Learning Evaporative Fraction with Memory

Wenli Zhao[1, 2, 3], Alexander J. Winkler[1, 3], Markus Reichstein[1, 3], Rene Orth[1, 4], Pierre Gentine[2]

[1] Max Planck Institute for Biogeochemistry, Jena, Germany.

[2] Columbia University, New York, USA.

[3] ELLIS Jena Unit, Jena, Germany.

[4] Faculty of Environment and Natural Resources, University of Freiburg, Germany

Corresponding author: Wenli Zhao (dr.wenli.zhao.pku@gmail.com)

Contents of this file

Introduction

This supplementary information contains two Texts.

Text S1.

Data Preprocessing Procedures of the Eddy-Covariance Dataset

The original sampling frequency of the data is half hourly. The data filter procedure can be summarized as follows: First, to reduce the noise in nighttime measurements, the original data is filtered with sensible heat flux > 5 W/m^2 and shortwave incoming radiation > 50 W/m^2 to select the daytime data only. Then, the original data is averaged to daily scale value (precipitation is calculated as the daily sum). Secondly, we only keep days with a fraction of good quality data > 0.8. The gaps in the time series for input features were interpolated using established methods(Reichstein et al., 2005; Vuichard and Papale, 2015). We also visually checked site by site to ensure that the signal-to-noise ratio is acceptable. Note that all the half hourly *LE* data from the eddy covariance sites is corrected to achieve energy balance closure using the Bowen ratio method(Twine et al., 2000). Due to the data limitation, only the shallowest soil moisture measurements were used for comparison with the evaporative fraction prediction dynamics during the dry-down periods.

Text S2.

Model Interpretations – Integrated Gradients (IG)

The integrated gradient is developed to interpret the trained models, which allows for obtaining the time-wise feature importance of the input features for each sample of the daily EF predictions(Jiang et al., 2022; Sundararajan et al., 2017). The IG method could unbox the LSTM-based machine learning model and trace back the specific contributions of the inputs and assign an importance score to each feature at each time prior to the predictions. A large positive IG score could indicate that the feature substantially increases the Evaporative Fraction predictions (e.g., that precipitation at the most proximate time may contribute more to current Evaporative Fraction projections than precipitation at an earlier time.). A large negative IG score indicates that the feature decreases the EF predictions. An IG score close to zero indicates little influence on the EF predictions. This way, our model could not only show the general feature importance but could also show the different feature importance at each time step prior the predictions. More specifically, it implies that temporal length of the input features will be considered for the EF predictions for different kinds of PFTs, in which hint the response of plant with different rooting depths during specific extreme events or environmental conditions, e.g., droughts with different severity level.

The IG score for the input feature *x* (e.g., the specific contribution of precipitation at the *i*th time step) is formulated as:

$$\emptyset_i(x) = (x_i - x_i') \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha$$

Where $\frac{\partial F(x'+\alpha\times(x-x'))}{\partial x_i}$ denotes the local gradient of the network *F* at a point interpolated from a baseline input (x' $when$ $\alpha = 0$), which is meant to represent the "absence" of the feature input, to the target input (x, when $\alpha = 1$). Note that the IG value is completeness and add up to the difference between the output of *F* at the target input *x* and the baseline input x', i.e., $\sum_i \quad \emptyset_i(x) = F(x) - F(x')$. Therefore, the model output can be decomposed into the sum of features' individual contributions, and it enables us to examine the contribution of a group of features by summing up their individual IG scores.

Due to the integral, the original definition of IG is incalculable. Therefore, the implementation of the method in practice uses approximated value by replacing the integral with the summation:

$$\emptyset_i^{approx}(x) = (x_i - x_i') \times \sum_{k=1}^{m} \frac{\partial F(x' + \frac{k}{m} \times (x - x'))}{\partial x_i} \times \frac{1}{m}$$

where m defines a number of interpolation steps.

In this study, we use the library captum (https://github.com/pytorch/captum) to obtain the IG scores for each feature at each time step.

References

Jiang, S., Bevacqua, E., and Zscheischler, J.: River flooding mechanisms and their changes in Europe revealed by explainable machine learning, Hydrol Earth Syst Sc, 26, 6339–6359, https://doi.org/10.5194/hess-26-6339-2022, 2022.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biol, 11, 1424–1439, https://doi.org/10.1111/j.1365-2486.2005.001002.x, 2005.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, http://arxiv.org/abs/1703.01365, 12 June 2017.

Twine, T. E., Kustas, W., Norman, J., Cook, D., Houser, Pr., Meyers, T., Prueger, J., Starks, P., and Wesely, M.: Correcting eddy-covariance flux underestimates over a grassland, Agricultural and forest meteorology, 103, 279–300, 2000.

Vuichard, N. and Papale, D.: Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis, Earth System Science Data, 7, 157–171, 2015.

Learning Evaporative Fraction with Memory

Wenli Zhao[1, 2, 3], Alexander J. Winkler[1, 3], Markus Reichstein[1, 3], Rene Orth[1, 4], Pierre Gentine[2]

[1] Max Planck Institute for Biogeochemistry, Jena, Germany.

