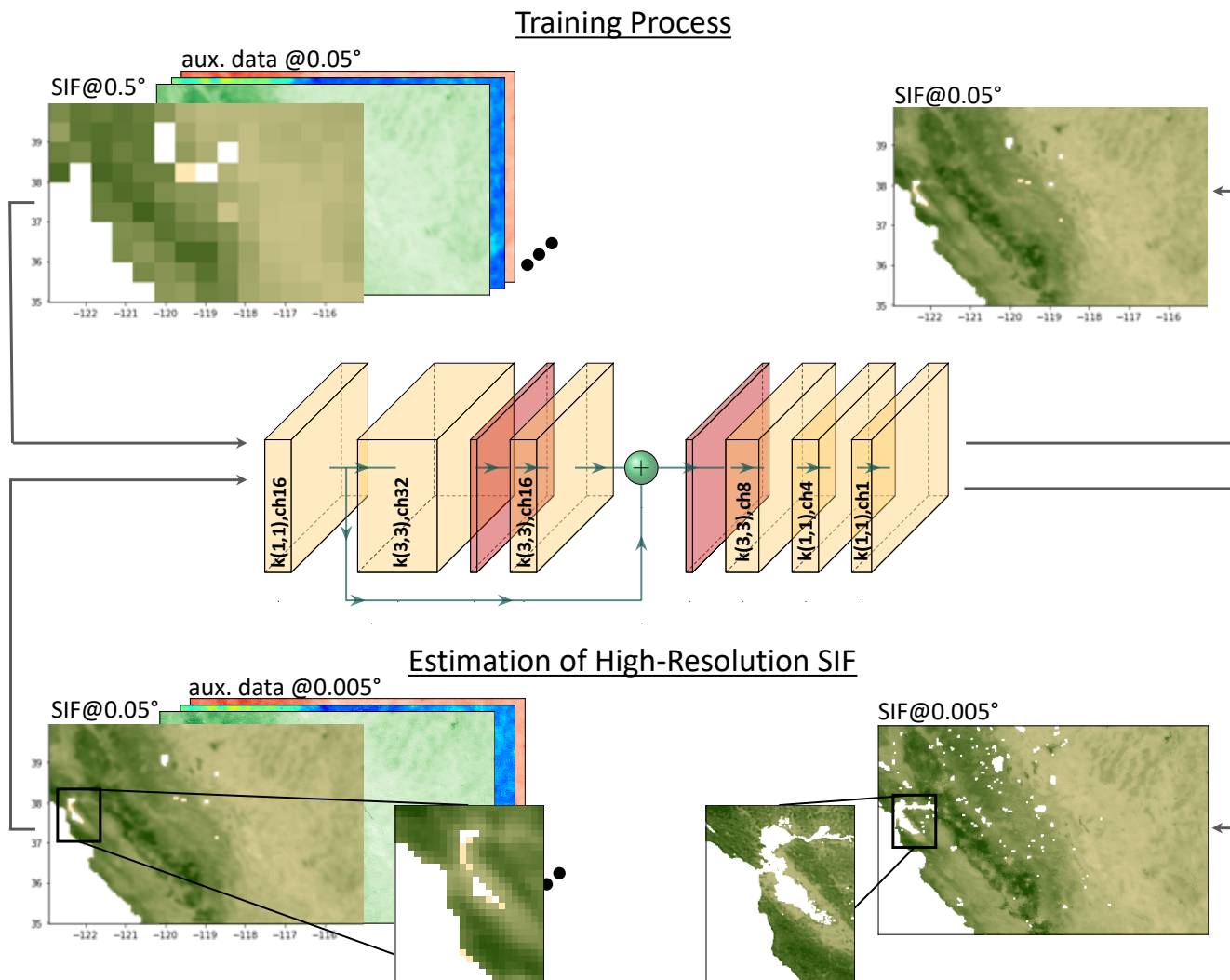


Figure 3. CNN model structure and training/estimation method. Yellow and red blocks are convolutional and ReLU layers, respectively. Notation of convolutional layers: $k(X1,X2)$: kernel sizes are $X1,X2$, chY : number of channels is Y . For training, the data is upsampled. We input auxiliary data at the target resolution and SIF data with a factor of 10 coarser.

140 3.1 Training and Optimization of the Neural Network

Convolutional neural networks (CNNs) are supervised machine learning methods that need matching feature and ground truth data pairs to compute the loss that is back propagated (Bishop, 2014). As such, we ~~being-begin~~ by coarsening SIF data to 0.5° and used ~~it~~ with auxiliary data at 0.05° as input to SIFnet, allowing us to estimate SIF at 0.05° . The model output is compared against the measured TROPOMI SIF at 0.05° . After optimizing the model it can resolve a scaling factor of ten between coarse resolution input SIF and model output SIF. Figure 3 visualizes this method. In the following step of estimating high resolution SIF, the feature SIF data has a resolution of 0.05° and auxiliary data of 0.005° , resulting in a model output of SIF at 0.005° .

Figure 3 shows our chosen CNN model structure for SIFnet. The model consists of convolutional and Rectified Linear Units (ReLU) layers that are arranged in a sequence. After the first convolutional block there is a residual connection that skips 2 convolutional and one ReLU layer. Convolutional kernel sizes are either (3,3) or (1,1). This structure is adapted from literature findings from e.g. Lim et al. (2017). Further, several model structures (Supplemental Section S4.3) with a different amount of layers, channels or residual blocks are compared. The chosen model structure represents the best trade-off between complexity and performance. The input feature collinearity and ~~PCA~~ Principal Component Analysis (PCA) presented in Supplemental Section S3 show that some input features have high correlations with each other. Nine out of the 19 PCs in the time variant and 13 out of the 15 PCs in the time invariant data carry above 99% of the variance. This suggests to use fewer channels in the CNN layers than the feature dimension (because some variables are similar). Therefore the number of channels in the first layer of SIFnet reduces the complexity from 34 to 16 channels (Figure 3). More complex model structures did not result in a notably improved loss metrics (Supplemental Section S4.3).

For training SIFnet we use three years of data (April 2018 - March 2021) in 16 day time steps. The study regions are shown in Supplemental Section S2. There are five folds used as training data: two folds over Asia, one over Europe, one over southern part of Africa, one over South America. Our validation region is North America (Figure S4). The hyperparameter tuning is done by training the model on the five folds and computing the loss of the validation data. The parameters are optimized to minimize the loss of the validation data set. Due to computational reasons and the size of the data set we do not apply a cross validation in the optimization process. The final product consists of high resolution SIF at 0.005° and is validated against independent SIF measurements of the instruments OCO-2 and OCO-3.

We center and scale each feature individually by subtracting the mean and normalizing by the standard deviation. For data augmentation of the training data, we use random crops and random flips. Each day of one fold has a matrix size of 1200×900 pixels. We ~~are~~ analyze 69 days in 16 day steps over 3 years. For each input during the training process we randomly crop a matrix with a size of 100×100 pixels. As some areas have a large fraction of missing values (e.g., due to water or clouds), we only use cropped matrices that consist of $>80\%$ valid pixels in the SIF product. Further, we randomly flip vertically and horizontally, both with a probability of 0.5. These data augmentation methods provide us a huge database that should avoid overfitting the network parameters. During training ~~water bodies and~~, all missing values in the data are set to zero. That mainly affects water regions, as the share of missing values in the used SIF data is caused by 91.2% by water. In case there is a missing value in the SIF training sample, all feature values of this pixel are also set to zero to ensure the network does not learn false

relationships between the predictors and the target variable (that also applies to vegetated regions). For the MODIS bands we applied the quality index value 0 (best quality only). This filtering also removes pixels that include clouds. To ensure a high coverage we interpolated in time for MODIS. Further, training and test folds are selected based on coverage, i.e. the regions near the equator (between $\pm 22.5^\circ$) are not included in the data cubes as MODIS reflectance is sensitive to clouds which appear frequently in these regions (compare Supplemental Figure S2 and S4). All static variables have full coverage on land. ERA5 data has full spatial and temporal coverage. We did not apply any further quality filtering on SMAP soil moisture data. The data is provided as a Level 3 product on Google Earth Engine.

Our individual loss function is comprised of two loss terms. We use the Mean Squared Error (MSE) loss in combination with the Structural Dissimilarity Index (DSSIM). The DSSIM is the counter-measure of the Structural Similarity Index (SSIM): $DSSIM = 1 - SSIM$ (Brunet et al., 2011). Therefore, we are not only optimizing the overall deviation of the estimated SIF to the measured SIF but also on structural patterns. Supplemental Section S4.4 shows the benefit of including both MSE and SSIM terms in the loss function. Equation 5 shows our loss function:

$$\mathcal{L} = a \cdot MSE + b \cdot DSSIM = a \cdot \frac{1}{n} \cdot \sum_{i=1}^n (y_i - \tilde{y}_i)^2 + b \cdot \left(1 - \frac{(2 \cdot \mu_Y \cdot \mu_{\tilde{Y}} \cdot (2 \cdot \sigma_{Y\tilde{Y}} + c_2))}{(\mu_Y^2 + \mu_{\tilde{Y}}^2 + c_1) \cdot (\sigma_Y^2 \cdot \sigma_{\tilde{Y}}^2 + c_2)} \frac{2 \cdot \mu_Y \cdot \mu_{\tilde{Y}} \cdot (2 \cdot \sigma_{Y\tilde{Y}} + c_2)}{(\mu_Y^2 + \mu_{\tilde{Y}}^2 + c_1) \cdot (\sigma_Y^2 \cdot \sigma_{\tilde{Y}}^2 + c_2)} \right) \quad (5)$$

where n is the number of datapoints; y_i is the datapoint i in measured (ground truth target variable) SIF; \tilde{y}_i is the datapoint i in estimated SIF; Y are all datapoints of measured (ground truth target variable) SIF; \tilde{Y} are all datapoints of estimated SIF; μ_Y is the mean of Y ; $\mu_{\tilde{Y}}$ is the mean of \tilde{Y} ; σ_Y is the variance of Y ; $\sigma_{\tilde{Y}}$ is the variance of \tilde{Y} ; $\sigma_{Y\tilde{Y}}$ is the covariance of Y and \tilde{Y} ; $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ being variables for stabilisation with $L = 2^{bits\ per\ pixel} - 1$, $k_1 = 0.01$ and $k_2 = 0.03$. The parameters a and b define the weights on the overall loss of the two individual losses. The overall model performance did not show a notable sensitivity to different a and b values. To approximately keep the individual losses in the same order of magnitude we set $a = 1$ and $b = 0.3$. DSSIM is in the range of 0 to 1, with 0 meaning structural similar and 1 structural dissimilar.

We use the optuna library for hyperparameter tuning of the learning rate, weight decay, and epoch of the CNN (Akiba et al., 2019). Here, a Tree-structured Parzen Estimator Sampler suggest the parameters of the next trial which is based on a Gaussian Mixture Model. Supplemental Section S4.1 provides more details on this hyperparameter tuning.

3.2 Results of Model Optimization

Figure 4 summarizes the results of the optimized model. We observe an overall r^2 of 0.92, $SSIM$ of 0.87, and $RMSE$ of $0.17 \text{ mWm}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ between the estimated SIF from SIFnet and retrieved SIF from TROPOMI at 0.05° (Fig 4e). $SSIM$ is calculated by comparing the average SIF signal of the three years under investigation. Figure 4e shows the three metrics for each month of the year. We observe the lowest r^2 values in January, February, and March. These are associated with low SIF values and, consequently, lower signal to noise ratios which drive the decreased performance. $SSIM$ also indicates reduced

performance during this time period. $RMSE$ values are correlated with overall productivity with the lowest $RMSE$ in winter; this is expected as this metric depends on the magnitude of the signal.

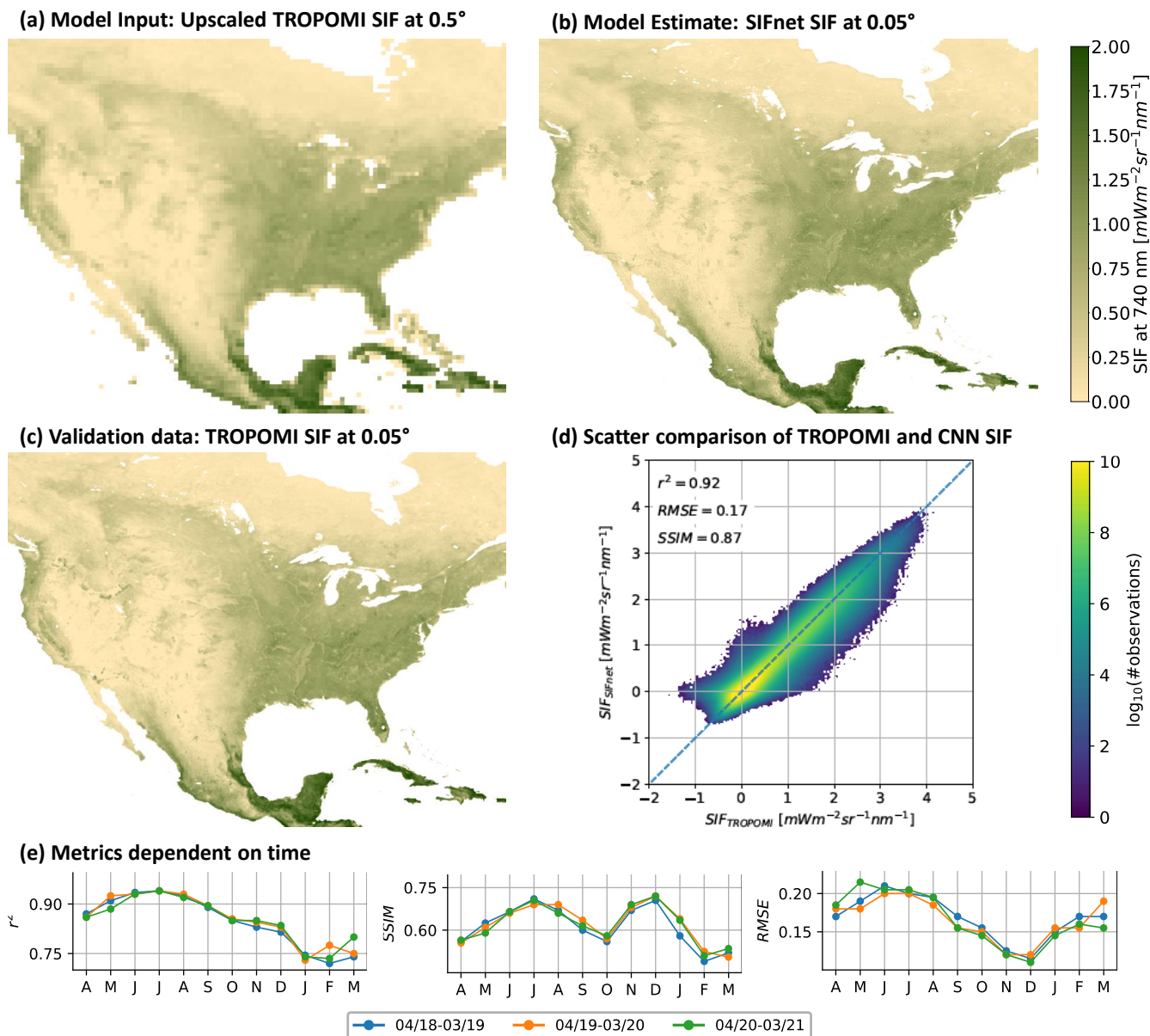


Figure 4. Test set results of CNN training at 0.05° . (a) shows low resolution SIF that is used as model input. (b) shows the estimated SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05° . (e) shows for each investigated month the metrics r^2 , $SSIM$, and $RMSE$. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards.

