Supplementary materials:

Data-based investigation of the effects of canopy structure and shadows on sun-induced chlorophyll fluorescence in a deciduous oak forest

Hamadou Balde[1,2,3,4], Gabriel Hmimina[1], Yves Goulas[1], Gwendal Latouche[2], Abderrahmane Ounis[1], Kamel Soudani[2]

[1]Laboratoire de Météorologie Dynamique, Sorbonne Université, IPSL, CNRS/L'École polytechnique, 91128, Palaiseau Cedex, France

[2]Ecologie Systématique et Evolution, Université Paris-Saclay, CNRS, AgroParisTech, 91190, Gif-sur-Yvette, France

[3]Centre national d'études spatiales (CNES), 18 av. Edouard Belin, 31400 Toulouse

[4]ACRI-ST, 260 Route du Pin Montard, BP 234, 06904 Sophia-Antipolis, France

Correspondence to: Hamadou Balde (hamadou.balde@lmd.ispl.fr)

Figure S1: The SIF3 and the active fluorometer instrument (LIF) setups at the top of the tower of the Fontainebleau-Barbeau forest ICOS site with dominant stand of sessile oak (height of about 25 m) and a hornbeam understory. SIF3 and LIF instruments were set above the canopy.
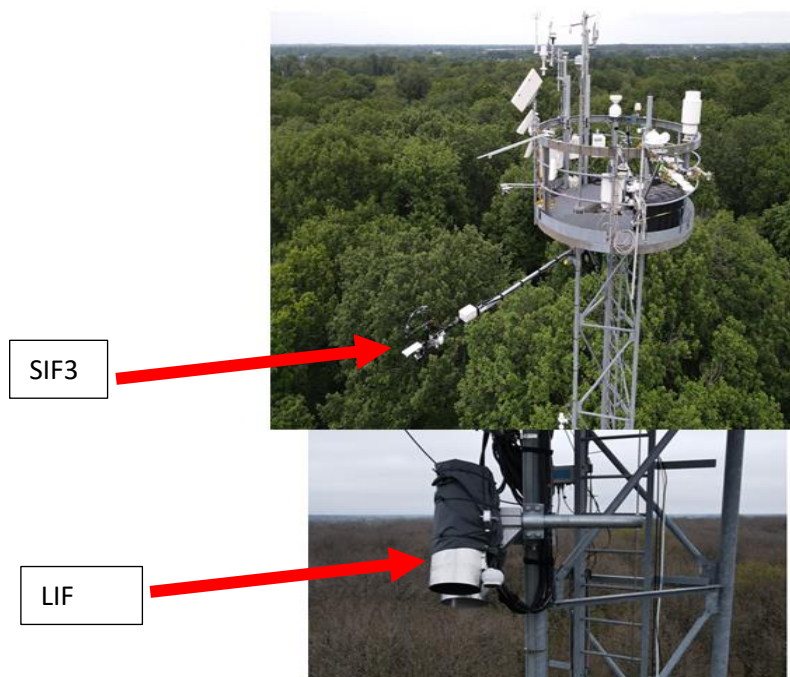
Figure S2. Correlation matrix between spectral reflectance bands of the canopy measured by SIF3. The data show strong relations between variables reflectance in the blue and the reflectance in the red, perhaps due to their strong light absorption, and between reflectance in the red-edge and reflectance in the near-infrared. Based on these observations all reflectance bands were not used for establishing the random forest models, but specific and low correlated bands were chosen in the blue, red, red-edge, and near-infrared regions as mentioned in Table 1.
