Sentinels4Carbon (Sense4Fire)

Sentinel-based fuel, fire and emissions products to constrain the changing role of vegetation fires in the global carbon cycle

ESA Contract Number: 4000134840/21/I-NB

Product Validation Report Version 2.1

(PVRv2.1)

7 May 2023, Version 2.1

Prepared by:

Matthias Forkel, Christine Wessollek, Daniel Kinalczyk, Christopher Marrs

Technische Universität Dresden, Faculty of Environmental Sciences, Dresden, Germany

Vincent Huijnen, Jos de Laat

Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

Niels Andela, Alfred Awotwi

Cardiff University, School of Earth and Environmental Sciences, Cardiff, Wales, UK

Contents

Figures

Tables

Acronyms

AAI	Absorbing Aerosol Index
AGB	Above Ground Biomass
AIC	Akaike's Information Criterion
ALH	Aerosol Layer Height
ATBD	Algorithm Theoretical Baseline Document
BAAD	Biomass and Allometry Database
CAMS	Copernicus Atmosphere Monitoring Service
CCI	Climate Change Initiative
DRC	Democratic Republic of Congo
ESA	European Space Agency
GAM	Generalised Additive Model
GEDI	Global Ecosystem Dynamics Investigation
GFAS	Global Fire Assimilation System
GFED	Global Fire Emission Database
IFS	Integrated Forecasting System
LAI	Leaf Area Index
LFMC	Live Fuel Moisture Content
MOPITT	Measurement of Pollution in the Troposphere
OMI	Ozone Monitoring Instrument
PVP	Product Validation Plan
PVR	Product Validation Report
SLA	Specific Leaf Area
SMAP	Soil Moisture Active Passive
TROPOMI	TROPOspheric Monitoring Instrument
VIIRS	Visible Infrared Imaging Radiometer Suite
VOD	Vegetation Optical Depth
VODCA	Vegetation Optical Depth Climate Archive
WP	Work Package

1 Introduction

This Product Validation Report version 2.1 (PVRv2.1) describes the strategy and results for product calibration and validation for methods and products that are developed within the ESA-funded Sense4Fire project: Sentinel-based fuel, fire and emissions products to constrain the changing role of vegetation fires in the global carbon cycle. The aim of Sense4Fire is to increase the scientific understanding of fire dynamics and their role in the carbon cycle by integrating observations from the Sentinels into new Earth observation products.

The validation of data products is crucial for the evaluation of the complex production chain underlying the Sense4Fire project: That is, algorithms that combined data from either Sentinel 1-2-3 into data products, algorithms that merge those data products into emissions, CAMS/IFS model simulations that use those emissions to simulate the spatio-temporal structure of the chemical composition of the atmosphere, and Sentinel-5p atmospheric composition data that is used to validate and re-calibrate those model simulation results. Differences between the CAMS/IFS model simulation and Sentinel-5p observations thus can be caused by one of the many processes in the Sense4Fire production chain. Hence the validation should not only focus on the comparison between model results and Sentinel-5p data, but also on the validation of individual data products further down the production chain. Understanding and quantifying errors and uncertainties is important to identify where and when the production chain can be improved. The validation may lead to immediate data product improvement, but equally to identification of data product limitations that cannot be immediately resolved and require additional action. The latter will find its way into a gap analysis that will be part of the PVR and final project report. The gap analysis can be used to initiate or support additional research and development activities beyond the Sense4Fire project end.

PVRv2.1 builds on the description of methods and algorithms from the Algorithm Theoretical Baseline Document version 2.1 (ATBDv2.1). ATBDv2.1 describes the used datasets and methods and hence PVRv2.1 describes calibration/validation strategy and initial results for selected test areas and for the Brazilian study region. An overview of the study regions and test areas is provided Chapter 2 in the ATBDv2.1.

The following chapters present validation and benchmarking results for the different approaches developed in Sense4Fire:

- **GFA-S4F** is based on the Global Fire Atlas (GFA) algorithm (Andela et al., 2019, 2022) and uses observations of active fires from the VIIRS and Sentinel-3 SLSTR instruments with a new fire type map to estimate fire emissions. Chapter 2 presents validation results for the classification of fire types.
- **TUD-S4F** is a new data-model fusion approach that combines several datasets from Sentinel-3 and other Earth observation products to estimate fuel loads, fuel moisture, fuel consumption, and fire emissions. Chapter 3 presents calibration, test and validation results of the S4F model with respect to tree height, biomass

compartments, fuel moisture content, litter and woody debris and fuel consumption, and emission factors.

- **KNMI-S5p** is based on observation from Sentinel-5p, whereby fire emissions of CO and NOx are estimated using a top-down approach. In Chapter 4, we then the first validation results of estimates of fire emissions from a top-down approach using Sentinel-5p observations.

2 GFA-S4F: Burned area and fire behaviour

This project component aims to deliver three products. First, we track individual fire perimeters as objects at a daily temporal resolution. Second, we classify these objects as specific fire types, for example forest, deforestation, or savannah fires. Third, we estimate, and scale burned area for each object, a critical step for estimating fire emissions. Each step requires training and quality assessment, training efforts are clarified in the ATBDv2 and details about quality assessment are provided here.

2.1 Tracking fire objects

Our aim is to identify individual wildfires and track the daily expansion of the fire perimeter. Defining a 'wildfire' itself can be difficult at times, as each continuous burned area can include multiple ignition locations sometimes driven by the fire itself (e.g., spotting). The precise definition of fire objects, and subsequent estimates of fire perimeters, has an indirect influence on our emissions estimates as it determines the aggregate fire properties and therefore the fire type and attributed burned area of each object. Here we use the Global Fire Atlas algorithm (Andela et al., 2019) to track individual wildfires at each time step.

In open cover types with rapid fire spread rates, active fire detections do not provide full coverage of the fire extent and fire events become fragmented in our approach. We do not account for this in our fire tracking algorithm, since the fire types for open cover are directly dependent on land cover type, but not fire behaviour or other fire attributes. Instead, we scale burned area for each object to provide robust estimates of burned area across larger scales. Given the large and long-term impacts of fires in forested systems on carbon emissions, our focus is on forest fires, which are generally well captured by active fire detections. Here we use our Sentinel-2 based burned area estimates (continuous burned area) in combination with visual interpretation of the burn date from the active fire detections to provide additional quality assessment of fire perimeters in closed cover types in addition to the assessment provided by Andela et al., 2019.

2.2 Fire type classification

Here we aim to differentiate eight different fire types (Table 1), with some fire types specific to one of the study regions and others more generic. Quality assessment of the fire type classification is based on pre- and post-fire Sentinel-2 pairs. To avoid issues related to class imbalance, we use a stratified random sample of relevant fire types for each region (~100 fires in each class) to assess the accuracy of our classification. Results

are presented both on a per fire-object basis and on a per active fire detection basis. Large fires are often better characterised and can therefore more accurately be attributed to a specific fire type, therefore the accuracy for active fire detections (closely related to carbon emissions) is often higher than that of individual events (including many small fires). Below, we provide an initial assessment of our capability to separate forest from deforestation fires in Brazil, the first two fire types in Table 1.

Table 1: Fire types by study region.

Fire type	Study region	Tree cover fraction (%)
Evergreen broadleaf forest fires	Brazil	≥50%
Deforestation fires	Brazil	≥50%
Small clearing and agricultural fires	Brazil	≥50%
Boreal forest ground fires	Russia	≥50%
Boreal forest crown fires	Russia	≥50%
Woodland fires	Africa	≥50%
Savannah and grassland fires	Brazil, Africa, Russia	<50%
Cropland burning	Brazil, Africa, Russia	<50%

Separating forest from deforestation fires in Brazil

We assessed the accuracy at which we can separate deforestation from understorey forest fires, two key Amazon fire types associated to forest degradation (Table 2 and Figure 1). Because of the class imbalance, with about five times more deforestation than forest fires, we took a stratified random sample of 100 deforestation and 100 understorey fires across the South American domain in 2019. We focused on fires that started in August, to avoid issues of cloud cover, and used nearest cloud free pre- and post-fire Sentinel-2 10m resolution images in time to interpret the reference fire type. In total, this resulted in a reference set of 194 fires across 118 image pairs after excluding one fire event due to cloud cover and five fire events due to inconclusive evidence of burning in the image pair. Overall, the accuracy of fire type classification improved with fire size. The overall accuracy was 66% (DICE coefficient of 0.66) for fire events and 92% for active fire detections (large fires include more active fire detections; Table 2). The largest difference in accuracy was observed for forest fires, with 55% User's accuracy (reflecting commission error) for fire events increasing to 93% User's accuracy for fire detections, likely indicating a more skewed distribution of fire size and associated fire detections compared to deforestation fires. Deforestation fire classification was more stable, with 78% User's and 63% Producer's accuracy for fire events, and 87% User's and 71% Producer's accuracy for fire detections.

Table 2: Accuracy assessment based on 100 Amazon deforestation and 100 Amazon forest fires. The algorithm accurately separated most deforestation from understorey forest fires. (a) Accuracy assessment of fire events and (b) accuracy assessment of active fire detections.

a) Fire events		Reference data			
		Deforestation	Forest	Total	User's Accuracy
Classification data	Deforestation	75	21	96	78%
	Forest	44	54	68	55%
	Total	119	75	194	
	Producer's accuracy	63%	72%		
	Overall Accuracy = 66%				

b) Fire detections		Reference data			
		Deforestation	Forest	Total	User's Accuracy
Classification data	Deforestation	4,309	667	4,976	87%
	Forest	1,721	22,997	24,718	93%
	Total	6,030	23,664	29,694	
	Producer's accuracy	71%	87%		
	Overall Accuracy = 92%				

Figure 1: Example highlighting how fire type can be interpreted from 10-m resolution pre- and post-fire Sentinel-2 image pairs. (a) Deforestation before fire; (b) Deforestation after fire; (c) Mixed forest and deforestation fire before (classified as forest fire)); (d) Mixed forest and deforestation fire (classified as forest fire); (e) Forest fire before (classified as forest fire); (f) Forest fire after (classified as forest fire)

2.3 Burned area estimates

To translate 'fire polygons' to fire emissions we quantify the burned area associated with each fire object. As can be seen in Figure 1, the 550 m (0.005°) spatial resolution of our approach can result in an overestimate of burned area for e.g., deforestation fires that often occur at smaller scales. In contrast, fast moving grassland and savannah fires often burn significant areas not captured with the three daily (1:30 am, 1:30 pm, and 10:00 pm) active fire observations. We therefore developed a strategy to scale burned area with large scaling factors (>1) required for open cover types and small scaling factors (<1) for some other fire types, like deforestation fires. Interestingly, we find that a scaling factor of one is appropriate for most forest fires, both in the tropics and at higher latitudes. Here we assess the quality of these scaled burned area estimates by comparison to Landsat based MapBiomas burned area estimates for Brazil (Figure 2). For the other regions, we create a Sentinel-2 based reference burned area dataset for 2020 using the BAMT approach developed by Roteta et al. (2021).

For Brazil, we classified the MapBiomas burned area into burned area from different fire types. Continuous clusters of burned area were overlayed with fire perimeters derived here, and their fire type was assigned on the basis of the fire type that had most overlap with the burned area clusters. Burned-area clusters that remained entirely undetected by our approach were assigned to the savanna class if average tree cover was below 50% (consistent with our approach) or to a "residual" class if tree cover was above 50%.

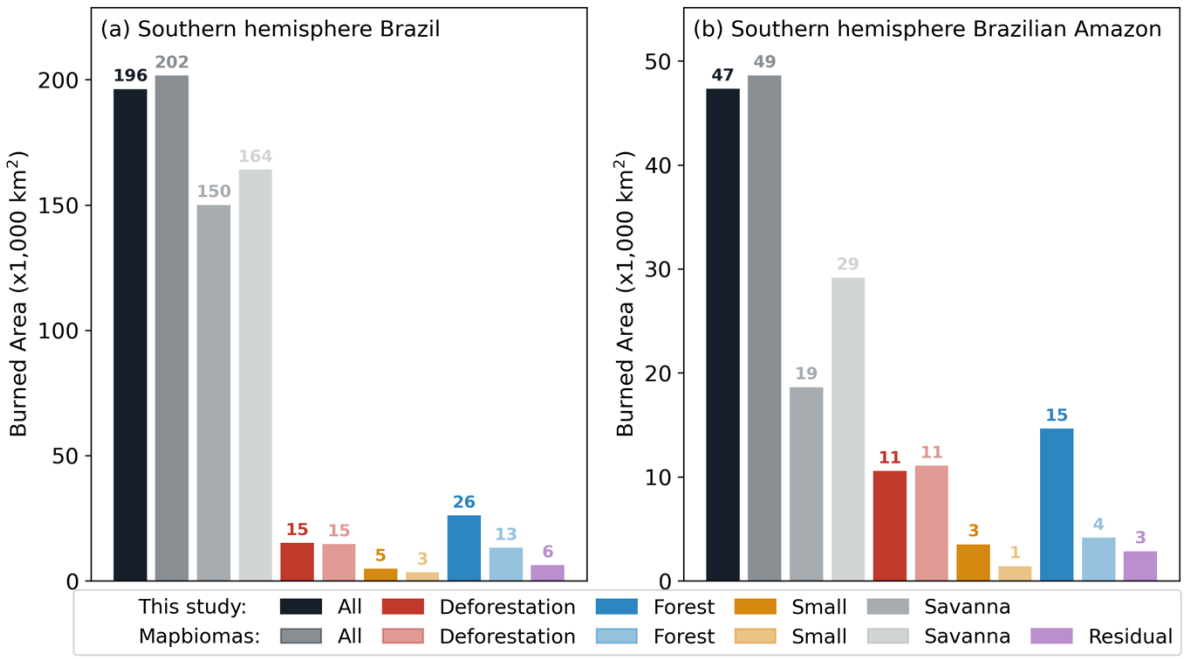

Figure 2: (a) Burned area estimates for southern hemisphere Brazil (0-25°S), and (b) burned area estimates for the southern-hemisphere Brazilian Amazon. Close comparison was found between regional burned area estimates from our approach and burned area from MapBiomas for most fire types, although we identified a larger area of forest fires.

For boreal forest fires, our approach provides a close match to the reference burned area data as assessed using standard statistical metrics (slope, r^2, RMSE) (Figure 3). For open

cover types we aim to quantify burned area at larger scales, in this case we provide standard metrics of quality assessment (User's/Producer's accuracy, precision and recall, and DICE coefficient) at relevant spatial scales.

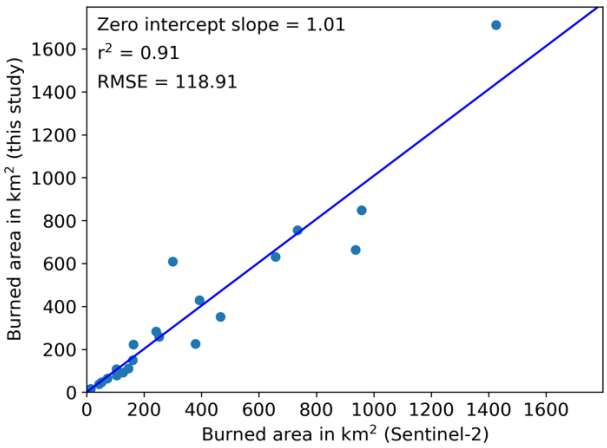

Figure 3: Burned area estimates for several clusters of fires in the boreal Russia highlight close correspondence between our burned area estimates and those from Sentinel-2 for forested systems in the boreal region.

3 TUD-S4F: Fuel loads, moisture content, fuel consumption and fire emissions

3.1 Conducted calibration/validation activities in PVRv2.1

3.1.1 Overview

Within Sense4Fire, two approaches are developed to estimate fuel loads. The first approach is a comprehensive but still simple **S**atellite data-based model of ecosystem **F**uel load, **F**uel moisture, **F**uel consumption and **F**ire emissions (hereinafter referred as **S4F model**). The S4F model is solely driven by satellite products and is constrained by various satellite and ground observations. The second approach is a **machine learning model (TUD-S4F)** to estimate fuel loads by combining the fuel load information from the North American Wildland Fuel Database with satellite datasets. Combustion completeness, fuel consumption and fire emissions are also estimated within the S4F model based on its own estimates of fuel loads. Details of both approaches are presented in the ATBD (Sections 4.3 and 4.4 of ATBDv2.1). In this PVR, calibration and validation results for the both approaches are presented. Specifically, calibration and validation tests were performed for the estimation of canopy height (Section 4.3.2 of ATBDv2.1), tree biomass components (Section 4.3.3 of ATBDv2.1), surface litter and woody debris (Section 4.3.5 of ATBDv2.1), fuel moisture and vegetation water content (Section 4.3.6 of ATBDv2.1), vegetation optical depth (Section 4.3.7 of ATBDv2.1), fuel loads, combustion completeness and fuel consumption (Sections 4.3.8 and 4.3.9 of ATBDv2.1), and of emission factors (Section 4.3.10 of ATBDv2.1)

3.1.2 Satellite-based input and calibration datasets

The S4F model requires input data on land cover (i.e. fractional cover of trees f_{tree} and herbaceous vegetation f_{herb}), leaf area index (LAI), fractional vegetation cover (fCOVER), soil moisture (here Soil Water Index, SWI was used) and burned area in order to compute fuel loads, fuel moisture, fuel consumption and hence fire emissions. In addition, the SF4 model can then calibrated at grid cell-level against further satellite datasets; here we used datasets of canopy height, above ground biomass, Ku-, X- and L-band Vegetation Optical Depth, Live-Fuel Moisture Content (LFMC), and Fire Radiative Energy (FRE). The processing and spatial resampling of the different resolutions of those datasets is explained in ATBDv2.1.

The following datasets were used as input to the S4F model:

- Land cover time series and of tree and herbaceous cover were taken from the PFT fractions derived from ESA CCI land cover map. All tree and shrub PFTs from the ESA CCI dataset were summed to calculate the total tree cover fraction (ftree). As the ESA CCI land cover map shows only little temporal variability at a resolution of 300 x 300 m and hence might underestimate land cover changes, we made one additional model run in which we used the annual land cover data set by Song et al. (2018) as alternative input to the model.
- LAI and fCOVER from Sentinel-3 OLCI and Proba-V Version 1.1 (Fuster et al., 2020) were used at the original resolution of 333 m and 10 day product. The LAI ranges from January 2014 to October 2021 (281 observations, 36 observations per full year). The spatial resolution of the LAI dataset was used as reference for all other datasets and hence all datasets have been resampled to the LAI data as described in ATBDv2.1 (section 4.2).
- SWI from Metop/ASCAT (Bauer-Marschallinger et al., 2018) was used at a resolution of 0.1° for the period since 2007. The SWI is used as proxy for surface fuel moisture. The SWI dataset is already available as a 10 day product, which uses the same 10 day interval as the LAI dataset. We use a temporal subset of the SWI time series from January 2014 to October 2021 (observation period of the LAI time series), which is resampled to the spatial resolution of the LAI data set using nearest neighbour interpolation.
- Burned area was taken from different products and tested. In the PVRv2.1, we show results based on the ESA CCI burned area dataset version 5.1 (Lizundia-Loiola et al., 2020), which was readily available for all study regions.

The following datasets were used to calibrate parameters of the S4F model and hence to retrieve estimates of fuel load, fuel moisture, fuel consumption and fire emissions at a grid-cell basis:

- Canopy Height: Since the first analysis in PVRv1 revealed that the map of forest canopy height from Potapov et al. (2021), in the comparison referred as GLAD, saturates at high biomass values in comparison to the above ground biomass map

from ESA CCI, we now used the GEDI L3 Gridded Land Surface Metrics product (see Section 4.2.4 of ATBDv2.1) for the study regions Amazon and southern Africa, and the Global Canopy Top height dataset from ETH (Lang et al., 2022) for Siberia. To correspond with the LAI data, the datasets were resampled to the same spatial resolution as the LAI data and then matched to the LAI pixel grid using nearest neighbour interpolation.