3.3 Which features drive SIFnet?

We are particularly interested in understanding which features drive our neural net. Here we evaluate the feature importance using the Permutation Feature importance method (Breiman, 2001; Fisher et al., 2019; Gregorutti et al., 2015, 2017) with our North American validation data at a target resolution of 0.05° . The method first computes the RMSE including all input features ($RMSE_{orig.}$). We then apply the following three steps:

1. Shuffle all pixels of one input feature randomly in time and space.
2. Compute the new RMSE of the estimation ($RMSE_{F,shuf.}$).
3. Compare the shuffled RMSE to the original: $d_{F,shuf.} = RMSE_{F,shuf.} / RMSE_{orig.}$.

Figure 5 shows the feature importance of clustered input classes and individual features to the overall estimation. Multiple applications of the feature permutation ~~did~~ yielded negligible differences in feature importance. Figure 5a shows the $RMSE$ share of shuffled data to the $RMSE$ of unshuffled data. SIFnet finds low resolution SIF (SIF_{LR}) to be the most important input variable, followed by the vegetation index NIR_v and the cosine of the solar zenith angle ($\cos(SZA)$). All other variables do not contribute notably to the model output. This result strengthens our findings from Figure 1 and Figure 2 that NIR_v is better correlated with SIF than $kNDVI$. Further, our feature importance is in line with Dechant et al. (2022) where they find a high correlation of SIF with NIR_v multiplied with Photosynthetic Active Radiation (PAR), of which the $\cos(SZA)$ can be used as a proxy. Figure 5b shows the spatial feature importance over the validation set in North America for the four most important features. We observe that SIF_{LR} has the biggest impact in the Eastern US, which corresponds strongly to the high vs low productivity regions in the US. NIR_v is a strong predictor in Southeastern US and in shrub regions in the Western US. The contribution of $\cos(SZA)$ is highest at high latitudes and weakens at lower latitudes. NIR_v is found to be less predictive of SIF at high latitudes. The land mask is the fourth most important input feature and contributes most in shrub regions. These four features consistently stand out as the strongest predictors. Other inputs such as fragmentation and soil moisture were not found to be strong predictors here. In supplemental Section S4.6 we test higher scaling factors between low and high resolution SIF. Even with scaling factors of 20 and 50 low resolution SIF stays the most and second most important input feature, respectively.

We also examined different combinations of inputs such as directly including the MODIS bands, as opposed to vegetation indices derived from MODIS bands (see Supplemental Figure S11). Low resolution SIF remains the most important feature, followed by the NIR band ρ_{NIR} . The land cover products increase in relevance. Interestingly, when low resolution SIF is omitted as input for the model, we observe contrasting results to Figure 5 where NIR_v is no longer a leading predictor. We find that $\cos(SZA)$, ρ_{NIR} , $kNDVI$, and $NDVI$ are the four most important features in this case (see Supplemental Figure S12). This may result from the collinearity between input features or suggests that another combination of ρ_{NIR} and ρ_{RED} is better correlated to SIF than e.g. NIR_v or $kNDVI$. This finding was robust to multiple optimizations and permutations.

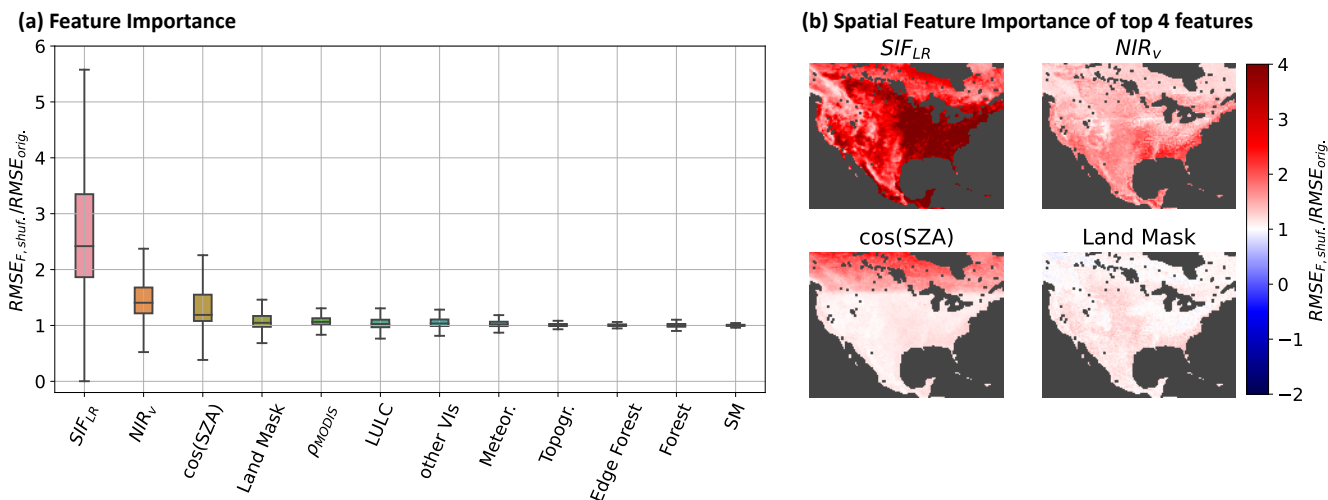


Figure 5. Feature Importance. (a) shows the total RMSE of the permuted feature divided by the RMSE without feature permutation (b) shows the RMSE of each pixel with permuted features divided by the RMSE without feature permutation. Some input variables are clustered and all variables of that class are permuted at the same time. ρ_{MODIS} : all 7 MODIS bands; LULC: all 11 land cover classes; other VIs: kNDVI, NDVI, EVI; Meteor.: temperature, precipitation, temperature with 16 day delay, precipitation with 16 days delay; SM: surface soil moisture and sub-surface soil moisture.

3.4 Comparison of SIFnet to Downscaled SIF

Figure 6 shows the 0.005° SIF estimated by SIFnet and SIF downscaled SIF from Turner et al. (2020). The difference between the two SIF estimates can be seen in Fig. 6c. SIFnet predicts lower SIF in the western US drylands and higher SIF over forested regions in the eastern US. This prediction of lower SIF in drylands is interesting because Turner et al. (2020) resorted to an ad hoc bias correction in these regions due to a low signal-to-noise ratio. Recent work from Wang et al. (2022) ~~found SIF to perform poorly in the~~ concluded SIF and NIR_V to capture complementary events in western US drylands as a proxy for GPP and that the linear correlation of SIF to GPP was substantially lower in these regions compared to other vegetation types. We also observe systematic differences in the predicted SIF in urban areas. These regions are further evaluated in Supplemental Figure S17. Notably, the SIFnet estimate is systematically lower than the downscaling estimate in most urban regions examined here, Seattle being a notable exception. Both SIFnet and the downscaling approach ~~do~~ allocate SIF to large urban parks and greenspaces, but SIFnet predicts little-to-no SIF over the rest of the urban area. In particular, SIFnet estimates nearly zero SIF in the urban core of Los Angeles and San Francisco. SIFnet and the downscaling method predict comparable SIF as we move away from the urban core. Fine-scale features in the urban region are visible in both SIF estimates such as the Schiller Woods in Chicago ($87.8^\circ W$, $42.0^\circ N$).

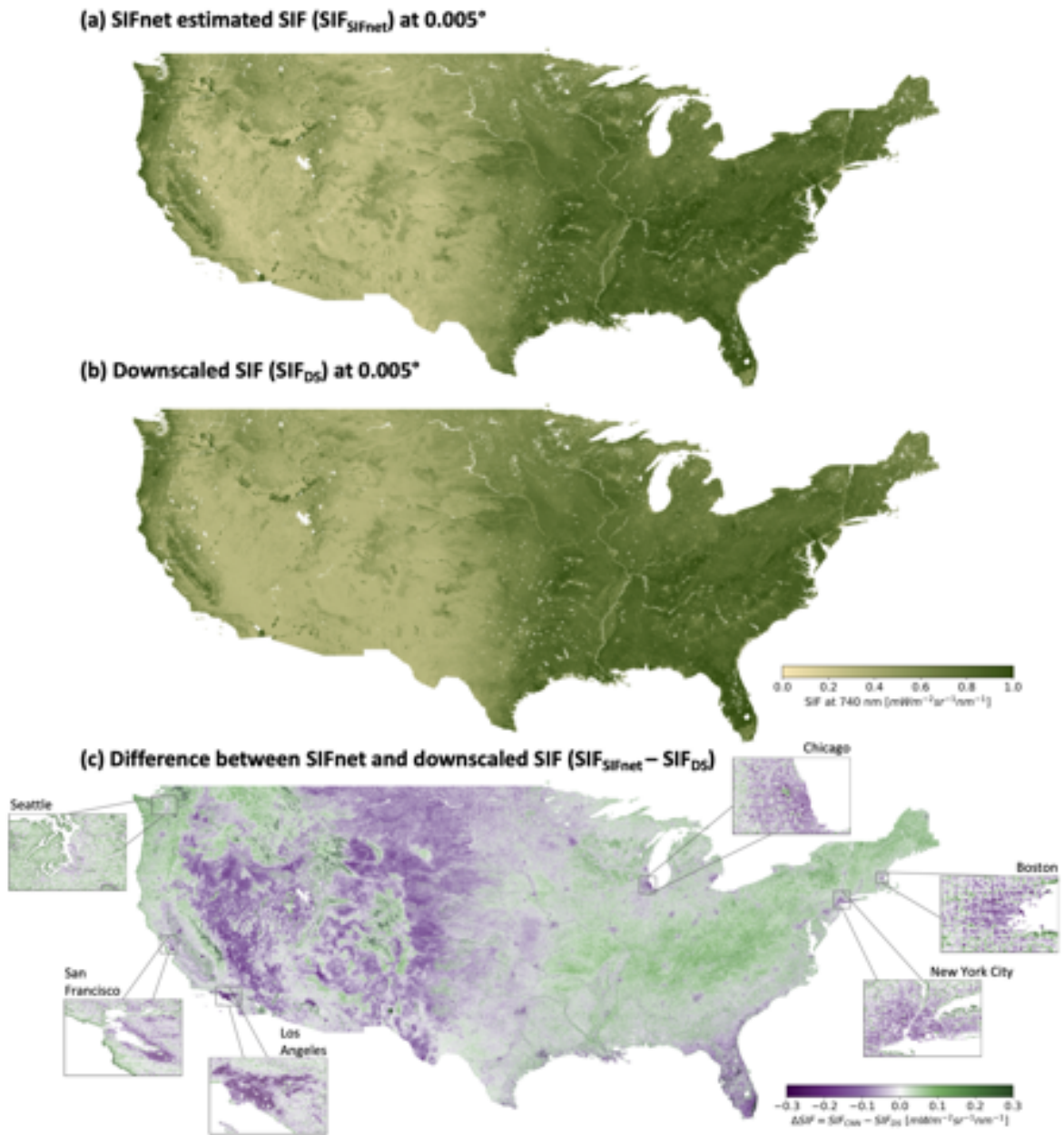


Figure 6. SIFnet estimated SIF at 0.005° for CONUS and its comparison to downscaled SIF. (a) shows the SIFnet estimated SIF at 0.005° . (b) shows the downscaled SIF from Turner et al. (2020). (c) shows the difference between SIFnet and downscaled SIF. Negative values imply a higher SIFnet SIF, positive values a higher downscaled SIF value.

The differences in SIF predicted from SIFnet and the downscaled SIF beg the question, “*which is correct?*” Here we evaluate both SIF products against independent SIF observations from OCO-2 and OCO-3 (OCO-2 Science Team/Michael Gunson, 2020; OCO-3 Science Team/Michael Gunson, 2020). These instruments have higher spatial resolution than TROPOMI and, as such, can be used to evaluate the high-resolution patterns predicted by both SIFnet and the downscaling approach. Specifically, OCO-2 and OCO-3 have nadir footprint sizes of 2.25 km x 1.29 km. However, OCO-2 and OCO-3 do not provide full spatial coverage. They observe narrow swaths that are ~10 km across-track. OCO-3 also provides also a scanning mode to observe urban areas. Here, we use quality-checked OCO-2 data from April 2018 until March 2021 and OCO-3 data from July 2019 until March 2021. To compare the ungridded OCO-2 and OCO-3 data against the SIF estimated from TROPOMI, we compute the weighted average of all 0.005° grid cells that fall within the bounds of an OCO footprint. Here, the TROPOMI estimates are subsampled to 0.0005° (appr. 50 m at equator) and the mean value is computed for all values which fall into the OCO footprint. For a quantitative comparison between OCO-2/3 and the SIFnet and downscaled estimate the metrics r , r^2 , and $RMSE$ are computed.

The high resolution SIF estimates from SIFnet, the downscaling, and OCO are instantaneous SIF measurements taken at a specific time of the day, while the time of TROPOMI observations can differ substantially. Here, we compute the daily average SIF by scaling with the cosine of the SZA (Frankenberg et al., 2011):

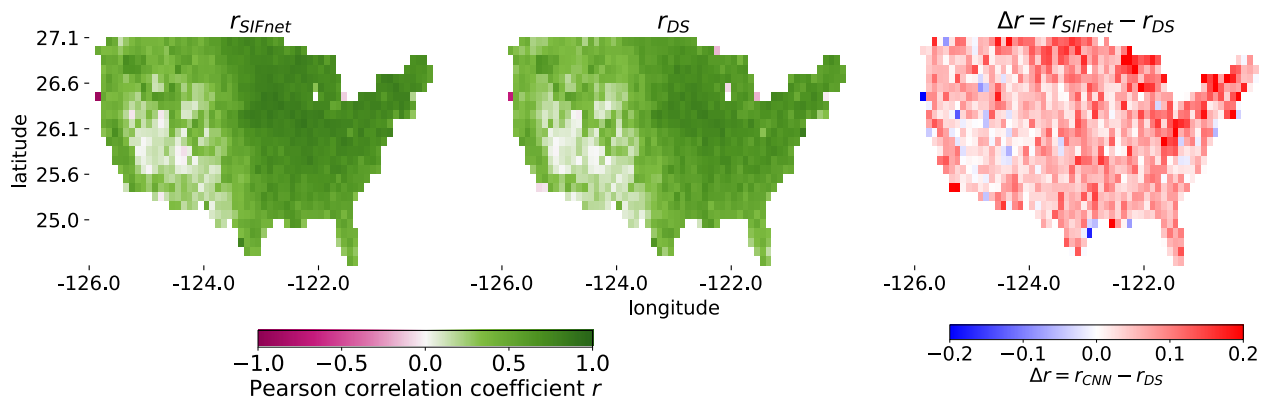
$$Daily\ SIF(x, y) = SIF(\tau_s, x, y) \cdot \frac{\int_{\tau_o}^{\tau_f} \cos[SZA(\tau_s, x, y)] d\tau}{\cos[SZA(\tau_s, x, y)]} \frac{\int_{\tau_o}^{\tau_f} \cos[SZA(\tau, x, y)] d\tau}{\cos[SZA(\tau_s, x, y)]} \quad (6)$$

where $Daily\ SIF(x, y)$ is the daily integrated SIF estimate, $SIF(\tau_s, x, y)$ is the instantaneous SIF at the individual measurement time, SZA is the solar zenith angle, τ_s is the time of the satellite measurement, τ_o is the time of sunrise, τ_f is the time of sunset. This implicitly assumes that both PAR and SIF scale with $\cos(SZA)$ under cloud-free conditions and we neglect Rayleigh scattering as well as gas absorption. Although this approach neglects several water or light conditions, it provides our best estimate of daily SIF and enables comparison between multiple SIF products with different measurement times (Turner et al., 2021; Köhler et al., 2018; Frankenberg et al., 2011). The method is equivalent to the daily correction scheme for OCO-2, OCO-3, and TROPOMI (OCO-2 Science Team/Michael Gunson, 2020; OCO-3 Science Team/Michael Gunson, 2020; Frankenberg et al., 2011). Additionally, we performed a sensitivity study where we trained SIFnet using daily-corrected SIF and found the results to be generally insensitive to the use of instantaneous vs daily-corrected SIF (see Supplemental Figure S18). Following this, we chose to apply the daily correction after deriving the high-resolution SIF .