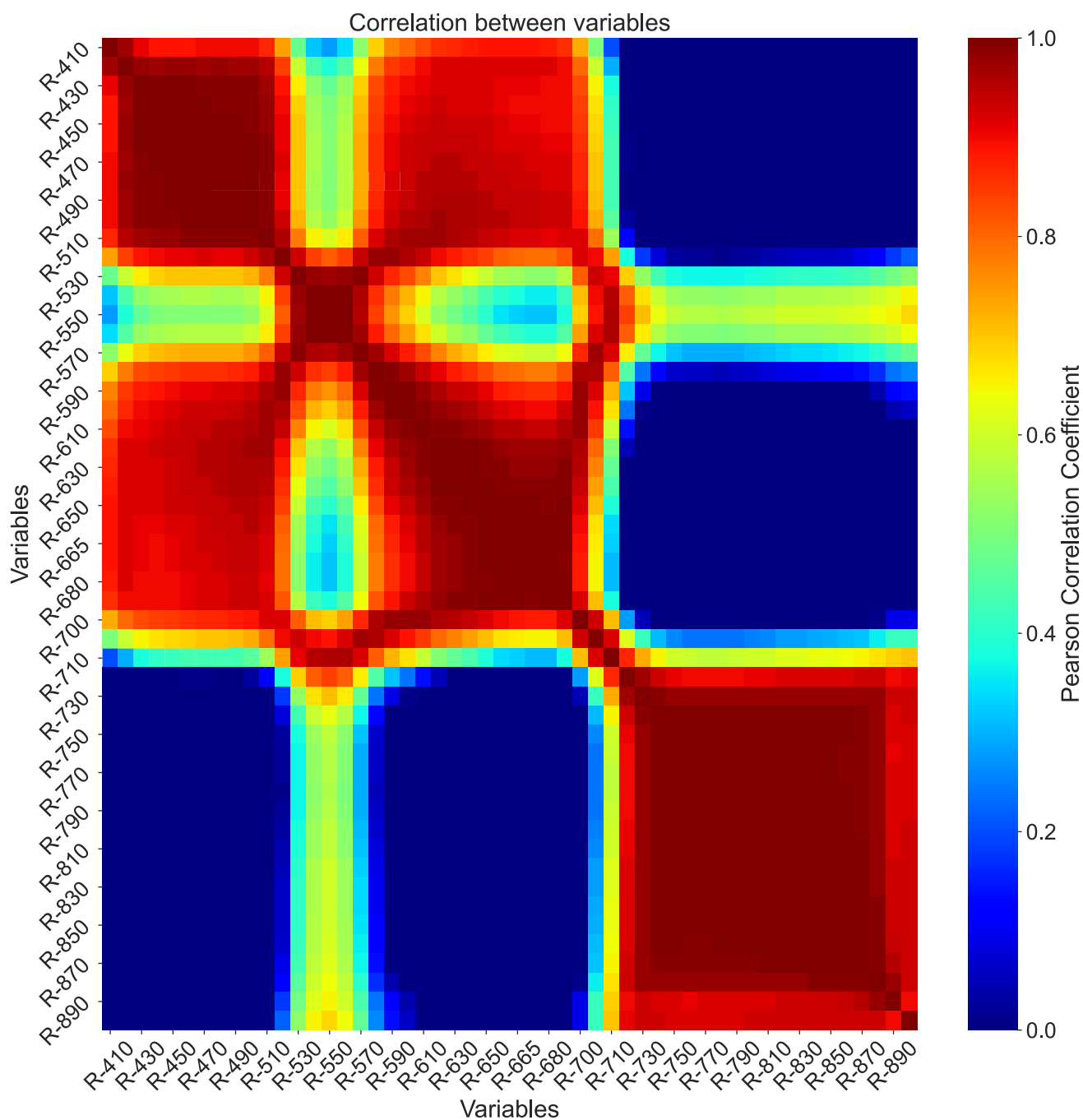
Figure S3. Correlation matrix between variables used for establishing the random forest models to predict $F_{yieldLIF}$ and Φ_k. The data show low correlations between variables than previously in Figure S2.