- The ESA CCI Biomass map was used for the years 2010, 2017 and 2018 (Santoro et al., 2021) and linearly interpolated to the 10-daily values of the S4F model. The spatial resolution of 100 m of this dataset was resampled to the resolution of the LAI data using nearest-neighbour resampling.
- FRE was taken from Sentinel-3 and VIIRS observations as described in chapter 3.3 of ATBDv2.1.
- LFMC was taken from the VOD2LFMC product (Forkel et al., 2023) to calibrate the parameters for the computation of LFMC.
- Ku-band and X-band Vegetation Optical Depth (Ku-VOD, X-VOD) from the VODCA dataset (Moesinger et al., 2020) was used for the calibration of vegetation water content (VWC) and VOD. The dataset was interpolated to the spatial and temporal resolution of the LAI data.
- L-VOD from the SMOS-LPRM retrieval dataset was used for the period 2010-2020 (van der Schalie et al., 2017). The dataset was interpolated to the spatial and temporal resolution of the LAI data.

3.1.3 Calibration and validation against databases of ground observations

All calibration and validation exercises are based on the S4F model run using the ESA CCI land cover map as input if not stated otherwise. The following databases of ground observations were used for the calibration and/or validation of the different components of the S4F model:

The Biomass and Allometry Database (BAAD) (Falster et al., 2015) compiles tree-level observations of different biomass components such as the total mass of stems and branches (variable m.st including hardwood, sapwood and bark), mass of branches (variable m.br), mass of leaves (m.lf), total above ground mass (m.so) as well as total plant height (h.t). We estimated stem mass as the difference between m.st and m.br. The estimated biomass per area (kg m^{-2}) was calculated by dividing the mass (kg) with the projected crown area (variable a.cp, m^2). The BAAD data was used to calibrate the allometry module of the S4F model as described in Section 4.3.3 of ATBDv2. Results of this calibration are presented in Section 3.2.2.

The global live-fuel moisture content database (Globe-LFMC) (Yebra et al., 2019) was used to calibrate S4F model parameters for the estimation of live-fuel moisture content. Therefore we selected data from 584 combinations of sites and plant species from the south-western United States. The south-western US has globally most LFMC measurements, which cover a wide environmental gradient from deserts, semi-arid regions, temperate forests, and montane forests and tundra. LFMC measurements are

not available for Siberia and the Amazon. Results of the calibration are presented in Section 3.3.3.

The Global Database of Litterfall Mass and Litter Pool Carbon (Holland et al., 2014) was used to validate estimates of surface fuel loads. Specifically, we extracted measurements of leaf litter weight (variable TLEAFW), fine woody debris (variable small wood weight, TSWW), total litter weight (Litter + FWD, variable TLTRW), and of large woody weight (CWD, variable LWW). As the entries of this database do not match spatially with our test areas, we selected data from sites with similar climate conditions (i.e. mean annual temperature and precipitation of the measurements sites have to be within the regional minimum and maximum of each study region. Results of this validation are presented in Section 3.3.6.

The field measurement database of biomass burning fuel consumption (originally by van Leeuwen et al., 2014; updated by van Wees et al., 2021) was used to validate the estimates of fuel loads, combustion completeness, and fuel consumption. The database was compiled to parametrise and evaluate fire emission inventories such as GFED (van der Werf et al., 2017). For the comparison we selected for each test area/study region data from neighbouring measurement sites (i.e. ±2° longitude/latitude around each test area). As the spatial coverage of the field measurements might not be representative for the spatial resolution of the S4F approach, we compared statistical distributions of fuel consumption and combustion completeness rather than performing direct spatial comparisons. Results of the validation are presented in Section 3.3.7.

The database by Andreae (2019) includes statistical distributions of combustion efficiencies and emission factors for different trace gases and biomes from various other databases and individual investigations. The database was used to calibrate and validate the simulated statistical distribution of emissions factors with measured emission factors. Results are presented in Section 3.3.8.

For the calibration and validation of the S4F model a stratified sampling of grid cells was conducted for each test area. Therefore each test area was first divided in areas with low, medium and high biomass based on the regional quantiles 0.25 and 0.75 of mean above ground biomass. Then each biomass class was further separated into pixels with low VOD or high VOD based on a threshold below or above the regional average of daily maximum Ku-VOD. Maximum Ku-VOD was chosen as the choice of VOD does not affect the sampling because of the high correlation with X- and C-VOD. L-VOD was not included in the stratification because of the correlation with biomass, which was included in the stratification as well. In the last step, only grid cells were selected that burned at least once.

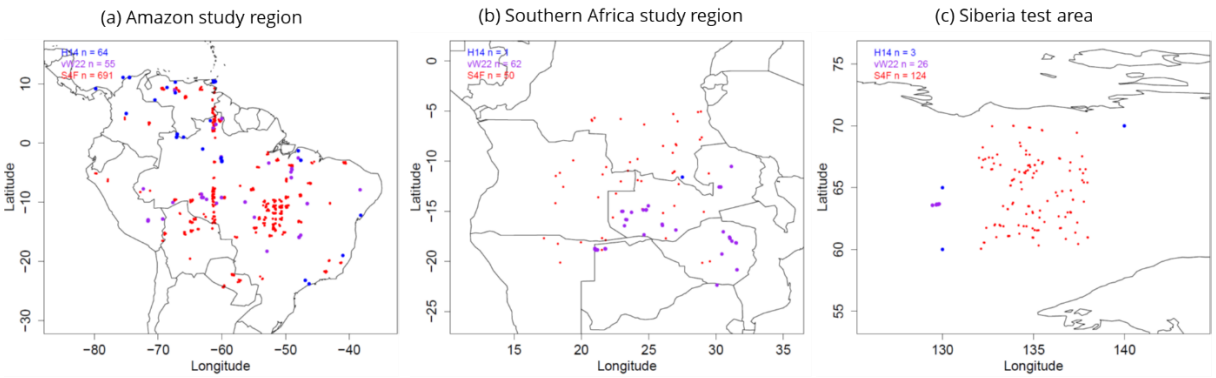

Figure 4: Overview of the sampled grid cells within study regions/test areas for the calibration and validation exercises of the S4F model in PVRv2.1. Red dots are grid cells for which the S4F model was calibrated. Blue and purple dots denote approximate locations of field measurements of surface litter from Holland et al. (2014) [blue] and of fuel loads and fuel consumption from van Wees et al. (2022) [purple].

3.2 Comparison of canopy height input datasets

The comparison of the canopy height datasets was carried out on the one hand by comparing the basic statistics and on the other hand by calculating the difference maps in pairs. The statistics of the three datasets for the Amazon test area reveal that the GLAD datasets has lower heights than the other two datasets (Table 3). The mean value is about 2.5 m lower, the maximum value even 4 m (ETH) and 5.5 m (GEDI) lower. In the ETH and GEDI data set, on the other hand, mean and maximum values are more similar.

Table 3: Statistics of the three canopy height datasets for the Amazon test area. Global Forest Canopy Height from Potapov et al. (2021) (GLAD), ETH Global Canopy Top Height (ETH) and GEDI L3 Gridded Land Surface Metrics (GEDI)

Statistic	GLAD [m]	ETH [m]	GEDI [m]
Minimum	0.00	0.00	1.10
Median	6.73	11.49	10.10
Mean	9.76	12.42	12.20
Maximum	30.18	34.36	35.90

The same tendency is reflected in the difference maps for the Amazon test area, especially in the corresponding histograms (Figure 5). The difference map between GLAD and GEDI shows that GEDI provides greater heights in areas with low tree cover (Figure 5 c, reddish colours), whereas GLAD has greater heights in areas with dense tree cover (Figure 5 c, blue colours).

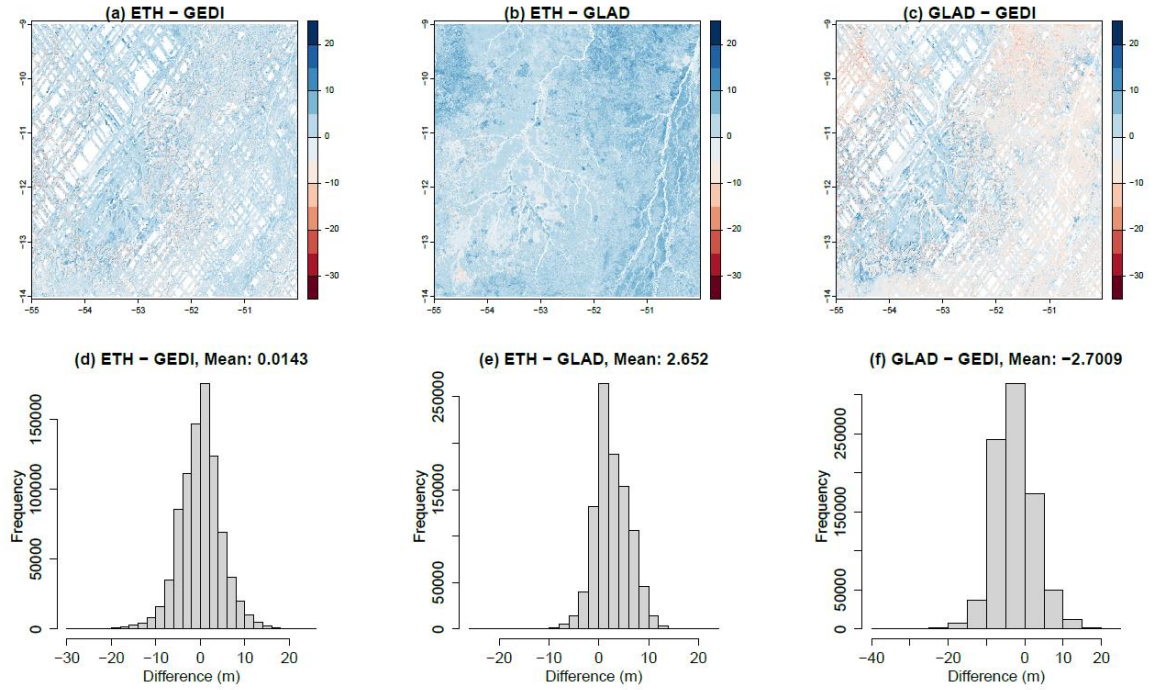

Figure 5: Pairwise comparison of canopy height datasets for the Amazon test area (top) difference maps, (below) corresponding histograms of differences.

The fact that GEDI and ETH datasets are very similar, but the ETH dataset is at a higher spatial resolution and has fewer data gaps, would lead to the conclusion that it is the most suitable for calibrating tree heights. To confirm this assumption, the same comparison of the datasets was carried out for the southern Africa test area.

For the southern Africa test area, however, the pattern is different. Here, the GEDI dataset shows the highest mean value more than 4 m above the mean value of the GLAD data set and approximately 3 m above the mean value of the ETH dataset (Table 4).

Table 4: Statistics of the three canopy height datasets for the southern Africa test area. Global Forest Canopy Height from Potapov et al. (2021) (GLAD), ETH Global Canopy Top Height (ETH) and GEDI L3 Gridded Land Surface Metrics (GEDI)

Statistic	GLAD [m]	ETH [m]	GEDI [m]
Minimum	0.00	0.00	1.33
Median	8.27	9.41	12.59
Mean	7.69	9.30	12.08
Maximum	24.22	33.02	27.49

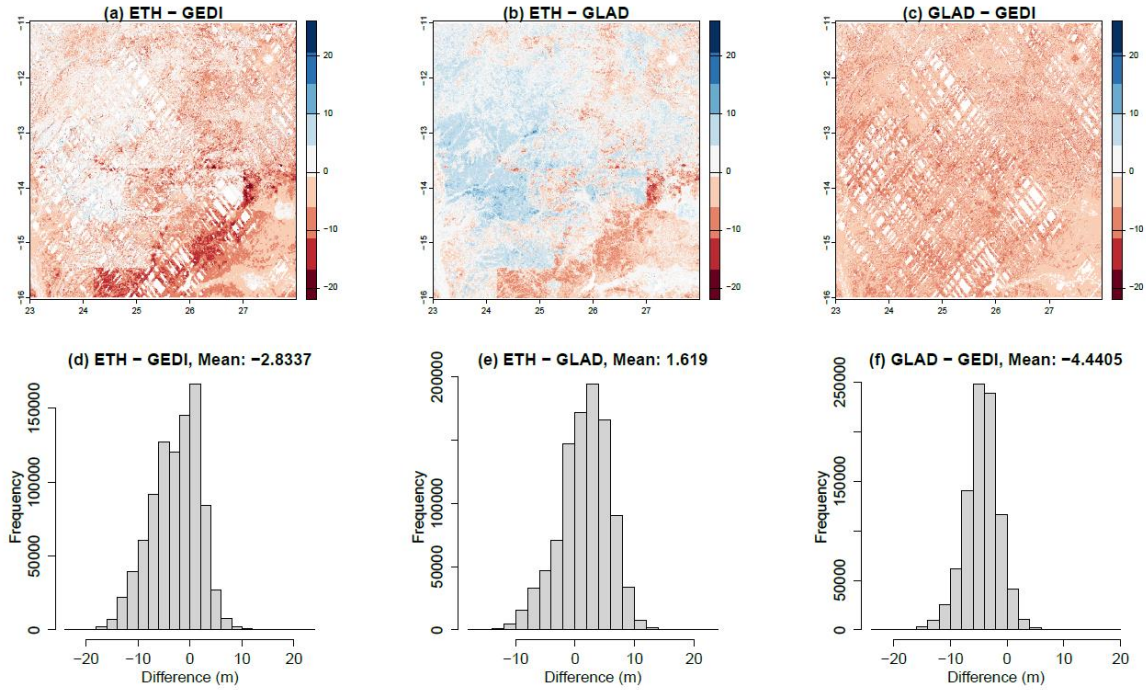

Figure 6: Pairwise comparison of canopy height datasets for the southern Africa test area. (top) difference maps, (below) corresponding histograms of differences

The difference maps between the ETH dataset and the other two datasets show tile-like patterns with abrupt height differences, which could originate from the combination of the GEDI heights with the Sentinel data (Figure 6 a and b). Due to these artefacts, the ETH dataset does not necessarily seem to be the best choice. Since the GLAD dataset is a combination of GEDI/Landsat and similar artefacts cannot be excluded, it was decided to initially use the GEDI L3 Gridded Land Surface Metrics product for the calibration, which is based exclusively on GEDI data.

The comparison of canopy height datasets was not performed for the Siberia test area, as for this region only the ETH dataset is available.

3.3 Calibration and validation of the S4F model

3.3.1 Training regression models for prior estimates for canopy height

The computation of canopy height in the S4F model is based on a regression between the predictor variables mean LAI (LAI_{mean}) and fractional tree cover (f_{tree}) and the target variable canopy height H (Section 4.3.2 of ATBDv2.1). The target variable for canopy height is taken from the GEDI L3 Gridded Land Surface Metrics product. Several regression models were tested to predict canopy height (Table 5). Linear regression models were tested with one predictor based on either mean LAI (M1 and M2). We then combined LAI_{mean} and f_{tree} in further regression models that account for non-linear relationships: Model 4 is a Generalised Additive Model (GAM), Model 5 is multi-variate linear regression with LAI_{mean} included as quadratic term, and Model 6 is a random forest. Model M7 combines the LAI_{mean} included as quadratic term with f_{tree} as a product. Model M5 achieved the best

performance in the Amazon study region, the GAM (M4) in southern Africa, and the GAM and RF in Siberia (Table 5). However, as the performance between the GAM and model M5 were similar in the Amazon and Southern Africa study region but M5 had a much faster computation time, we selected M5 for those two study regions. For Siberia, we selected model M8, despite having a lower correlation and higher RMSE than some of the other models. This is because all other models could not predict canopy heights > 12 m in Siberia while the GEDI L3 datasets suggests canopy height of up to 20 m. Model M8 was the only model that could predict such canopy height values.

Table 5: Performance statistics of ten regression models to estimate canopy height from mean or maximum LAI and fractional tree cover. Correlation (r) and root mean squared error (RMSE) between canopy height from the GEDI L3 Gridded Land Surface Metrics product and the regression model.

	Amazon study region		**Southern Africa study region**		**Siberia test area**	
Model	**r**	**RMSE (m)**	**r**	**RMSE (m)**	**r**	**RMSE (m)**
M1) $H = a \times \sqrt{LAI_{mean}}$	0.75	5.76	0.66	3.12	0.63	3.52
M2) $H = a \times LAI_{mean} + b$	0.79	5.07	0.68	3.05	0.64	3.20
M4) $[GAM]\ H = s(LAI_{mean}) + s(f_{tree})$	0.8	4.94	0.73	2.85	0.66	3.12
M5) $H = h1 \times LAI_{mean}^2 + h2 \times f_{tree} + h3$	0.81	4.90	0.71	2.92	0.62	3.24
M6) $[RF]\ H = RF(LAI_{mean}, f_{tree})$	0.79	5.06	0.72	2.86	0.66	3.13
M8) $H = a \times LAI^2{}_{mean} \cdot ftree$	0.79	8.05	0.66	7.50	0.62	3.82

Model 5 was then applied to the entire Amazon and Southern Africa study regions, and Model 8 to the Siberia test area and compared with the full spatial distribution of the GEDI L3 Gridded Land Surface Metrics product (Figure 7). The S4F model shows a strong agreement with the GEDI L3 canopy height product but a slightly under- or overestimated in certain parts of the test area. The error between the estimated canopy height and canopy height dataset is on average 0.1 m (Figure 7).

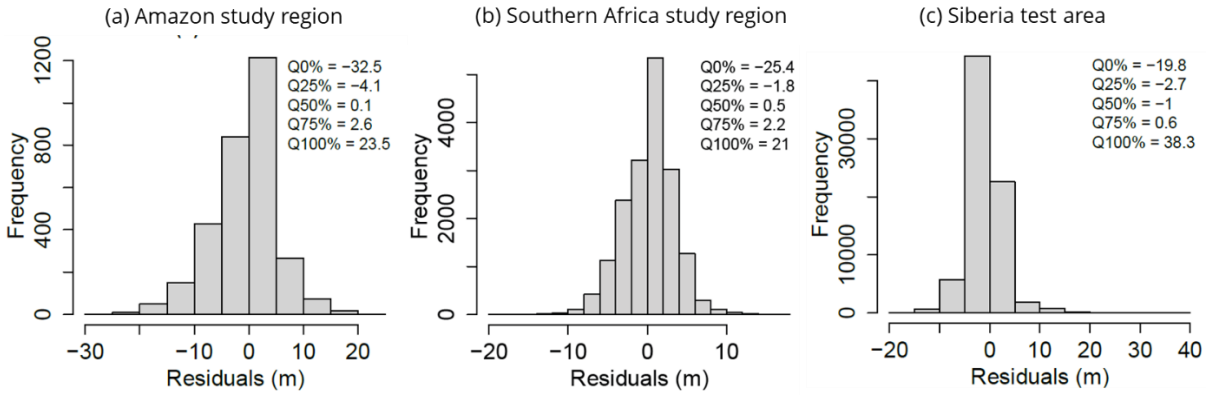

Figure 7: Comparison of residuals between estimated canopy height from the S4F model and from the GEDI L3 Gridded Land Surface Metrics product.

3.3.2 Calibration against BAAD for biomass components

The estimated tree canopy height is used in the S4F model to estimate the biomass of tree stems, branches and leaves based on allometric relationships (Section 4.3.3 of ATBDv2.1). The parameters *a1* to *a6* of these allometric relationships are calibrated against the BAAD measurements. Thereby, measurements in BAAD from tropical broad-leaved evergreen and deciduous trees were taken for the Amazon and Southern Africa study regions with using prior parameter values for tropical broad-leaved evergreen trees in the Amazon and for tropical broad-leaved deciduous trees in Southern Africa. For Siberia, measurements for temperate and boreal needle-leaved evergreen and deciduous trees were used. To calibrate the parameters of the allometric relationships, we use the measured total tree height from BAAD as input to the allometry module of the S4F model. Thereby, we assume that the phenology status of the leaves is 1, i.e. all measured trees in BAAD are at full seasonal leaf cover. The calibration approaches uses a cost function based on the Kling-Gupta efficiency and a genetic optimisation algorithm as described in Section 4.3.11 of ATBDv2.1. We then compute the total above-ground biomass, the biomass of wood, stem, branches and leaves. The cost function is then computed jointly against the corresponding measurements from BAAD.