Figure 7 shows a comparison of both SIFnet and the downscaled SIF to OCO-2 and OCO-3. Specifically, Fig 7a shows the correlation of SIFnet and the downscaled estimate with OCO-2/3 for every 1° pixel over CONUS. Both SIFnet and the downscaled SIF generally show good agreement with r in excess of 0.7 for most of the high-productivity regions. We observe weaker correlations in the Western drylands due, in part, to a lower signal-to-noise ratio. Overall, we find SIFnet to perform systematically better than the downscaled SIF, as shown in the difference plot. Fig. 7b summarizes these spatial patterns in a scatterplot comparison. SIFnet again shows better performance than the downscaled SIF against OCO-2, OCO-3, and OCO-2/3.

The Pearson correlation coefficient r is 0.78 and 0.72 for the SIFnet and downscaled estimate, respectively, when comparing to all OCO data (right column in Figure 7b). The generally high $RMSE$ indicates different scales and variability in the data sets. These also appear in the comparison of TROPOMI SIF at 0.05° resolution against OCO-2/3 (Supplemental Figure S19).
285 However, the errors generally increase at 0.005° compared to the 0.05° resolution.

(a) Spatial comparison of the correlation of SIFnet and downscaled SIF to OCO-2 and OCO-3



(b) Validation of SIFnet and downscaled SIF to OCO-2 and OCO-3 over CONUS

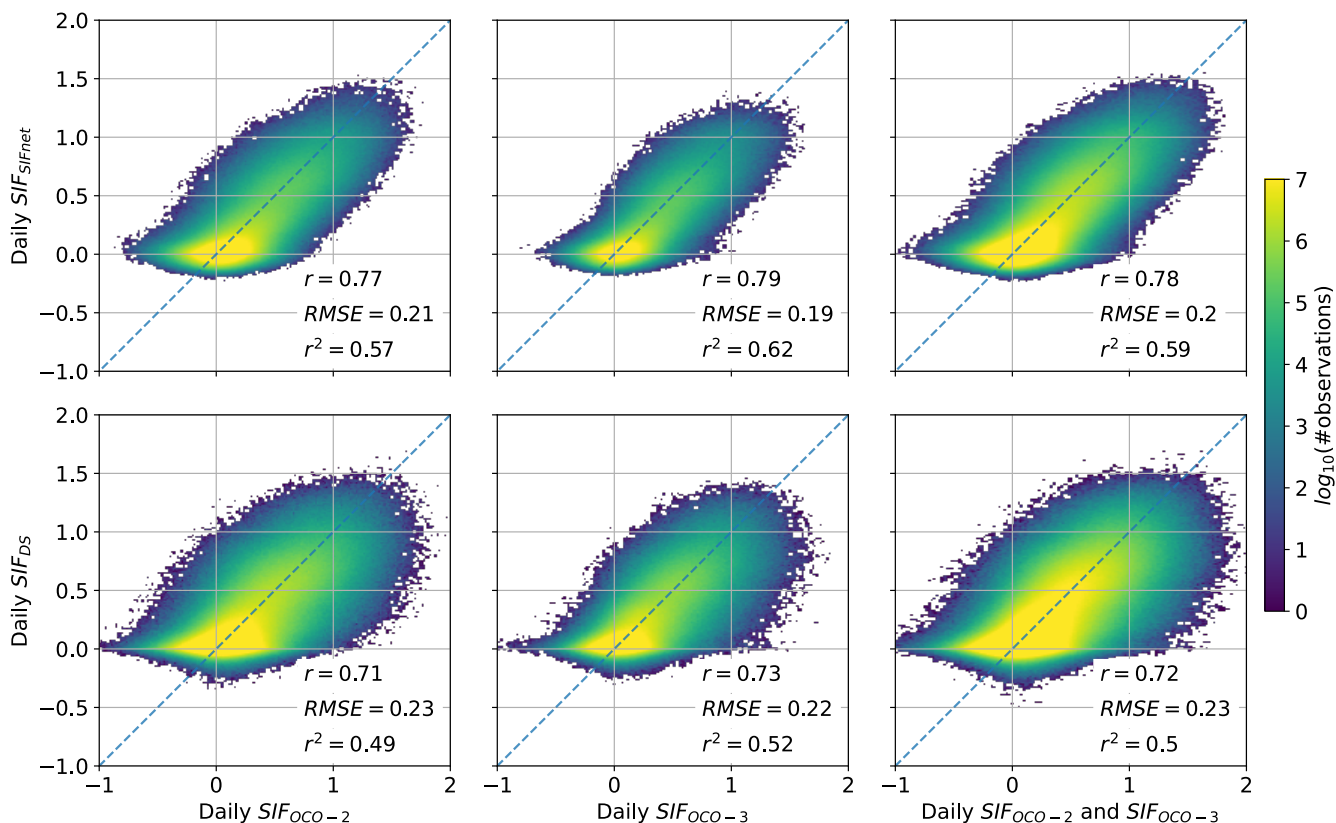


Figure 7. Validation of SIFnet and Downscaled SIF to OCO-2 and OCO-3 SIF over CONUS. Comparison from April 2018 until March 2021 in 16 day time steps. Daily OCO-2 and OCO-3 data is assigned to the closest 16 day time step. (a) shows the gridded correlation of the two products against the combined data of OCO-2 and OCO-3. We first use compute the SIF data from SIFnet and downscaling estimate that falls into the OCO footprint. Then we assign every OCO footprint to the closest grid point on the 1° grid dependent on the center location of that footprint and compute the Pearson correlation coefficient. (b) shows the scatter comparison of the weighted average of all grid cells on the 0.005° estimated SIFnet SIF (ours) and downscaled SIF (Turner et al., 2020) that fall into the OCO-2 or OCO-3 footprint.

Figure 8 presents a detailed comparison of SIFnet and the downscaled SIF in four US cities. The first column shows the SIFnet estimate, the second the downscaled SIF from Turner et al. (2020), the third the difference between SIFnet and downscaled SIF, and the last column the difference in correlation of the high resolution SIF estimates to combined OCO-2 and OCO-3 data on a 0.02° grid multiplied with the L_1L^1 -norm between the SIFnet and the downscaled SIF. The column on the right highlights regions where both the differences in predicted SIF are large and which product is performing better. As such, the right column will show white in areas where the difference in predicted SIF is small or the correlation with OCO is similar. While we observe large differences in predicted SIF for the urban areas (column 3), we don't find one product to perform systematically better in urban areas. This likely indicates the complexity in the SIF signal arising from urban areas. Additionally, urban areas make up a small fraction of the overall land mass and, as such, do not represent a large share of the training data in SIFnet. These factors likely contribute to the ~~heterogenous~~ heterogeneous performance observed in the right column of Fig 8.

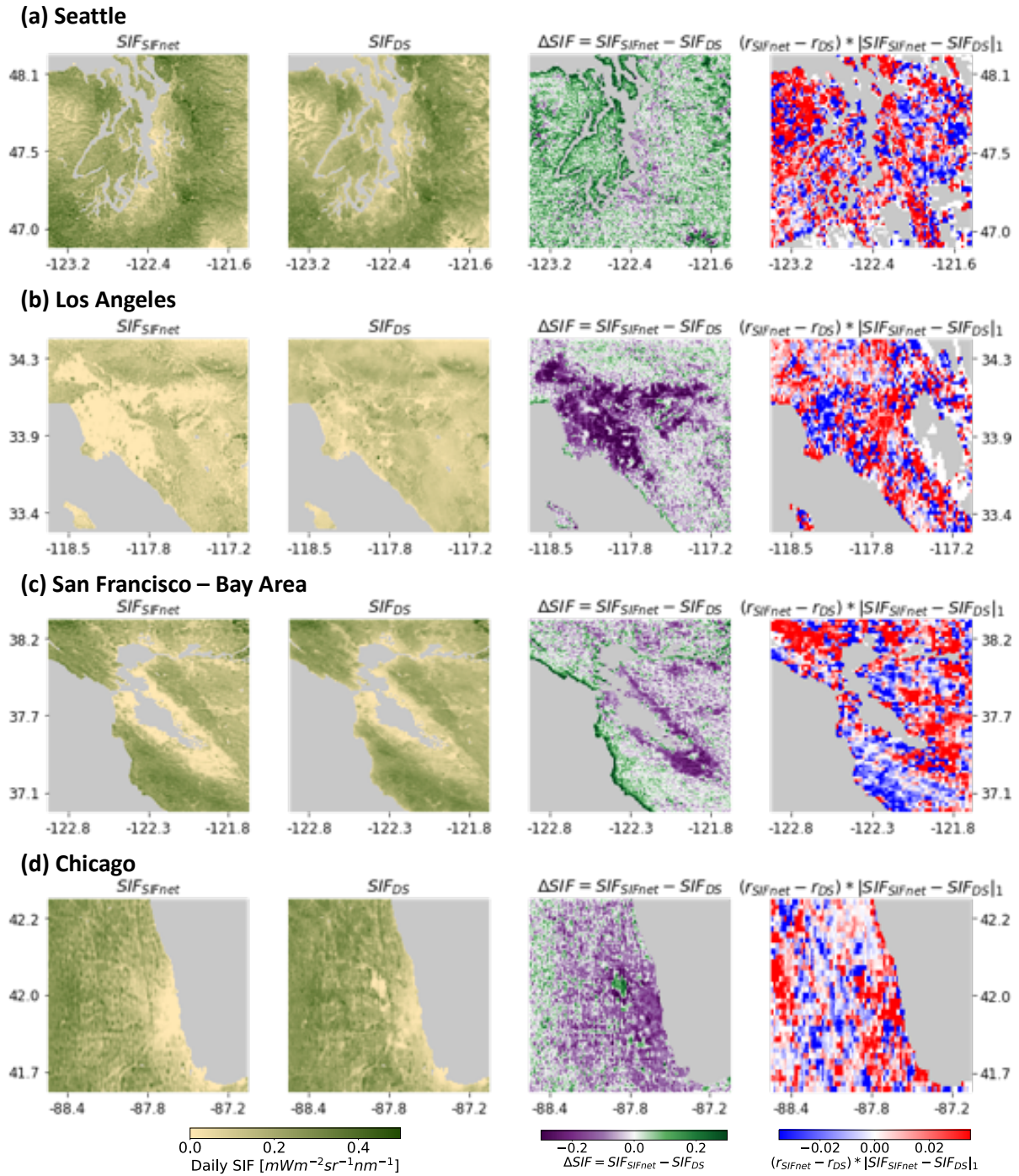


Figure 8. SIFnet and Downscaled SIF, the Difference between these, and the Difference in Correlation to OCO-2 and OCO-3 for **Four Urban Regions**. First column shows the SIFnet estimate, second the downscaled SIF from Turner et al. (2020), third the difference between SIFnet and downscaled SIF and the last the difference in correlation of the high resolution SIF estimates to combined OCO-2 and OCO-3 data on a 0.02° grid multiplied with the $\ell^1\mathcal{L}^1$ -norm between the SIFnet and downscaled estimate.

However, there are some ~~notbale~~notable successes of SIFnet in urban areas that can be mapped directly to features in the urban area. Figure 9 shows both SIFnet and the downscaled SIF along with NIR_v from MODIS and a true color image of Chicago. A feature clearly stands out in both the downscaled SIF and NIR_v image. This is a region with missing NIR_v and effectively no downscaled SIF. However, SIFnet does not show a strong gradient here. This region corresponds to the Chicago airport. In the MODIS NIR_v image it is visible that there is no valid data available for that region for the three investigated years. The downscaling method from Turner et al. (2020) relies only on NIR_v in the weighting function. If there is no data available the region it is interpolated in space and time. Here, it shows that the method seem to fail in urban regions where no MODIS NIR_v signal is available. SIFnet handles this region better and seems to rely on other auxiliary data if there is no MODIS NIR_v available. In Figure 8 it is also visible that the SIFnet estimate correlates better with the OCO-X data than the downscaled SIF for the region of the Chicago airport.

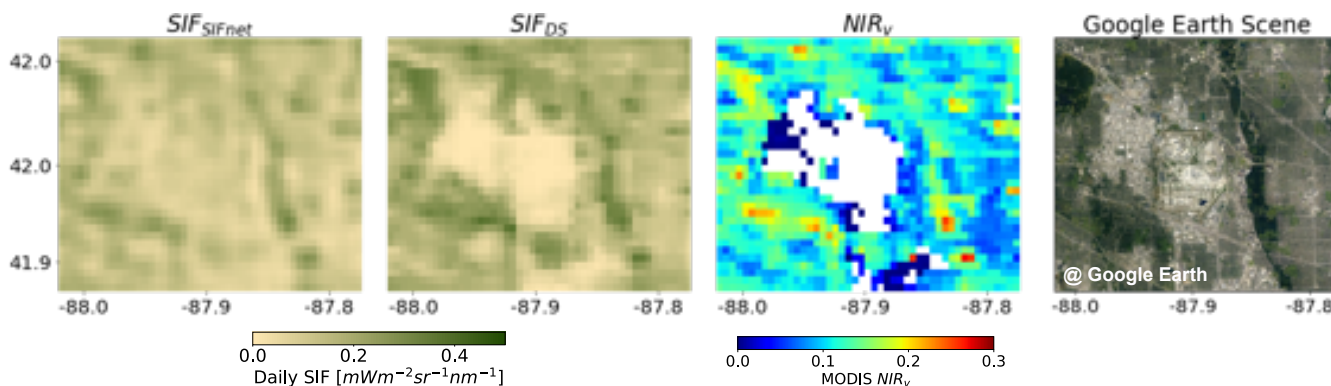


Figure 9. SIFnet SIF, Downscaled SIF, MODIS NIR_v , and a Google Earth Cut-Out for a part of Chicago Left panel shows the SIFnet estimate, second panel shows the downscaled estimate from Turner et al. (2020), third panel shows MODIS NIR_v for Chicago, and last column shows the Google Earth cut-out (Google LLC, 2021). For panel 1-3 the average data for April 2018 until March 2021 is shown.

5 Conclusions

Here, we develop a convolutional neural network (CNN) model named SIFnet to increase the resolution of TROPOMI SIF by a factor of 10. The novelty of our method consists of using coarse resolution SIF measurements together with high resolution auxiliary data as model input to estimate high resolution SIF. After optimization and hyperparameter tuning of SIFnet, the estimated SIF at 500 m resolution yields an r^2 and $RMSE$ of 0.92 and 0.17, respectively, when compared against validation data (Figure 4). We further compare the output of SIFnet against a recently developed downscaling method to estimate high-resolution SIF (Turner et al., 2020) and evaluate both methods against independent observations from the Orbiting Carbon Observatory 2 and 3 (OCO-2/3). SIFnet is found to perform systematically better than the downscaling approach when compared against independent measurements. Through interpretable machine learning methods, we identify the key features that SIFnet

utilizes to accurately predict high-resolution spatial patterns of SIF. We find that SIFnet relies heavily on the low resolution SIF feature (SIF_{LR}) and the vegetation index NIR_v (Figure 5).