[2] Columbia University, New York, USA.

[3] ELLIS Jena Unit, Jena, Germany.

[4] Faculty of Environment and Natural Resources, University of Freiburg, Germany

Corresponding author: Wenli Zhao (dr.wenli.zhao.pku@gmail.com)

Contents of this file

Figure S1 to S4

Introduction

This supplementary information contains one Text and Four Figures.

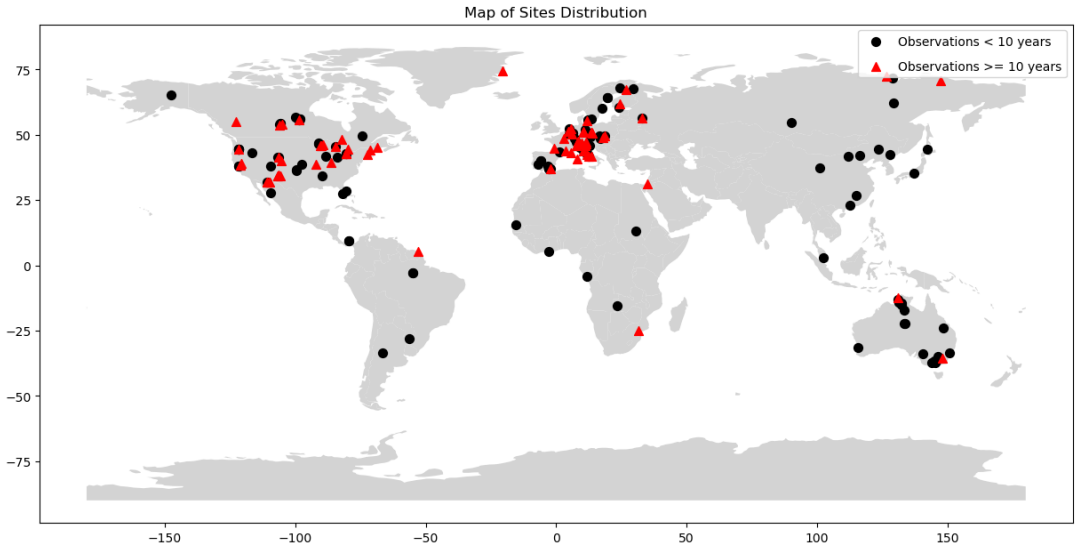

Figure S1. Map of Sites Distribution of the combined Fluxnet2015, AmeriFlux, ICOS networks. The sites which have more than 10 years' observations are used for model training.

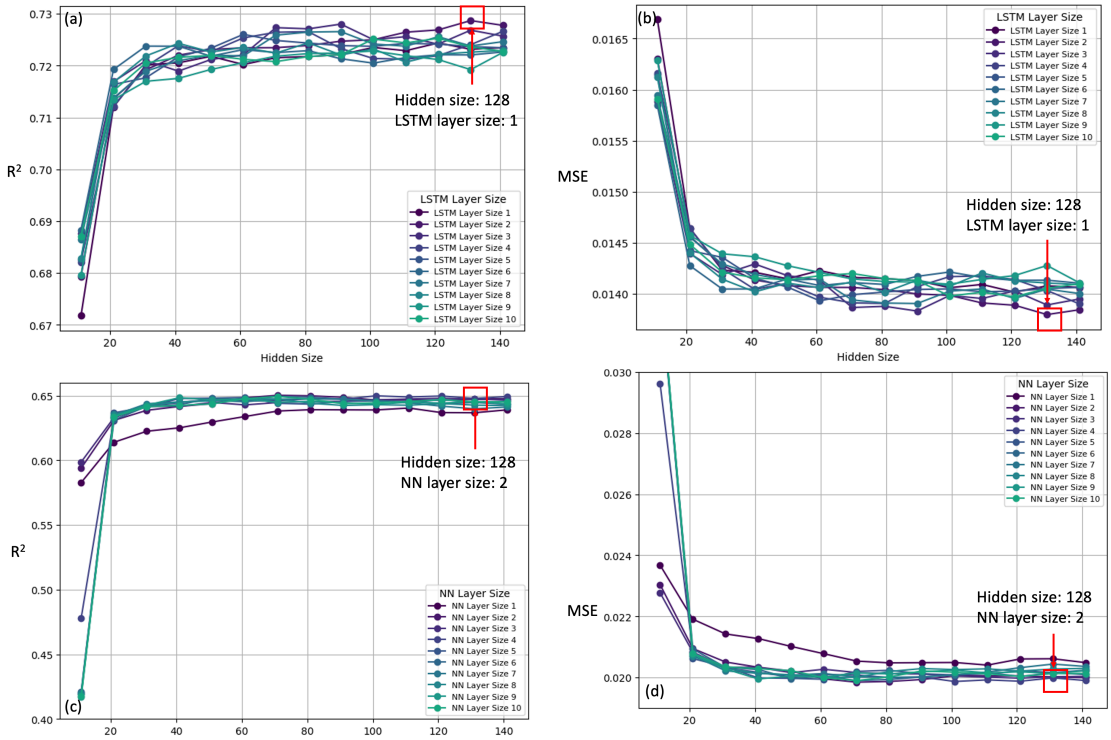

Figure S2. Optimal architecture of machine learning model. Note we sampled the entire dataset to speed up identifying the optimal model structure.

