Supplement of Biogeosciences, 23, 7973–7991, 2026
https://doi.org/10.5194/bg-23-7973-2026-supplement
© Author(s) 2026. CC BY 4.0 License.



Biogeosciences Open Access EGU

Supplement of

Spatiotemporal dynamics and drivers of bare soil albedo in European croplands

Ke Yu et al.

Correspondence to: Yang Su (yang.su@ens.fr)

The copyright of individual parts of the supplement might differ from the article licence.

Table of contents

S1. The 10-fold cross-validation in the random forest model 2
S2. Supplementary Figures 3
Figure S1 3
Figure S2 4
Figure S3 5
Figure S4 6
Figure S5 7
Figure S6 8
Figure S7 9
Figure S8 10
S3. Supplementary Tables 11
Table S1 11
References 17

S1. The 10-fold cross-validation in the random forest model

The 10-fold cross-validation divided the training dataset into 10 subsets: 90% for the training model with specified hyperparameters and 10% for validation in each subset. The trained model then predicted the dependent variable using selected features in the 10% validation dataset, and both the dependent variable and the selected features were recorded. During the process, the importance of each feature is calculated based on how much each feature contributes to decreasing the impurity (measured by Gini index) in the tree nodes across all the trees in the forest. Higher scores indicate more important predictors (Strobl et al., 2008). After processing all 10 subsets, the overall hyperparameter-specific coefficient of determination (R^2) and root mean square error (RMSE) were calculated based on the aggregated dependent variable and the selected features from all subsets. The above procedure was repeated for all pairs of hyperparameters and the ones with the highest R^2 were selected to define the final random forest model. Note that the RMSE values for all hyperparameter pairs remain consistently around 0.02.

S2. Supplementary Figures

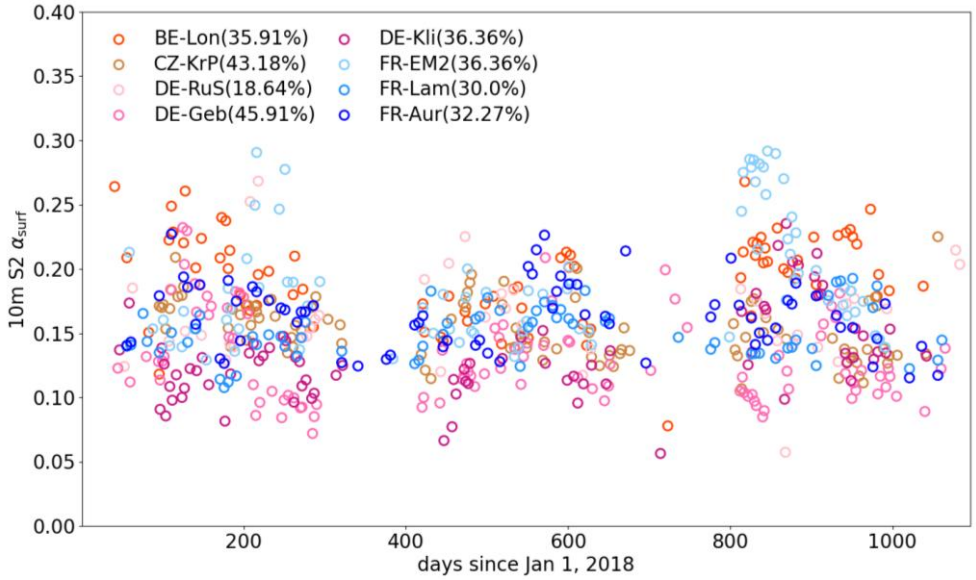


Figure S1 The dynamics of 10m 5-day Sentinel-2 surface albedo at 8 ICOS sites from 2018 to 2020. The numbers in legend brackets represent the percentage of available data (%) in three years.