SIFnet is a multi-layer CNN that increases the spatial resolution of the TROPOMI SIF by a factor of 10. Our model uses auxiliary datasets related to gross primary productivity and SIF as inputs and yields a high-resolution SIF estimate. The model is trained using three years of data from Asia, Europe, Africa, and South America. North America is used as the validation data set. Our loss function is comprised of two terms: the mean squared error and the structural dissimilarity index. The combination of these two terms improved the performance of our model.

SIFnet was further compared to the recent downscaled SIF product developed by Turner et al. (2020). The two high-resolution estimates showed pronounced differences across the western US drylands. This difference is particularly interesting because these drylands tend to be low-productivity regions and traditionally have been difficult for SIF to accurately capture due to the low signal-to-noise ratio. Both high-resolution SIF estimates were compared to independent observations from OCO-2/3. SIFnet performed systematically better than the downscaled SIF ($r = 0.78$ for SIFnet, $r = 0.72$ for downscaling). SIFnet and the downscaling method also yielded differences in urban regions. However, there was substantial heterogeneity in the performance of SIFnet and downscaling in urban areas. One product did not perform systematically better than the other within urban areas. The mixed results in urban areas likely relates to both the complexity of the photosynthetic activity in urban areas as well as the lack of training data, as urban areas represent a small fraction of the total landmass.

We adapted techniques from the area of interpretable machine learning to assess the key features driving SIFnet. Specifically, we conducted random permutations to input datasets and assessed the impact on the resulting RMSE. From this, we found that SIFnet relies most heavily on the low-resolution SIF feature (SIF_{LR}). The second most important factor is the MODIS vegetation index NIR_v . NIR_v is also found to outperform the recently proposed $kNDVI$ vegetation index, in contrast to Camps-Valls et al. (2021). The interpretable machine learning approach also allowed us to identify spatial regions of importance for the different parameters. Interestingly, SIFnet relies more heavily on NIR_v in the western drylands where the SIF signal-to-noise ratio is low. This implies that SIFnet is picking up on key physics that leads to the improved performance relative to the downscaling method. ~~Urban areas represent a region where SIFnet and the downscaling method performed comparatively well. Urban areas, while important in carbon cycle, represent a small percent of the total land mass. This means there is relatively less training data for SIFnet in urban areas. Further investigation into the processes controlling high-resolution GPP and SIF in urban areas is warranted.~~ Overall, SIFnet represents a robust method to infer ~~continuous~~ continuous high spatial resolution information about processes related to gross primary productivity.

Data availability. The high resolution SIF data is temporally available here: <https://syncandshare.lrz.de/getlink/fiXHgheNbMsbMBDeWVV3DpzR/>. It will be moved to a permanent data storage after acceptance.

Author contributions. JG, AJT, and JC conceived the study. JG compiled data sets, conducted data analysis, and generated figures. JG, AJT, and JC wrote the manuscript with feedback from all authors. JG did the literature research. JC provided project guidance. All authors contributed to the discussion. All authors have read and agreed to the published version of the manuscript.

350 *Competing interests.* The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements. We thank Xiaojing Tang, Luca Lloyd, and Lucy Hutyra from Boston University, USA for providing us their valuable global data about forest fragmentation. JG and JC are supported in part by the Technical University of Munich-Institute for Advanced Study through the German Excellence Initiative and the European Union Seventh Framework Program under the Grant Nr. 291763 and, in part, by the German Research Foundation (DFG) under the Grant Nr. 419317138. AJT was supported through the NASA Early Career Faculty program (grant 80NSSC21K1808) and the NASA Carbon Cycle Science program (grant 80HQTR21T0101).

355

References

- Copernicus Climate Change Service (C3S), ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS). <https://cds.climate.copernicus.eu/cdsapp#!/home>, accessed: 2021-08-13, 2017.
- 360 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.
- Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, *Science advances*, 3, e1602 244, 2017.
- Benesty, J., Chen, J., Huang, Y., and Cohen, I.: Pearson correlation coefficient, in: *Noise reduction in speech processing*, pp. 37–40, Springer, 2009.
- 365 Bishop, C.: *Bishop-Pattern Recognition and Machine Learning*-Springer 2006, *Antimicrob. Agents Chemother.*, pp. 03 728–14, 2014.
- Breiman, L.: Random forests, *Machine learning*, 45, 5–32, 2001.
- Brunet, D., Vrscay, E. R., and Wang, Z.: On the mathematical properties of the structural similarity index, *IEEE Transactions on Image Processing*, 21, 1488–1499, 2011.
- Buchhorn, M., Smets, B., Bertels, L., Roo, B. D., Lesiv, M., Tsendbazar, N.-E., Herold, M., and Fritz, S.: Copernicus Global Land Service: 370 Land Cover 100m: collection 3: epoch 2018: Globe, <https://doi.org/10.5281/zenodo.3518038>, 2020.
- Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., et al.: A unified vegetation index for quantifying the terrestrial biosphere, *Science Advances*, 7, eabc7447, 2021.
- Castro, A. O., Chen, J., Zang, C. S., Shekhar, A., Jimenez, J. C., Bhattacharjee, S., Kindu, M., Morales, V. H., and Rammig, A.: OCO-2 375 solar-induced chlorophyll fluorescence variability across ecoregions of the Amazon basin and the extreme drought effects of El Niño (2015–2016), *Remote Sensing*, 12, 1202, 2020.
- Chen, S., Liu, L., He, X., Liu, Z., and Peng, D.: Upscaling from Instantaneous to Daily Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) for Satellite Products, *Remote Sensing*, 12, 2083, 2020.
- Danielson, J. J. and Gesch, D. B.: Global multi-resolution terrain elevation data 2010 (GMTED2010), US Department of the Interior, US 380 Geological Survey, 2011.
- 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.
- Dechant, B., Ryu, Y., Badgley, G., Köhler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., Rossini, M., et al.: NIRVP: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales, *Remote Sensing of Environment*, 385 268, 112 763, 2022.
- Duveiller, G., Filipponi, F., Walther, S., Köhler, P., Frankenberg, C., Guanter, L., and Cescatti, A.: A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity, *Earth System Science Data*, 12, 1101–1116, 2020.
- Entekhabi, D., Njoku, E. G., O’Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., et al.: The soil moisture active passive (SMAP) mission, *Proceedings of the IEEE*, 98, 704–716, 2010.
- 390 Fisher, A., Rudin, C., and Dominici, F.: All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously., *J. Mach. Learn. Res.*, 20, 1–81, 2019.

- 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.
- 395 Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, 400 G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, *Earth System Science Data*, 11, 1783–1838, <https://doi.org/10.5194/essd-11-1783-2019>, 2019.
- Gentine, P. and Alemohammad, S.: Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS 405 surface reflectance to reproduce GOME-2 solar-induced fluorescence, *Geophysical research letters*, 45, 3136–3146, 2018.
- Google LLC: Google earth. Chicago, USA. 41.895°N - 42.061°N, 88.016°W - 87.7856°, <http://www.earth.google.com>, 2021.
- Gregorutti, B., Michel, B., and Saint-Pierre, P.: Grouped variable importance with random forests and application to multiple functional data analysis, *Computational Statistics & Data Analysis*, 90, 15–35, 2015.
- Gregorutti, B., Michel, B., and Saint-Pierre, P.: Correlation and variable importance in random forests, *Statistics and Computing*, 27, 659– 410 678, 2017.
- Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Lovejoy, T. E., Sexton, J. O., Austin, M. P., Collins, C. D., et al.: Habitat fragmentation and its lasting impact on Earth's ecosystems, *Science advances*, 1, e1500052, 2015.
- Hanes, J.: *Biophysical applications of satellite remote sensing*, Springer Science & Business Media, 2013.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G.: Overview of the radiometric and biophysical performance of 415 the MODIS vegetation indices, *Remote sensing of environment*, 83, 195–213, 2002.
- Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyrá, L. R., Yang, W., Nemani, R. R., and Myneni, R.: Amazon rainforests green-up with sunlight in dry season, *Geophysical research letters*, 33, 2006.
- Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S. I., and Frankenberg, C.: Soil moisture–atmosphere feedback dominates land carbon uptake variability, *Nature*, 592, 65–69, 2021.
- 420 Joiner, J., Yoshida, Y., Vasilkov, A., Middleton, E., et al.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, *Biogeosciences*, 8, 637–651, 2011.
- Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A., Middleton, E., Huemmrich, K., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, *Atmospheric Measurement Techniques*, 6, 2803–2823, 2013.
- 425 Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, *Scientific data*, 6, 1–14, 2019.
- 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.

- Li, X. and Xiao, J.: A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS, and reanalysis data, *Remote Sensing*, 11, 517, 2019.
- 430 Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K.: Enhanced deep residual networks for single image super-resolution, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 136–144, 2017.
- Magney, T. S., Bowling, D. R., Logan, B. A., Grossmann, K., Stutz, J., Blanken, P. D., Burns, S. P., Cheng, R., Garcia, M. A., Khler, P., et al.: Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence, *Proceedings of the National Academy of Sciences*, 116, 11 640–11 645, 2019.
- 435 Magney, T. S., Barnes, M. L., and Yang, X.: On the covariation of chlorophyll fluorescence and photosynthesis across scales, *Geophysical Research Letters*, 47, e2020GL091 098, 2020.
- Mahadevan, P., Wofsy, S. C., Matross, D. M., Xiao, X., Dunn, A. L., Lin, J. C., Gerbig, C., Munger, J. W., Chow, V. Y., and Gottlieb, E. W.: A satellite-based biosphere parameterization for net ecosystem CO₂ exchange: Vegetation Photosynthesis and Respiration Model (VPRM), *Global Biogeochemical Cycles*, 22, <https://doi.org/10.1029/2006GB002735>, 2008.
- 440 Miller, J. B., Lehman, S. J., Verhulst, K. R., Miller, C. E., Duren, R. M., Yadav, V., Newman, S., and Sloop, C. D.: Large and seasonally varying biospheric CO₂ fluxes in the Los Angeles megacity revealed by atmospheric radiocarbon, *Proceedings of the National Academy of Sciences*, 117, 26 681–26 687, 2020.
- Morreale, L., Thompson, J., Tang, X., Reinmann, A., and L.R, H.: Fragmentation impacts on temperate forest productivity: reversal of the tropical edge paradigm., *Nature Communications*, in press., 2021.
- 445 OCO-2 Science Team/Michael Gunson, A. E.: OCO-2 Level 2 bias-corrected solar-induced fluorescence and other select fields from the IMAP-DOAS algorithm aggregated as daily files, in: Retrospective processing V10r, Goddard Earth Sciences Data and Information Services Center (GES DISC), <https://doi.org/10>, 2020.
- OCO-3 Science Team/Michael Gunson, A. E.: OCO-3 Level 2 bias-corrected solar-induced fluorescence and other select fields from the IMAP-DOAS algorithm aggregated as daily files https://disc.gsfc.nasa.gov/datacollection/OCO3_L2_Lite_SIF_EarlyR.html, in: Retrospective processing VEarlyR, Goddard Earth Sciences Data and Information Services Center (GES DISC), accessed: 2021-08-20, 2020.
- 450 PySolar: Python library PySolar, <https://github.com/pingswept/pysolar>, 2021.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global change biology*, 11, 1424–1439, 2005.
- 455 Reinmann, A. B. and Hutyrá, L. R.: Edge effects enhance carbon uptake and its vulnerability to climate change in temperate broadleaf forests, *Proceedings of the National Academy of Sciences*, 114, 107–112, 2017.
- Schaaf, C., Wang, Z.: *MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global - 500m V006* [Data set] NASA EOSDIS Land Processes DAAC., <https://doi.org/10.5067/MODIS/MCD43A4.006>, 2015.
- 460 Sellers, P. J.: Canopy reflectance, photosynthesis and transpiration, *International journal of remote sensing*, 6, 1335–1372, 1985.
- Shekhar, A., Chen, J., Bhattacharjee, S., Buras, A., Castro, A. O., Zang, C. S., and Rammig, A.: Capturing the impact of the 2018 european drought and heat across different vegetation types using OCO-2 solar-induced fluorescence, *Remote Sensing*, 12, 3249, 2020a.
- Shekhar, A., Chen, J., Paetzold, J. C., Dietrich, F., Zhao, X., Bhattacharjee, S., Ruisinger, V., and Wofsy, S. C.: Anthropogenic CO₂ emissions assessment of Nile Delta using XCO₂ and SIF data from OCO-2 satellite, *Environmental Research Letters*, 15, 095 010, 2020b.

- 465 Siegmann, B., Cendrero-Mateo, M. P., Cogliati, S., Damm, A., Gamon, J., Herrera, D., Jedmowski, C., Junker-Frohn, L. V., Kraska, T.,
Muller, O., et al.: Downscaling of far-red solar-induced chlorophyll fluorescence of different crops from canopy to leaf level using a
diurnal data set acquired by the airborne imaging spectrometer HyPlant, *Remote Sensing of Environment*, 264, 112 609, 2021.
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Flanagan, L. B., Goldstein, A. H., Hollinger, D. Y., Misson,
L., Monson, R. K., et al.: On the use of MODIS EVI to assess gross primary productivity of North American ecosystems, *Journal of*
470 *Geophysical Research: Biogeosciences*, 111, 2006.
- Smith, I. A., Hutyrá, L. R., Reinmann, A. B., Marrs, J. K., and Thompson, J. R.: Piecing together the fragments: elucidating edge effects on
forest carbon dynamics, *Frontiers in Ecology and the Environment*, 16, 213–221, 2018.
- Tucker, C. J.: Red and photographic infrared linear combinations for monitoring vegetation, *Remote sensing of Environment*, 8, 127–150,
1979.
- 475 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.
- Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: Extreme events driving year-to-year differences in gross
primary productivity across the US, *Biogeosciences Discussions*, pp. 1–13, 2021.
- Wang, X., Biederman, J. A., Knowles, J. F., Scott, R. L., Turner, A. J., Dannenberg, M. P., Köhler, P., Frankenberg, C., Litvak, M. E.,
480 Flerchinger, G. N., Law, B. E., Kwon, H., Reed, S. C., Parton, W. J., Barron-Gafford, G. A., and Smith, W. K.: Satellite solar-induced
chlorophyll fluorescence and near-infrared reflectance capture complementary aspects of dryland vegetation productivity dynamics, *Re-
mote Sensing of Environment*, 270, 112 858, <https://doi.org/https://doi.org/10.1016/j.rse.2021.112858>, 2022.
- Wu, D., Lin, J. C., Duarte, H. F., Yadav, V., Parazoo, N. C., Oda, T., and Kort, E. A.: A model for urban biogenic CO₂ fluxes: Solar-Induced
Fluorescence for Modeling Urban biogenic Fluxes (SMUrF v1), *Geoscientific Model Development*, 14, 3633–3661, 2021.
- 485 Yu, L., Wen, J., Chang, C., Frankenberg, C., and Sun, Y.: High-resolution global contiguous SIF of OCO-2, *Geophysical Research Letters*,
46, 1449–1458, 2019.
- Zeng, J., Matsunaga, T., Tan, Z.-H., Saigusa, N., Shirai, T., Tang, Y., Peng, S., and Fukuda, Y.: Global terrestrial carbon fluxes of 1999–2019
estimated by upscaling eddy covariance data with a random forest, *Scientific data*, 7, 1–11, 2020.
- Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., and Gentine, P.: A global spatially contiguous solar-induced fluorescence (CSIF)
490 dataset using neural networks, *Biogeosciences*, 15, 5779–5800, 2018.
- Zhang, Z., Xu, W., Qin, Q., and Long, Z.: Downscaling Solar-Induced Chlorophyll Fluorescence Based on Convolutional Neu-
ral Network Method to Monitor Agricultural Drought, *IEEE Transactions on Geoscience and Remote Sensing*, 59, 1012–1028,
<https://doi.org/10.1109/TGRS.2020.2999371>, 2021.