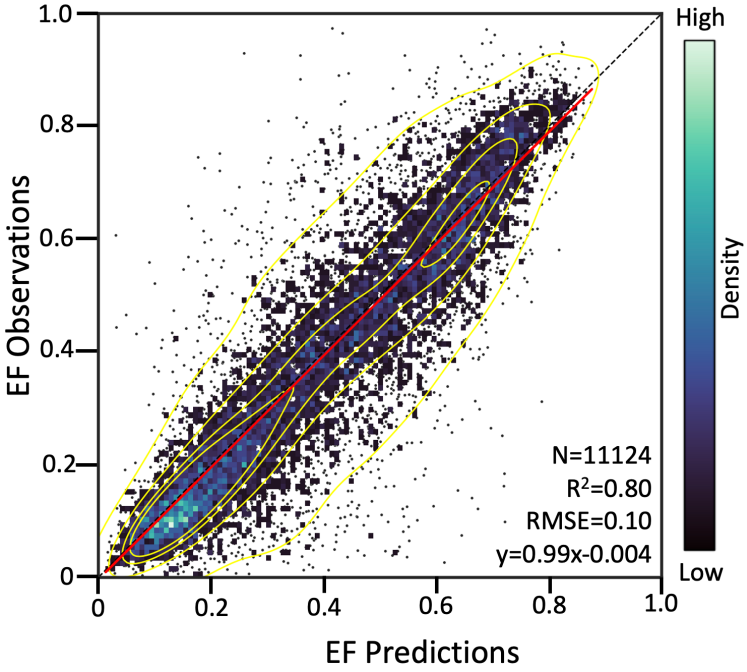

Figure S3. Model performance. The model performance with 365 days' time series data as input. The EF predictions shown here are based on the ensemble mean EF predictions of 20 repeated models. The density of scatter points is represented using shading colors. The diagonal black dashed line depicts the 1:1 line and the red solid line depicts the linear regression line. Note that N, R^2, and RMSE represent the number of points, coefficient of determination, and root mean square error, respectively. All the metrics are calculated using the test set data. The dataset is split into training, validation, and test sets based on time periods: the last year of each site is used for testing, the second-to-last year for validation, and the remaining data for training.

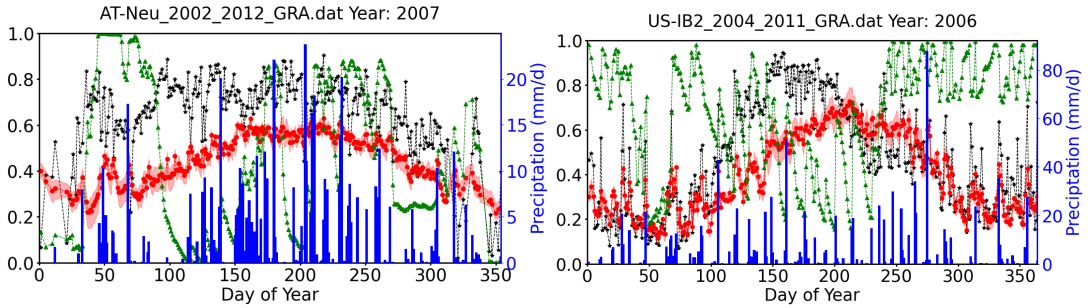

Figure S4. Grassland sites receiving sufficient precipitation. Blue bars show the observed daily sum precipitation (P), black curves show EF observations, red curves show EF predictions, green curves show soil moisture percentiles of the shallowest soil depth. The x-axis represents the day of year (DOY) of the whole year. The daily volumetric water content values are converted into percentiles, indicating the fraction of daily values lower than a specific value. Thus, the 100th percentile (or a percentile value of 1) represents the wettest soil conditions observed at a specific site throughout the study period, and the 0 percentile signifies the driest soil conditions. The memory length is set as 365 days for the machine learning model. The EF predictions here are using the ensemble mean EF predictions of 20 models with different initializations. Shaded areas represent regions of predictions uncertainty in the 25%-75% quartiles of these 20 repeat training models.

Learning Evaporative Fraction with Memory

Wenli Zhao[1, 2, 3], Alexander J. Winkler[1, 3], Markus Reichstein[1, 3], Rene Orth[1, 4], Pierre Gentine[2]

[1] Max Planck Institute for Biogeochemistry, Jena, Germany.

[2] Columbia University, New York, USA.

[3] ELLIS Jena Unit, Jena, Germany.