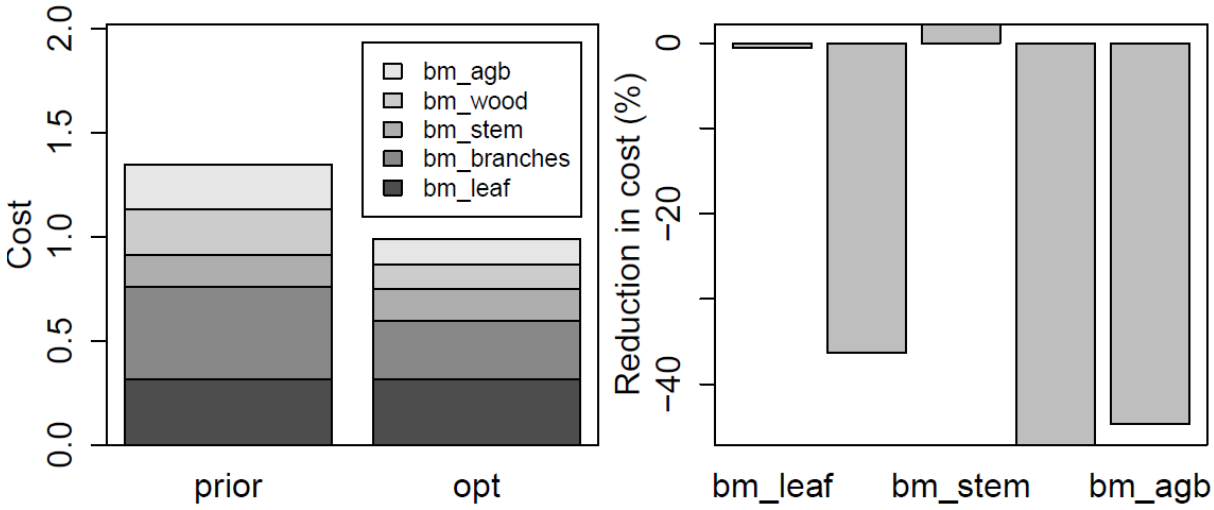

Figure 8: Comparison of the cost of the S4F model for the Amazon study region prior and after calibration against the BAAD database of total above-ground biomass (BMagb), woody, stem and branches biomass (BMwood = BMstem + BMbranches), and leaf biomass (BMleaf).

The change of the cost function between the prior and the optimised parameter set is shown for the Amazon study region in Figure 8. Prior to calibration, the cost was dominated by the error in leaf biomass. After calibration, the cost was reduced for all biomass components but stem biomass (although for total woody biomass it was strongly reduced). Similar changes in the cost occurred in the other study regions.

The corresponding scatterplots of the calibration results are shown for leaf and woody biomass for all study regions in Figure 9. The allometry model provides an acceptable fit to all biomass components with medium to high correlations and low to medium RMSE.

The lowest performance is thereby achieved in the estimation of leaf biomass, which also shows a strong scatter in the observations and weak relationships with stem or woody biomass or tree height. Please note that most measurements in the BAAD database originate from measurements of rather small trees with low biomass, hence the allometric relationships are poorly constrained for high-biomass trees.

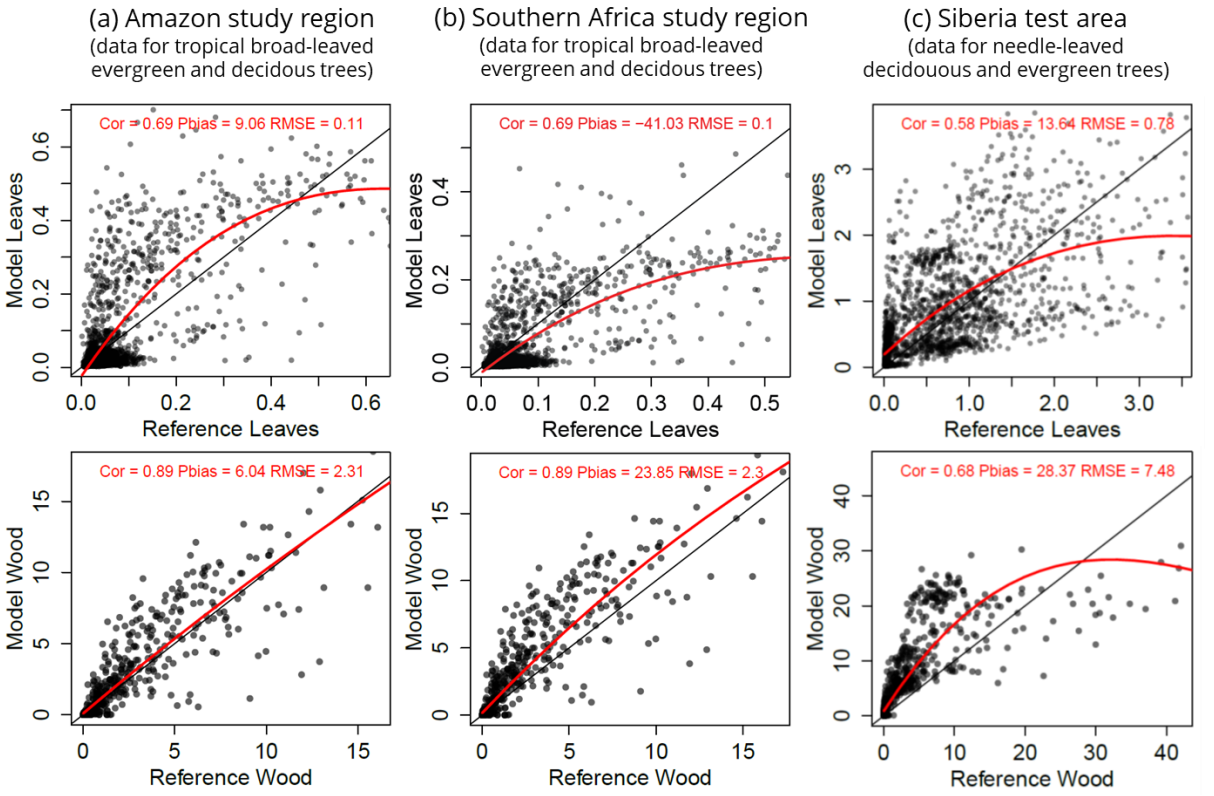

Figure 9: Scatterplots of estimated biomass compartments (kg m-2) from the S4F model (y-axis) against measurements of the BAAD database (x-axis) after optimisation for leaf biomass (upper row) and woody biomass (lower row).

3.3.3 Calibration for prior estimates of regional above-ground woody biomass

We then applied the pre-calibrated height and allometry modules to calibrate the S4F jointly against BAAD (for leaf, stem and branches biomass) and against the ESA CCI biomass map (for total woody biomass) for the selected grid cells in each study region. The S4F model reproduces the overall spatial patterns of woody biomass (Table 6). However, locally large residuals can occur (Figure 10), which often occurs along edges of forests or rivers likely caused by inconsistencies between the land cover, biomass and LAI datasets in such transitional areas.

Table 6: Performance statistics of the prior S4F model to estimate above-ground woody biomass in comparison to the ESA CCI biomass map for the year 2018.

	Amazon study region		Southern Africa study region		Siberia test area	
Model	**r**	**RMSE (kg/m²)**	**r**	**RMSE (kg/m²)**	**r**	**RMSE (kg/m²)**

S4F before calibration against BAAD+ESACCI	0.82	9.20	0.8	3.44	0.55	4.98
S4F after calibration against BAAD+ESACCI	0.85	7.11	0.8	3.32	0.55	2.60

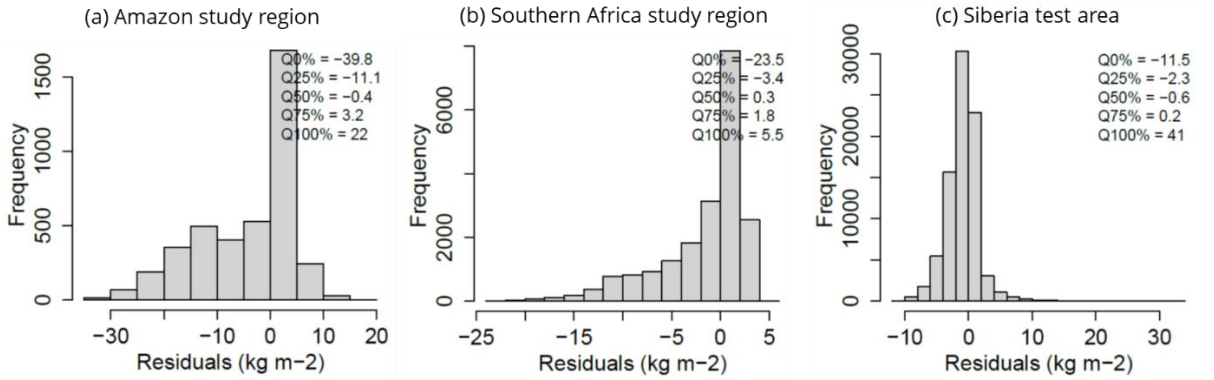

Figure 10: Comparison of residuals between estimated above ground woody biomass from the S4F model and from the ESA CCI biomass dataset (year 2018).

3.3.4 Prior estimates for live-fuel moisture content

The module to estimate live-fuel moisture content (LFMC) in the S4F model was calibrated against ground measurements of LFMC from 584 combinations of sites and plant species from the Globe-LFMC database. For all site-species combinations, the S4F achieved in most cases medium to high temporal correlations between observed and estimated LFMC (Figure 11). The highest correlations were found for shrublands (median r = 0.75) and the lowest correlations for broad- and needle-leaved forests (median r = 0.5). Root-mean squared error was for all vegetation in average between 20 and 30 %-LFMC.

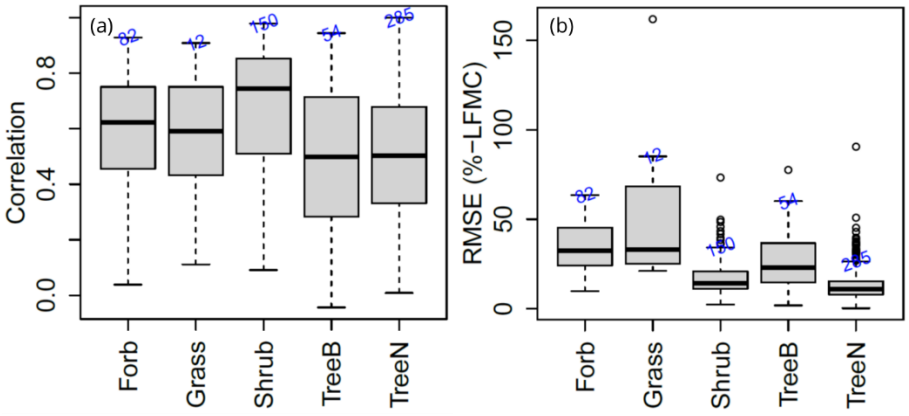

Figure 11: Performance of the S4F model in reproducing LFMC from site observations of forbs, grass, shrubs, broad-leaved and needle-leaved trees in the south-western United States.

3.3.5 Fire event-based retrieval of fuel and fire dynamics

After the initial calibrations of the S4F model modules for canopy height, tree biomass, and LFMC, the S4F model was then calibrated for individual fire events jointly against the datasets of GEDI canopy height (2020), ESA CCI above-ground biomass (2010, 2017, 2018 interpolated), LFMC from the VOD2LFMC dataset, VOD in Ku-, X and L-band, and FRE. This calibration allows finding a best solution for each local fire event that is constrained by the different datasets and hence allows to retrieve estimates of pre-fire biomass compartments, surface fuel loads, fuel moisture, vegetation water content, and finally fuel consumption and fire emissions. In the following we illustrate the approach based on one example and then describe the calibration performance across several fire events.

As an example, the estimated temporal dynamics of biomass, fuel moisture, fuel loads, fuel consumption, FRP and fire emissions are shown for the fire event in 2020 at 53.96726°W, 10.654762°S in the Amazon test area (Figure 12). Based on the input data, this example grid cell is dominated by a high tree cover (85%), medium-high canopy height (18.5 m) and biomass (24 kg/m²). The site shows a strong reduction in LAI in 2020, which is caused by the observed fire with a fire radiative energy of around 45 MJ/m² (Figure 12 a-d).

The calibration of the S4F model for this grid cell results in an unbiased canopy height (Figure 12 b), a slightly under-estimation of total woody biomass (Bias = 7%, RMSE = 1.9 kg/m², Figure 12 c), an unbiased and highly correlated LFMC (r = 0.7, Figure 12 e), and an appropriate representation of the VOD across the different bands (e.g. r = 0.5 with Ku-VOD, Bias = -15%, -3% and -55% for Ku-, X- and L-VOD, Figure 12 f-h). The initial overestimation of FRE (prior FRE = 120 MJ/m²) was clearly reduced (optimised FRE = 80 MJ/m²).

From this calibration, we retrieve a strong reduction of woody biomass and of coarse woody debris (CWD) during 2020 (Figure 12 i and j). Fuel consumption is dominated by the burning of CWD (Figure 12 k), which causes emissions of around 4000 g/m² of CO_2 and around 300 g/m² of CO.

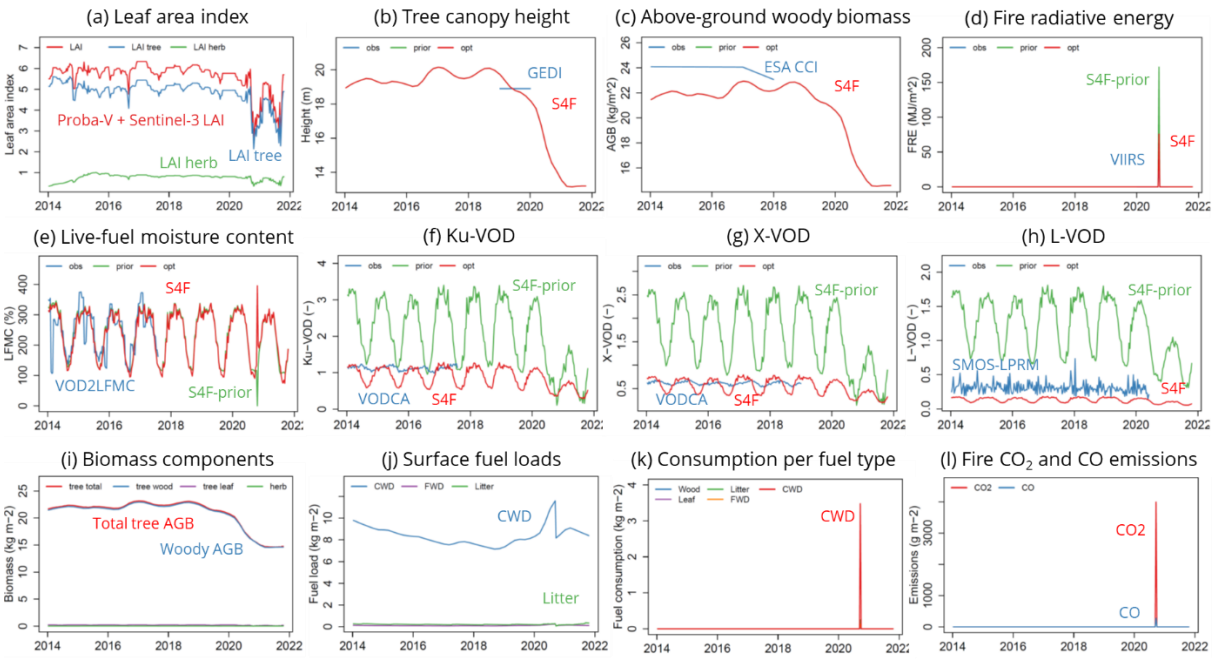

Figure 12: Example of the estimated fuel and fire dynamics from the S4F model for one example grid cell in the Amazon test area (53.96726°W, 10.654762°S).

The calibration of the S4F model was conducted for several individual fire events in all study regions (Figure 4). In the Amazon, the calibration was performed for 691 single fire events because of the complexity of the different fire types. In souther Africa, the calibration was performed for only 50 fire events because of the large similarity in biomass stocks and land cover. In Siberia, the calibration was performed for 124 fire events.

The evaluation for tree height and woody biomass after calibration across the fire events is shown in Figure 13. For most events, the S4F estimates agree well with the GEDI canopy height and with the ESA CCI woody biomass. Although the calibration results show no bias in tree canopy height, there is a tendency to underestimate the woody biomass at sites with large biomass. In all study regions, grid cells occur where the S4F model predicts no canopy height or woody biomass although the two datasets show medium to high values. This suggests inconsistencies between the ESA CCI land cover data sets (input to the S4F model) and canopy height and/or woody biomass, which often occurs in transitional areas.

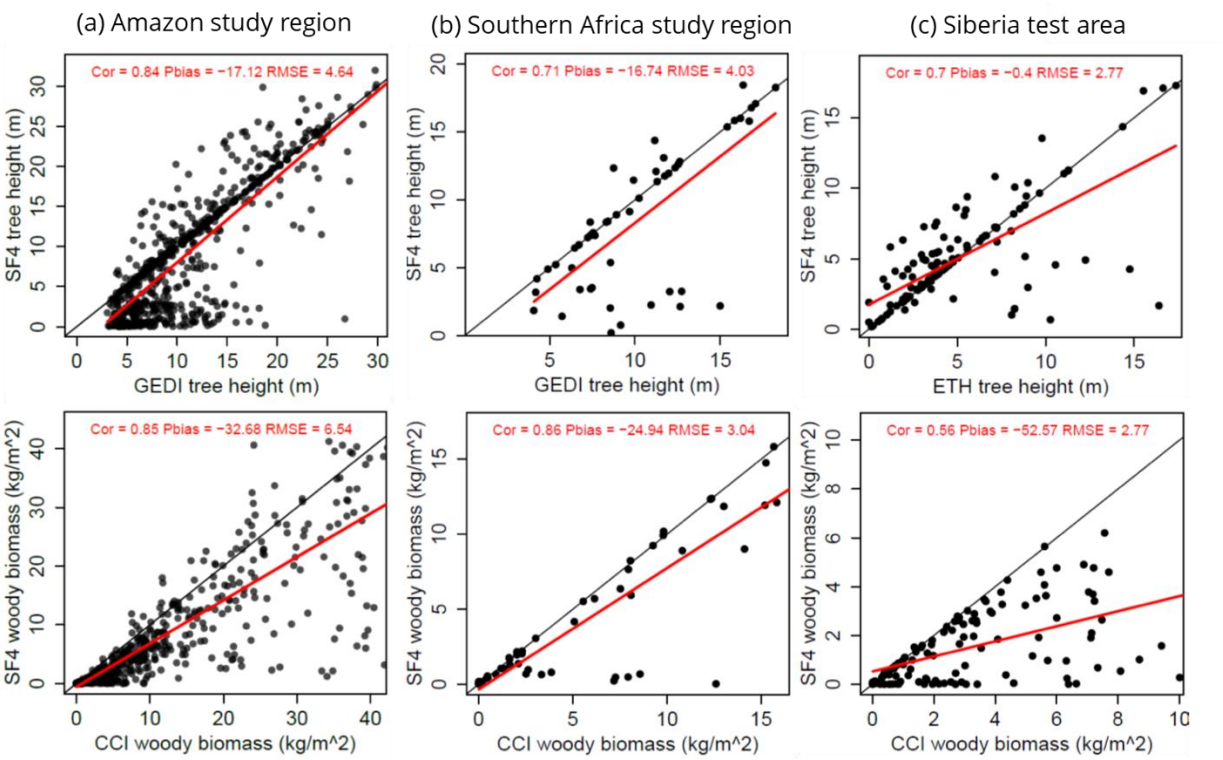

Figure 13: Performance of the S4F model across 353 fire events in the Amazon test area based on event-level calibration against GEDI tree canopy height, ESA CCI woody biomass, VOD2LFMC, Ku-, X- and L-VOD and FRE. (a) Relationship between tree canopy height in 2019 from GEDI and estimated from the S4F model. (b) Relationship between mean above-ground woody biomass in 2017 and 2018 from ESA CCI and the S4F model.

The calibration resulted in the majority of fire events in medium to high positive correlations of the S4F-retrieved LFMC with the VOD2LFMC dataset (with slightly lower correlations in the Amazon) and with VOD from the VODCA dataset (Ku- and X-band) and from SMOS (LPRM L-band) (Figure 14).

