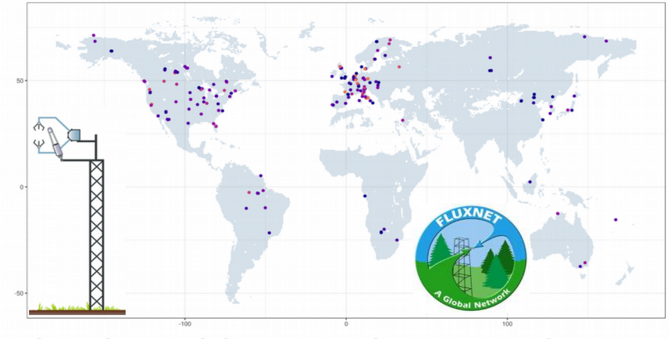
On the role of complementary EO data sets in data-driven estimates of terrestrial carbon fluxes

Sophia Walther, Jacob A. Nelson, Martin Jung,
Fabian Gans, Basil Kraft et Fluxcom al.,
Mirco Migliavacca, Gregory Duveiller, Sofia L. Ermida,
Darren Ghent, Karen L. Veal

Data-driven modelling of terrestrial fluxes



in-situ eddy-covariance carbon fluxes & meteorology

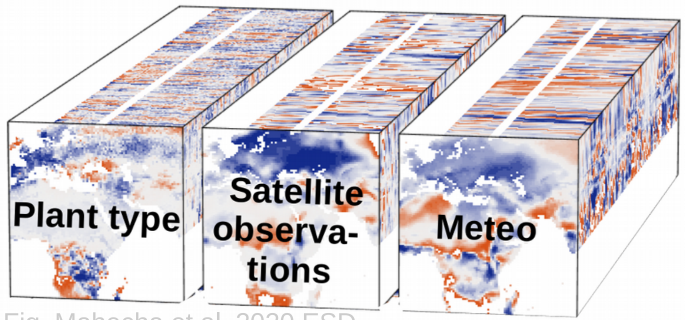
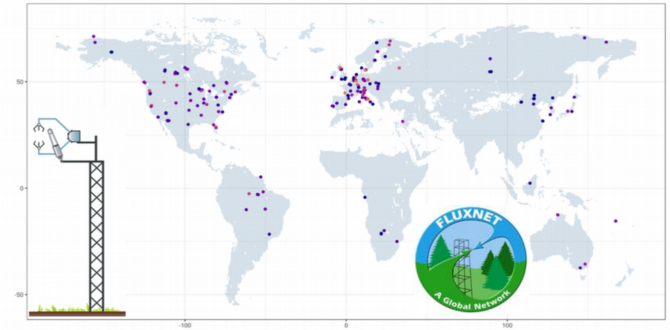


Fig. Mahecha et al. 2020 ESD

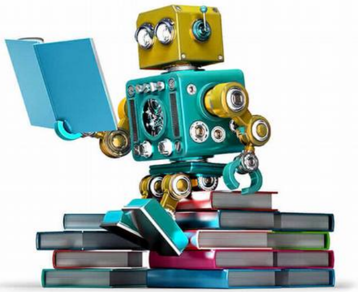
predictor data sets, at site & global

Data-driven modelling of terrestrial fluxes



in-situ eddy-covariance carbon fluxes & meteorology

+



machine learning

+

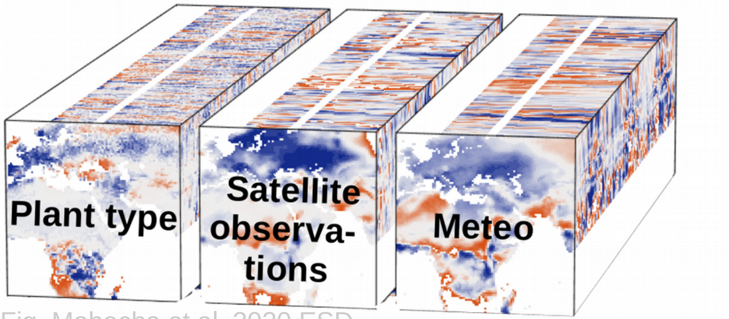
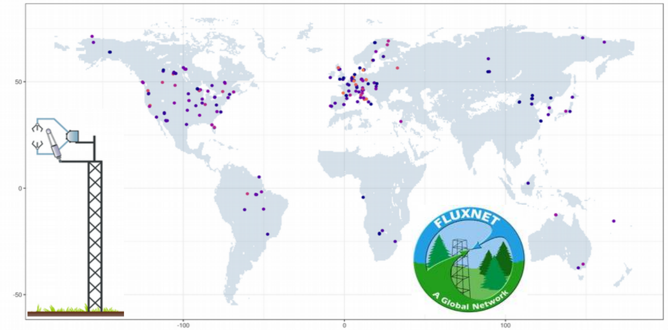