[4] Faculty of Environment and Natural Resources, University of Freiburg, Germany

Corresponding author: Wenli Zhao (dr.wenli.zhao.pku@gmail.com)

Contents of this file

Introduction

This supplementary information contains one Table.

Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
AR-SLu	MF	-33,46	-66,46	2009-2011	https://doi.org/10.18140/FLX/1440191
AR-Vir	ENF	-28,24	-56,19	2009-2012	https://doi.org/10.18140/FLX/1440192
AT-Neu	GRA	47,12	11,32	2002-2012	https://doi.org/10.18140/FLX/1440121
AU-ASM	SAV	-22,28	133,25	2010-2014	https://doi.org/10.18140/FLX/1440194
AU-Ade	WSA	-13,08	131,12	2007-2009	https://doi.org/10.18140/FLX/1440193
AU-Cpr	SAV	-34,00	140,59	2010-2014	https://doi.org/10.18140/FLX/1440195
AU-Cum	EBF	-33,62	150,72	2012-2014	https://doi.org/10.18140/FLX/1440196
AU-DaP	GRA	-14,06	131,32	2007-2013	https://doi.org/10.18140/FLX/1440123
AU-DaS	SAV	-14,16	131,39	2008-2014	https://doi.org/10.18140/FLX/1440122
AU-Dry	SAV	-15,26	132,37	2008-2014	https://doi.org/10.18140/FLX/1440197
AU-Emr	GRA	-23,86	148,47	2011-2013	https://doi.org/10.18140/FLX/1440198
AU-Gin	WSA	-31,38	115,71	2011-2014	https://doi.org/10.18140/FLX/1440199
AU-How	WSA	-12,49	131,15	2001-2014	https://doi.org/10.18140/FLX/1440125
AU-RDF	WSA	-14,56	132,48	2011-2013	https://doi.org/10.18140/FLX/1440201
AU-Rig	GRA	-36,65	145,58	2011-2014	https://doi.org/10.18140/FLX/1440202
AU-Stp	GRA	-17,15	133,35	2008-2014	https://doi.org/10.18140/FLX/1440204
AU-TTE	GRA	-22,29	133,64	2012-2014	https://doi.org/10.18140/FLX/1440205
AU-Tum	EBF	-35,66	148,15	2001-2014	https://doi.org/10.18140/FLX/1440126
AU-Wac	EBF	-37,43	145,19	2005-2008	https://doi.org/10.18140/FLX/1440127
AU-Whr	EBF	-36,67	145,03	2011-2014	https://doi.org/10.18140/FLX/1440206
AU-Wom	EBF	-37,42	144,09	2010-2014	https://doi.org/10.18140/FLX/1440207
AU-Ync	GRA	-34,99	146,29	2012-2014	https://doi.org/10.18140/FLX/1440208
BE-Bra	MF	51,31	4,52	1996-2020	https://doi.org/10.18160/2G60-ZHAK
BE-Dor	GRA	50,31	4,97	2011-2020	https://doi.org/10.18160/2G60-ZHAK
BE-Lon	CRO	50,55	4,75	2004-2020	https://doi.org/10.18160/2G60-ZHAK
BE-Maa	CSH	50,98	5,63	2016-2020	https://doi.org/10.18160/2G60-ZHAK
BE-Vie	MF	50,31	6,00	1996-2020	https://doi.org/10.18160/2G60-ZHAK
BR-Sa1	EBF	-2,86	-54,96	2002-2011	https://doi.org/10.18140/FLX/1440032
BR-Sa3	EBF	-3,02	-54,97	2000-2004	https://doi.org/10.18140/FLX/1440033
CA-Cbo	DBF	44,32	-79,93	1994-2020	https://doi.org/10.17190/AMF/1854365
CA-ER1	CRO	43,64	-80,41	2015-2020	https://doi.org/10.17190/AMF/1832154
CA-Gro	MF	48,22	-82,16	2003-2014	https://doi.org/10.18140/FLX/1440034
CA-LP1	ENF	55,11	-122,84	2007-2020	https://doi.org/10.17190/AMF/1832155
CA-Man	ENF	55,88	-98,48	1994-2008	https://doi.org/10.18140/FLX/1440035
CA-NS1	ENF	55,88	-98,48	2001-2005	https://doi.org/10.18140/FLX/1440036
CA-NS2	ENF	55,91	-98,52	2001-2005	https://doi.org/10.18140/FLX/1440037
CA-NS3	ENF	55,91	-98,38	2001-2005	https://doi.org/10.18140/FLX/1440038
CA-NS4	ENF	55,91	-98,38	2002-2005	https://doi.org/10.18140/FLX/1440039
CA-NS5	ENF	55,86	-98,49	2001-2005	https://doi.org/10.18140/FLX/1440040
CA-NS6	OSH	55,92	-98,96	2001-2005	https://doi.org/10.18140/FLX/1440041
CA-NS7	OSH	56,64	-99,95	2002-2005	https://doi.org/10.18140/FLX/1440042
CA-Oas	DBF	53,63	-106,20	1996-2010	https://doi.org/10.18140/FLX/1440043