Supplemental Information for "A Convolutional Neural Network for Spatial Downscaling of Space Derived Solar-Induced Chlorophyll Fluorescence (SIFnet)"

Johannes Gensheimer^{1,2}, Alexander J. Turner³, Philipp Köhler⁴, Christian Frankenberg⁴, and Jia Chen¹

¹Environmental Sensing and Modeling, Technical University of Munich (TUM), Munich, Germany

²Max Plank Institute for Biogeochemistry, 07745 Jena, Germany

³Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA

⁴Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

Correspondence: Johannes Gensheimer (johannes.gensheimer@bgc-jena.mpg.de), Jia Chen (jia.chen@tum.de), 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 $+22.5^\circ$ 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.

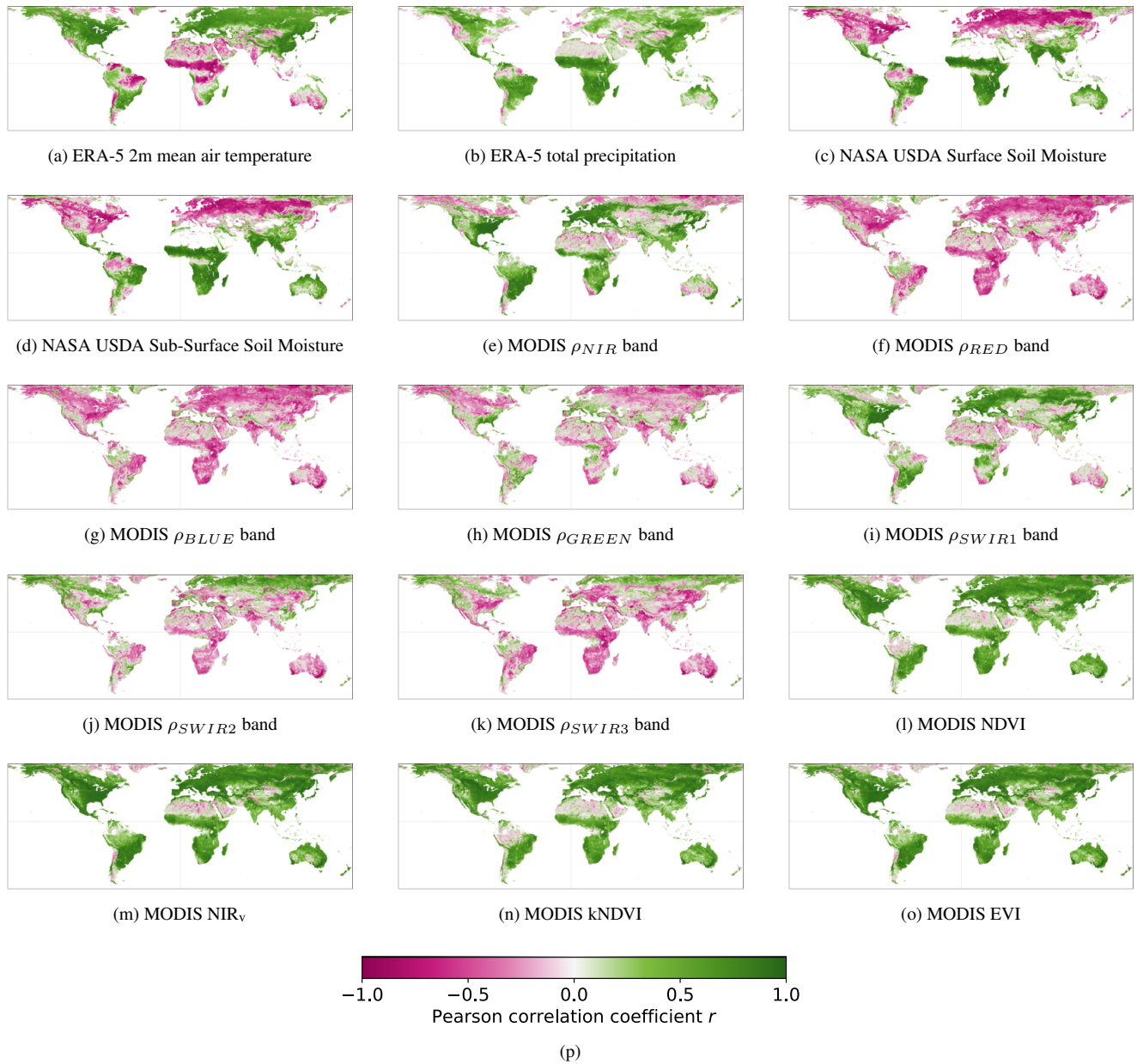
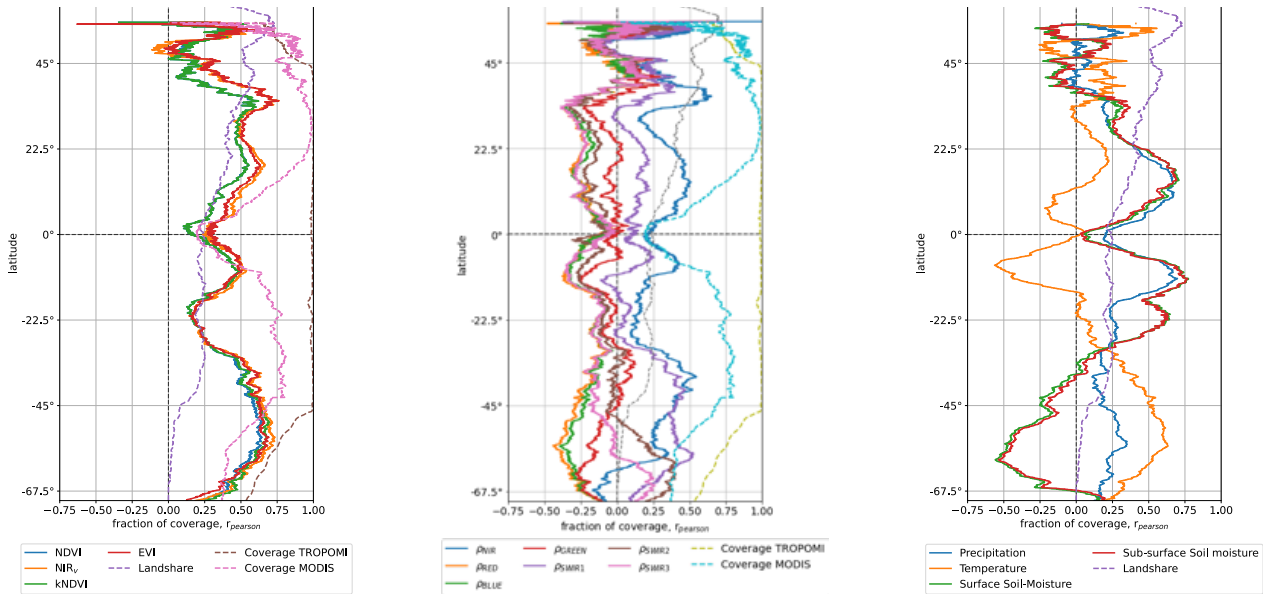


Figure S1. Pearson correlation coefficient of auxiliary data to TROPOMI SIF. Data is compared in 16-day resolution starting in April 2018 until March 2021. Comparison in lowest resolution of the two products: For ERA-5 at 0.25° , for USDA at 0.1° and for all MODIS data at 0.05° .



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

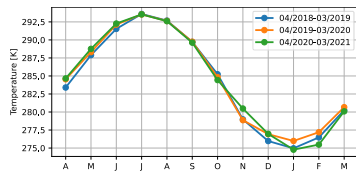
(b) Latitude dependent r_{pearson} of MODIS bands and TROPOMI SIF at 0.05° resolution

(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

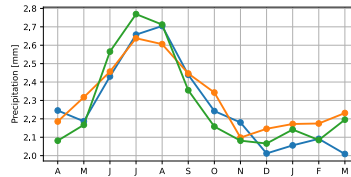
Figure S2. Latitudinal pearson correlation coefficient of auxiliary data to TROPOMI SIF. Data is compared in 16-day resolution starting in April 2018 until March 2020 and March 2021 for ERA-5 data and all others, respectively. Comparison in lowest resolution of the two products: For ERA-5 and USDA soil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

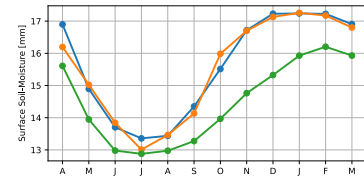
20 The vegetation indices therefore show a very contrary behavior than the SIF measurements.



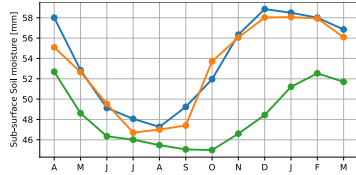
(a) ERA-5 2m mean air temperature



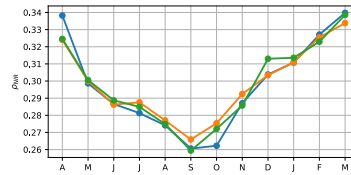
(b) ERA-5 total precipitation



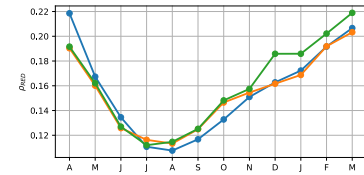
(c) NASA USDA Surface Soil Moisture



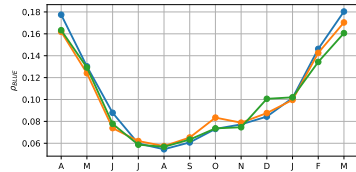
(d) NASA USDA Sub-Surface Soil Moisture



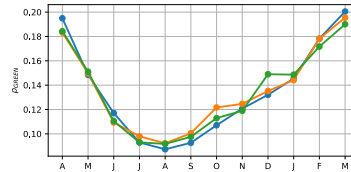
(e) MODIS ρ_{NIR} band



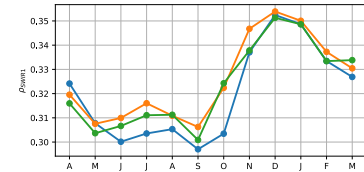
(f) MODIS ρ_{RED} band



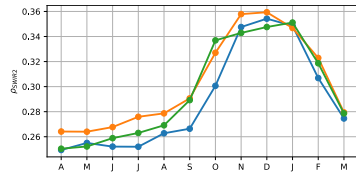
(g) MODIS ρ_{BLUE} band



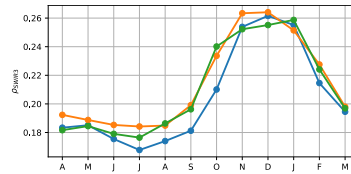
(h) MODIS ρ_{GREEN} band



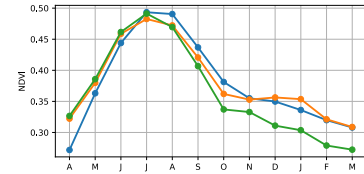
(i) MODIS ρ_{SWIR1} band



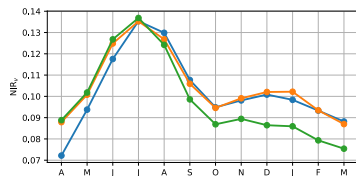
(j) MODIS ρ_{SWIR2} band



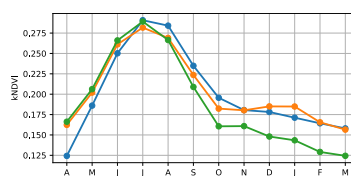
(k) MODIS ρ_{SWIR3} band



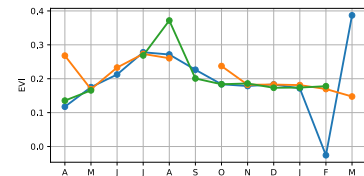
(l) MODIS NDVI



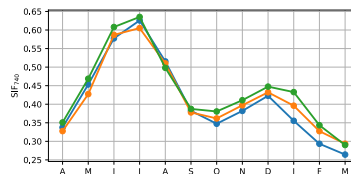
(m) MODIS NIR_v



(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.

- 25 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 is 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.
- 30

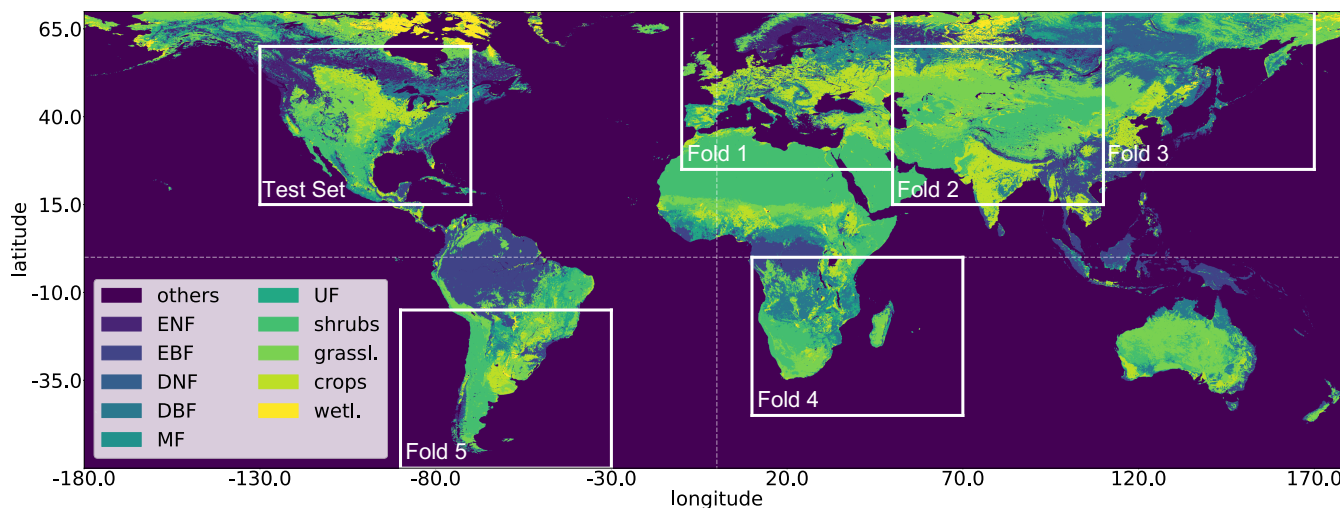
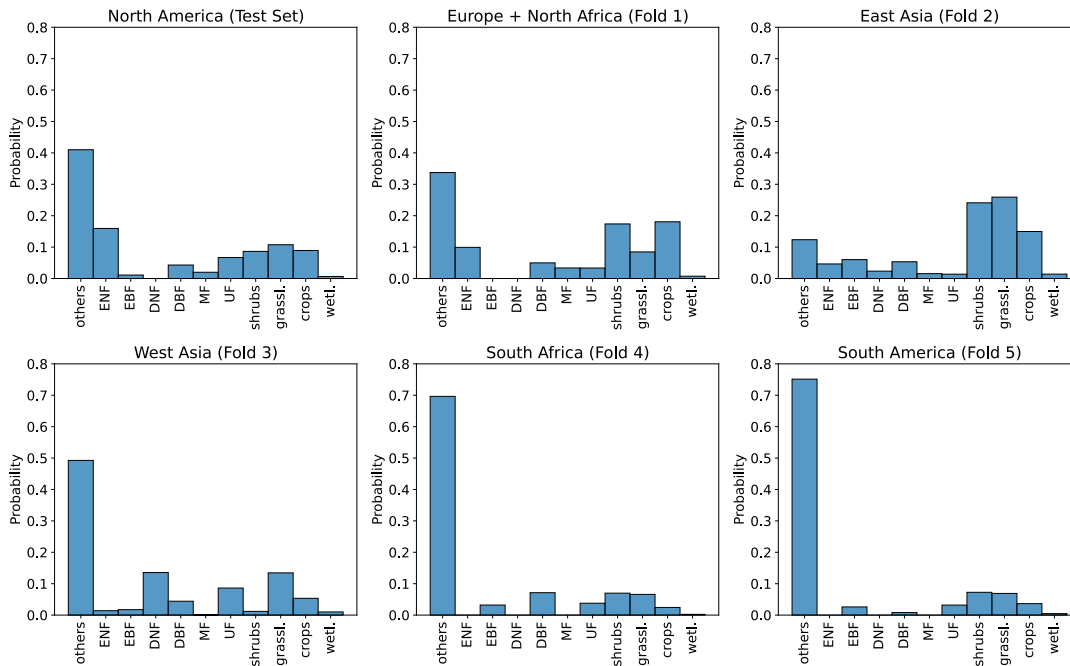


Figure S4. Regions of 5-fold cross validation and test set over dominant land cover type. The land cover type is at a resolution of 0.05° from Corine LULC.

