

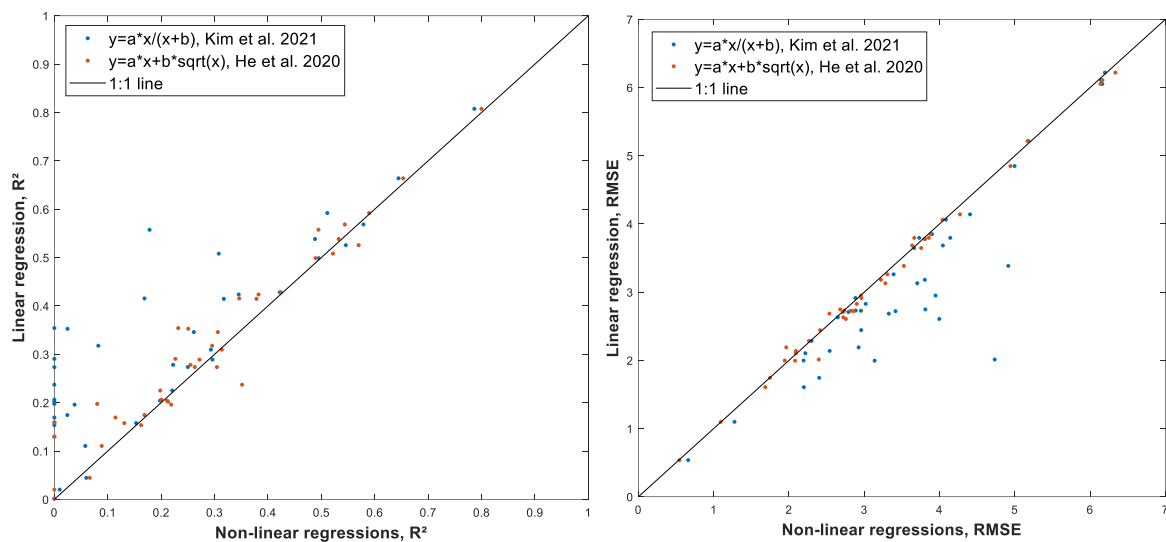
Dear reviewer,

We would like to thank you for taking your time to evaluate our work and foremostly for your interesting and useful comments, questions, remarks and suggestions.

We will try to answer your questions and we will integrate your suggestions into the manuscript (all answers are in blue color).

The authors claim there is a linear correlation between SIFd and GPP both in the site and the PFT levels. However, a quick look at the figures (2-4) shows that in most cases at some point the SIFd-GPP relation reaches saturation. The authors did not mention this even once in their manuscript. Several works are demonstrating this relation and discuss its meaning (see He et al., 2020 for example), however, the authors here ignore it and refer to it as a linear relation. Moreover, in many cities and PFT, the linear correlation is also low for the same reason.

While non-linear relationships have been shown at canopy scale (He et al. 2020, Kim et al. 2021), when using satellite data, the added noise and inherent linearization at larger scales (3.5*7.5 km in our case) makes it hard to fit non-linear model across a diverse set of sites. In our case, non-linear models do not show a clear improvement in performance, as shown in the following figure and table :



	linear regression		Hyperbolic model (Kim et al. 2021)		linear+square root model (He et al. 2020)	
Site name	R ²	RMSE	R ²	RMSE	R ²	RMSE
BE-Bra	0.592	2.00	0.511	2.19	0.590	1.94
BE-Lcr	0.204	4.06	0.198	4.09	0.211	4.04
BE-Lon	0.274	6.05	0.250	6.16	0.263	6.14
BE-Vie	0.202	2.44	-0.160	2.96	0.213	2.41
CH-Dav	0.153	1.74	-0.597	2.40	0.163	1.75
CZ-BK1	0.354	3.38	-0.345	4.91	0.232	3.53
CZ-Lnz	0.416	3.18	0.169	3.80	0.346	3.22
CZ-Wet	0.196	3.68	0.038	4.04	0.219	3.64
DE-Geb	0.526	3.80	0.546	3.73	0.570	3.66
DE-Gri	0.278	3.26	0.222	3.39	0.255	3.31
DE-Hai	0.569	2.91	0.579	2.88	0.544	2.96
DE-HoH	0.346	4.14	0.261	4.40	0.306	4.27
DE-Hzd	0.415	2.73	0.317	2.96	0.379	2.82
DE-Kli	0.310	3.85	0.293	3.90	0.314	3.88

DE-RuR	0.174	3.80	0.024	4.14	0.169	3.86
DE-RuS	0.289	6.22	0.296	6.20	0.272	6.34
DE-RuW	0.159	2.61	-0.959	4.00	0.000	2.76
DE-Tha	0.558	1.61	0.178	2.20	0.495	1.68
DK-Sor	0.808	2.73	0.787	2.89	0.800	2.74
FI-Hyy	0.274	2.68	-0.111	3.33	0.304	2.54
FI-Sii	0.318	1.10	0.082	1.28	0.296	1.09
FI-Var	0.130	2.00	-1.146	3.14	-0.008	2.08
FR-Aur	0.225	3.65	0.221	3.66	0.198	3.76
FR-Bil	0.169	3.13	-0.160	3.70	0.114	3.28
FR-EM2	0.111	4.85	0.058	5.00	0.089	4.94
FR-Fon	0.664	2.71	0.645	2.79	0.653	2.72
FR-Hes	0.499	3.78	0.495	3.80	0.489	3.81
FR-LGt	0.428	2.29	0.422	2.30	0.424	2.27
FR-Mej	0.045	5.22	0.060	5.18	0.066	5.17
GF-Guy	0.020	2.63	0.010	2.65	-0.045	2.72
IT-BCi	0.158	6.11	0.001	6.67	0.131	6.16
IT-Cp2	0.001	2.01	-4.463	4.73	-0.397	2.40
IT-Lsn	0.538	2.10	0.488	2.22	0.532	2.09
IT-SR2	0.206	2.95	-0.991	4.68	0.201	2.96
IT-Tor	0.424	2.83	0.345	3.02	0.383	2.90
NL-Loo	0.237	2.19	-0.349	2.93	0.352	1.96
SE-Deg	0.353	0.54	0.025	0.66	0.251	0.54
SE-Htm	0.291	2.72	-0.114	3.41	0.227	2.86
SE-Nor	0.508	2.14	0.308	2.54	0.522	2.09
SE-Svb	0.198	2.75	-0.542	3.81	0.080	2.68