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)

Continued Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
CA-Obs	ENF	53,99	-105,12	1997-2010	https://doi.org/10.18140/FLX/1440044
CA-Qfo	ENF	49,69	-74,34	2003-2010	https://doi.org/10.18140/FLX/1440045
CA-SF1	ENF	54,49	-105,82	2003-2006	https://doi.org/10.18140/FLX/1440046
CA-SF2	ENF	54,25	-105,88	2001-2005	https://doi.org/10.18140/FLX/1440047
CA-SF3	OSH	54,09	-106,01	2001-2006	https://doi.org/10.18140/FLX/1440048
CA-TP1	ENF	42,66	-80,56	2002-2014	https://doi.org/10.18140/FLX/1440050
CA-TP2	ENF	42,77	-80,46	2002-2007	https://doi.org/10.18140/FLX/1440051
CA-TP3	ENF	42,71	-80,35	2002-2014	https://doi.org/10.18140/FLX/1440052
CA-TP4	ENF	42,71	-80,36	2002-2014	https://doi.org/10.18140/FLX/1440053
CA-TPD	DBF	42,64	-80,56	2012-2014	https://doi.org/10.18140/FLX/1440112
CG-Tch	SAV	-4,29	11,66	2006-2009	https://doi.org/10.18140/FLX/1440142
CH-Aws	GRA	46,58	9,79	2006-2020	https://doi.org/10.18160/2G60-ZHAK
CH-Cha	GRA	47,21	8,41	2005-2020	https://doi.org/10.18160/2G60-ZHAK
CH-Dav	ENF	46,82	9,86	1997-2020	https://doi.org/10.18160/2G60-ZHAK
CH-Fru	GRA	47,12	8,54	2005-2020	https://doi.org/10.18160/2G60-ZHAK
CH-Lae	MF	47,48	8,37	2004-2020	https://doi.org/10.18160/2G60-ZHAK
CH-Oe1	GRA	47,29	7,73	2002-2008	https://doi.org/10.18140/FLX/1440135
CH-Oe2	CRO	47,29	7,73	2004-2020	https://doi.org/10.18140/FLX/1440136
CN-Cha	MF	42,40	128,10	2003-2005	https://doi.org/10.18140/FLX/1440137
CN-Cng	GRA	44,59	123,51	2007-2010	https://doi.org/10.18140/FLX/1440209
CN-Din	EBF	23,17	112,54	2003-2005	https://doi.org/10.18140/FLX/1440139
CN-Du2	GRA	42,05	116,28	2006-2008	https://doi.org/10.18140/FLX/1440140
CN-HaM	GRA	37,37	101,18	2002-2004	https://doi.org/10.18140/FLX/1440190
CN-Qia	ENF	26,74	115,06	2003-2005	https://doi.org/10.18140/FLX/1440141
CN-Sw2	GRA	41,79	111,90	2010-2012	https://doi.org/10.18140/FLX/1440212
CZ-BK1	ENF	49,50	18,54	2004-2020	https://doi.org/10.18140/FLX/1440143
CZ-BK2	GRA	49,49	18,54	2004-2012	https://doi.org/10.18140/FLX/1440144
CZ-KrP	CRO	49,57	15,08	2014-2020	https://doi.org/10.18160/2G60-ZHAK
CZ-Lnz	MF	48,68	16,95	2015-2020	https://doi.org/10.18160/2G60-ZHAK
CZ-RAJ	ENF	49,44	16,70	2012-2020	https://doi.org/10.18160/2G60-ZHAK
CZ-Stn	DBF	49,04	17,97	2010-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Geb	CRO	51,10	10,91	2001-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Gri	GRA	50,95	13,51	2004-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Hai	DBF	51,08	10,45	2000-2020	https://doi.org/10.18160/2G60-ZHAK
DE-HoH	DBF	52,09	11,22	2015-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Hzd	DBF	50,96	13,49	2010-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Kli	CRO	50,89	13,52	2004-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Lkb	ENF	49,10	13,30	2009-2013	https://doi.org/10.18140/FLX/1440214
DE-Lnf	DBF	51,33	10,37	2002-2012	https://doi.org/10.18140/FLX/1440150
DE-Obe	ENF	50,78	13,72	2008-2020	https://doi.org/10.18160/2G60-ZHAK
DE-RuR	GRA	50,62	6,30	2011-2020	https://doi.org/10.18160/2G60-ZHAK
DE-RuS	CRO	50,87	6,45	2011-2020	https://doi.org/10.18160/2G60-ZHAK

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)