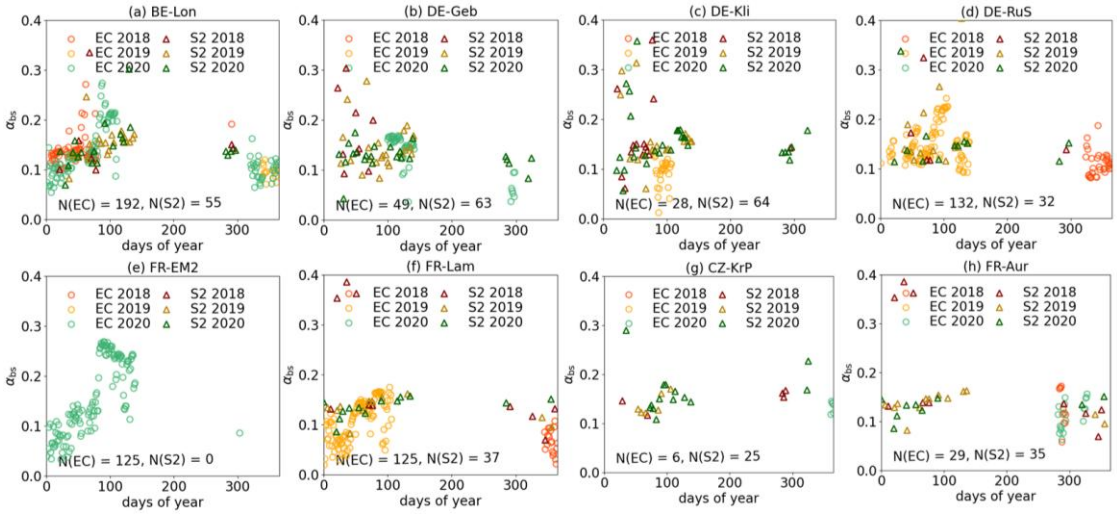
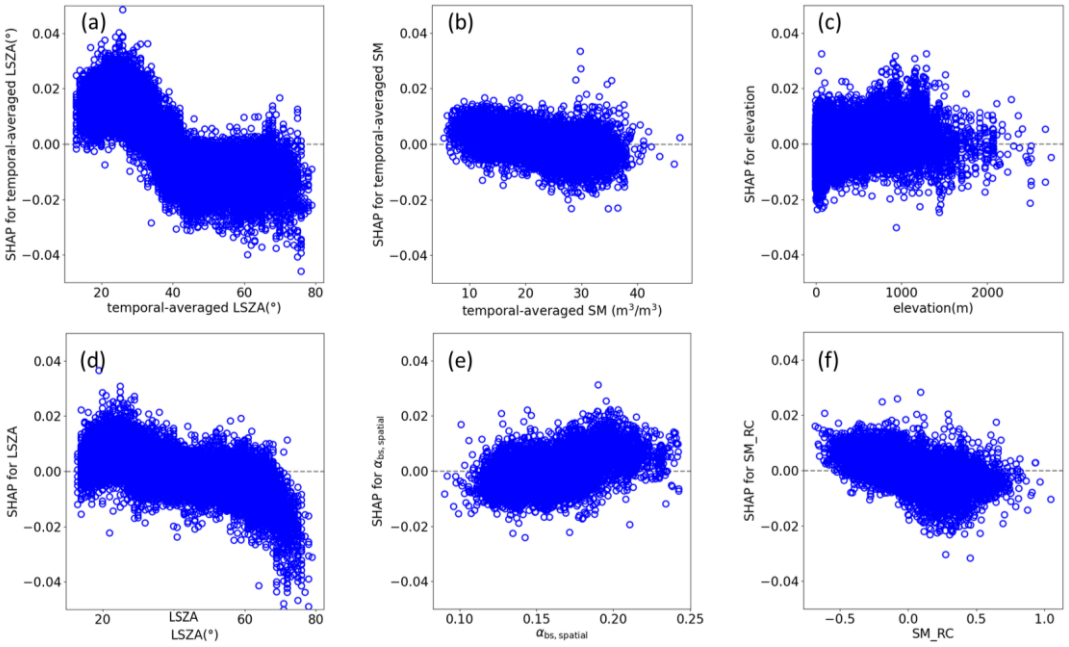


Figure S2 The bare soil albedo extracted from Sentinel-2 (S2) surface albedo product ($\alpha_{bs,\ S2}$, triangles) and computed by site eddy covariance (EC) measurements ($\alpha_{bs,\ site}$, circles) during the same bare soil periods detected by field photos instead of satellite vegetation indices at 8 ICOS sites from 2018 to 2020 (Table 3). 'EC' and 'S2' in each plot mean eddy covariance and Sentinel-2, respectively. "N" is the number of dots