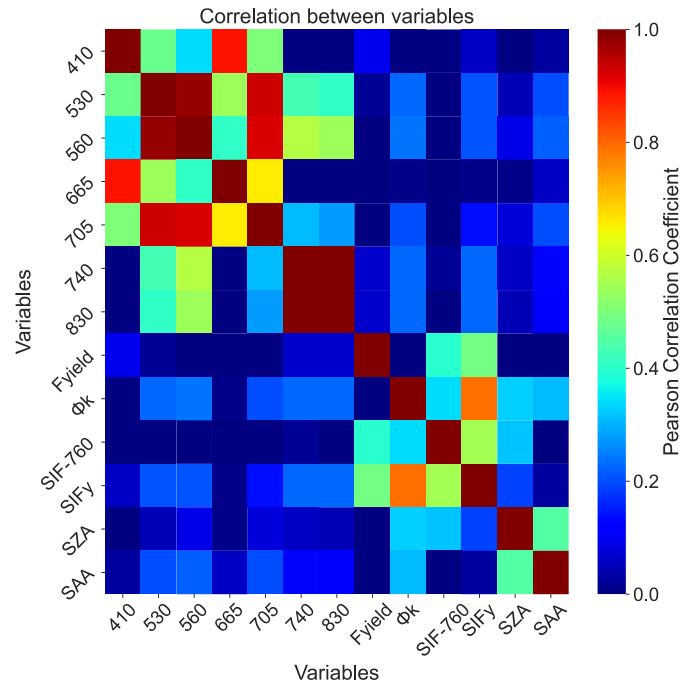
Figure S4. shows the coefficient of determination (R^2) of the relationship between SIF_y and $F_{yieldLIF}$ at hourly timescale. The results show improved correlations between SIF_y and $F_{yieldLIF}$ compared to the same analysis at the daily timescale (Figure 1).
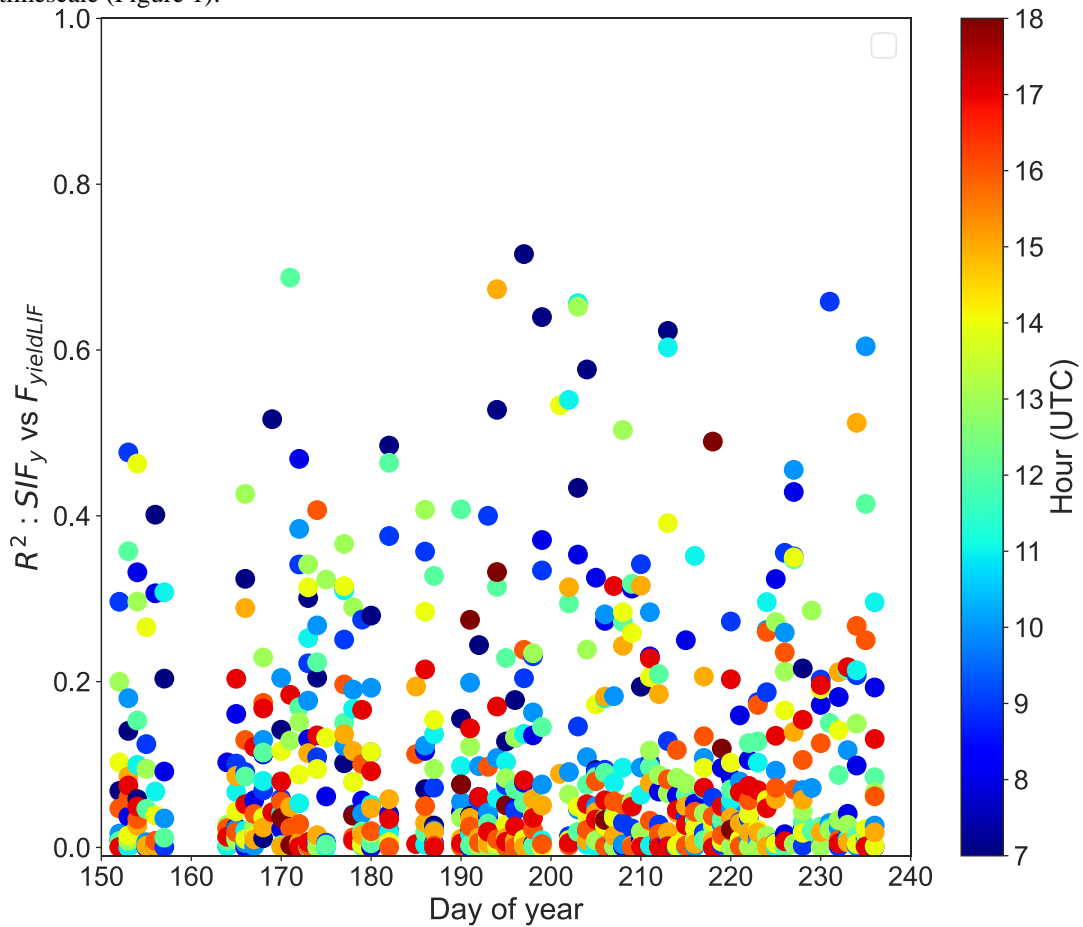
Figure S5. Depicts, during the clear sky day, the sun fraction of the field of view (FOV) of SIF3 and the sun elevation angle as observed in the temperate deciduous forest of Barbeau. The sun fraction was estimated from randomly taken RGB images of the FOV of SIF3 from July to September 2022. The sun elevation angle showed a well-known trend, with its peaks at noon and low values recorded in the morning and late afternoon hours. As expected, the sun fraction revealed its lowest values in the morning and late afternoon and peak values at noon. However, the decline observed in the sun fraction between 9 and around 12 am shows the impact of the vegetation structure and sun-canopy geometry.