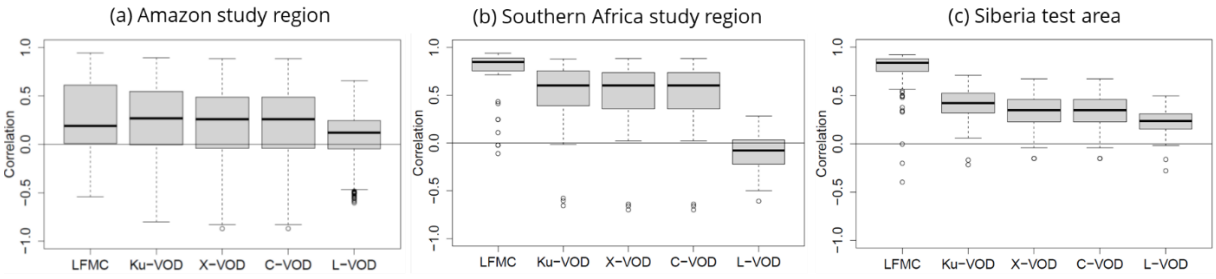

Figure 14: Temporal correlation between estimated LFMC and VOD from the S4F model and the corresponding estimates from the VOD2LFMC, VODCA (Ku-/X-/C-VOD) and SMOS-LPRMS (L-VOD) datasets based on a calibration of the S4F model against individual fire events in each study region.

The estimated fire radiative energy is unbiased in comparison to the Sentinel-3/VIIRS-derived FRE (Figure 15). FRE is overestimated for some fire events in the Amazon and in Siberia (Figure 15 a and c) but underestimated for some fire events in Southern Africa (Figure 15 b).

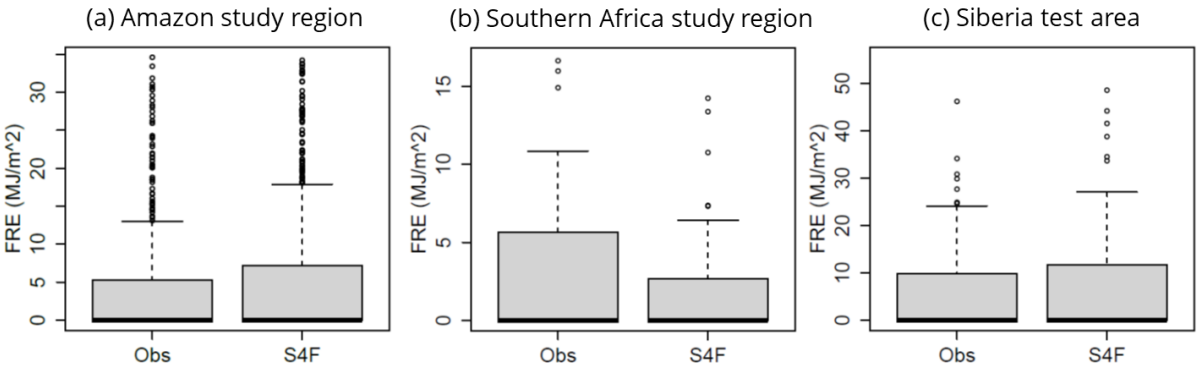

Figure 15: Comparison of estimated and observed FRE.

3.3.6 Validation of fuel loads

The S4F model computes surface fuel loads (litter, FWD and CWD) based on the turnover of biomass from living vegetation components to the surface which is driven by long-term changes prescribed from the LAI and land cover datasets (Section 4.3.5. of ATBDv2.1) and by taking typical decomposition rates for litter and woody debris (Harmon et al., 2020). We validated the estimated fuel loads against measurements from the global database of litter fall mass and litter pool sizes (Holland et al., 2014) and against the updated fuel database (van Wees et al., 2022) (Figure 16). Especially in the Amazon region, estimated fuel loads agree well with the statistical distribution from the databases. In southern Africa, the S4F model tends to overestimate fuel loads (Figure 16 b). However, this needs further investigation as most of the field observations come from the more drier southern part of the study region while the S4F model was mostly calibrated for the wetter central and northern parts (Figure 4). Total fuel loads and litter agree are comparable with the databases in Siberia but CWD is overestimated (Figure 16 c).

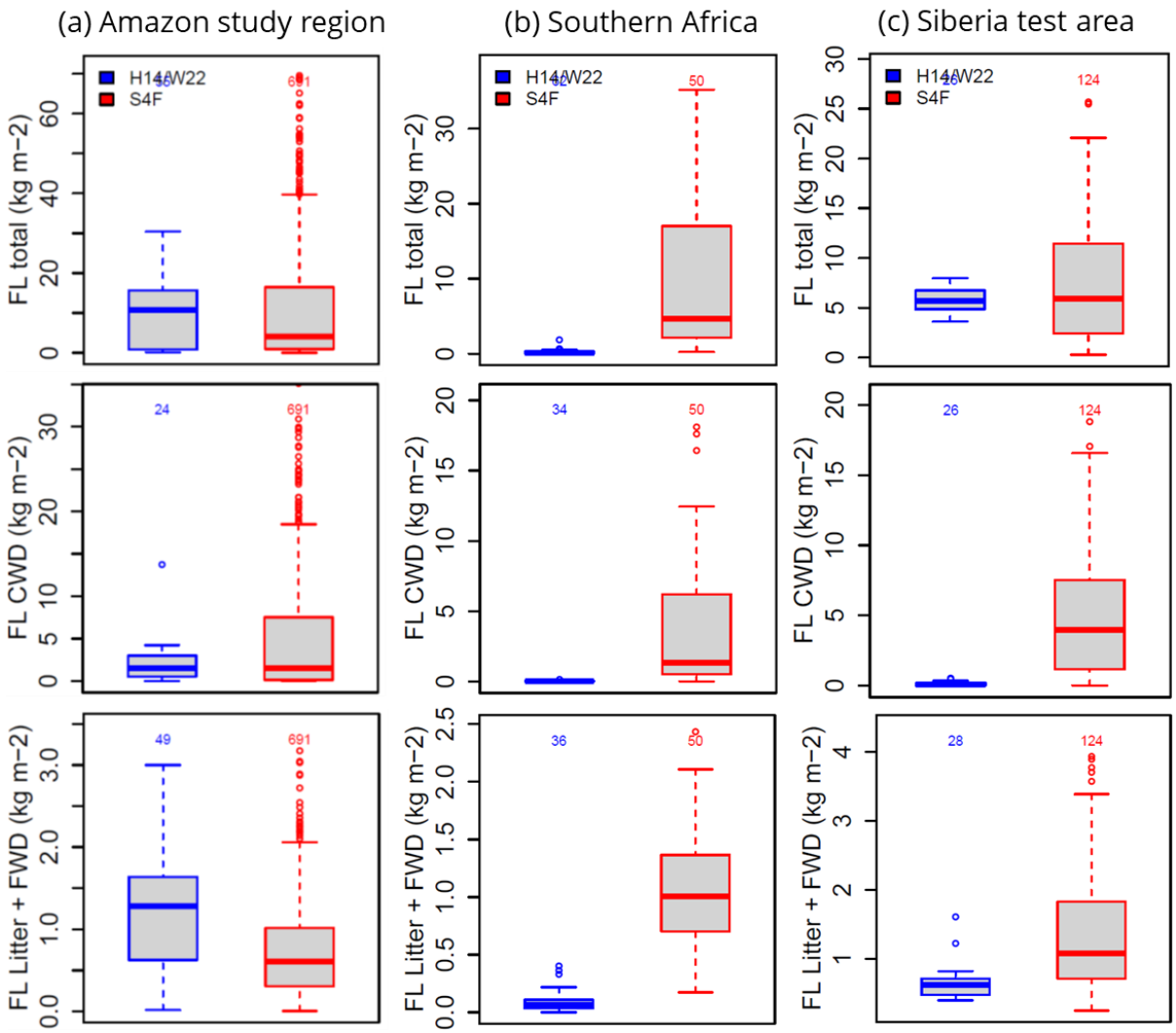

Figure 16: Validation of the estimated fuel loads from the S4F model against measurements from the global database of litter fall masses and from the fuel consumption database. Numbers above each box indicate the number of observations from the fuel database and of the sampled locations of S4F fuel model, respectively. Note that locations in measurements in S4F grid cells do not match but only represent regional distributions.

For the Amazon study region, we further investigated how the used land cover datasets (i.e. default ESA CCI land cover or as alternative Song et al. (2018)) as input to the S4F model affect the estimated fuel loads (Figure 17). As above-ground biomass is constrained by the ESA CCI biomass map, both datasets result in a similar total fuel load, which is dominated by woody biomass. However, the Song dataset causes lower herbaceous biomass, FWD and litter but much larger CWD than the ESA CCI dataset. This is likely because of the stronger temporal dynamic in the Song dataset, which causes a more frequent turnover of woody biomass to CWD in the S4F model. However, the estimated fuel loads derived from the ESA CCI land cover map agree better with statistical distributions of fuel loads for different fuel types than derived from the Song dataset.

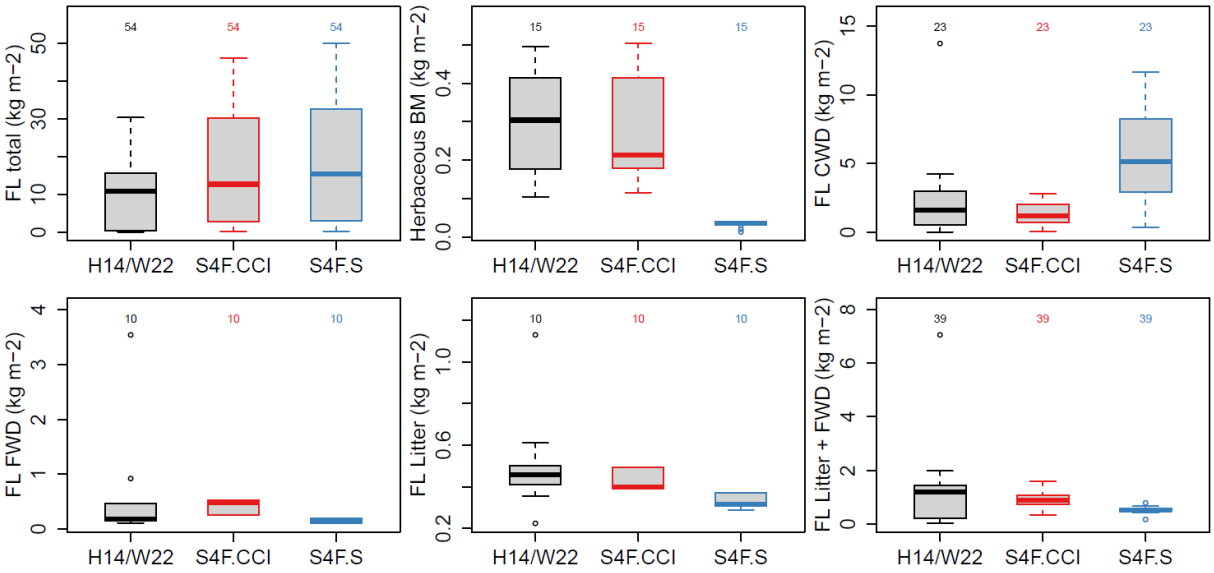

Figure 17: Distribution of fuel loads for the Amazon study region for the same measurement locations from the global database of litter fall masses (H14) and from the fuel consumption database (W22) and grid cells from the S4F model using the ESA CCI land cover map (S4F.CCI) and the land cover map by Song et al. (S4F.S) as input.

3.3.7 Validation of fuel consumption

The estimated combustion completeness and fuel consumption from the S4F model (using ESA CCI land cover as input) were then compared against measurements from the updated fuel consumption database (van Leeuwen et al., 2014; van Wees et al., 2022). In the Amazon and Siberia study regions, total fuel consumption agree well with the database (Figure 18). In both regions, the total fuel consumption mostly originates from the burning of wood while the S4F model attributes the fuel consumption more to CWD than to live wood (Figure 18, second and third row). In southern Africa, total fuel consumption seems to be over-estimated, which however might be related to higher fuel loads in the wetter parts of the study region than represented by the field measurements from the drier regions.

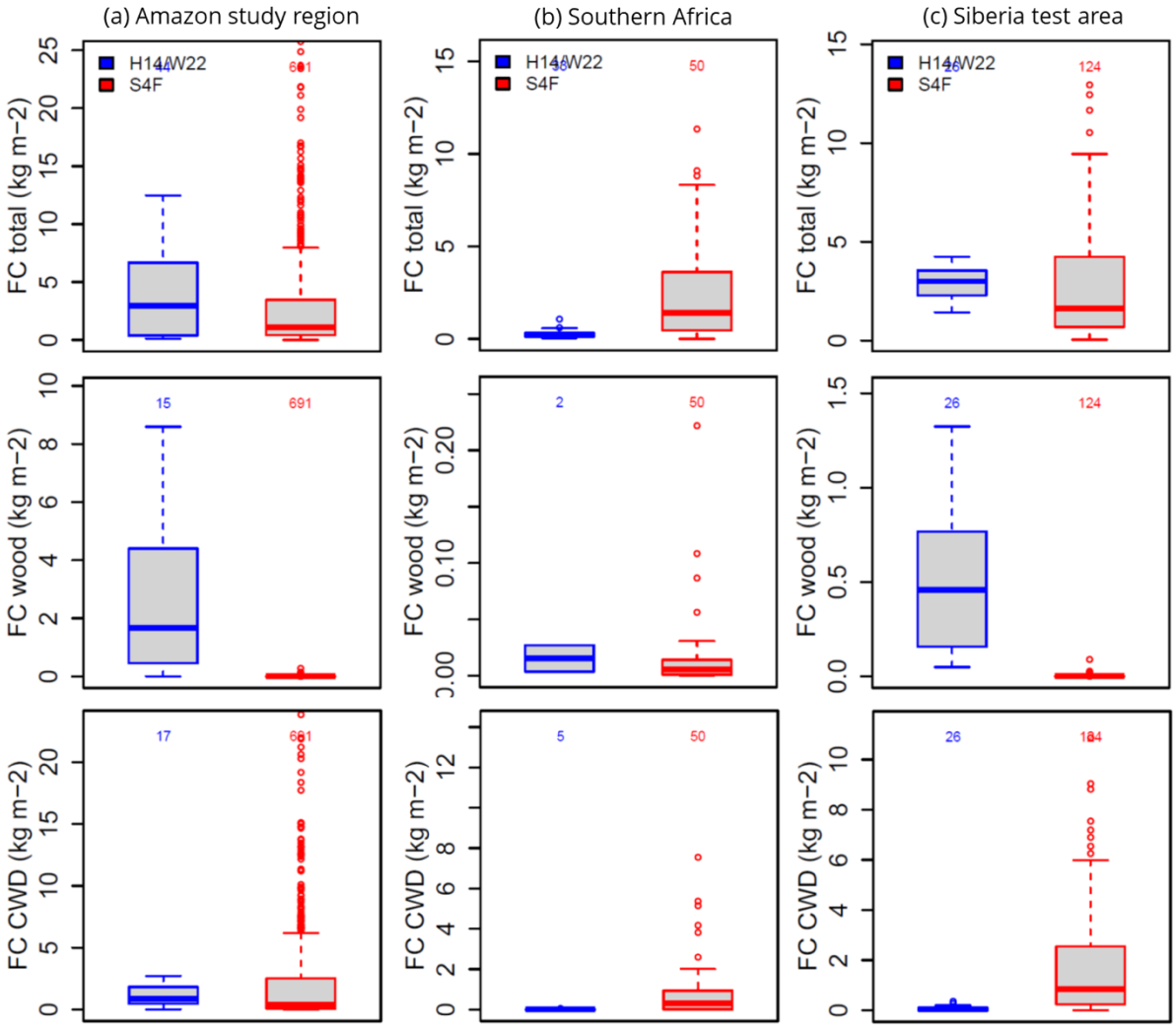

Figure 18: Validation of fuel consumption from the S4F model against the updated fuel database by van Wees et al. (2022). Numbers above each box indicate the number of observations from the fuel database and of sampled locations of S4F fuel model, respectively.

3.3.8 Combustion efficiency and emission factors

After the calibration of the S4F model for individual fire events in each test area, the regional statistical distribution of the estimated Modified Combustion Efficiency (MCE) and of the estimated emission factors and of the spatial distribution of FRE was then calibrated and compared against the statistical distribution obtained from the compilation of emission factors by Andreae (2019). The Equivalent Oxygen to Fuel Ratio (EOFR) for all fuel types and the fraction of cellulose of leaves, herbaceous vegetation and wood, and the fraction of volatiles of leaves and herbaceous vegetation was calibrated by jointly fitting the estimated quantiles against the observed quantiles of MCE and emission factors.

The distribution of MCE and of emission factors for CO_2, CO and NO_x are shown in Figure 19. The S4F model reproduces the observed statistical distributions of emission factors. Only in Siberia, the S4F model produces a lower median MCE, resulting in a higher median

emission factor for CO. Prior to the optimization, the emission factor for NO_x was overestimated in all study regions but was reproduced after optimization.

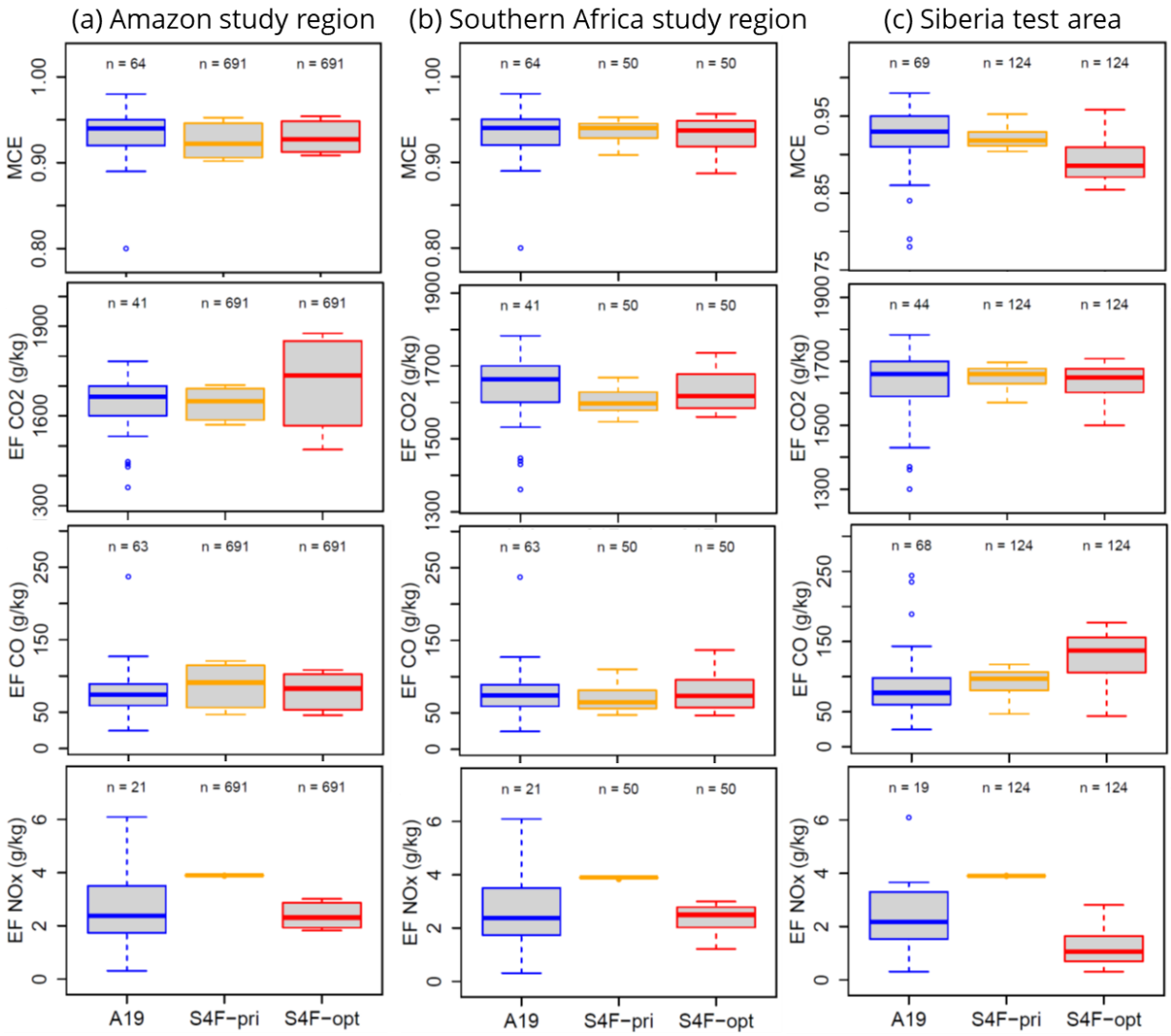

Figure 19: Validation of the modified combustion efficiency (MCE) and of the emission factors of CO2, CO and NOx from S4F model before and after optimisation against observed emission factor values [A19] (Andreae, 2019). For the Amazon test area, all values from tropical forests and from savannah and grasslands were taken from A19. For the southern Africa study region the values for savannahs and grasslands were taken from A19. For Siberia, values for grasslands and boreal forests were taken.