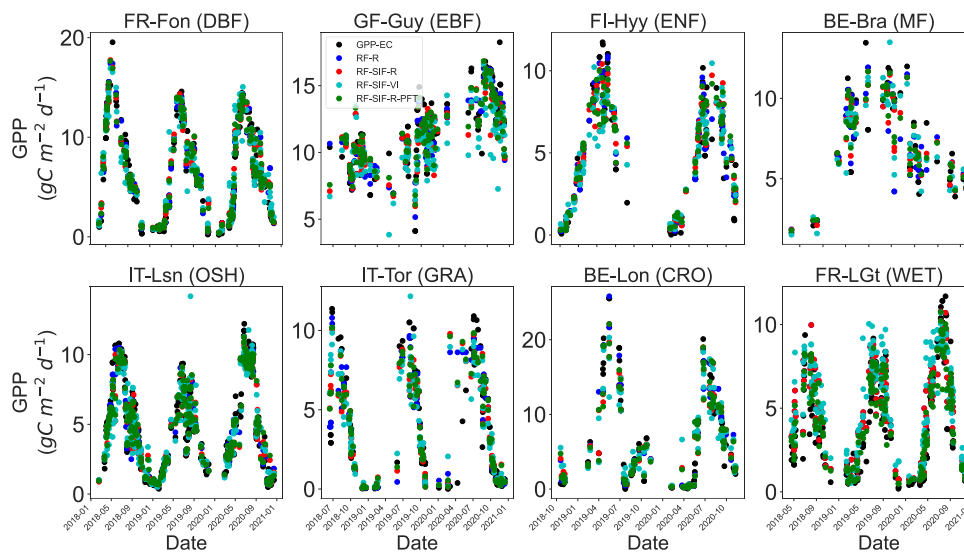
While the two tested non-linear models do not show significantly lower RMSE overall, they exhibit a strong instability, and cannot be accurately fitted on all sites. As our goal is to compare relationships between sites and considering the level of noise in the TROPOMI SIF data, we see no clear benefit in using a non-linear model which only brings in marginal improvement over a few sites at the expense of a loss in genericity.

I'm not sure what is the added value of the pooled graph of all PFTs vs. SIFd (Fig. 4)

Within this figure, we would like to evaluate the genericity of the relationship between GPP and SIFd across the study sites, which is demonstrated by the low R^2 value found on data pooled across all sites.

It is very hard to estimate the performance of the different models vs. EC GPP in Figure 8. Please consider reducing the size of the dots and making them transparent.

The size of the dots has been changed and we made them transparent.



Line 399: Please try to explain why the high correlation in the DBF and OSH PFT's

We briefly tried to explain this high correlation in line 399. The main explanation is that in DBF and OSH (one sample of vineyard plantation) biomes, there are explicitly marked seasonal and phenological changes compared to EBF or ENF forest where there is greenness all time. Thus, in DBF and OSH biomes SIF signal may easily capture the variations in LAI and APAR and consequently display a high correlation between GPP and SIF_d (added to: line 405-408).

Line 410: This is not clear to me, the authors mention in the methods section that they took out the cloudy day data. Line 414: So, why not do that in your data?

The representativity of satellite SIF data which needs to be filtered for cloud coverage is indeed a limitation of the current study. This limitation is inherent to the use of satellite data, and can only be lifted through the collection of ground-based SIF data across diverse ecosystems.

Line 436: it is problematic to say there is a difference in the models while earlier you mentioned there was no statistical difference between them (line 296). Line 447: same comment as above.

The line 296 was indeed unclear. We edited it as follows:

“A paired t-test realized between the four models based on the adjusted R^2 performance revealed that the difference in mean **adjusted R^2** between RF-R and RF-SIF-R, RF-R and RF-SIF-R-PFT, and RF-SIF-R and RF-SIF-R-PFT models was not statistically significant. In other words, these three FR models have the same performance.”

We have reframed the line 436 as follows to clarify it:

“The current results show that the RF-R (surface spectral reflectance alone), RF-SIF-R (SIF_d plus surface spectral reflectance) and RF-SIF-R-PFT (SIF_d plus surface spectral reflectance plus PFT) models explain a non-significant different percentage of the variance in tower-based GPP at the daily time scale (82~86%), whereas the RF-SIF-VI (SIF_d plus reflectance based-indices) explains 75% of the interannual variabilities in GPP across all sites.”

As for line 447, it is based on a difference in relative importance rather than a difference in R^2 (see Figure 9). Unlike the differences in R^2 which aren't statistically different and cannot be interpreted, the relative importance depends on input variables and can be interpreted. This sentence was edited as follows to clarify it:

“SIF remains a better predictor of GPP than each reflectance band individually (Fig. 9).”