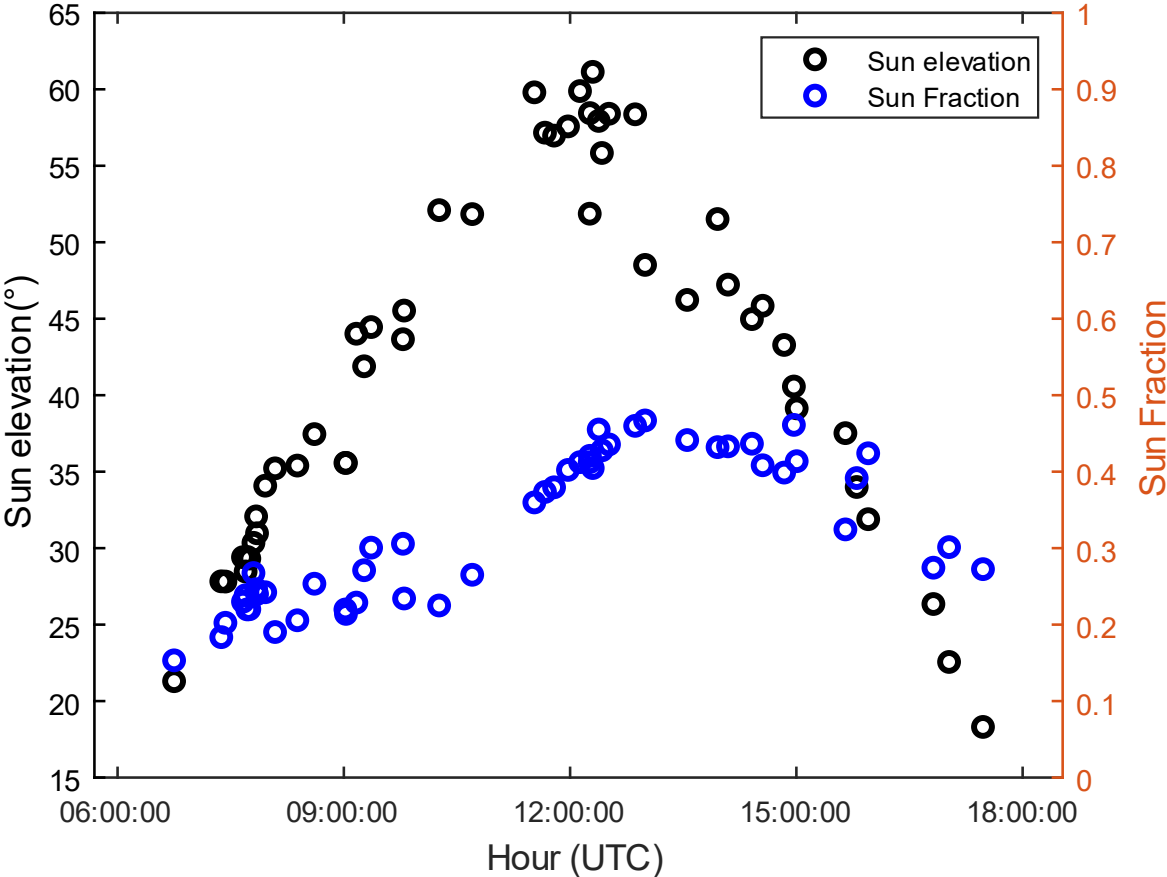
Figure S6. Depicts the shadows on the FOV of SIF3 observed on an RGB image captured on August, 6th, 2022, at 10:15 am (UTC). The cycle inside the left pink image showed the FOV of SIF3, which showed a shadowed canopy with only a small portion in the sun. The black and white image on the right showed the sun fraction (white), and the value of the sun fraction in the FOV of SIF3: 22.91%. This result indicates that more than 75% of the FOV of SIF3 was in the shade at this time of the day.