3.3.9 Example results from the S4F data-model fusion approach

As final outputs of the S4F model, temporal dynamics and spatial patterns of total fuel consumption and of fire emissions are shown for one example area in the Amazon test area in Figure 20 and for the entire Amazon study region in Figure 21. The example area in Figure 20 experienced large fires in 2016 and 2020. High fuel consumption in this area occurs mainly because of fires in forest stands that mainly burn woody debris and show a smouldering combustion (e.g. fire at 51.85°W, 12.1°S in Figure 20 a, b and e). Fires in

grasslands show a higher consumption of litter with more flaming combustion (e.g. fires east of 51.7°W in Figure 20 c and d).

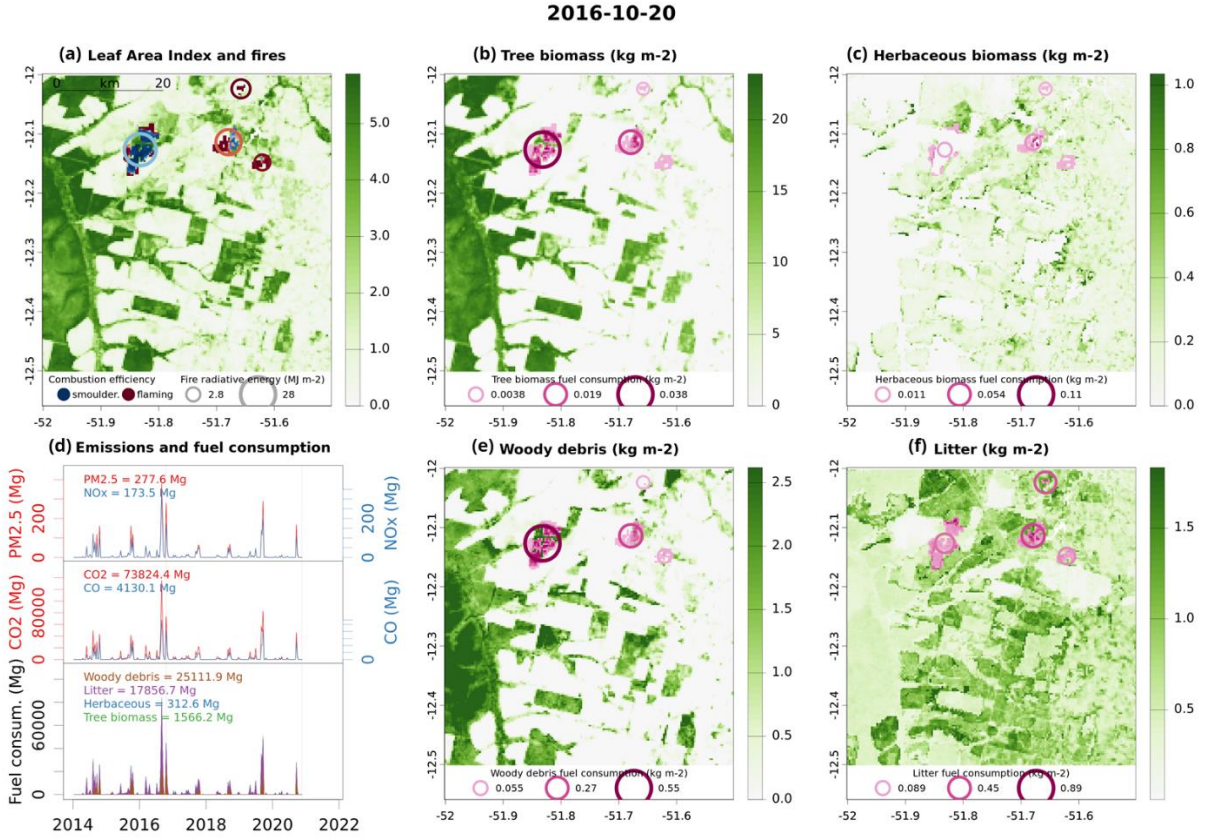

Figure 20: Example of fuel consumption and fire emissions derived with the S4F data-model fusion approach in a part of the Amazon test area at a spatial resolution of 333 x 333 m for a period between 11th and 20th October 2016. (a) Leaf area index (CGLS Proba-V) and detected fires (ESA CCI v5.1). Blue and red colours identify smouldering and flaming combustion (based on MCE), respectively. Circle size represent the fire radiative energy. (b) Tree biomass (wood + leaves) and tree biomass fuel consumption. (c) Herbaceous biomass and herbaceous fuel consumption. (e) Sum of fine and coarse woody debris and fuel consumption of woody debris. (f) Litter and litter fuel consumption. (d) The time series show the 10-daily emissions of PM2.5, NOx, CO2, CO and the fuel consumption aggregated for different fuel types. Numbers in (d) are totals of fuel consumption and emissions for the period between 11th and 20th October 2016. A full multi-temporal animation of this figure can be found at https://vimeo.com/820320850

The continental maps in Figure 21 show high loads of woody biomass in the Amazon forest and higher loads of litter in some savannah regions outside of the tropical forests (Figure 21 a and b). The highest fuel consumption in the study region occurred between 10th and 20th September 2020 with up to 1.2 kg/m² fuel consumption in some regions (Figure 21 c). There is a clear contrast in combustion behaviour between fires along the rivers in the Amazon basin with lower combustion efficiencies (e.g. more smouldering combustion) and higher combustion efficiency in grasslands and savannahs (Figure 21 d). This results in higher emission factors for e.g. CO in tropical forests than in Savannahs (Figure 21 e). Nevertheless, total CO emissions are mainly dominated by the spatial distribution of total fuel consumption (Figure 21 f).

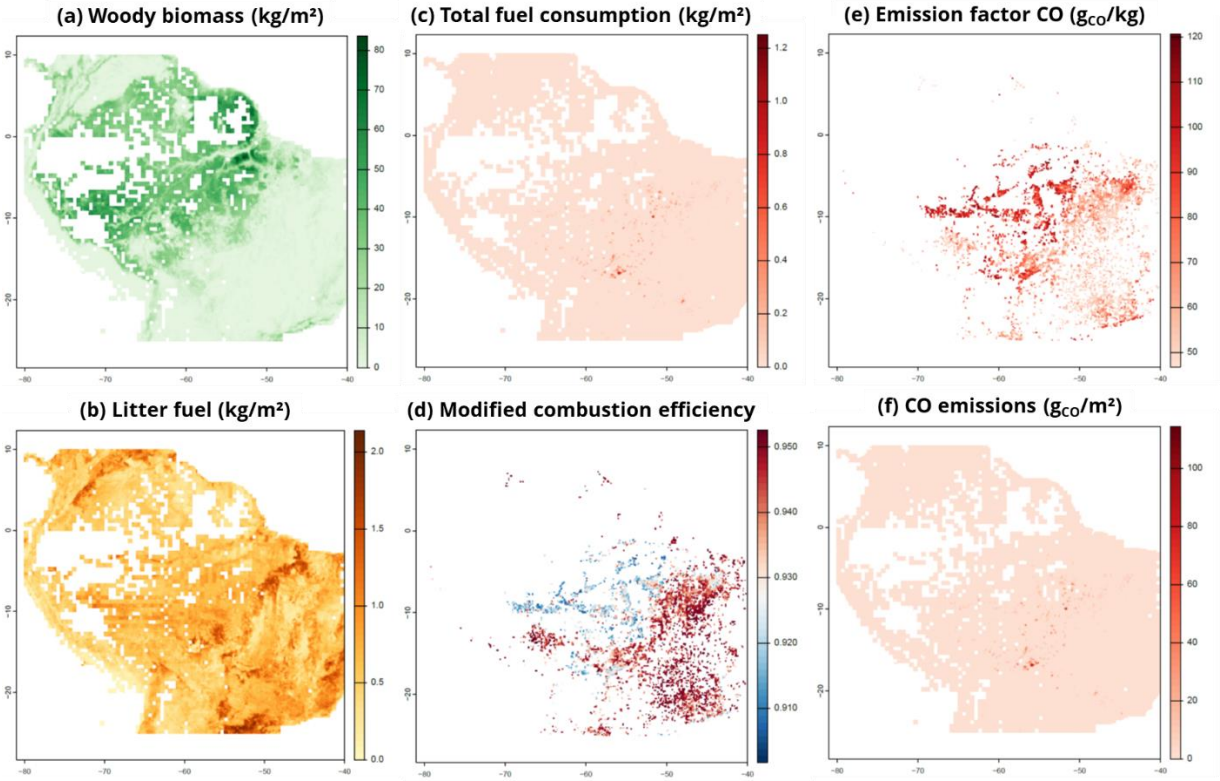

Figure 21: Continental results from the S4F data-model fusion approach for the Amazon study region for the period 10th-20th September 2020. In white areas in (a-c, f) fires never occurred and where hence not computed. In white areas in (e) and (d) fires did not occur in the period in September 2020. Results were aggregated and then resampled from the original spatial resolution (333 x 333 m) to a resolution of 0.1° x 0.1°.

Finally, we compared the estimated CO emissions for the Amazon study region in 2020 with emissions estimates from GFED using the near-real time estimates from the Amazon dashboard (available at http://www.globalfiredata.org/analysis.html) (Figure 22). While the default S4F model setup with the ESA CCI land cover as input shows lower CO emissions than the GFED VIIRS-NRT dataset, S4F with using the Song et al. land cover dataset shows comparable fire CO emissions. The much lower CO emissions for the S4F.CCI results are surprising because the derived fuel loads, fuel consumption, and emission factors in this experiment agree well with observations (Figure 16, Figure 17, Figure 18, Figure 19). We hypothesize that the underestimation of CO emission hence results from an underestimation of small burned areas in the used ESA Fire_CCI5.1 burned area dataset. On the other hand, the better agreement of the S4F.S model experiment using the Song et al. land cover dataset likely originates from the overestimation in CWD and hence an overestimation of fuel consumption of CWD. The use of an alternative burned area datasets than ESA Fire_CCI5.1 together with the land cover dataset by Song et al. might hence result in much higher CO emissions. These results suggest to conduct a more detailed analysis and validation of the S4F data-model fusion approach by using different input datasets of burned area and land cover.

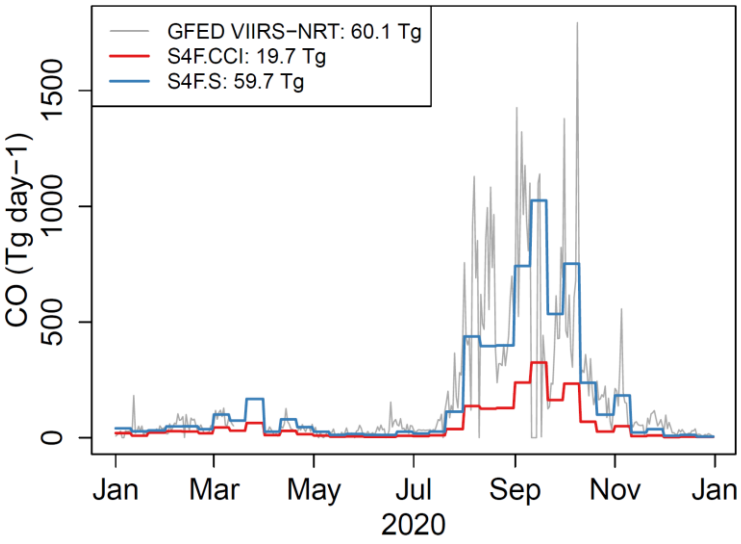

Figure 22: Time series of CO emissions for the year 2020 in the Amazon study region from the GFED VIIRS NRT product from globalfiredata.org and derived with the S4F model with using the ESA CCI land cover as input (S4F.CCI) and the land cover dataset by Song (S4F.S).

4 Top-down constraints on fire emissions

4.1 Theoretical baseline

Current state-of-the-art satellite observation-based fire emission databases such as GFAS and GFED combine information of burned area and active fire detections with associated radiative power to derive trace gas emissions that are further used for estimating fire total carbon emissions. However, total carbon emission using existing methods are highly uncertain. Fire emissions – and in particular ratios of emissions between various trace gases - also depend on for example the type of vegetation and the hydrological conditions of the vegetation. Such information has now become available from the Sentinel 1-2-3 satellite missions but that information is not yet used to upgrade and improve state-of-the-art fire total carbon emission estimates. The Sense4Fire project aims to bridge this gap.

Satellite-based assessment and evaluation of fire emissions to date have mostly relied on satellite measurements of CO and HCHO. Model simulations of atmospheric chemistry and the atmospheric composition are used to bridge the gap between emissions and satellite observations by using the fire emissions as boundary conditions for simulating the atmospheric chemical composition. Although useful to estimate the continental and global effects of fire emissions, the satellite measurements of CO and HCHO that have been used so far, e.g. based on MOPITT and OMI instruments, are less suitable for detailed regional to local evaluation of emissions due to the need to average satellite data in time and/or space, and the relatively coarse spatial resolution of satellite measurements.

The Sentinel-5p satellite mission (TROPOMI) provides joint measurements of fire-relevant trace-gases CO, HCHO, NO_2, as well as aerosol properties, with unprecedented accuracy and spatial resolution. This opens hitherto unavailable possibilities to assess fire emission estimates. The accuracy and spatial resolution Sentinel-5p trace gas measurements allow for zooming in on much smaller regions, and spatiotemporal averaging is for CO and NO_2 not needed, *i.e.* Sentinel-5p is capable of monitoring individual fire emission plumes on a daily basis. Furthermore, the accuracy also allows for advanced use of trace gas ratios, which provides additional information about the fire process as especially NO_2 and CO emission dependencies differ significantly. The combined focus on NO_2, CO (HCHO) and aerosol parameters (AAI, ALH) thereby allows for better constraining total fire carbon emissions. The emissions of NO_2 and CO (HCHO) depend on different fire characteristics. NO_2 is typically emitted at higher burning temperatures (Zeldovich mechanism), whereas CO (HCHO) emissions are the result of incomplete combustion. Both depend on fuel type and fuel characteristics such as mass and (soil) moisture, but in a different way. The more incomplete the combustion, the more carbon that remains unburned and less NO_2 emission is to be expected. Sentinel-5p trace gas ratios thus can provide information about the burning efficiency and thereby constraining burned and emitted carbon. Aerosol information allows for estimating emission mass determination of plume extents and volumes but also to characterise fire types: more incomplete combustion will result in a larger amount of unburned (scarred) fire material and higher AAI values.

Feeding the Sentinel 1-2-3 based high resolution emission estimates as boundary conditions into the IFS/CAMS model simulations then allows for a detailed comparison with Sentinel-5p data to make optimal use of the information richness of Sentinel-5p data. Sentinels -2 and -3 observe surface characteristics at much higher spatial resolution than Sentinel 5p (~300 metre for Sentinel-3; ~10-20 metre for Sentinel-2 depending on wavelength). The Sentinel 1-2-3 observations are then used to quantify fire behaviour, burned area and fuel information and combined to identify fire types and estimated fire emissions (see Sense4Fire ATBD for details). Feeding all the information of the comparison back into the algorithms and data used by the algorithms has the potential to validate and improve emission estimates of these parameters and thereby – via the dependence of emissions on fire characteristics as described above – also improve associated emissions of total carbon.

4.2 Input data

For constraining the emissions top down using Sentinel-5p data the analyses are performed for four regions and time periods defined in Chapter 2 in the ATBDv2.

The Sentinel-5p data being used is tropospheric NO_2, CO total columns, HCHO total columns, the AAI, and the ALH. The accuracy of tropospheric NO_2 and CO is sufficient to be used in a daily basis at the Sentinel-5p spatial resolution.

The short atmospheric lifetime of NO_2 (hours) ensures that local enhancements of NO_2 near fires can almost uniquely be attributed to those fires unless there are non-fire NO_2

sources in the immediate vicinity of the fires. The regions were chosen such that there are few other sources like cities or industrial activities. Due to its decreasing vertical sensitivity with altitude, the Sentinel-5p NO_2 averaging kernel needs to be applied for the comparison with model simulations. Also, for sufficiently thick aerosol clouds the Sentinel-5p NO_2 quality flag values are lower, indicating that the tropospheric NO_2 date product actually may not have sensed the entire tropospheric column. Depending on the chosen data quality threshold such measurements then may be discarded, even though they still can provide valuable information of the emitted NO_2. Although it is beyond the scope of the Sense4Fire project to explore individual fire plume behaviour, it is explored to some extent whether there is added value in considering such lower quality tropospheric NO_2 measurements.

CO has a much longer atmospheric lifetime (weeks to months) than NO_2, resulting in seasonal changes in background CO levels due to accumulation of widespread emissions and/or advection. These seasonal changes are considered and accounted for in analyses, especially when averaging over longer periods of time as background CO levels vary seasonally. An advantage of Sentinel-5p CO is that its vertical sensitivity is nearly uniform so that to first order the effect of the averaging kernel is limited (even though it is applied).

HCHO has a lifetime more similar to that of NO_2 but due to the lower accuracy of Sentinel-5p HCHO measurements time averaging for HCHO time averaging is required, which limits the possibility to explore HCHO variations on a daily to multi-day basis. Similar to NO_2, the HCHO averaging kernel must be considered for the model comparison.

The AAI – being a qualitative indicator of absorbing aerosols - is used strictly qualitatively for identifying the extent of emission plumes and/or air masses being affected by fire emissions.

The ALH can be used for estimating the height of aerosols. However, the ALH measures a centroid aerosol layer height, not the top of the aerosol layer height. Also, the ALH is more accurate for optically thinner aerosol plumes. This, however, strongly reduces the number of useful ALH observations. For optically thick aerosol plumes the ALH quality flag values are lower, and depending on the chosen quality flag thresholds, may be disregarded. However, for optically thick aerosol plumes the atmospheric oxygen-based cloud top height can be used as a cloud top height proxy [see (de Laat et al., 2020)]. Because of the reduced capacity to assess aerosol layer heights by standard use of the ALH, the use of cloud top heights for optically thick aerosol plumes as proxy for the ALH is explored. For identifying optically thick aerosol plumes the AAI is used. There is some ambiguity of the Sentinel-5p AAI values to identify optically thick aerosol clouds. For a strict filtering AAI values > 4 are used, but it has been suggested that AAI values > 2 can also be used as long as no bright clouds are present below the aerosol plume (de Laat et al., 2020). In case of the regions and fires considered in Sense4Fire, emission plumes are expected mostly in the boundary layer and/or the lowermost atmosphere (*e.g.* (Val Martin et al., 2010); (Veira et al., 2015); (Rémy et al., 2017)). Hence it is generally not expected that many scenes with

clouds below the aerosol plumes occur, but it is something to keep in mind when performing the Sense4Fire analyses.

The IFS/CAMS global atmospheric composition modelling system is used to simulate the atmospheric conditions during the periods and for the regions under consideration (see Section 4.1). The model results serve to bridge the gap between bottom-up emission estimates (based on fire types/section 2.2 or based on the S4F model/section 3.3, GFAS, GFED) and Sentinel-5p observations. GFAS and GFED are the state-of-the-art and widely used and referred to satellite-observation-based fire emission databases (GFAS in near-real-time and GFED as a post-processing data product). As the aim of Sense4Fire is to improve satellite-observation-based fire emission estimates, results should be shown to improve on the current fire emission standards. The model results can be compared with Sentinel-5p observations after applying satellite observation operators.

The IFS/CAMS model is using target CY48R1 tropospheric chemistry and aerosol modules. It is driven by ECMWF operational meteorology, and using emissions (including anthropogenic and biogenic) as prepared for the operational model configuration. It is setup to run at an approximately 40 km spatial resolution (T511 reduced Gaussian grid), with 137 vertical model levels (approximately 60 layers in the lowest 10 km of the atmosphere). This implies that the horizontal model resolution (0.25 degrees at best, which is approximately 25 km) is much less refined than the Sentinel 1-2-3 data (sub-km scale), emission estimates (km scale) and the Sentinel-5p spatial resolution (3.5x5.5 km at best). This may have a considerable effect on the model-observation comparison, especially as localised sub-model scale non-linear emission plume chemistry may result in changes in atmospheric composition that differ from direct injection of fire emissions at the native model resolution, particularly for short-lived trace gases. This is a well-known phenomenon from for example transport (aviation, shipping, road traffic) but also stationary sources such as power plants. Such impacts traditionally have been mitigated in modelling by applying resolution-dependent correction factors to emission values. To what extent such sub-model-grid processes also play a role for fire emissions is not well established but will be considered within the context of Sense4Fire study and results.