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 upsampled 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.

40 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
45 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)~~(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
55 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.

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.

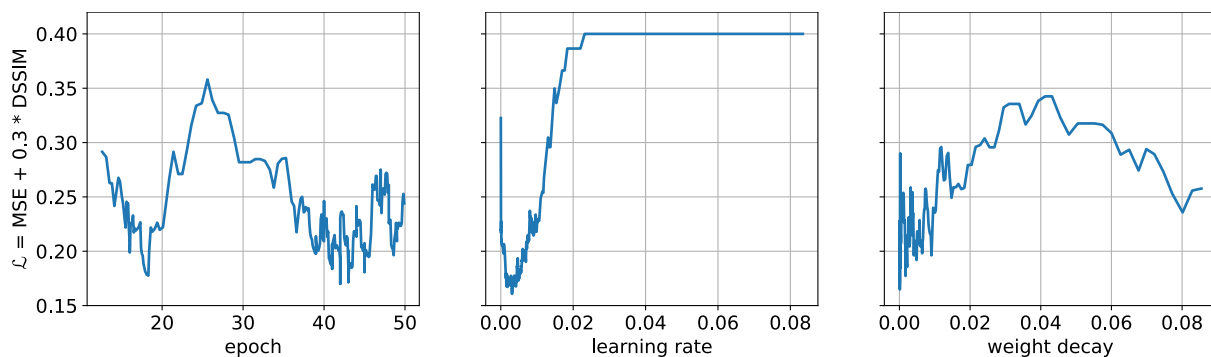


Figure S7. Loss dependent on hyperparameters. y axis shows the losses, x axis the investigated values for the hyperparameters. For visualization a 10-value rolling mean is applied to the data.

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

65 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.

70 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.

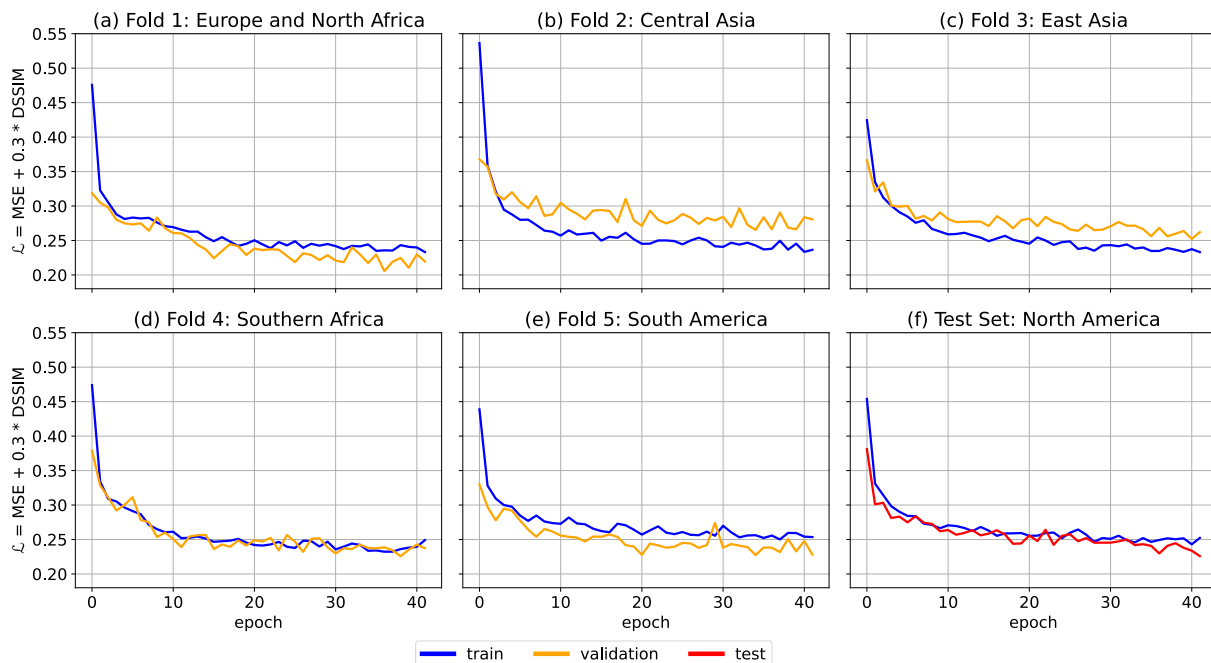
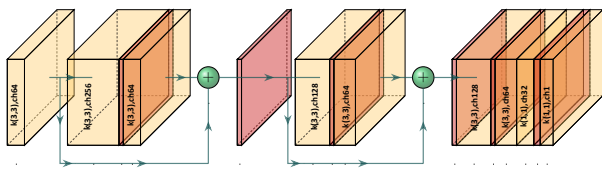


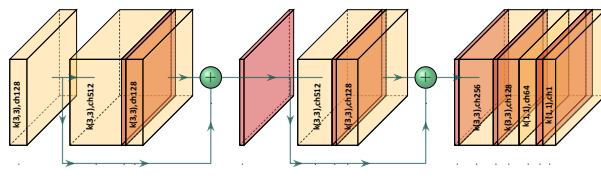
Figure S8. Loss of cross validation. (a) to (e) shows the losses of the 5-fold cross validation. The region in the title of the subfigure refers to the validation fold. All other four folds are used for training. (f) shows the losses of training on all five training folds and testing on the test set.

S4.3 Comparison of Model Structures

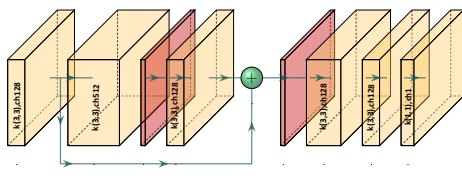
Figure S9 visualizes eight investigated CNN model structures. Model complexity is decreasing from a to h. All models perform ~~similar~~ similarly well (Table S3). Model structure 5 (e) is chosen as the best trade off between complexity and performance. Further, the input feature collinearity and PCA 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).



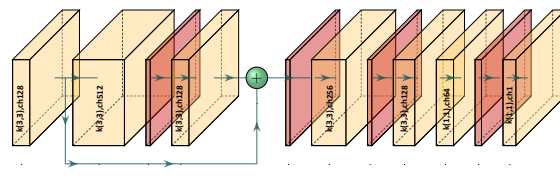
(a) Model Structure 1



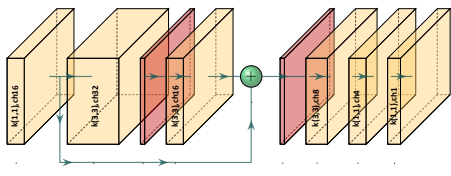
(b) Model Structure 2



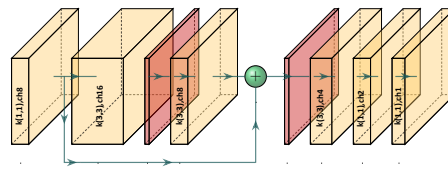
(c) Model Structure 3



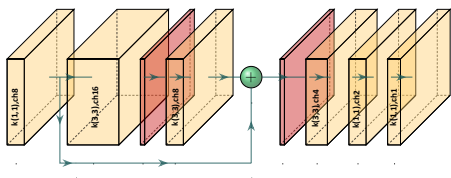
(d) Model Structure 4



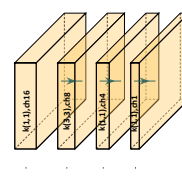
(e) Model Structure 5 (chosen for SIFnet)



(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

85 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
 90 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.

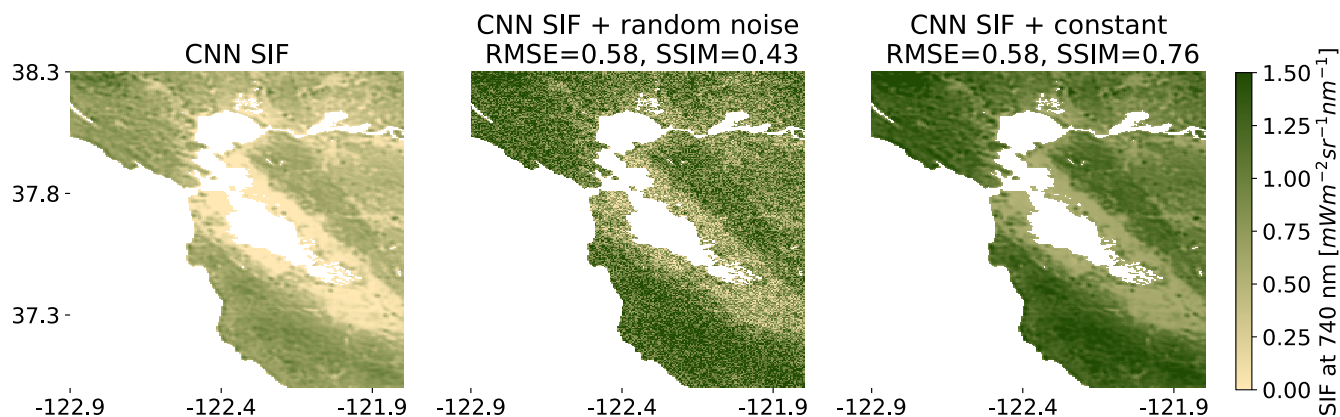
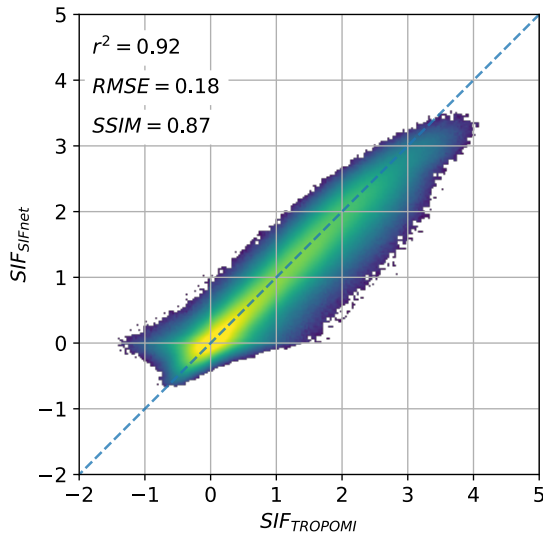


Figure S10. Advantage of own loss function

S4.5 Varying Input Data Combinations

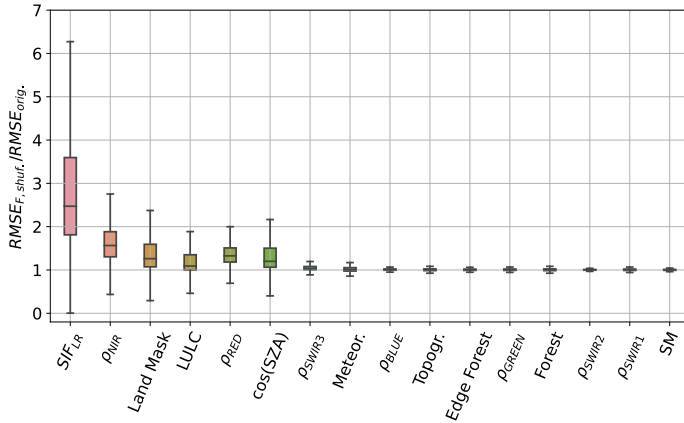
Three further input feature combinations are evaluated in Figures S11, S12, and S14. In the results from Figure S11 the MODIS vegetation indices (NIR_v , $NDVI$, EVI , and $kNDVI$) are not used as features. Therefore, the network needs to get
95 the vegetation information directly from the MODIS bands. 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

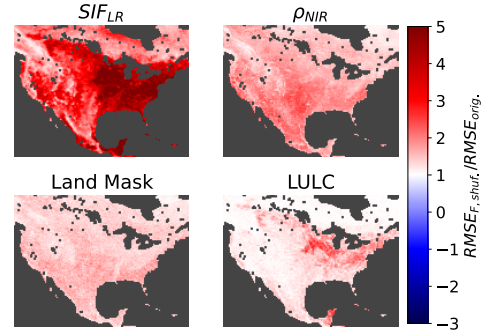
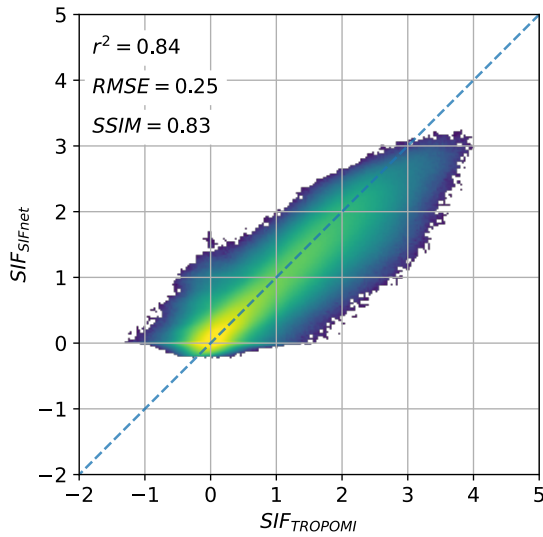


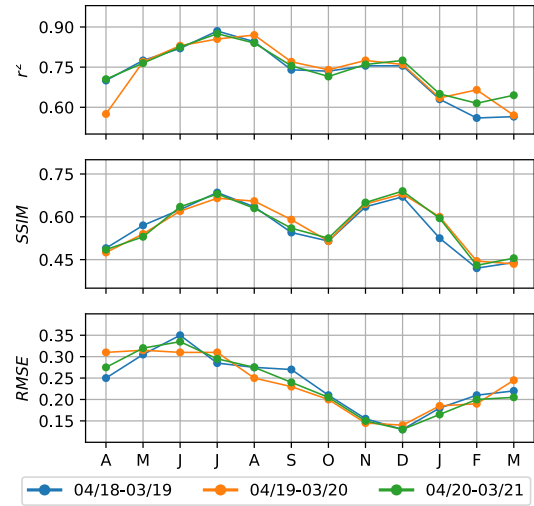
Figure S11. Results of Data Option without any MODIS vegetation indices. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05° . (b) shows the metrics r^2 , $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 day resolution and averaged to monthly 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).