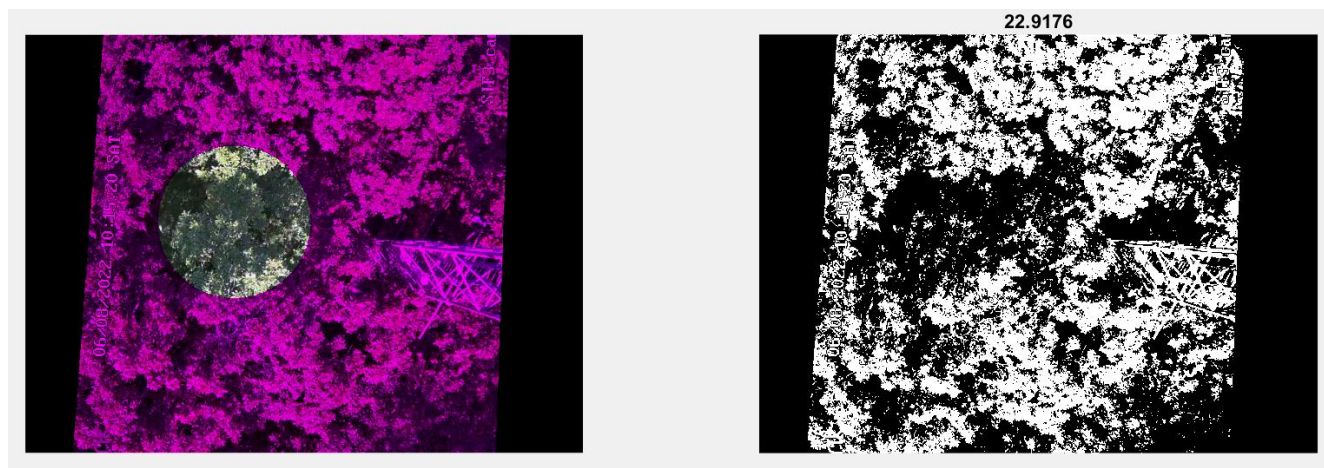
Figure S7. show the full diurnal cycles of $F_{yieldLIF}$ and photosynthetically active radiation (PAR) measured in June 17th, and 18th, 2022 in the deciduous oak temperate forest of Barbeau. The diurnal cycles of PAR showed that the two days were totally clear and sunny. It can be seen that when the first light beams hit the canopy a peak of $F_{yieldLIF}$ appeared, followed by a rapid decrease. It was the case for all recorded diurnal cycles. Further, $F_{yieldLIF}$ showed different diurnal patterns and responds to the incident PAR. For clear sunny days, like here, $F_{yieldLIF}$ significantly increased after sunrise and reached a maximum value when the PAR reached around 600 µmol photons m^{-2} s^{-1}. Afterwards, $F_{yieldLIF}$ drastically declined while PAR continued to increase until $F_{yieldLIF}$ reached a minimum value in the afternoon, from where $F_{yieldLIF}$ slightly increased during the night. This suggests that the decrease of $F_{yieldLIF}$ can be attributed to the activation of the non-photochemical quenching for the dissipation of the excess light energy induced by the high level of incoming radiation and hence avoiding the photodamage of the plant.