Continued Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
DE-RuW	ENF	50,50	6,33	2012-2020	https://doi.org/10.18160/2G60-ZHAK
DE-Seh	CRO	50,87	6,45	2007-2010	https://doi.org/10.18140/FLX/1440217
DE-Tha	ENF	50,96	13,57	1996-2020	https://doi.org/10.18160/2G60-ZHAK
DK-Eng	GRA	55,69	12,19	2005-2008	https://doi.org/10.18140/FLX/1440153
DK-Sor	DBF	55,49	11,64	1996-2014	https://doi.org/10.18140/FLX/1440155
ES-Abr	SAV	38,70	-6,79	2015-2020	https://doi.org/10.18160/2G60-ZHAK
ES-Agu	OSH	36,94	-2,03	2006-2020	https://doi.org/10.18160/2G60-ZHAK
ES-Amo	OSH	36,83	-2,25	2007-2012	https://doi.org/10.18140/FLX/1440156
ES-Cnd	EBF	37,91	-3,23	2014-2020	https://doi.org/10.18160/2G60-ZHAK
ES-LJu	OSH	36,93	-2,75	2004-2013	https://doi.org/10.18140/FLX/1440157
ES-LM1	SAV	39,94	-5,78	2014-2020	https://doi.org/10.18160/2G60-ZHAK
ES-LM2	SAV	39,93	-5,78	2014-2020	https://doi.org/10.18160/2G60-ZHAK
ES-LgS	OSH	37,10	-2,97	2007-2009	https://doi.org/10.18140/FLX/1440225
FI-Hyy	ENF	61,85	24,30	1996-2020	https://doi.org/10.18160/2G60-ZHAK
FI-Jok	CRO	60,90	23,51	2000-2003	https://doi.org/10.18140/FLX/1440159
FI-Ken	ENF	67,99	24,24	2018-2020	https://doi.org/10.18160/2G60-ZHAK
FI-Let	ENF	60,64	23,96	2009-2012	https://doi.org/10.18140/FLX/1440227
FI-Qvd	CRO	60,30	22,39	2018-2020	https://doi.org/10.18160/2G60-ZHAK
FI-Sod	ENF	67,36	26,64	2001-2014	https://doi.org/10.18140/FLX/1440160
FI-Var	ENF	67,75	29,61	2016-2020	https://doi.org/10.18160/2G60-ZHAK
FR-Aur	CRO	43,55	1,11	2005-2020	https://doi.org/10.18160/2G60-ZHAK
FR-FBn	MF	43,24	5,68	2008-2020	https://doi.org/10.18160/2G60-ZHAK
FR-Fon	DBF	48,48	2,78	2005-2020	https://doi.org/10.18160/2G60-ZHAK
FR-Gri	CRO	48,84	1,95	2004-2020	https://doi.org/10.18160/2G60-ZHAK
FR-Hes	DBF	48,67	7,06	2014-2020	https://doi.org/10.18160/2G60-ZHAK
FR-LBr	ENF	44,72	-0,77	1996-2008	https://doi.org/10.18140/FLX/1440163
FR-Lam	CRO	43,49	1,24	2005-2020	https://doi.org/10.18160/2G60-ZHAK
FR-Pue	EBF	43,74	3,60	2000-2014	https://doi.org/10.18140/FLX/1440164
FR-Tou	GRA	43,57	1,37	2018-2020	https://doi.org/10.18160/2G60-ZHAK
GF-Guy	EBF	5,28	-52,93	2004-2020	https://doi.org/10.18160/2G60-ZHAK
GH-Ank	EBF	5,27	-2,69	2011-2014	https://doi.org/10.18140/FLX/1440229
GL-ZaH	GRA	74,47	-20,55	2000-2014	https://doi.org/10.18140/FLX/1440224
IL-Yat	ENF	31,34	35,05	2000-2020	https://doi.org/10.18160/2G60-ZHAK
IT-BCi	CRO	40,52	14,96	2004-2014	https://doi.org/10.18140/FLX/1440166
IT-CA1	DBF	42,38	12,03	2011-2014	https://doi.org/10.18140/FLX/1440230
IT-CA2	CRO	42,38	12,03	2011-2014	https://doi.org/10.18140/FLX/1440231
IT-CA3	DBF	42,38	12,02	2011-2014	https://doi.org/10.18140/FLX/1440232
IT-Col	DBF	41,85	13,59	1996-2014	https://doi.org/10.18140/FLX/1440167
IT-Cp2	EBF	41,70	12,36	2012-2020	https://doi.org/10.18160/2G60-ZHAK
IT-Cpz	EBF	41,71	12,38	1997-2009	https://doi.org/10.18140/FLX/1440233
IT-La2	ENF	45,95	11,29	2000-2002	https://doi.org/10.18140/FLX/1440235
IT-Lav	ENF	45,96	11,28	2003-2020	https://doi.org/10.18160/2G60-ZHAK

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)