To compare Sentinel-5p observations with IFS/CAMS output we take into account all relevant aspects that are required when matching observation data to model data. Only observations with quality assurance threshold above 0.75 are used, as recommended by the NO2 product user manual. This concerns observations with cloud radiance fraction of less than 0.5, and excludes retrievals with ground pixels covered with snow/ice, as well as problematic retrievals.

The model fields are interpolated in time to match with local overpass time of TROPOMI. The collocated model-observation pairs are gridded on a common 0.5x0.5 degree output field (or different resolution, which is configuration setting), and only written to NetCDF if a threshold coverage of 50% of the grid cell is reached.

The averaging is done by an area-weighted approach, hence taking into account the area of the TROPOMI-pixel that is within the model grid box (Douros et al., 2022).

4.3 Methods

To enable constraining fire emission using a top-down approach the following topics need to be addressed to validate this method:

1. analysis of sensitivity to model parameters and model input affecting assessment of fire emissions
2. comparison of model simulations with Sentinel-5p data to derive conclusions on input fire emissions
3. optimising the bottom-up emission estimates based on results above

These topics are evaluated based on the following set of Sentinel-5p parameters relevant for fire emission monitoring:

i. Nitrogen Dioxide (NO_2) tropospheric column
ii. Carbon Monoxide (CO) total column
iii. Formaldehyde (HCHO) total column
iv. Absorbing Aerosol Index (AAI)
v. Aerosol Layer height (ALH)

Note that Sentinel-5p is only capable of measuring the total column amounts of CO, HCHO and tropospheric NO_2. However, the accuracy and precision of Sentinel-5p measurements of CO tropospheric NO_2 is sufficient to allow measuring daily single satellite pixel atmospheric enhancements of both due to fires and near fires. Furthermore, free tropospheric abundances are small most of the time (tropospheric NO_2 and HCHO) and/or free tropospheric abundances vary more slowly (CO) which allows for monitoring single fire emission plumes. For longer lived CO slowly varying background CO levels need to be kept in mind. These include seasonal cycles due to larger scale to regional to continental accumulation of CO in the free troposphere in case of widespread persistent fires such as over the Amazon and equatorial Africa. However, such slow varying background CO levels and especially spatial gradients are well represented in the IFS model. For HCHO the lower Sentinel-5p single pixel accuracy and precision means some averaging is needed, making it less useful for fire monitoring.

Furthermore, the following observed and modelled trace gas enhancement ratios in fire-emission affected regions are explored. Note that these ratios are compared with IFS modelled ratios in which the averaging kernels are taken into account, rather than studying the ratios themselves. By comparing with IFS model results differences in vertical sensitivity is initially irrelevant. However, if needed, the availability of IFS model results allow to explore the sensitivity of analysis results to differences in Sentinel-5p vertical sensitivity for different parameters:

vi. NO_2/CO
vii. HCHO/CO

viii. $HCHO/NO_2$

Model experiments varying the following model parameters and input data are performed and analysed to identify, characterise, and rank their impact on the model results.

A. model resolution (0.25, 0.5, 1 degree)
B. emissions amounts (GFAS, GFED; variations in emission factors)
C. emission heights
D. diurnal cycle of emissions
E. model chemistry
F. observation filtering selection criteria

The baseline simulation is then compared with Sentinel-5p observations for the Sentinel-5p parameters listed above, taking observation operators into account. This provides understanding on the use of Sentinel-5p satellite observations in the interpretation of emission biases, and in particular understanding on the interpretation of the optimal way to optimise Sentinel 1-2-3 based emissions.

The diurnal cycle of emissions is based on fire radiative power climatologies.

4.4 Validation approach

Validation of the Sense4Fire emission estimates can subsequently take place using the IFS/CAMS modelling system. Specifically, the IFS/CAMS reference configuration is run with different variants of the emission estimates based on fire type classifications (denoted S4F-Cardiff) or based on the results from S4F model approach (denoted S4F-Dresden) and analysed to identify the best-performing configuration with respect to Sentinel-5p data, taking observation operators into account. The comparison results will then be further analysed by evaluation of differences with Sentinel 1-2-3 based emission parameters such as burned area, fuel type, fuel moisture content, and soil moisture content.

Disagreement is further explored by tracing the disagreement back to specific fires and/or fire locations. The specific emission characteristics of fire emissions at the identified location is evaluated, including the results from the sensitivity analyses performed before: given the results of the sensitivity analyses, what is the most likely explanation of the differences found, and do these differences fall within estimated uncertainty limits or are updates of the emissions needed. These findings are then used to update and optimise the Sentinel 1-2-3 emission estimates.

The results of the comparison of IFS/CAMS model simulation with updated Sentinel 1-2-3 emissions and Sentinel-5p observations are evaluated against the uncertainty ranges derived from the model sensitivity tests.

In addition, these results are also evaluated against limitations of the sensitivity analysis. Certain sensitivities may be difficult to capture and/or constrained with the existing IFS/CAMS model setup up. Certain processes may not be captured or resolved at the

desired scales, may be difficult to constrain due to limited or missing information and data, or may be difficult to tailor as a whole (*e.g.* optimise an entire atmospheric chemistry scheme).

Finally, other uncertainties are explored and characterised on a best-effort-basis, for example assumptions about parameters and parameter values *(e.g.* emission factors), observations and methods. This is by no means a simple and straightforward exercise and may result in an incomplete analysis but it is important to identify how fire emission estimates can be further improved.

4.5 IFS baseline results with GFAS and Sense4Fire emissions

Figure 23 shows a comparison of IFS simulated tropospheric NO_2 columns and Sentinel - 5p when using the GFAS biomass burning emissions on 10 September 2020. The comparison is shown for a larger region over the Amazon than originally defined within the Sense4Fire project (black box in the plots).

The scatter plot shows that there is a reasonable agreement between simulated and observed tropospheric NO_2 columns but that there are some fire hotspots where IFS overestimates tropospheric NO_2.

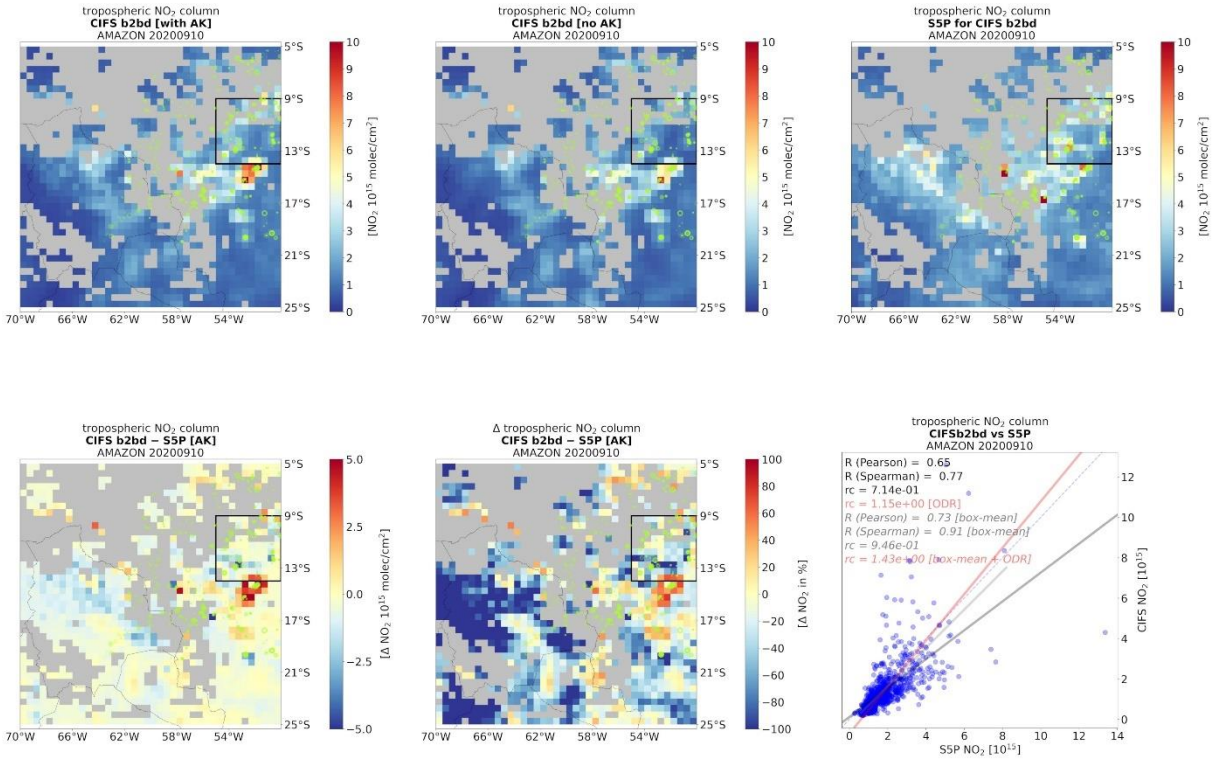

Figure 23: Comparison of IFS simulated tropospheric NO_2 columns with GFAS emissions. The upper left panel represents applying the Sentinel-5p tropospheric NO_2 columns averaging kernel, upper middle panel without applying the averaging kernel) and Sentinel-5p tropospheric NO_2 column measurements (upper right). The differences between IFS and Sentinel-5p tropospheric NO_2 columns are shown in the lower left plot (absolute differences) and lower middle plot (relative differences). The lower right plot shows the scatter plot and several statistics and linear fits. Results are just for illustration purposes.

These data can also be accumulated into a probability distribution as shown in Figure 24, in this case for the originally defined Amazon test area (black box in Figure 23) but for all data in August and September 2020. Figure 25 show the same results but for the entire Amazon region shown in Figure 23. Results reveal a fair agreement between modelled and observed tropospheric NO_2 columns but again some significant overestimation of NO_2 emissions for certain grid points. These overestimations of tropospheric NO_2 columns does not disappear or diminish for the S4F-Cardiff emission estimates compared to the GFAS emissions. If anything, there are a few IFS grid points for which the bias is even larger with Sense4Fire emissions compared to GFAS emissions.

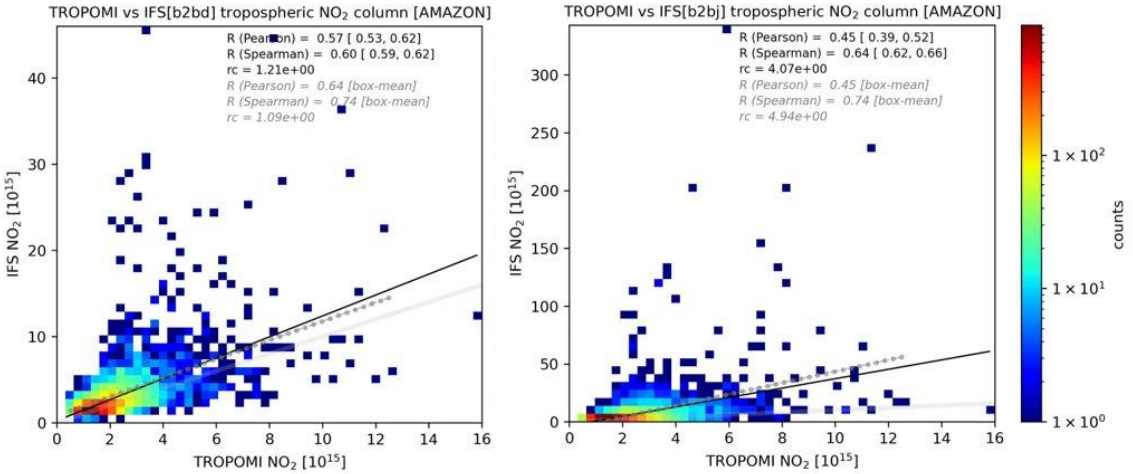

Figure 24: Probability distribution IFS simulated and Sentinel-5p observed tropospheric NO_2 columns for the originally selected Sense4Fire Amazon test area (black box in Figure 23) for all data in the period August-September 2020. Results are for illustration purposes.

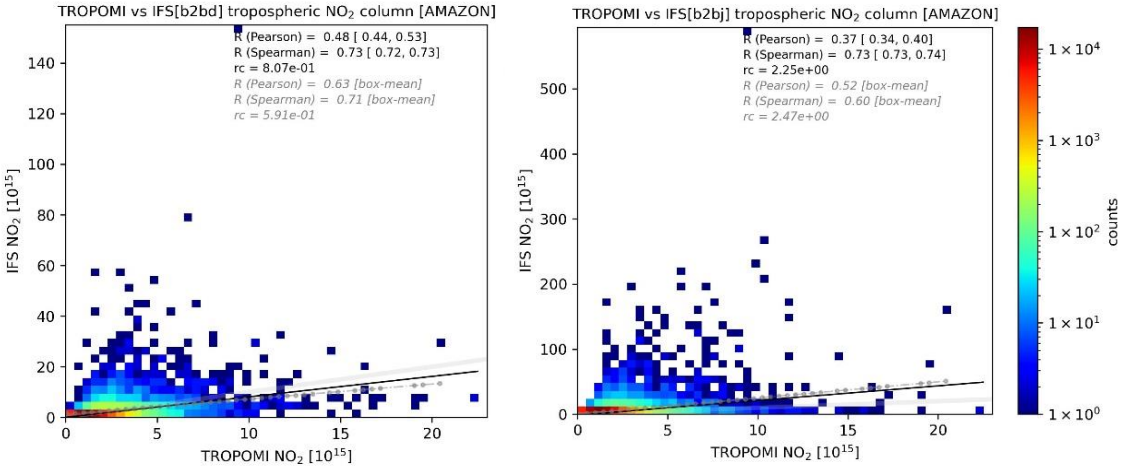

Figure 25: As Figure 24 but for the larger Amazon region as shown in Figure 23. Note the different axis ranges. Results are for illustration purposes.

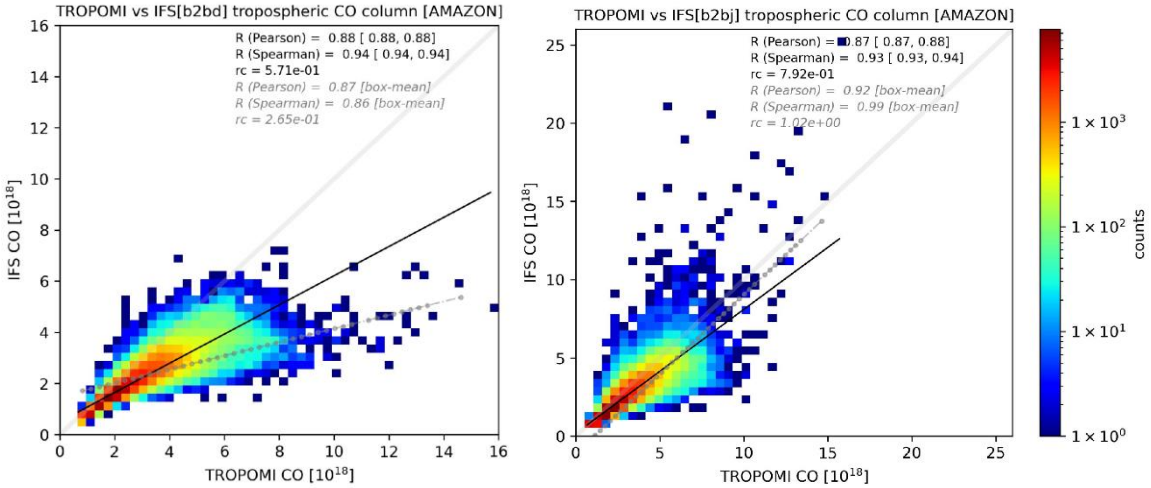

Figure 26: As Figure 25 but for carbon monoxide total columns and for the larger Amazon region shown in Figure 23

Figure 26 shows the comparison of model CO with Sentinel5p observations, using GFAS and S4F-Cardiff emissions respectively but for the entire Amazon region as shown in

Figure 23. Indicated are also some statistics of the comparison and linear regression fits. There is a thin grey line indicating the 1:1 line. Results are for illustration purposes. The comparison with GFAS emissions reveals a significant low bias (GFAS emissions too small) by more than 40 % and for very large CO columns by more than 70 %. In this case the Sense4Fire emissions significantly improve the comparison between model simulations and observations compared to the simulations with GFAS emissions. With the Sense4Fire emissions this bias is significantly reduced (approximately 20 %; including the large CO column bias of simulation b2bd) and the bias falls well within the data spread. The Sense4Fire emissions thus significantly improve the comparison with S5p CO total column data.

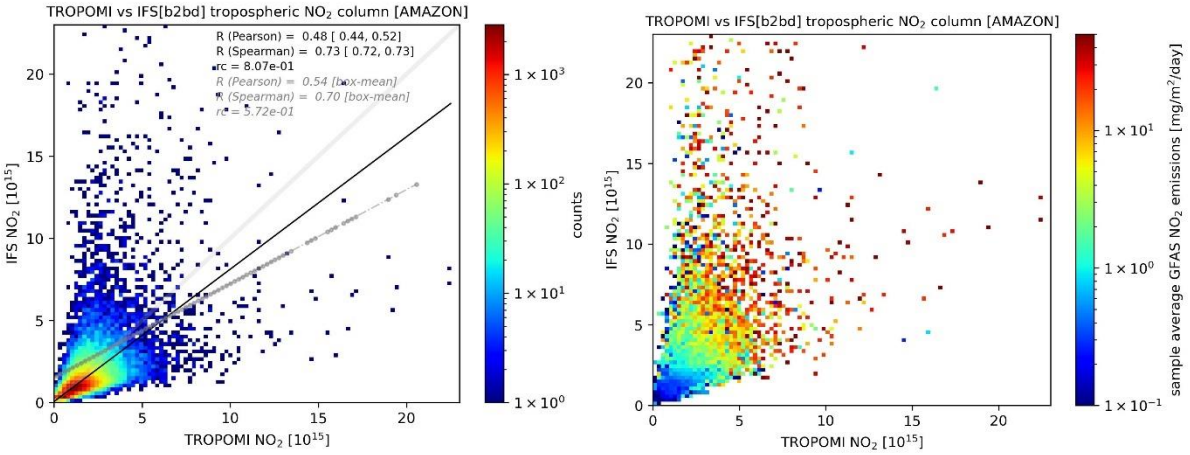

Figure 27: Left plot as in Figure 24 for the baseline GFAS emission but for NO_2 tropospheric columns and for the larger Amazon region shown in Figure 25. The right plot histogram colour coding displays the average GFAS NO_2 emission rate for each histogram interval. Figure based on all daily data in the period August-September 2020.

Further exploration of the GFAS NO_2 emissions reveals that for the smaller NO_2 columns (< 5×10^{15} molecules cm^{-2}) there is quite a reasonable agreement with Sentinel-5P observations (Figure 19). The distribution is confined around the 1:1 line although with a large spread. This relatively large spread is possibly related to spatial misrepresentation of exact emission locations due to the large IFS grid size relative to the actual fire sizes and S5p pixel sizes. Detailed exploration of IFS simulations, emission data, and Sentinel-5p observations has started aiming to shed more light on how well daily small scale fires and fire emissions are presented in the IFS simulations and emission databases.

For the comparatively large modelled NO_2 columns (>$1x10^{16}$ molecules cm^{-2}) the average GFAS emission rate is also larger (Figure 25 right panel). The deviating IFS NO_2 columns thus are associated with larger GFAS fire emissions. Note the logarithmic scale in the right panel in Figure 25, suggesting that the apparent model NO_2 column bias is associated with emissions of one and up to two orders of magnitude larger than the emission domain where Sentinel-5p and IFS agree on NO_2 column amounts.

4.6 IFS sensitivity simulations results

Section 4.5 revealed that the GFAS fire emission database underestimated CO emissions and overestimated NO_2 emissions, in particular for large fires. Although the S4F-Cardiff emissions resulted in a significant improvement in IFS simulated CO columns, the same was not the case for IFS simulated NO_2 columns. On the contrary, results are even slightly worse for Sense4Fire emissions.