45

46 **Figure S3** The SHapley Additive exPlanations (SHAP) Values of the top 3 selected features in
47 the spatial random forest model ((a)-(c)), and in the temporal random forest model ((d)-(f)).
48 'LSZA', 'SM' and 'SM_RC' represent the local solar zenith angle, soil moisture and relative
49 SM anomaly, respectively. $\alpha_{bs,\ spatial}$ is the spatial bare soil albedo distribution predicted by the
50 spatial random forest model. The blue dots represent the contribution of each feature to bare
51 soil albedo variation.

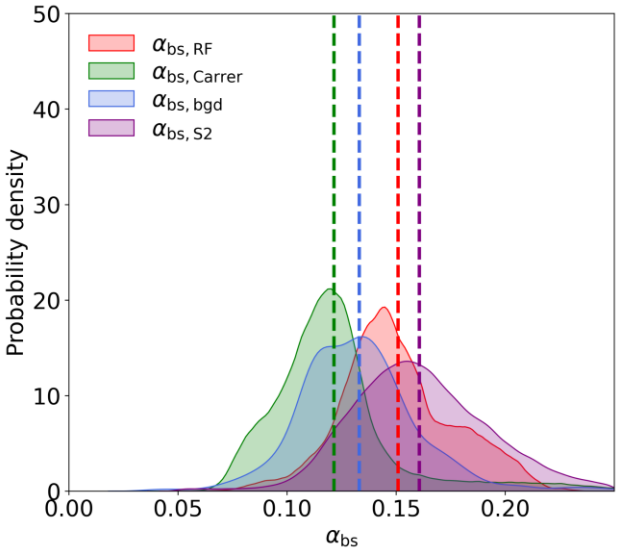


Figure S4 The probability density of different bare soil albedo (α_{bs}) products within the European cropland regions. The $\alpha_{bs, RF}$ is produced by the trained random forest models from 2018 to 2020, with 300 m spatial resolution and 5-day intervals. $\alpha_{bs, Carrer}$ and $\alpha_{bs, bgd}$ are derived from 8-day 1 km α_{bs} map (Carrer et al., 2014) and a monthly 50 km α_{bs} map (Pinty et al., 2011), respectively. $\alpha_{bs, S2}$ is the α_{bs} datasets trained in random forest models obtained from Sentinel-2 surface albedo observation (Lin et al., 2023) using vegetation indices (see Methods).

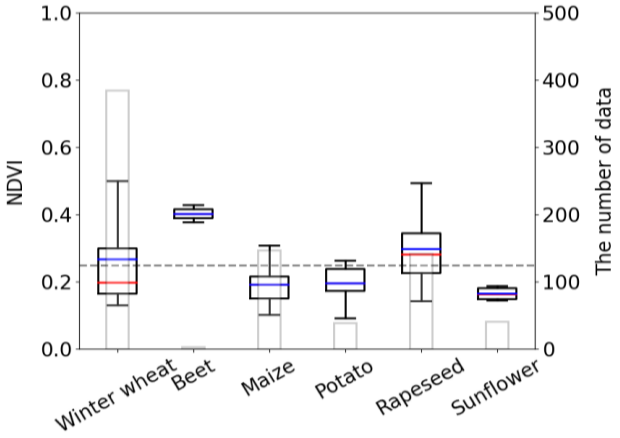


Figure S5 The NDVI distribution of crop residues at 6 European cropland sites from 2018 to 2020. The NDVI datasets were derived from 5-day Sentinel-2 reflectance product with a spatial resolution of 300 m on the Google Earth engine platform. The residue periods were identified by daily or weekly site photos. The red and blue lines in each figure represent the median and mean NDVI, respectively. The gray line represents the NDVI threshold of 0.25 used to extract bare soil periods and the corresponding Sentinel-2 bare soil albedo samples.