Continued Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
IT-Lsn	OSH	45,74	12,75	2016-2020	https://doi.org/10.18160/2G60-ZHAK
IT-MBo	GRA	46,01	11,05	2003-2020	https://doi.org/10.18160/2G60-ZHAK
IT-Noe	CSH	40,61	8,15	2004-2014	https://doi.org/10.18140/FLX/1440171
IT-PT1	DBF	45,20	9,06	2002-2004	https://doi.org/10.18140/FLX/1440172
IT-Ren	ENF	46,59	11,43	1999-2020	https://doi.org/10.18160/2G60-ZHAK
IT-Ro1	DBF	42,41	11,93	2000-2008	https://doi.org/10.18140/FLX/1440174
IT-Ro2	DBF	42,39	11,92	2002-2012	https://doi.org/10.18140/FLX/1440175
IT-SR2	ENF	43,73	10,29	2013-2020	https://doi.org/10.18160/2G60-ZHAK
IT-SRo	ENF	43,73	10,28	1999-2012	https://doi.org/10.18140/FLX/1440176
IT-Tor	GRA	45,84	7,58	2008-2020	https://doi.org/10.18160/2G60-ZHAK
JP-MBF	DBF	44,39	142,32	2003-2005	https://doi.org/10.18140/FLX/1440238
JP-SMF	MF	35,26	137,08	2002-2006	https://doi.org/10.18140/FLX/1440239
MX-Tes	DBF	27,84	-109,30	2004-2009	https://doi.org/10.17190/AMF/1832156
MY-PSO	EBF	2,97	102,31	2003-2009	https://doi.org/10.18140/FLX/1440240
NL-Hor	GRA	52,24	5,07	2004-2011	https://doi.org/10.18140/FLX/1440177
NL-Loo	ENF	52,17	5,74	1996-2014	https://doi.org/10.18140/FLX/1440178
PA-SPn	DBF	9,32	-79,63	2007-2009	https://doi.org/10.18140/FLX/1440180
PA-SPs	GRA	9,31	-79,63	2007-2009	https://doi.org/10.18140/FLX/1440179
RU-Cok	OSH	70,83	147,49	2003-2014	https://doi.org/10.18140/FLX/1440182
RU-Fy2	ENF	56,45	32,90	2015-2020	https://doi.org/10.18160/2G60-ZHAK
RU-Fyo	ENF	56,46	32,92	1998-2020	https://doi.org/10.18160/2G60-ZHAK
RU-Ha1	GRA	54,73	90,00	2002-2004	https://doi.org/10.18140/FLX/1440184
RU-Sam	GRA	72,37	126,50	2002-2014	https://doi.org/10.18140/FLX/1440185
RU-SkP	DNF	62,26	129,17	2012-2014	https://doi.org/10.18140/FLX/1440243
RU-Tks	GRA	71,59	128,89	2010-2014	https://doi.org/10.18140/FLX/1440244
SD-Dem	SAV	13,28	30,48	2005-2009	https://doi.org/10.18140/FLX/1440186
SE-Htm	ENF	56,10	13,42	2015-2020	https://doi.org/10.18160/2G60-ZHAK
SE-Nor	ENF	60,09	17,48	2014-2020	https://doi.org/10.18160/2G60-ZHAK
SE-Ros	ENF	64,17	19,74	2014-2020	https://doi.org/10.18160/2G60-ZHAK
SE-Svb	ENF	64,26	19,77	2014-2020	https://doi.org/10.18160/2G60-ZHAK
SN-Dhr	SAV	15,40	-15,43	2010-2013	https://doi.org/10.18140/FLX/1440246
US-AR1	GRA	36,43	-99,42	2009-2012	https://doi.org/10.18140/FLX/1440103
US-AR2	GRA	36,64	-99,60	2009-2012	https://doi.org/10.18140/FLX/1440104
US-ARM	CRO	36,61	-97,49	2003-2020	https://doi.org/10.17190/AMF/1854366
US-Bar	DBF	44,06	-71,29	2004-2021	https://doi.org/10.17190/AMF/2006969
US-Bi1	CRO	38,10	-121,50	2016-2021	https://doi.org/10.17190/AMF/1871134
US-Bi2	CRO	38,11	-121,54	2017-2021	https://doi.org/10.17190/AMF/1871135
US-Blo	ENF	38,90	-120,63	1997-2007	https://doi.org/10.18140/FLX/1440068
US-CF1	CRO	46,78	-117,08	2017-2020	https://doi.org/10.17190/AMF/1832158
US-CRT	CRO	41,63	-83,35	2011-2013	https://doi.org/10.18140/FLX/1440117
US-Cop	GRA	38,09	-109,39	2001-2007	https://doi.org/10.18140/FLX/1440100
US-GBT	ENF	41,37	-106,24	1999-2006	https://doi.org/10.18140/FLX/1440118

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)