Here we further explore this GFAS/ Sense4Fire NO_2 emission bias. Although spatially large fires cover only a relatively small area most fires are small, the amount of emissions is dominated by those large fires. For example, although fires with GFAS emission rates larger than 1×10^{-10} kg m^{-2} s^{-1} cover less than 10 % of the fire area in the "larger" Amazon region, they emit more than 40 % of the total fire emitted NO_2 in this region. Similarly, GFAS fires emitting an order of magnitude smaller (< 1×10^{-11} kg m^{-2} s^{-1}) only emit 5% of the total GFAS NO_2 emissions in this region but cover almost 60 % of the fire area. Since fire NO_2 emissions are indicative of fire type and the burning process, a significant GFAS NO_2 emission bias may also translate into a significant GFAS carbon emission bias (regardless of CO, as Sentinel-5p observations of CO and NO_2 over fires are not well correlated).

To further explore the GFAS/S4F-Cardiff NO_2 emission bias a number of IFS sensitivity experiments were performed:

- *T1. Sense4Fire emissions with a IFS sub-grid emission plume parameterization (expid b2bk)*
- *T2. no GFAS fire emissions (NO_2 and CO) (expid b2bl)*
- *T3. GFAS NO_2 emissions limited to a maximum of 1×10^{-10} kg m^{-2} s^{-1} (expid b2by)*
- *T4. GFAS NO_2 emissions limited to a maximum of 3×10^{-10} kg m^{-2} s^{-1} (expid b2c4)*
- *T5. Sense4Fire NO_2 emissions limited to a maximum of 3×10^{-10} kg m^{-2} s^{-1} (expid b2c6)*

The NO_2 emission limits for T3, T4 and T5 were based visually estimating where in Figure 20 IFS NO_2 columns start to deviate from Sentinel-5p NO_2 columns. The T1 results are not shown and discussed as the simulation did not result into a significant change in the IFS NO_2 column bias here. For T2 – no GFAS fire emissions – IFS obviously underestimates most Sentinel-5p NO_2 columns. Only for column amounts smaller than 2×10^{15} molecules cm^{-2} IFS and Sentinel-5p are consistent. This does indicate that there some "background" NO_2 not associated with fires, possibly biogenic and/or anthropogenic of origin but whose emissions are small in magnitude relative to the fire emissions in the regions. For the three "upper limited emission" simulations T3-T5 the large fire bias is significantly reduced, as intended. Note that based on this analysis it is not possible to assess which upper limit threshold and which emission database (GFAS or Sense4Fire) performs better. Nevertheless, limiting large fire NO_2 emissions in IFS simulations does provide directions on the type of fires to analyse in more detail and on a case-by-case basis.

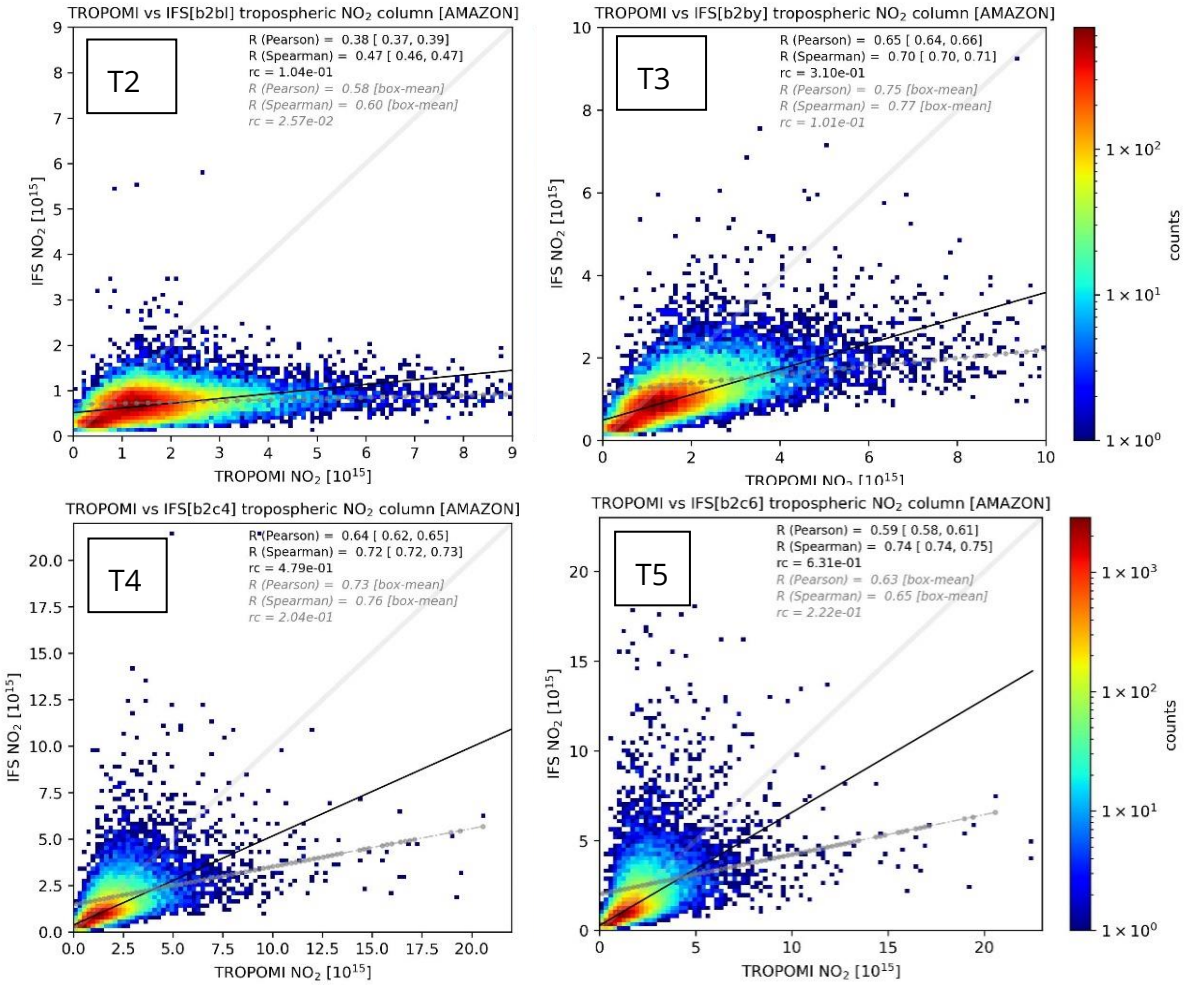

Figure 28: as in Figure 25 for four IFS sensitivity simulations and for the Amazon study region. All daily data in the period August-September 2020.

Figure 29 shows an example of a single day IFS T3 simulation over the larger Amazon region (as in Figure 23). In this simulation S4F-Cardiff NO_2 emissions were limited to a maximum of 3×10^{-10} kg m^{-2} s^{-1}. The "NO_2 hot spots" that were present in the baseline IFS GFAS simulation have thereby vanished and the simulation visually appears more similar to what was observed by Sentinel-5p. The large outliers have disappeared from the scatter plot (lower right panel) and from difference plot (lower left panel). Note that the statistics are dominated by grid points without fire emissions, hence not all statistics automatically improve.

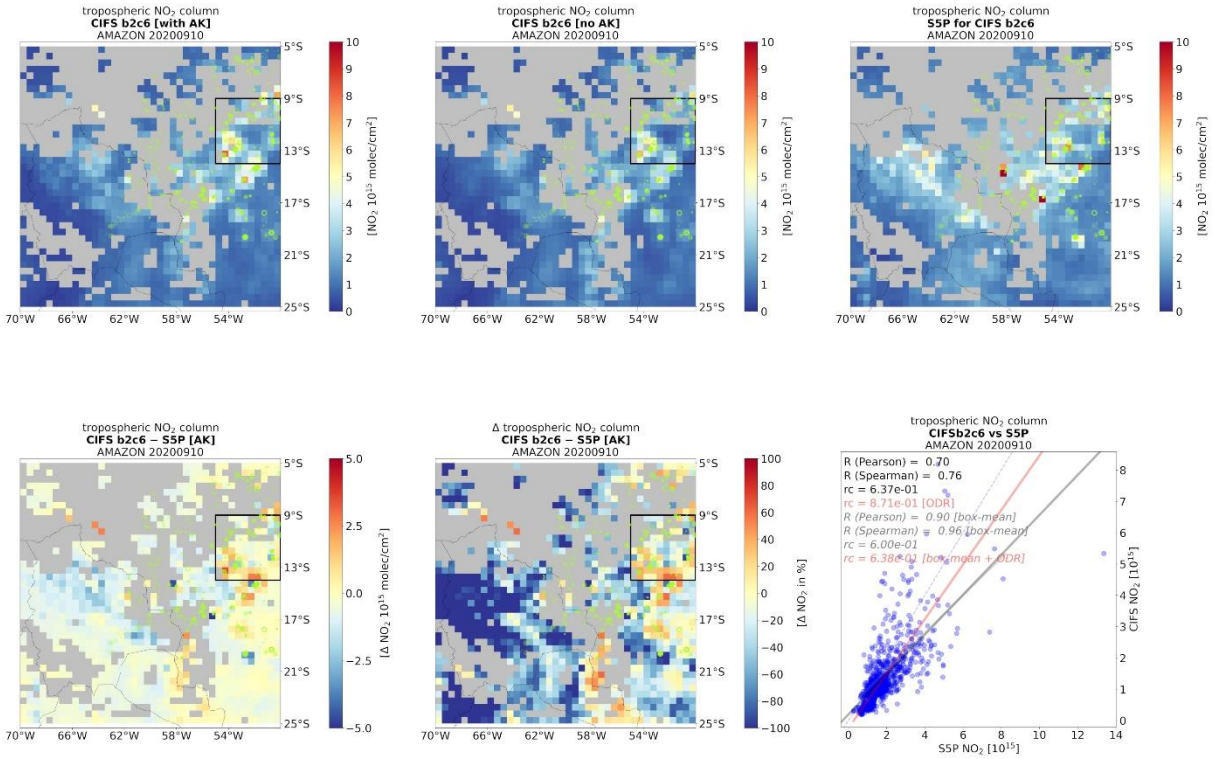

Figure 29: As Figure 23 (10 September 2020) but for the IFS sensitivity simulation T5 with S4F NO_2 emissions limited to a maximum of 3×10^{-10} kg m^{-2} s^{-1} (expid b2c6).

4.7 Intercomparison and optimization of fire emissions

Here we present an intercomparison of the various emission estimates developed and used in Sense4Fire. Figure 28 shows that over the Amazon region the S4F-Cardiff emissions have increased substantially compared to those estimated from GFAS, although the timing is well matched. This explains the higher model CO columns, which are in better agreement than those using the GFAS emissions, providing evidence that the S4F-Cardiff carbon emissions are realistic. Also the emission estimates using the S4F model described in Section 3 are included for reference (TUD-S4F.S using Song et al. land cover). These emissions have not been used in an actual IFS model experiment, but an intercomparison to the GFAS and S4F-Cardiff emissions indicates that the CO emission totals are in a comparable range.

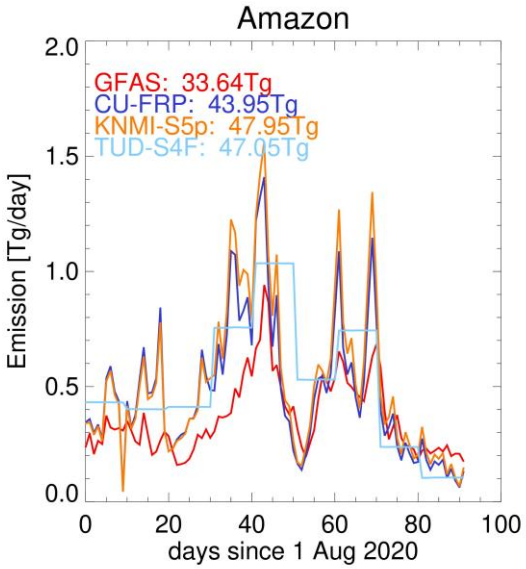

Figure 30: Intercomparison of GFAS, CU-FRP KNMI-S5p and TUD-S4F (TUD-S4F.S using Song et al. land cover as input) estimates of CO emissions over the Amazon, August-October 2020.

An experiment using the standard GFAS fire NO_x emissions globally, but replaced with CU-FRP emissions over the Amazon study region, has been evaluated against Sentinel-5p observations,Figure 31. We find considerable positive model biases over the Amazon, as also described in the previous section, while negative model biases are seen over the southern Africa study region.

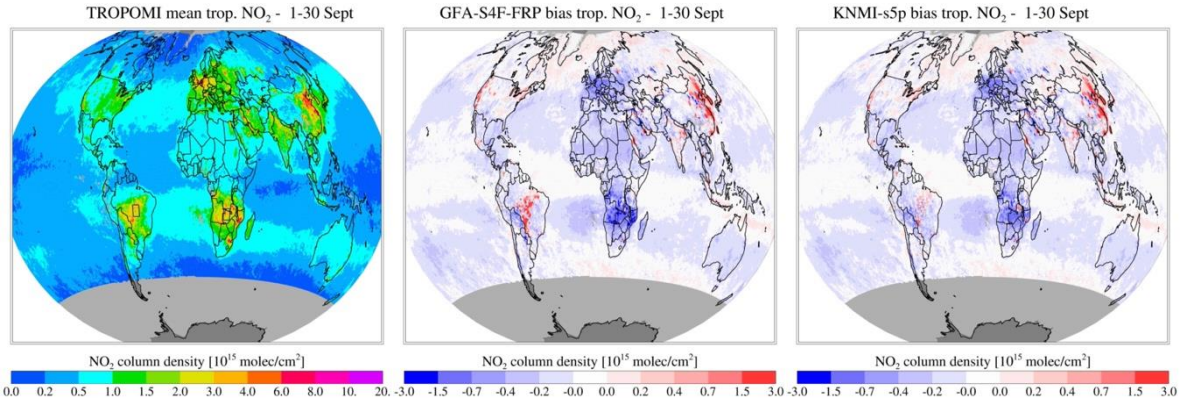

Figure 31: Left: monthly mean Sentinel-5p observations of tropospheric NO_2 for September 2020. Middle: Model bias using standard GFAS andGFA-S4F-FRP emissions over the Amazon. Right: model bias when using KNMI-S5p-based optimised fire emissions.

Where a-priori fire emissions are non-zero we attribute model biases to uncertainty in the fire emissions, and optimise them following Castellanos et al (2014). The resulting optimised emissions are presented in Figure 31. Based on this the GFA-S4F-FRP emissions over the Amazon appear in reasonable agreement, although a various peaks with high emission values benefit from a reduction. The same is apparent for the GFAS emission estimate over Siberia, beginning of August (not shown here), while over the southern

Africa study region the GFAS-based fire NO_x emissions are under-estimated. An evaluation of the KNMI-S5p updated NO_x emissions is also presented in this figure. This indeed illustrates that hotspots with strong positive model biases over the Amazon have disappeared, while the negative model bias over the southern Africa region is also improved. Time series showing the quantitative differences between the various emission estimates are given in Figure 32.

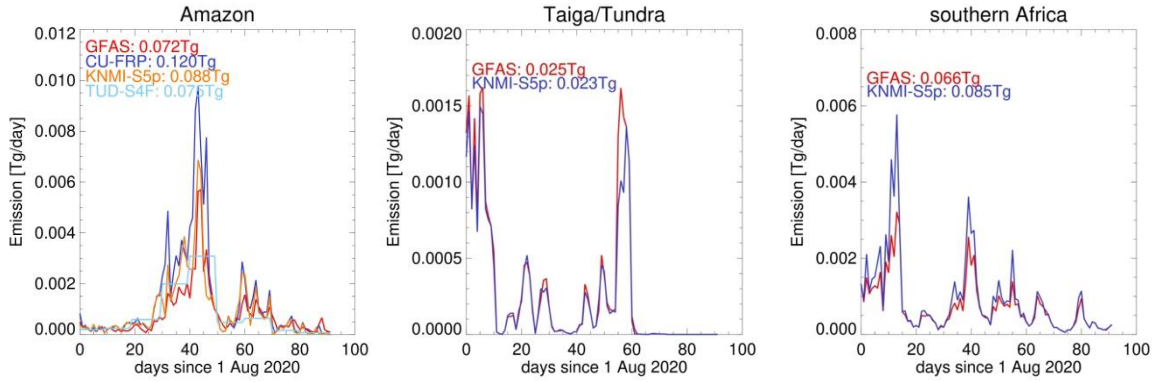

Figure 32: Intercomparison of GFAS, TUD-S4F (using Song land cover) and GFA-S4F-FRP (CU-FRP) estimates of NOx emissions (Tg NO/day) over the Amazon, along with KNMI-S5p optimised S4F emissions based on Sentinel-5p observations. Also the GFAS and KNMI-S5p optimized emissions over the Taiga/Tundra study region in Sibera, and southern Africa are shown.

4.8 Discussion

4.8.1 Limitations and gap analysis

The validation results are analysed from a data product improvement framework perspective, and some results can be used immediately within the Sense4Fire product development, in an iterative process, to improve its quality. More fundamental limitations, which cannot be resolved within the Sense4Fire project duration, will also be identified and described in terms of their complexity, and implementation requirements, which requires follow-up activities beyond the scope of this project. Finally, some of the improvements needed may fall outside of the project consortium capacity – for instance fundamental model refinement or improved quality of satellite data products – which would result in recommendations for additional research and/or development. Combined, the validation analysis and evaluation results should provide clear pathways towards improvement.

4.8.2 Case-based analysis

The evaluation so far has focussed on statistical analyses of large databases: months of daily data at high spatial resolution over large areas.

This resulted in the identification of a significant large-fire bias in bottom-up NO_2 emissions. To better understand where this bias originates from a pilot study has started with a more detailed daily-case-based evaluation of fire emissions and simulation results. A first test case was identified for 8 September 2020 (Figure 33). The IFS results indicate

an NO_2 hot spot between 52°W-52.5°W and 11.5°S-10.5°S which was not observed in Sentinel-5p measurements. The corresponding VIIRS image shows no clouds and a number of well discernible smoke plumes.

Figure 34 shows the corresponding vegetation type from the ESA CCI land cover map as well as fires derived from VIIRS radiative power. The smoke plumes are clearly associated with VIIRS fires but predominantly occur in the western half of the area. The IFS/GFAS NO_2 hot spot occurs over a region where not many fires were identified by VIIRS. It is, however, a region where the vegetation type changes from rain forest ("tree broadleaved evergreen") to shrubland, which can also be seen in the VIIRS image in Figure 33.

Similarly, Figure 35 shows that the Sentinel-5p orbit-level tropospheric NO_2 columns do not provide indications of large amounts of NO_2 while further west and east NO_2 hot spots are visible in the Sentinel-5p NO_2 data.

Interestingly, Figure 36 reveals that the IFS simulated NO_2 columns using S4F-Cardiff emissions shows NO_2 hot spots that spatially appear to better align with the VIIRS fire radiative power hot spots, although the simulated NO_2 columns are still significantly overestimated compared to the sentinel-5p measurements.

The results from the two IFS simulations using GFAS and S4F-Cardiff NO_2 emissions lead to significantly different results. This Indicates that not only the absolute amounts of fire NO_2 emissions are uncertain - especially for large fires – but that also the spatial distribution may differ largely from one fire emissions database to the other on the spatial scales of the actual fires.

Finally, the GFAS emissions used for the IFS simulation results in figures 34 and 35 are in part based on MODIS FRP, which has a mid-morning overpass time that is approximately three hours earlier than the Sentinel-5p observations. This is not the case for the S4F-Cardiff emissions shown in figure 37 that rely on VIIRS FRP, which is flying in what is called a "loose constellation" with Sentinel-5p (approximately 5 minute difference). It is imaginable that MODIS FRP and thus GFAS emissions may locally not be fully representative for the Sentinel-5p observations. On the other hand, comparison between VIIRS and MODIS FRP (Li et al., 2017) does not result in systematic biases large enough to explain systematic S5P-IFS NO_2 differences as shown in figure 33. Also, simulations using S4F emissions based on the GFA algorithm which uses VIIRS FRP rather than MODIS FRP result in similar S5p-IFS NO_2 differences as when using GFAS in IFS. Comparison of VIIRS and MODIS FRP-based NO_2 emission estimates (Fu et al., 2022) do not reveal spatio-temporal differences large enough to affect the main findings here. Hence although the use of MODIS FRP in GFAS definitely will cause some differences compared to using VIIRS-FRP based emissions, they are insufficient to explain the larger IFS-Sentinel-5p NO_2 differences. Nevertheless, the importance of FRP as cause of inconsistencies between IFS simulations and Sentinel-5p observations will be made part of the trace back of these differences to individual bottom-up emission input parameters using the beta-method results for directly estimating NO_2 emission from Sentinel-5p NO_2 measurements.