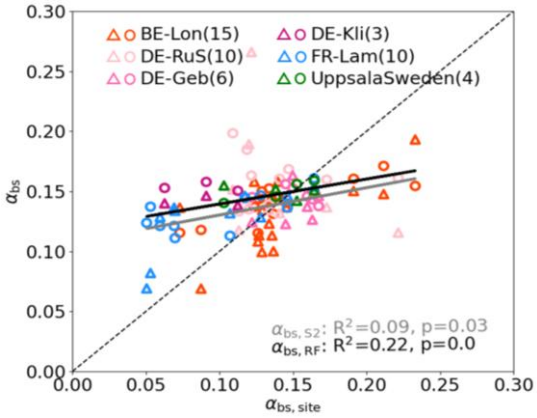


Figure S6 The linear relationship between $\alpha_{bs, RF}$ and $\alpha_{bs, site}$ (circles, black line), and between $\alpha_{bs, S2}$ and $\alpha_{bs, site}$ (triangles, gray line) at 4 sites during the same bare soil periods from 2018 to 2020. The bare soil periods are detected by field photos instead of satellite vegetation indices. The $\alpha_{bs, RF}$ is produced by the trained random forest models. The $\alpha_{bs, S2}$ is extracted from the Sentinel-2 surface albedo products, while $\alpha_{bs, site}$ is computed by site eddy covariance measurements (Table 3). Different colors represent different sites. The gray and black lines are the fitting trends between $\alpha_{bs, site}$ and $\alpha_{bs, S2}$, and between $\alpha_{bs, site}$ and $\alpha_{bs, RF}$, respectively. The numbers in the legend brackets represent the sample size. The black dotted line is 1:1 line. R^2 and p are the coefficient of determination and significance level.

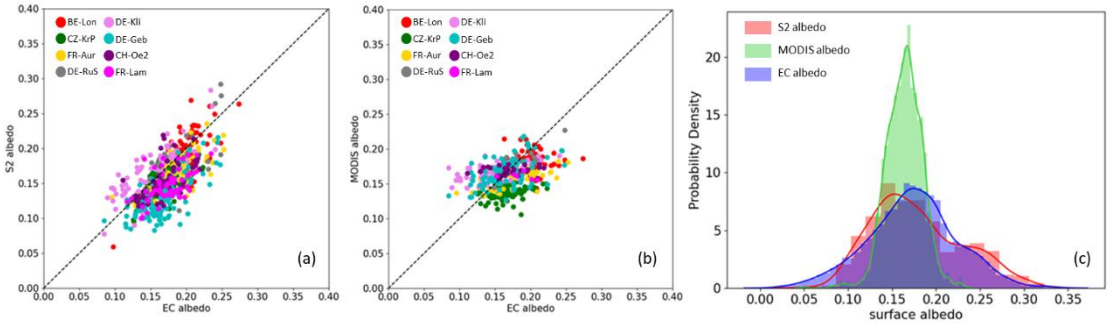
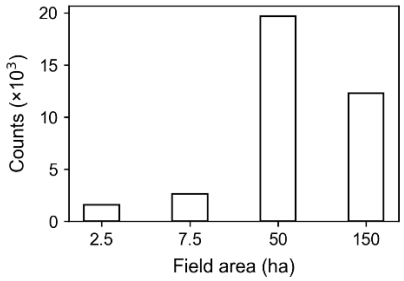


Figure S7 The comparison of surface albedo at 8 cropland sites during the same periods between (a) Sentinel-2 (300m, 5-day) and site measurements; (b) MODIS surface albedo product (500m, daily) and site measurements. (c) represent the probability density of the surface albedo from Sentinel-2 data, MODIS product and site measurements.



85

86 **Figure S8** The count of field sizes of LUCAS 2018 farms.

87

S3. Supplementary Tables

Table S1. Basic contextual information for the validation sites

Site	Latitude (°)	Longitude (°)	Mean annual air temperature (°C)	Mean annual incoming shortwave solar radiation (seasonal maximum and minimum range) (W m^{-2})	Mean annual rainfall (mm)	Mean annual soil water content (seasonal maximum and minimum range) (%)	Mean annual cloud cover (%)	Soil type	Silt, clay and sand contents (%)	Crop rotations	Averaged length of fallow periods
BE-Lon	50.55	4.75	10.2	110.2 (22.8-267.9)	742.6	29.5 (13.4-40.3) (3-layer soil with depth of 25 cm)	63.2	Luvisol	78.1, 14.7, 7.2	4-year rotation: sugar beet-winter wheat-potato-winter wheat	60-90 days
FR-Lam	43.50	1.24	13.4	157.0 (43.8-283.4)	677.0	32.8 (11.1-51.1) (3-layer soil with depth of 30 cm)	55.1	Calcaric Fluvisol	31.8, 52.9, 15.3	2-year rotation: winter wheat-irrigated maize	90-180 days