Continued Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
US-GLE	ENF	41,37	-106,24	2005-2020	https://doi.org/10.17190/AMF/1871136
US-Goo	GRA	34,25	-89,87	2002-2006	https://doi.org/10.18140/FLX/1440070
US-Ha1	DBF	42,54	-72,17	1991-2020	https://doi.org/10.17190/AMF/1871137
US-Ho2	ENF	45,21	-68,75	1999-2020	https://doi.org/10.17190/AMF/1881581
US-IB2	GRA	41,84	-88,24	2004-2011	https://doi.org/10.18140/FLX/1440072
US-KLS	GRA	38,77	-97,57	2012-2019	https://doi.org/10.17190/AMF/1854367
US-KS2	CSH	28,61	-80,67	2003-2006	https://doi.org/10.18140/FLX/1440075
US-MMS	DBF	39,32	-86,41	1999-2020	https://doi.org/10.17190/AMF/1854369
US-MOz	DBF	38,74	-92,20	2004-2019	https://doi.org/10.17190/AMF/1854370
US-Me2	ENF	44,45	-121,56	2002-2020	https://doi.org/10.17190/AMF/1854368
US-Me3	ENF	44,32	-121,61	2004-2009	https://doi.org/10.18140/FLX/1440080
US-Me4	ENF	44,50	-121,62	1996-2000	https://doi.org/10.18140/FLX/1440081
US-Me5	ENF	44,44	-121,57	2000-2002	https://doi.org/10.18140/FLX/1440082
US-Me6	ENF	44,32	-121,61	2010-2014	https://doi.org/10.18140/FLX/1440099
US-Mpj	WSA	34,44	-106,24	2008-2020	https://doi.org/10.17190/AMF/1832161
US-NR1	ENF	40,03	-105,55	1998-2016	https://doi.org/10.17190/AMF/1871141
US-Ne1	CRO	41,17	-96,48	2001-2020	https://doi.org/10.17190/AMF/1871140
US-Ne2	CRO	41,16	-96,47	2001-2013	https://doi.org/10.18140/FLX/1440085
US-Ne3	CRO	41,18	-96,44	2001-2013	https://doi.org/10.18140/FLX/1440086
US-ONA	GRA	27,38	-81,95	2015-2020	https://doi.org/10.17190/AMF/1832163
US-Oho	DBF	41,55	-83,84	2004-2013	https://doi.org/10.18140/FLX/1440088
US-PFa	MF	45,95	-90,27	1995-2014	https://doi.org/10.18140/FLX/1440089
US-Prr	ENF	65,12	-147,49	2010-2014	https://doi.org/10.18140/FLX/1440113
US-Ro5	CRO	44,69	-93,06	2017-2020	https://doi.org/10.17190/AMF/1818371
US-Rwe	CSH	43,07	-116,76	2003-2007	https://doi.org/10.17190/AMF/1871143
US-SRC	OSH	31,91	-110,84	2008-2014	https://doi.org/10.18140/FLX/1440098
US-SRG	GRA	31,79	-110,83	2008-2021	https://doi.org/10.17190/AMF/2204877
US-SRM	WSA	31,82	-110,87	2004-2014	https://doi.org/10.18140/FLX/1440090
US-Seg	GRA	34,36	-106,70	2007-2021	https://doi.org/10.17190/AMF/1984572
US-Ses	OSH	34,33	-106,74	2007-2021	https://doi.org/10.17190/AMF/1984573
US-Sne	GRA	38,04	-121,75	2016-2020	https://doi.org/10.17190/AMF/1871144
US-Snf	GRA	38,04	-121,73	2018-2020	https://doi.org/10.17190/AMF/1854371
US-Sta	OSH	41,40	-106,80	2005-2009	https://doi.org/10.18140/FLX/1440115
US-Syv	MF	46,24	-89,35	2001-2014	https://doi.org/10.18140/FLX/1440091
US-Ton	WSA	38,43	-120,97	2001-2014	https://doi.org/10.18140/FLX/1440092
US-Twt	CRO	38,11	-121,65	2009-2014	https://doi.org/10.18140/FLX/1440106
US-UMB	DBF	45,56	-84,71	2000-2014	https://doi.org/10.18140/FLX/1440093
US-UMd	DBF	45,56	-84,70	2007-2014	https://doi.org/10.18140/FLX/1440101
US-Var	GRA	38,41	-120,95	2000-2021	https://doi.org/10.17190/AMF/1993904
US-WCr	DBF	45,81	-90,08	1999-2014	https://doi.org/10.18140/FLX/1440095
US-Whs	OSH	31,74	-110,05	2007-2020	https://doi.org/10.17190/AMF/1984574
US-Wi3	DBF	46,63	-91,10	2002-2004	https://doi.org/10.18140/FLX/1440057

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)

Continued Table S1. Information on the eddy covariance sites from the combined ICOS, AmeriFlux, FLUXNET 2015 tier 1 dataset

site	IGBP	latitude	longitude	period	DOI
US-Wi4	ENF	46,74	-91,17	2002-2005	https://doi.org/10.18140/FLX/1440058
US-Wjs	SAV	34,43	-105,86	2007-2021	https://doi.org/10.17190/AMF/1871146
US-Wkg	GRA	31,74	-109,94	2004-2021	https://doi.org/10.17190/AMF/1984575
ZA-Kru	SAV	-25,02	31,50	2000-2013	https://doi.org/10.18140/FLX/1440188
ZM-Mon	DBF	-15,44	23,25	2000-2009	https://doi.org/10.18140/FLX/1440189

IGBP: The land cover classification defined by the International Geosphere Biosphere Programme (IGBP)