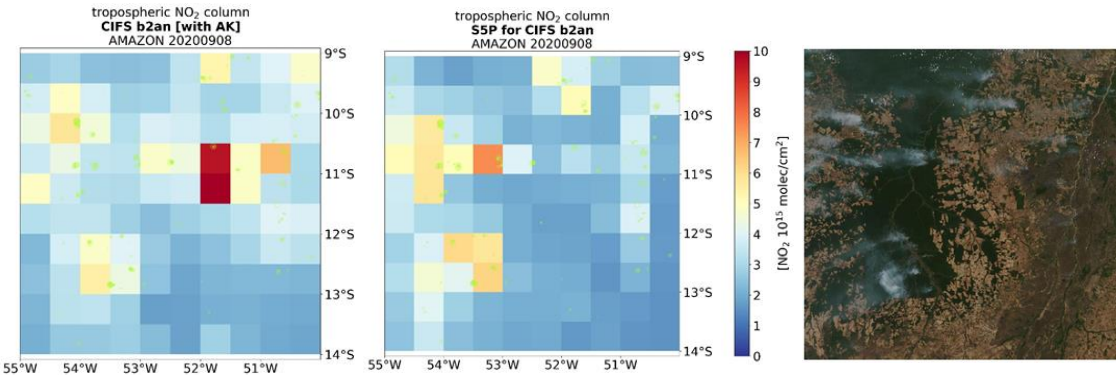

Figure 33: IFS/GFAS simulated tropospheric NO_2 columns (left) and corresponding gridded Sentinel-5p tropospheric NO_2 columns with applying the averaging kernel (middle) and associated VIIRS RGB image (right) for the Sense4Fire Amazon test area on 8 September 2020.

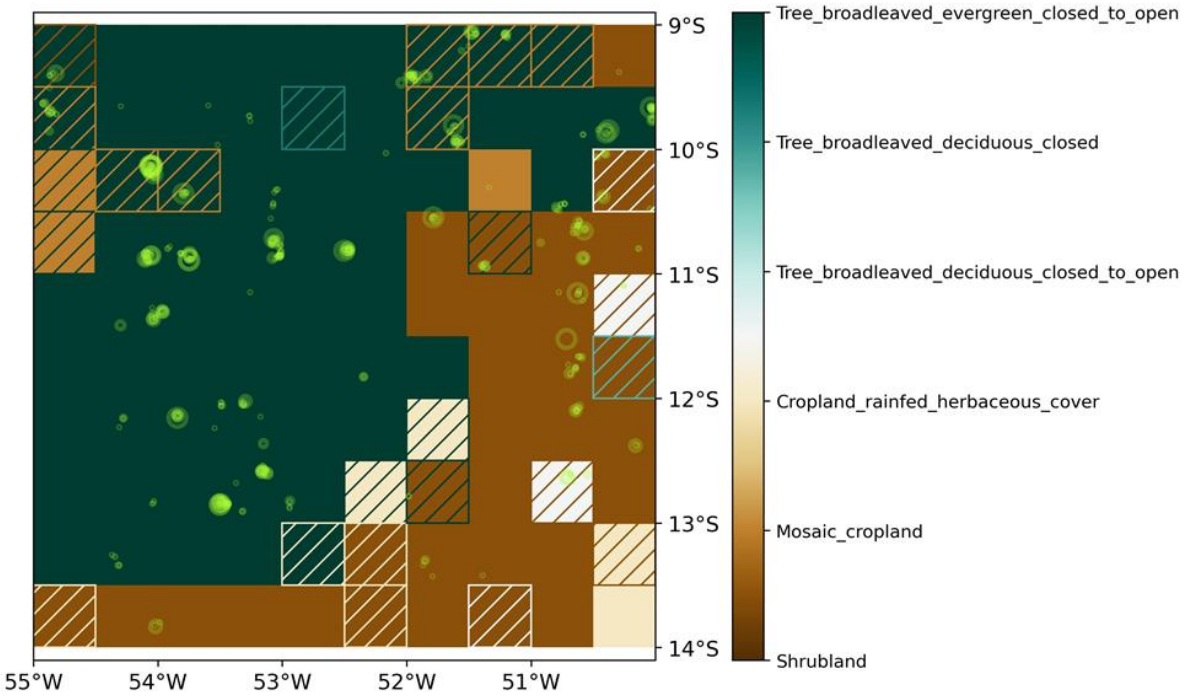

Figure 34: ESA CCI Vegetation type for the area and date presented in Figure 33.
The colors indicate the dominant vegetation type, the shading indicates grids where there is a significant second type vegetation (shading colours indicating the vegetation type). The bright green dots represent the associated VIIRS fire radiative power hot spots. Note that ESA CCI Vegetation type data is not explicitly used in GFAS.

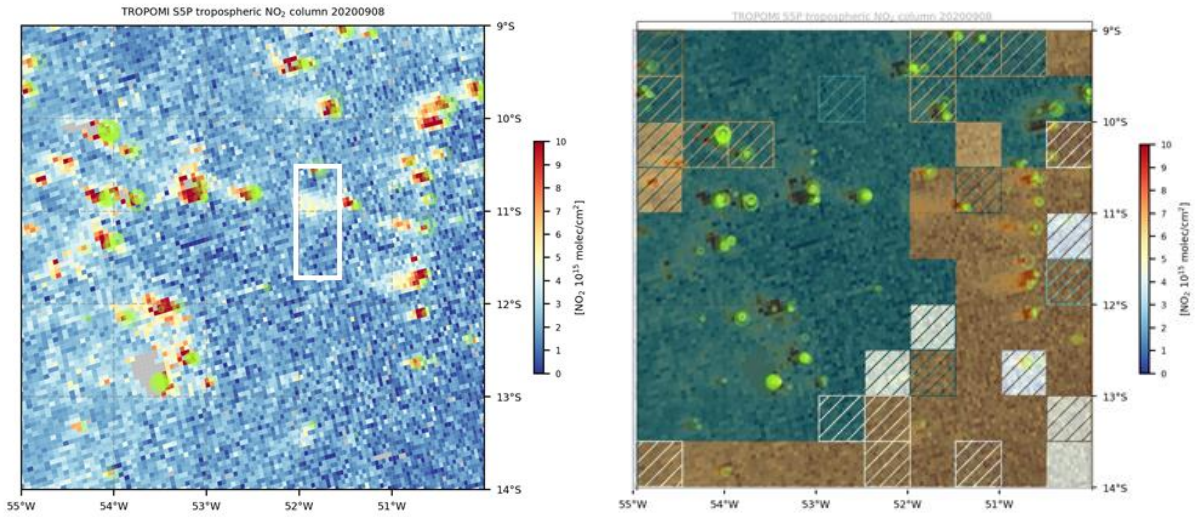

Figure 35: Sentinel-5p orbit-level tropospheric NO_2 columns (left) and overlaid on top of the vegetation type map of Figure 34 (right). The white rectangle in the left plot indicates the pixels with high NO_2 in the IFS/GFAS experiment (see Figure 33 left).

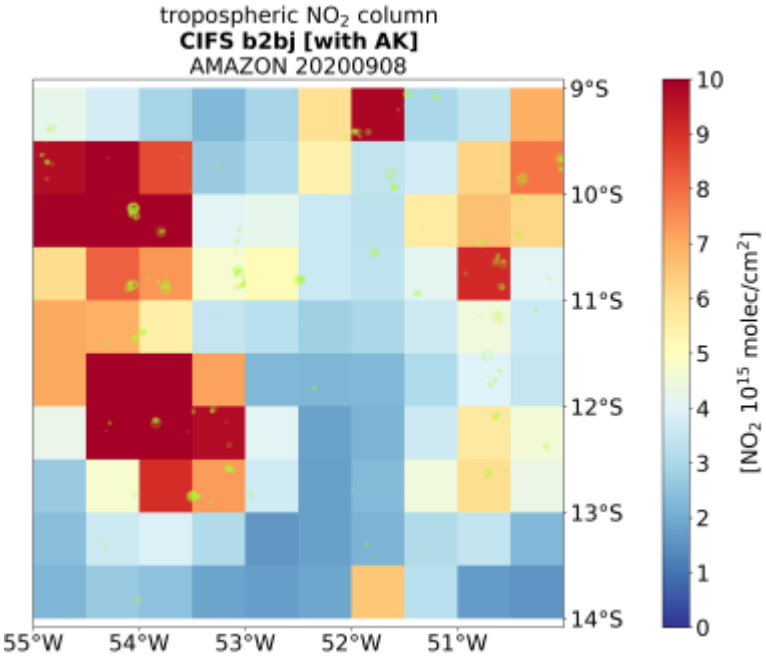

Figure 36: IFS simulation for the same date and region as Figure 33 but with Sense4Fire (University Cardiff) NO_2 emissions.

5 Summary and future developments

The results of the evaluation of satellite-based fire emissions with Sentinel-5p CO and NO_2 column measurements has uncovered significant emission biases in state-of-the-art fire emission databases like GFAS. Some of these appear to be resolvable by including more (satellite)-based information, and the bottom-up fire emissions of CO based on the fire type classification (GFA-S4F) appear an improvement. The TUD-S4F data-model fusion approach allows to trace-back biases in estimated fire emissions to the used land cover or burned area datasets and how they affect the estimated fuel loads, fuel consumtpion and hence fire emissions.

Biases in NO_2 were not yet resolved. NOx emissions are overestimated by both GFAS and GFA-S4F in comparison to Sentinel-5p especially because of single extreme fires. The TUD-S4F NOx emission estimates of extreme fires are lower which might not result in an overestimation of NOx fields. However, the comparison of TUD-S4F estimates with Sentinel-5p was not yet performed because of the much coarser temporal (10-daily) resolution. Those initial joint validation results of all components of the Sense4Fire project highlight the potential of obtaining novel information on fire dynamics and the need to consolidate and integrate the three different estimates.

References

Andela, N., Morton, D. C., Giglio, L., Paugam, R., Chen, Y., Hantson, S., Van Der Werf, G. R., and Randerson, J. T.: The Global Fire Atlas of individual fire size, duration, speed and direction, Earth Syst. Sci. Data, 11, 529–552, 2019.

Andela, N., Morton, D. C., Schroeder, W., Chen, Y., Brando, P. M., and Randerson, J. T.: Tracking and classifying Amazon fire events in near real time, Sci. Adv., 8, eabd2713, https://doi.org/10.1126/sciadv.abd2713, 2022.

Andreae, M. O.: Emission of trace gases and aerosols from biomass burning–an updated assessment, Atmospheric Chem. Phys., 19, 8523–8546, 2019.

Bauer-Marschallinger, B., Paulik, C., Hochstöger, S., Mistelbauer, T., Modanesi, S., Ciabatta, L., Massari, C., Brocca, L., and Wagner, W.: Soil Moisture from Fusion of Scatterometer and SAR: Closing the Scale Gap with Temporal Filtering, Remote Sens., 10, 1030, https://doi.org/10.3390/rs10071030, 2018.

Falster, D. S., Duursma, R. A., Ishihara, M. I., Barneche, D. R., FitzJohn, R. G., Vårhammar, A., Aiba, M., Ando, M., Anten, N., Aspinwall, M. J., Baltzer, J. L., Baraloto, C., Battaglia, M., Battles, J. J., Bond-Lamberty, B., van Breugel, M., Camac, J., Claveau, Y., Coll, L., Dannoura, M., Delagrange, S., Domec, J.-C., Fatemi, F., Feng, W., Gargaglione, V., Goto, Y., Hagihara, A., Hall, J. S., Hamilton, S., Harja, D., Hiura, T., Holdaway, R., Hutley, L. S., Ichie, T., Jokela, E. J., Kantola, A., Kelly, J. W. G., Kenzo, T., King, D., Kloeppel, B. D., Kohyama, T., Komiyama, A., Laclau, J.-P., Lusk, C. H., Maguire, D. A., le Maire, G., Mäkelä, A., Markesteijn, L., Marshall, J., McCulloh, K., Miyata, I., Mokany, K., Mori, S., Myster, R. W., Nagano, M., Naidu, S. L., Nouvellon, Y., O'Grady, A. P., O'Hara, K. L., Ohtsuka, T., Osada, N., Osunkoya, O. O., Peri, P. L., Petritan, A. M., Poorter, L., Portsmuth, A., Potvin, C., Ransijn, J., Reid, D., Ribeiro, S. C., Roberts, S. D., Rodríguez,

R., Saldaña-Acosta, A., Santa-Regina, I., Sasa, K., Selaya, N. G., Sillett, S. C., Sterck, F., Takagi, K., Tange, T., Tanouchi, H., Tissue, D., Umehara, T., Utsugi, H., Vadeboncoeur, M. A., Valladares, F., Vanninen, P., Wang, J. R., Wenk, E., Williams, R., de Aquino Ximenes, F., Yamaba, A., Yamada, T., Yamakura, T., Yanai, R. D., and York, R. A.: BAAD: a Biomass And Allometry Database for woody plants, Ecology, 96, 1445–1445, https://doi.org/10.1890/14-1889.1, 2015.

Forkel, M., Schmidt, L., Zotta, R.-M., Dorigo, W., and Yebra, M.: Estimating leaf moisture content at global scale from passive microwave satellite observations of vegetation optical depth, Hydrol. Earth Syst. Sci., 27, 39–68, https://doi.org/10.5194/hess-27-39-2023, 2023.

Fuster, B., Sánchez-Zapero, J., Camacho, F., García-Santos, V., Verger, A., Lacaze, R., Weiss, M., Baret, F., and Smets, B.: Quality Assessment of PROBA-V LAI, fAPAR and fCOVER Collection 300 m Products of Copernicus Global Land Service, Remote Sens., 12, 1017, https://doi.org/10.3390/rs12061017, 2020.

Harmon, M. E., Fasth, B. G., Yatskov, M., Kastendick, D., Rock, J., and Woodall, C. W.: Release of coarse woody detritus-related carbon: a synthesis across forest biomes, Carbon Balance Manag., 15, 1, https://doi.org/10.1186/s13021-019-0136-6, 2020.

Holland, E. A., Post, W. M., Matthews, E. G., Sulzman, J. M., Staufer, R., and Krankina, O. N.: A Global Database of Litterfall Mass and Litter Pool Carbon and Nutrients, , https://doi.org/10.3334/ORNLDAAC/1244, 2014.

de Laat, A., Vazquez-Navarro, M., Theys, N., and Stammes, P.: Analysis of properties of the 19 February 2018 volcanic eruption of Mount Sinabung in S5P/TROPOMI and Himawari-8 satellite data, Nat. Hazards Earth Syst. Sci., 20, 1203–1217, https://doi.org/10.5194/nhess-20-1203-2020, 2020.

Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of the Earth, https://doi.org/10.48550/arXiv.2204.08322, 13 April 2022.

van Leeuwen, T. T., van der Werf, G. R., Hoffmann, A. A., Detmers, R. G., Rücker, G., French, N. H. F., Archibald, S., Carvalho Jr., J. A., Cook, G. D., de Groot, W. J., Hély, C., Kasischke, E. S., Kloster, S., McCarty, J. L., Pettinari, M. L., Savadogo, P., Alvarado, E. C., Boschetti, L., Manuri, S., Meyer, C. P., Siegert, F., Trollope, L. A., and Trollope, W. S. W.: Biomass burning fuel consumption rates: a field measurement database, Biogeosciences, 11, 7305–7329, https://doi.org/10.5194/bg-11-7305-2014, 2014.

Lizundia-Loiola, J., Otón, G., Ramo, R., and Chuvieco, E.: A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data, Remote Sens. Environ., 236, 111493, https://doi.org/10.1016/j.rse.2019.111493, 2020.

Moesinger, L., Dorigo, W., Jeu, R. de, Schalie, R. van der, Scanlon, T., Teubner, I., and Forkel, M.: The global long-term microwave Vegetation Optical Depth Climate Archive (VODCA), Earth Syst. Sci. Data, 12, 177–196, https://doi.org/10.5194/essd-12-177-2020, 2020.

Rémy, S., Veira, A., Paugam, R., Sofiev, M., Kaiser, J. W., Marenco, F., Burton, S. P., Benedetti, A., Engelen, R. J., Ferrare, R., and others: Two global data sets of daily fire emission injection heights since 2003, Atmospheric Chem. Phys., 17, 2921–2942, 2017.

Roteta, E., Bastarrika, A., Franquesa, M., and Chuvieco, E.: Landsat and Sentinel-2 Based Burned Area Mapping Tools in Google Earth Engine, Remote Sens., 13, 816, https://doi.org/10.3390/rs13040816, 2021.

Santoro, M., Cartus, O., Carvalhais, N., Rozendaal, D. M. A., Avitabile, V., Araza, A., de Bruin, S., Herold, M., Quegan, S., Rodríguez-Veiga, P., Balzter, H., Carreiras, J., Schepaschenko, D., Korets, M., Shimada, M., Itoh, T., Moreno Martínez, Á., Cavlovic, J., Cazzolla Gatti, R., da Conceição Bispo, P., Dewnath, N., Labrière, N., Liang, J., Lindsell, J., Mitchard, E. T. A., Morel, A., Pacheco Pascagaza, A. M., Ryan, C. M., Slik, F., Vaglio Laurin, G., Verbeeck, H., Wijaya, A., and Willcock, S.: The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations, Earth Syst. Sci. Data, 13, 3927–3950, https://doi.org/10.5194/essd-13-3927-2021, 2021.

van der Schalie, R., de Jeu, R. A. M., Kerr, Y. H., Wigneron, J. P., Rodríguez-Fernández, N. J., Al-Yaari, A., Parinussa, R. M., Mecklenburg, S., and Drusch, M.: The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E, Remote Sens. Environ., 189, 180–193, https://doi.org/10.1016/j.rse.2016.11.026, 2017.

Song, X.-P., Hansen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., and Townshend, J. R.: Global land change from 1982 to 2016, Nature, 560, 639–643, https://doi.org/10.1038/s41586-018-0411-9, 2018.

Val Martin, M., Logan, J. A., Kahn, R. A., Leung, F.-Y., Nelson, D. L., and Diner, D. J.: Smoke injection heights from fires in North America: analysis of 5 years of satellite observations, Atmospheric Chem. Phys., 10, 1491–1510, https://doi.org/10.5194/acp-10-1491-2010, 2010.

Veira, A., Kloster, S., Wilkenskjeld, S., and Remy, S.: Fire emission heights in the climate system – Part 1: Global plume height patterns simulated by ECHAM6-HAM2, Atmospheric Chem. Phys., 15, 7155–7171, https://doi.org/10.5194/acp-15-7155-2015, 2015.

van Wees, D., van der Werf, G., Randerson, J. T., Andela, N., Chen, Y., and Morton, D. C.: The role of fire in global forest loss dynamics, Glob. Change Biol., 27, 2377–2391, https://doi.org/10.1111/gcb.15591, 2021.

van Wees, D., van der Werf, G. R., Randerson, J. T., Rogers, B. M., Chen, Y., Veraverbeke, S., Giglio, L., and Morton, D. C.: Global biomass burning fuel consumption and emissions at 500 m spatial resolution based on the Global Fire Emissions Database (GFED), Geosci. Model Dev., 15, 8411–8437, https://doi.org/10.5194/gmd-15-8411-2022, 2022.

van der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., Collatz, G. J., and others: Global fire emissions estimates during 1997-2016, Earth Syst. Sci. Data, 9, 697–720, 2017.

Yebra, M., Scortechini, G., Badi, A., Beget, M. E., Boer, M. M., Bradstock, R., Chuvieco, E., Danson, F. M., Dennison, P., Dios, V. R. de, Bella, C. M. D., Forsyth, G., Frost, P., Garcia, M., Hamdi, A., He, B., Jolly, M., Kraaij, T., Martín, M. P., Mouillot, F., Newnham, G., Nolan, R. H., Pellizzaro, G., Qi, Y., Quan, X., Riaño, D., Roberts, D., Sow, M., and Ustin, S.: Globe-LFMC, a global plant water status database for vegetation ecophysiology and wildfire applications, Sci. Data, 6, 1–8, https://doi.org/10.1038/s41597-019-0164-9, 2019.