FR-Aur	43.55	1.11	13.1	161.3 (40.3-280.6)	652.1	27.4 (13.0-39.1) (3-layer soil with depth of 30 cm)	55.1	Calcaric Cambisol	42.5, 39.4, 18.1	4-year rotation: winter wheat - rapeseed - barley - sunflower	60-120 days
FR-EM2	49.87	3.02	11.2	137.3 (24.1-269.5)	616.7	26.6 (18.6-33.3) (3-layer soil with depth of 75 cm)	60.0	Luvisol	74.4, 18.7, 6.9	6-year rotation: spring pea-rapeseed-winter wheat - spring barley - maize (or sugar beet)-winter wheat	60-90 days

DE-RuS	50.87	6.45	11.2	130.0 (25.0-276.8)	548.9	20.6 (8.2-36.7) (1-layer soil with depth of 5 cm)	60.1	Luvisol	32.0, 53.0, 15.0	5-year rotation: sugar beet-winter wheat-winter barley-rapeseed-potatoes	60-120 days
DE-Geb	51.10	10.91	9.2	126.0 (20.7-275.3)	521.2	20.1 (8.4-31.0) (3-layer soil with depth of 10 cm)	60.9	Chernozem / Phaeozem	32.0, 53.0, 15.0	8-year crop rotation: winter wheat-potatoes-rape-winter barley-sugar beet-peppermint-durum-spring wheat	30-90 days

DE-Kli	50.89	13.52	8.1	130.6 (17.2-290.2)	782.5	25.3 (8.7-36.4) (1-layer soil with depth of 5 cm)	59.4	Cambisol	26.6, 46.9, 26.5	5-year crop rotation: rapeseed-winter wheat-forage maize-spring barley-winter barley-occasional catch crop	30-60 days
CZ-KrP	49.57	15.08	7.1	133.5 (26.4-291.0)	620.0	17.4 (8.2-24.2) (3-layer soil with depth of 20 cm)	57.7	Cambisol	44.0, 18.0, 38.0	winter wheat-spring barley-rapeseed-temporary leys	60-90 days

Ultuna in Uppsala in Sweden	59.82	17.65	8.4	119.0 (6.1-290.2)	519.0	-	62.6	Eutric Cambisol	40.0, 42.0, 18.0	winter barley-spring wheat-oats-rye	117 days (the year of 2019); 145 days (the year of 2020)

90

References

Carrer, D., Meurey, C., Ceamanos, X., Roujean, J. L., Calvet, J. C., and Liu, S. L.: Dynamic mapping of snow-free vegetation and bare soil albedos at global 1 km scale from 10-year analysis of MODIS satellite products, Remote Sens. Environ., 140, 420-432, doi: 10.1016/j.rse.2013.09.005, 2014.

Lin, X., Wu, S., Chen, B., Lin, Z., Yan, Z., Chen, X., Yin, G., You, D., Wen, J., Liu, Q., Xiao, Q., Liu, Q., and Lafortezza, R.: Estimating 10-m land surface albedo from Sentinel-2 satellite observations using a direct estimation approach with Google Earth Engine, ISPRS J. Photogramm. Remote Sens., 194, 1-20, doi: 10.1016/j.isprsjprs.2022.10.001, 2022.

Pinty, B., Andredakis, I., Clerici, M., Kaminski, T., Taberner, M., Verstraete, M. M., Gobron, N., Plummer, S., and Widlowski, J.-L.: Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties, J. Geophys. Res., 116, D09105, doi: 10.1029/2010JD015372, 2011.

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., and Zeileis, A.: Conditional variable importance for Random Forests, BMC Bioinformatics, 9, 307, doi: 10.1186/1471-2105-9-307, 2008.