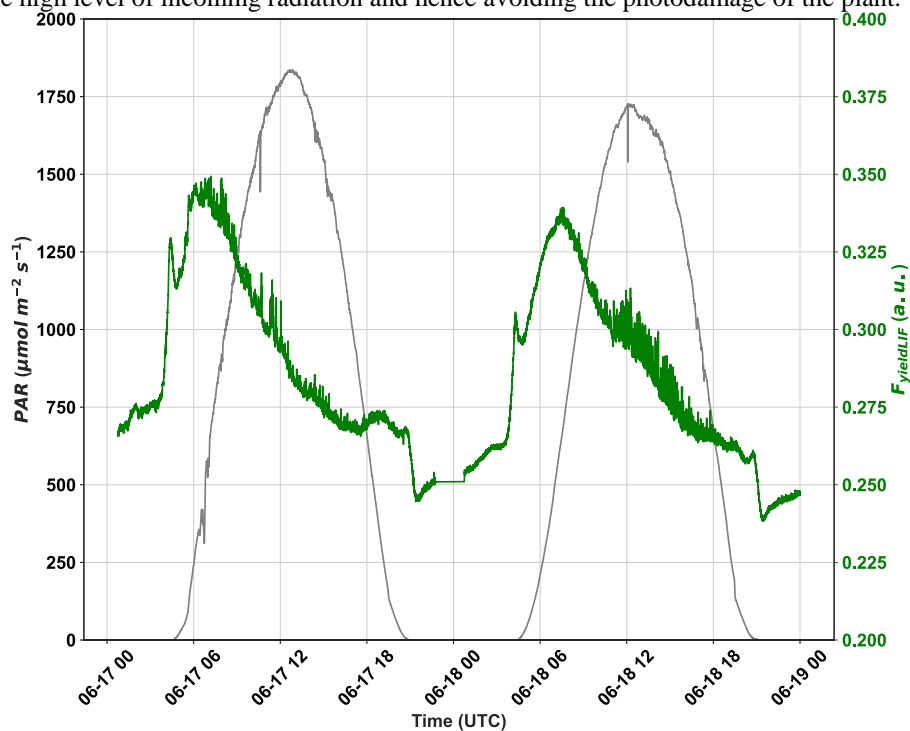
Figure S8. Figure S8a exhibits the inter-daily variations of the coefficient of determination (R^2) of the relationships between the reflectance at near-infrared band (R-NIR) and the Φ_k at instantaneous scale, as a function of the ratio between diffuse and total PAR. Figure S8b presents the seasonal relationships between daily means $\overline{R-NIR}$ and $\overline{\Phi_k}$ as a function of the ratio between diffuse and total PAR. And Figure S8c shows the seasonal dynamics in $\overline{R-NIR}$ and $\overline{\Phi_k}$. The shaded area indicates the 95% confidence interval. The asterisks stand for the statistical significance level (** = P≤0.01). The data show some strong correlations between R-NIR and Φ_k-760 at the daily timescale, while this relation was weak at the seasonal scale. These results are similar to the ones we obtained by studying the relationship between NIR_v and Φ_k, suggesting that spectral reflectance at the near-infrared had the same relation with Φ_k, and hence may be also relevant in capturing changes in fAPAR and f_{esc} at canopy scale.
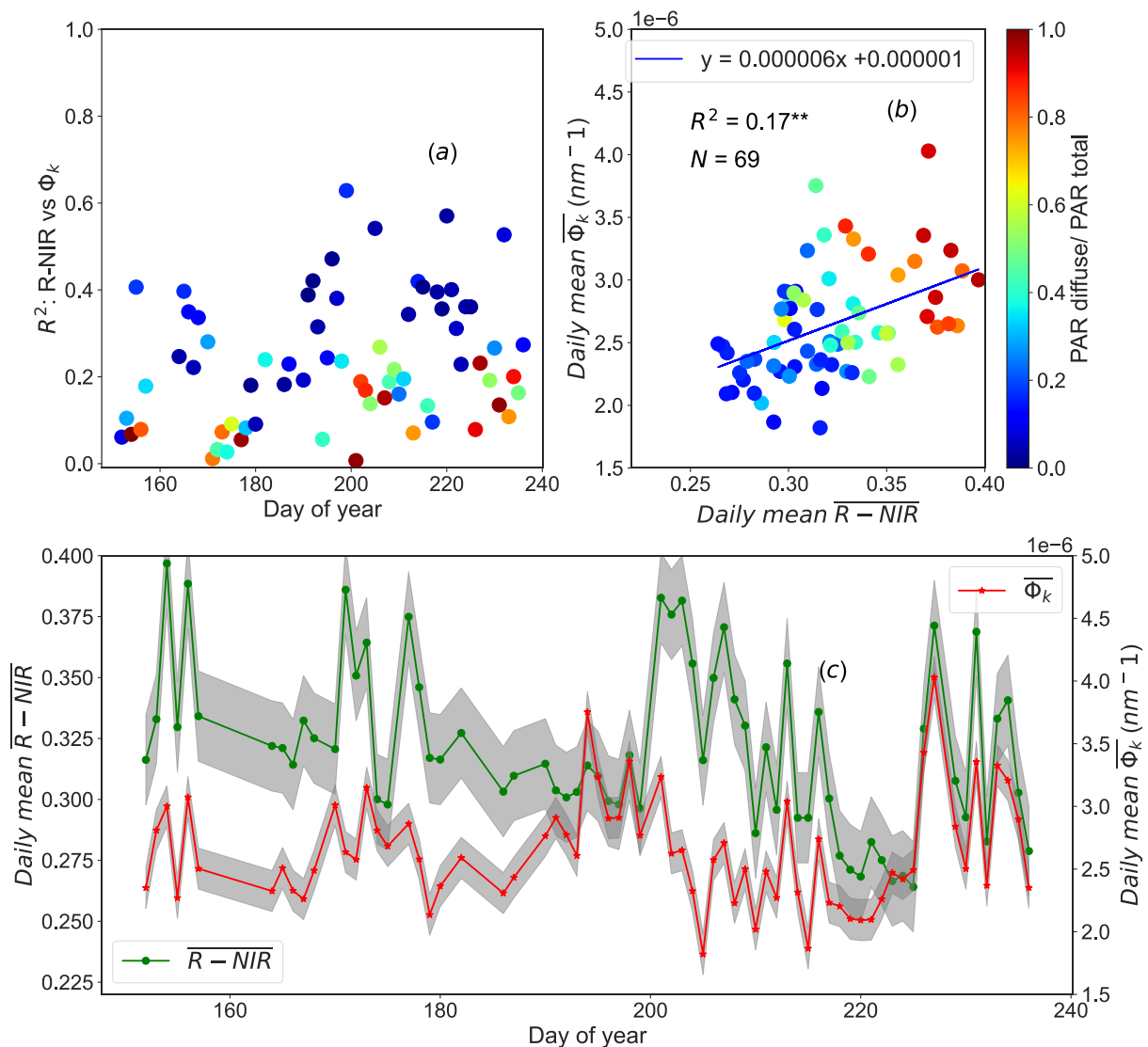
Figure S9. Random forest (RF) model outputs: Figure S9a presents FY-R model performance between observed and predicted $F_{yieldLIF}$, Figure S9b shows predictor importance estimates for FY-R model, Figure S9c represents FY-R-SIF model performance between observed and predicted $F_{yieldLIF}$, Figure S9d shows predictor importance estimates for FY-R-SIF model, and Figure S9e depicts FY-R-SIF$_y$ model performance between observed and predicted $F_{yieldLIF}$, Figure S9f shows predictor importance estimates for FY-R-SIF$_y$ model. *N* denotes the number of data points used for the RF model's testing, adj. R^2 represents the adjusted coefficient of determination of the relationship between the test dataset $F_{yieldLIF}$ and predicted $F_{yieldLIF}$, OOB R^2 is the model accuracy on the validation dataset (1/3 of the training set), and the RMSE is the root mean square error between observed $F_{yieldLIF}$ and RF model predicted $F_{yieldLIF}$. The dashed diagonal line depicts the 1:1 line. FY-*R* denotes $F_{yieldLIF}$ prediction using R, as inputs to predict $F_{yieldLIF}$; and FY-R-SIF includes *R* and SIF to predict $F_{yieldLIF}$, and FY-R-SIFy uses R and SIF$_y$ to estimate $F_{yieldLIF}$. The data revealed that adding SIF or SIF$_y$ as predictors did not improve much the model performance compared to FY-R model. But the predictor importance estimates showed that SIF and SIF$_y$ provide useful and impactful information in determining $F_{yieldLIF}$. This result indicates that even at high temporal resolution the contribution of SIF or SIF$_y$ is important compared to each R band individually, but the combined effect of R bands could mitigate or hide the use of SIF as vegetation physiological trait.