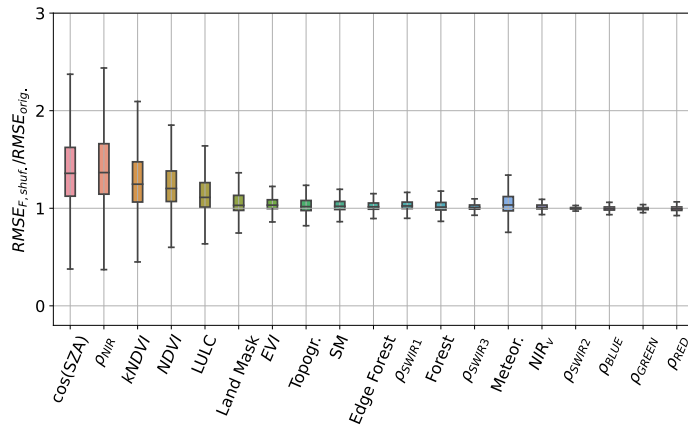
These findings suggests that there are several solutions to the optimization problem that depend on the input feature combination. Reasons for this can be the high collinearity between the features (Section S3). Further, CNNs can process non-linear relationships and can therefore compute the vegetation 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

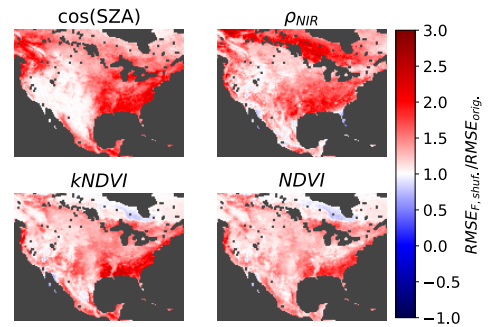
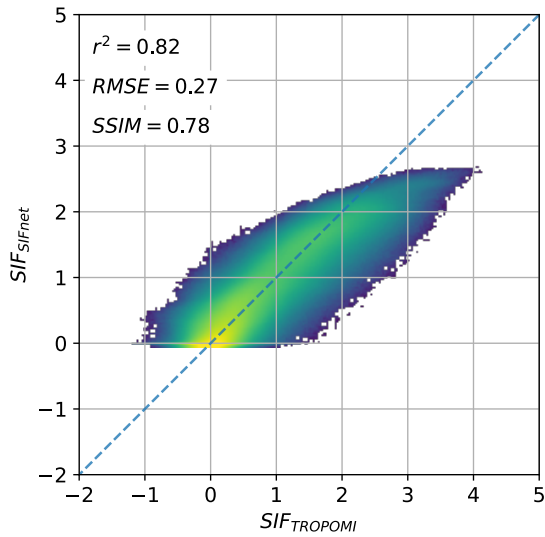
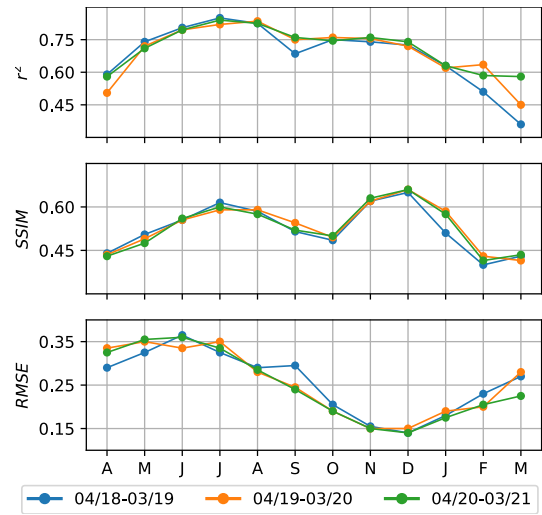


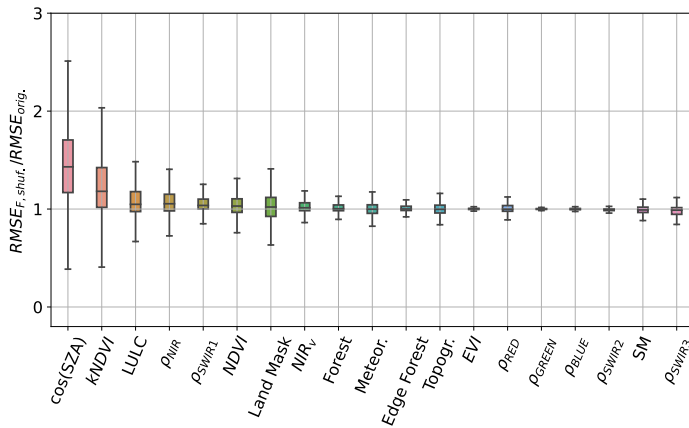
Figure S12. Results of Data Option without low resolution SIF as feature. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05° . (b) shows the metrics r^2 , $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 day resolution and averaged to monthly 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; Meteor.: 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

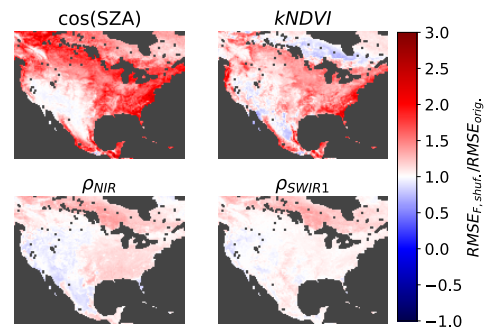


Figure S13. Results of Data Option without low resolution SIF as feature with model structure 2 from Figure S9 (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05° . (b) shows the metrics r^2 , $SSIM$, and $RMSE$ of each investigated month. Metrics is calculated in 16 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; Meteor.: 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).

110 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).

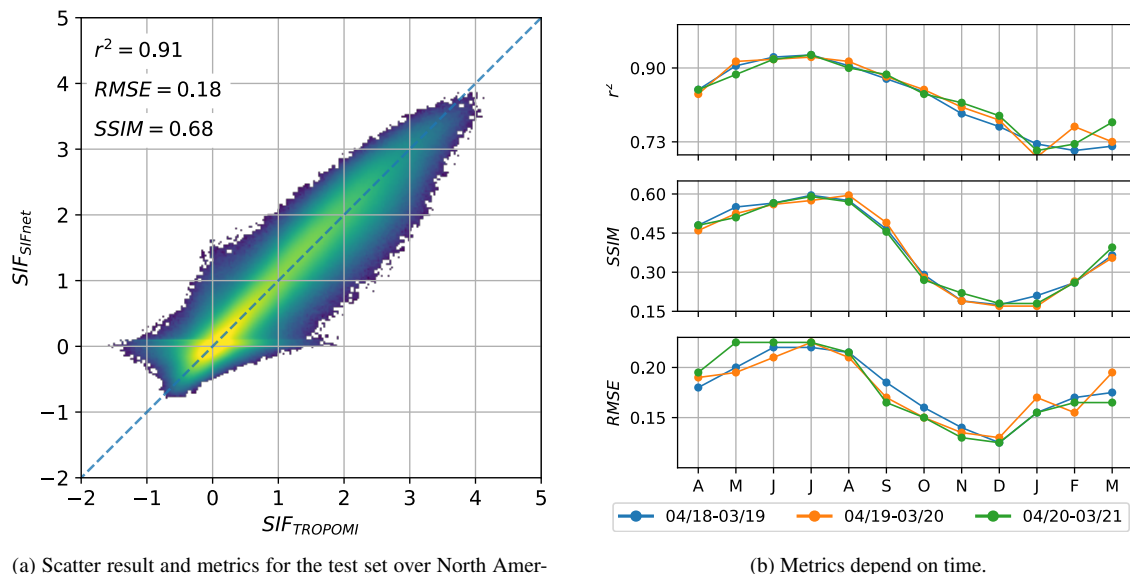


Figure S14. Results of Data Option with only SIF_{LR} and NIR_v as features. (a) shows the scatter comparison between TROPOMI SIF and the SIFnet estimate at 0.05° . (b) shows the shows of each investigated month the metrics r^2 , SSIM, and RMSE. Metrics is calculated in 16 day resolution and averaged to monthly values afterwards.

115 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 an 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
 120 (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.

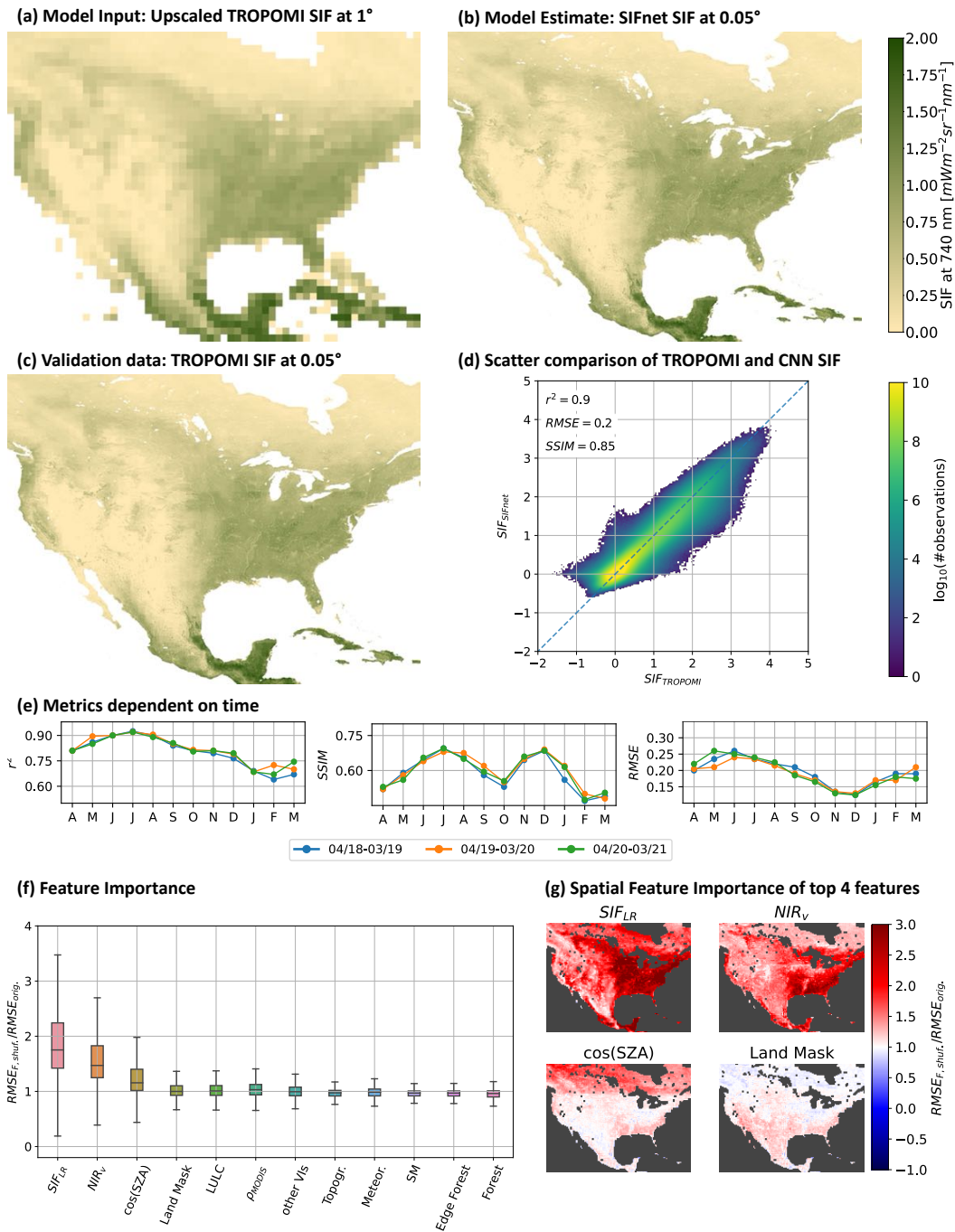


Figure S15. Test set results of SIFnet training at 0.05° with a scaling factor of 20. (a) shows low resolution SIF that is used as model input. (b) shows the estimated SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between 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.

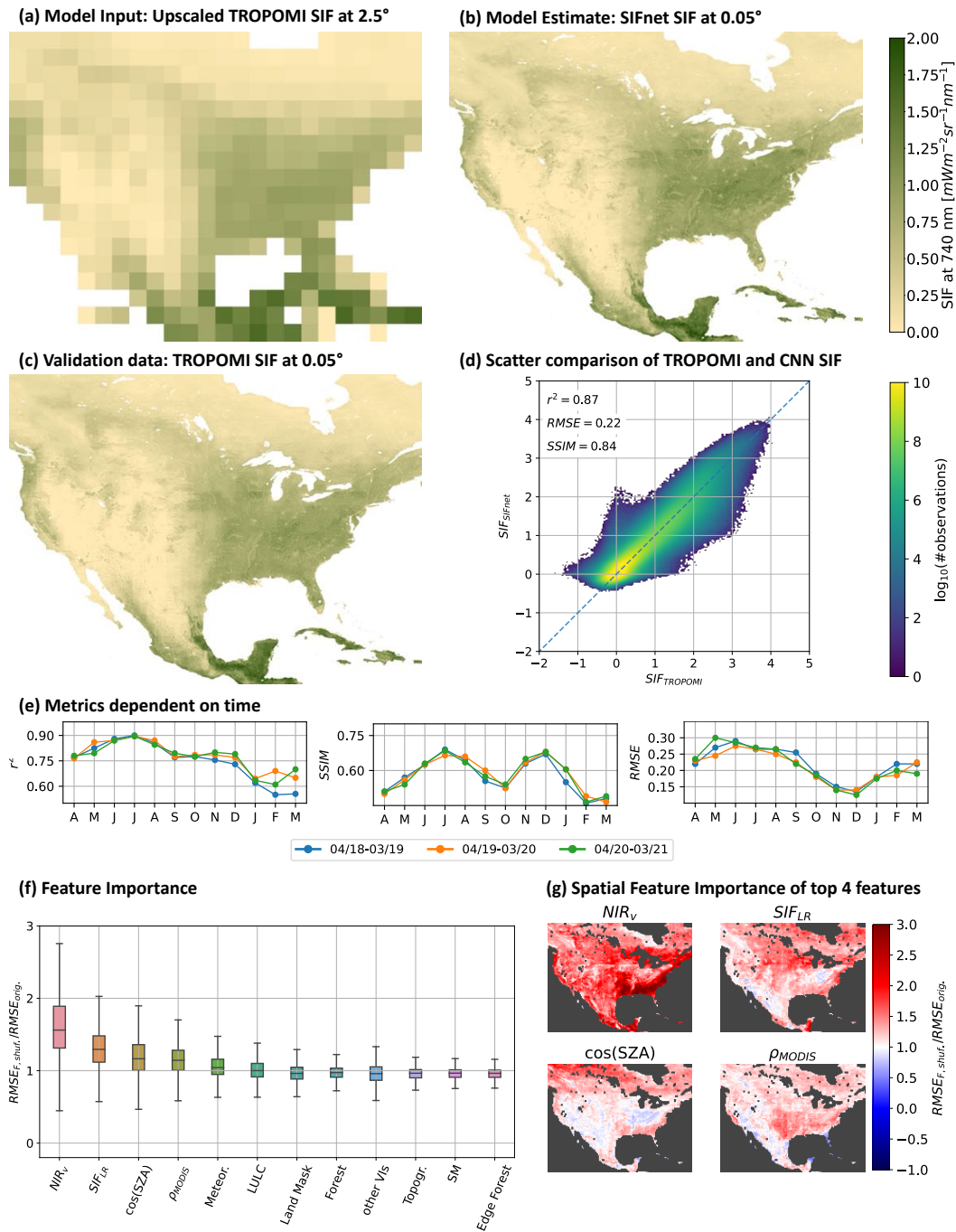


Figure S16. Test set results of CNN training at 0.05° with a scaling factor of 50. (a) shows low resolution SIF that is used as model input. (b) shows the estimated SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison between 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.

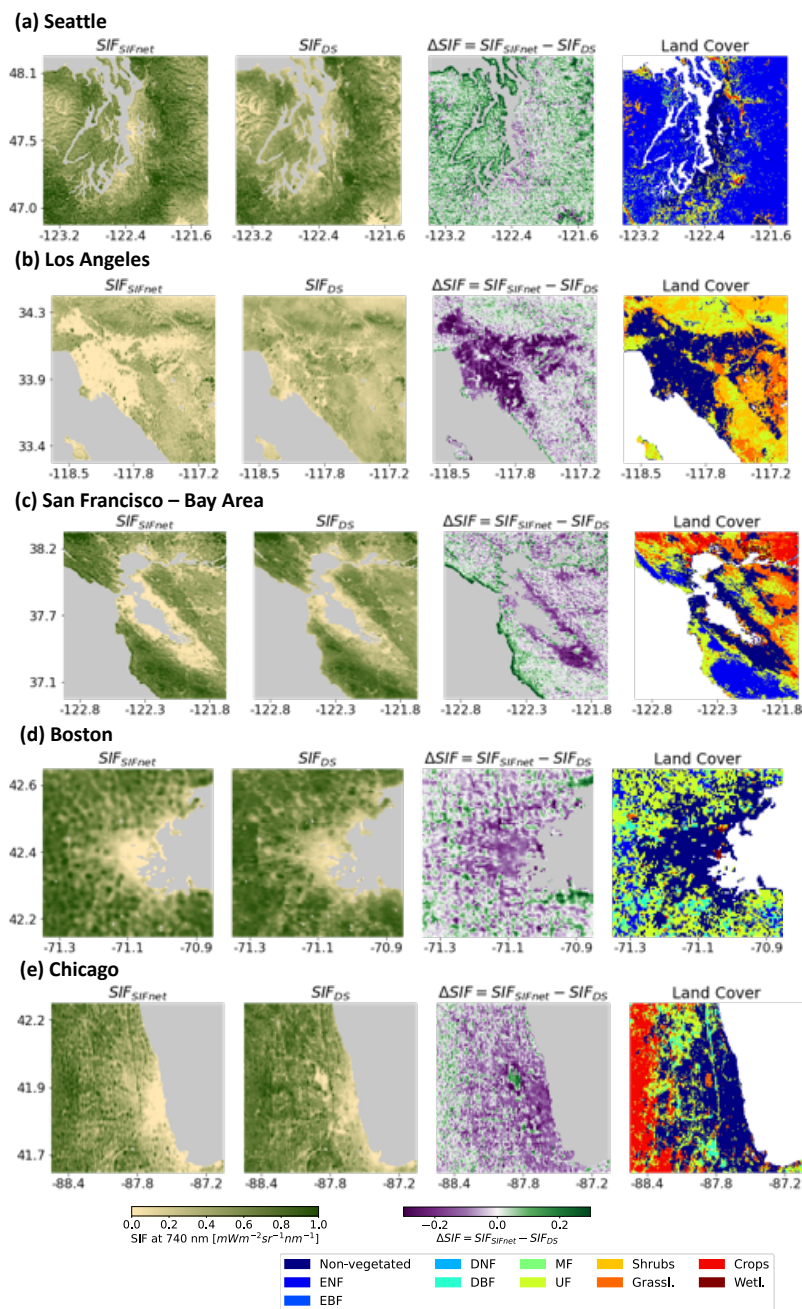


Figure S17. SIFnet and downscaled SIF, the difference between these, and the dominant land cover type for five urban regions across CONUS. First column shows the SIFnet estimate, second the downscaled SIF from Turner et al. (2020), third the difference between SIFnet and downscaled SIF and the last shows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

S6 Training SIFnet on daily integrated SIF estimates

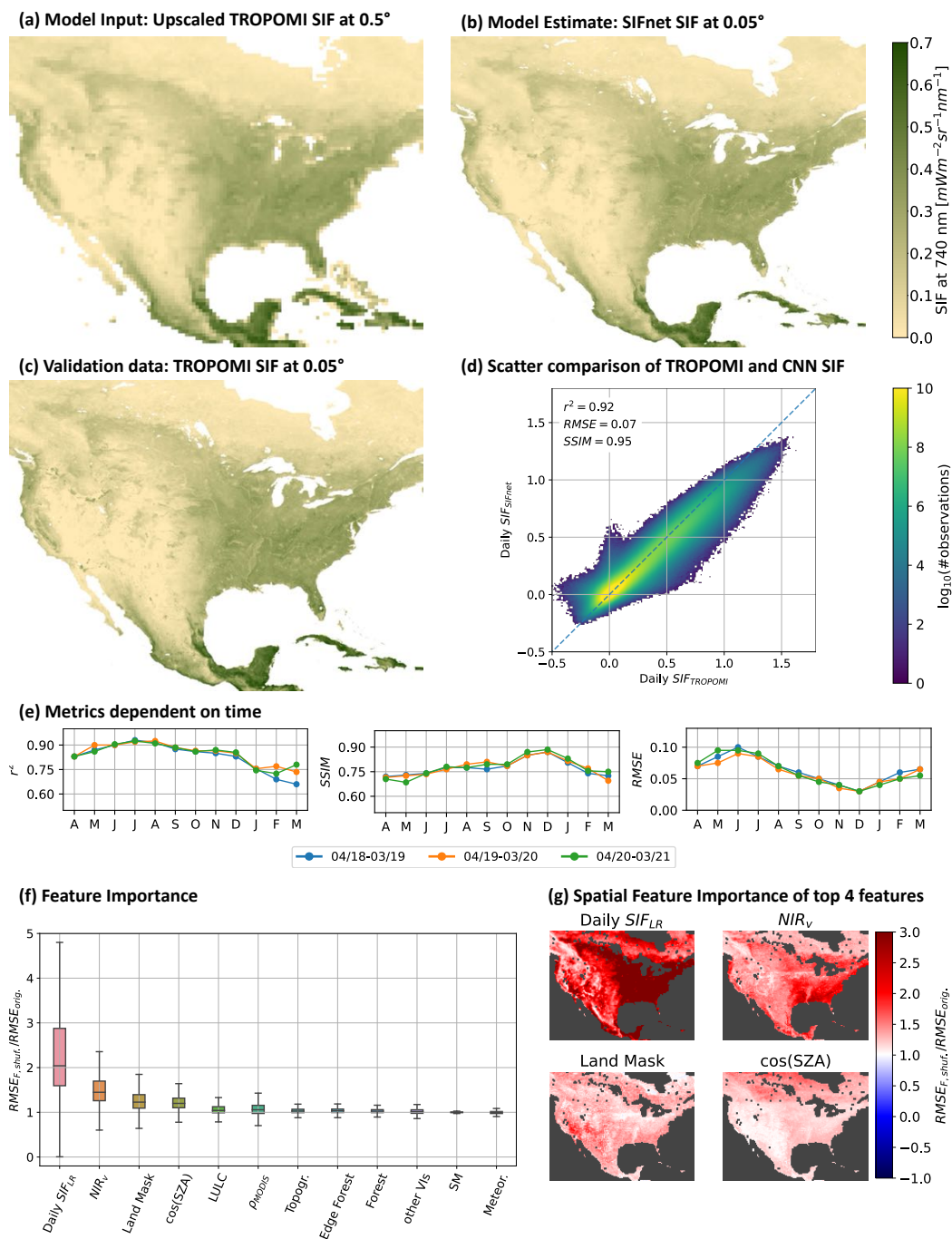
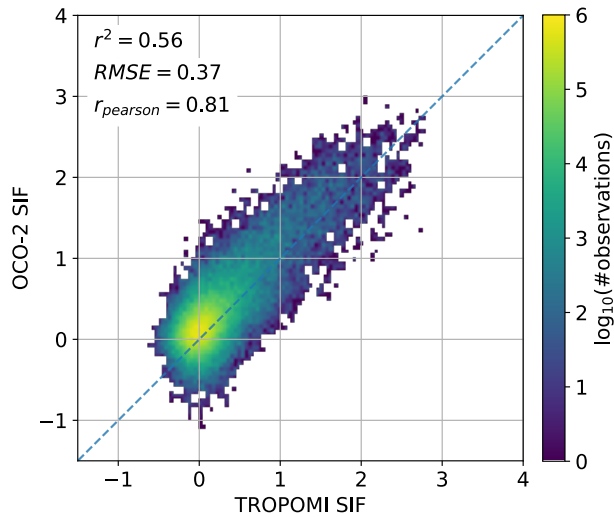


Figure S18. Test set results of SIFnet training at 0.05° using daily SIF estimates (a) shows low resolution Daily SIF that is used as model input. (b) shows the estimated Daily SIF at 0.05° by SIFnet. (c) shows the measured TROPOMI Daily SIF at 0.05° from Köhler et al. (2018). (d) shows the scatter comparison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

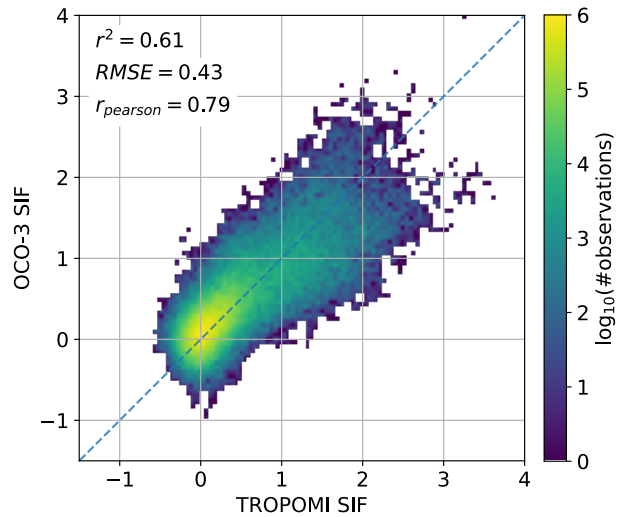
Figure S18 shows the result of training on daily SIF estimates. For estimating daily corrected SIF the instantaneous measurement is scaled by the cosine of the solar zenith angle (Equation 7 main text). r^2 is similar to the approach of instantaneous SIF and the *SSIM* shows a better result. RMSE is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

130 S7 Validation against OCO-2 and OCO-3 data

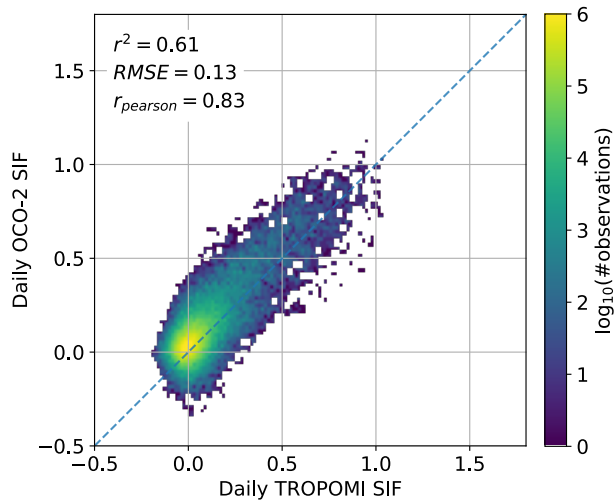
~~Figure S19a-~~



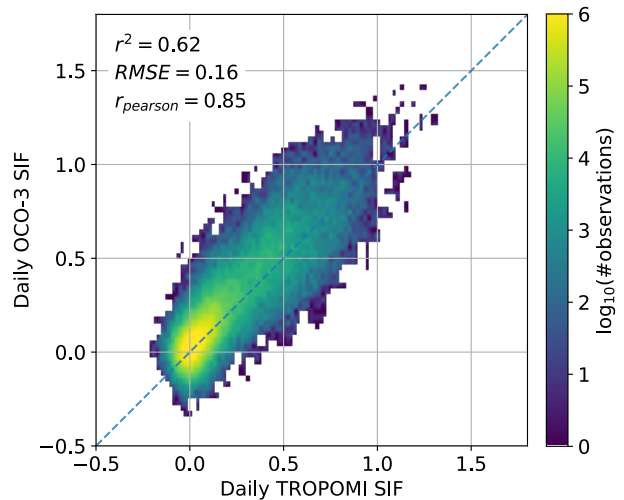
(a) OCO-2 SIF vs. TROPOMI SIF



(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

135 are from 25° to 50° and longitudinal border are from -126° to -66. shows a better agreement with-

140 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 ~~than with~~ 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 ~~in the metrics~~ 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 ~~but a weaker correlation in the person correlation coefficient~~. Generally, the data sets deviate with a high *RMSE* relative to the value range. *RMSE* is 0.37 and 0.67, only additional aggregation lead 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 ~~for~~ $mWm^{-2}sr^{-1}nm^{-1}$ for the CONUS. That error is close to the *RMSE* between instantaneous TROPOMI SIF and instantaneous OCO-2 ~~and~~ 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 ~~, respectively,~~ 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 measurements decreases.

150 **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.

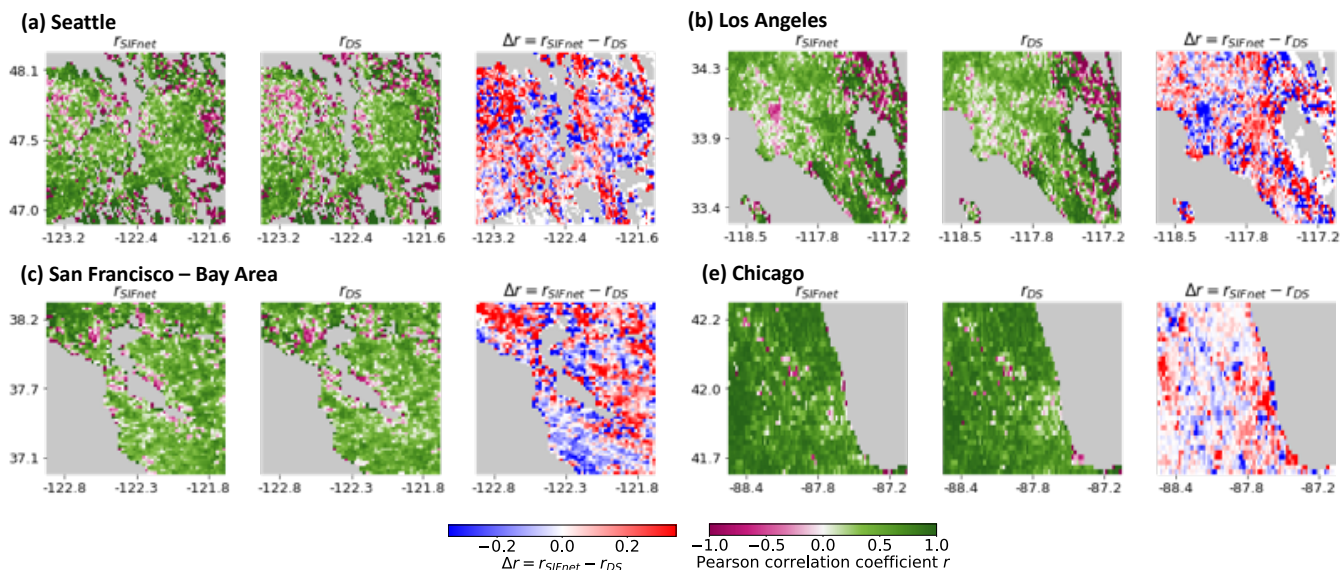


Figure S20. SIFnet and downscaled SIF, the difference between these, and the difference in correlation to OCO-2 and OCO-3 for four urban regions. First column shows the correlation of SIFnet SIF to OCO-X, second the correlation of downscaled SIF from Turner et al. (2020) to OCO-X, third the difference 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 deciduous 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

155 **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.
- 160 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
165 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
170 photosynthesis as observed from space, *Biogeosciences*, 17, 405–422, 2020.