Fig. Mahecha et al. 2020 ESD

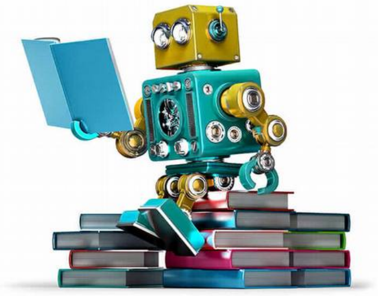
predictor data sets, at site & global

Data-driven modelling of terrestrial fluxes



in-situ eddy-covariance carbon fluxes & meteorology

+



machine learning

+

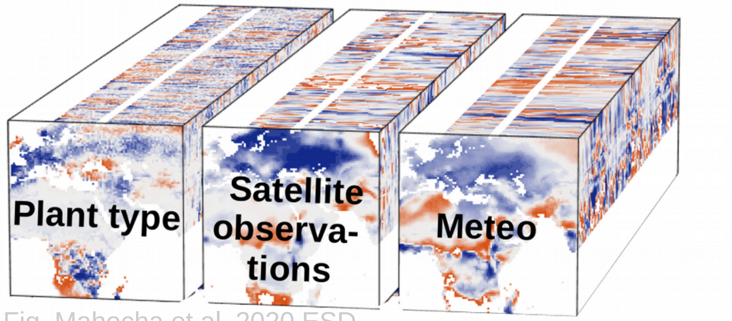
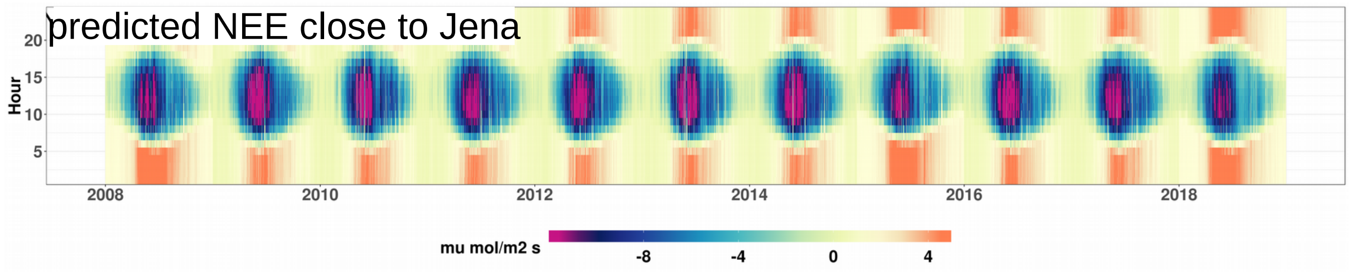
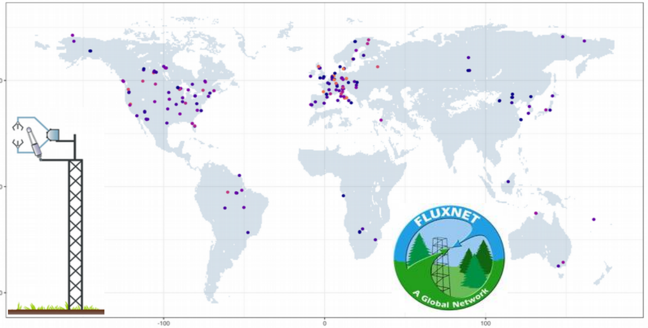


Fig. Mahecha et al. 2020 ESD

predictor data sets, at site & global

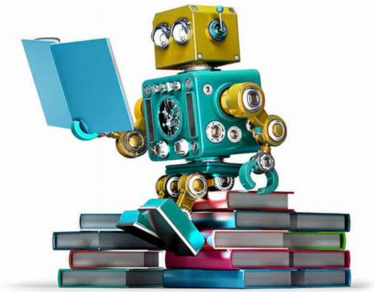


Data-driven modelling of terrestrial fluxes



in-situ eddy-covariance carbon fluxes & meteorology

+



machine learning

+

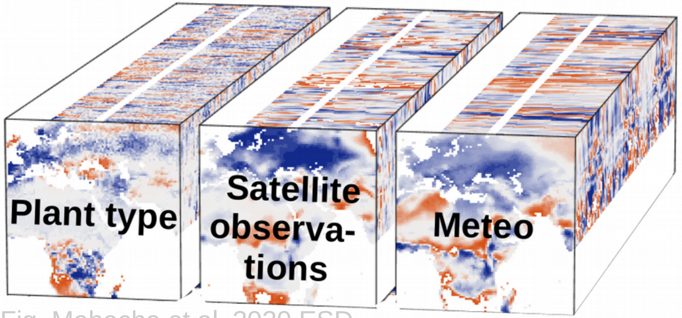
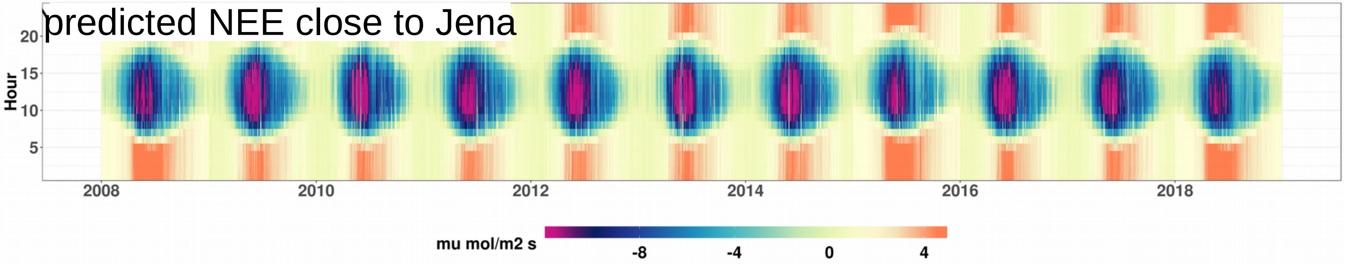


Fig. Mahecha et al. 2020 ESD

predictor data sets, at site & global



`Fluxcom-X':
poster by M. Jung et al.
today, A4.01 5.20pm

On the role of ...

Physiological vegetation responses

...complementary EO

for biogenic C fluxes

Morphological vegetation responses

Diagnosis of soil moisture status

On the role of ...

SIF
changes in photosynthetic activity

VOD
vegetation water content, biomass, functional diversity

Physiological vegetation responses

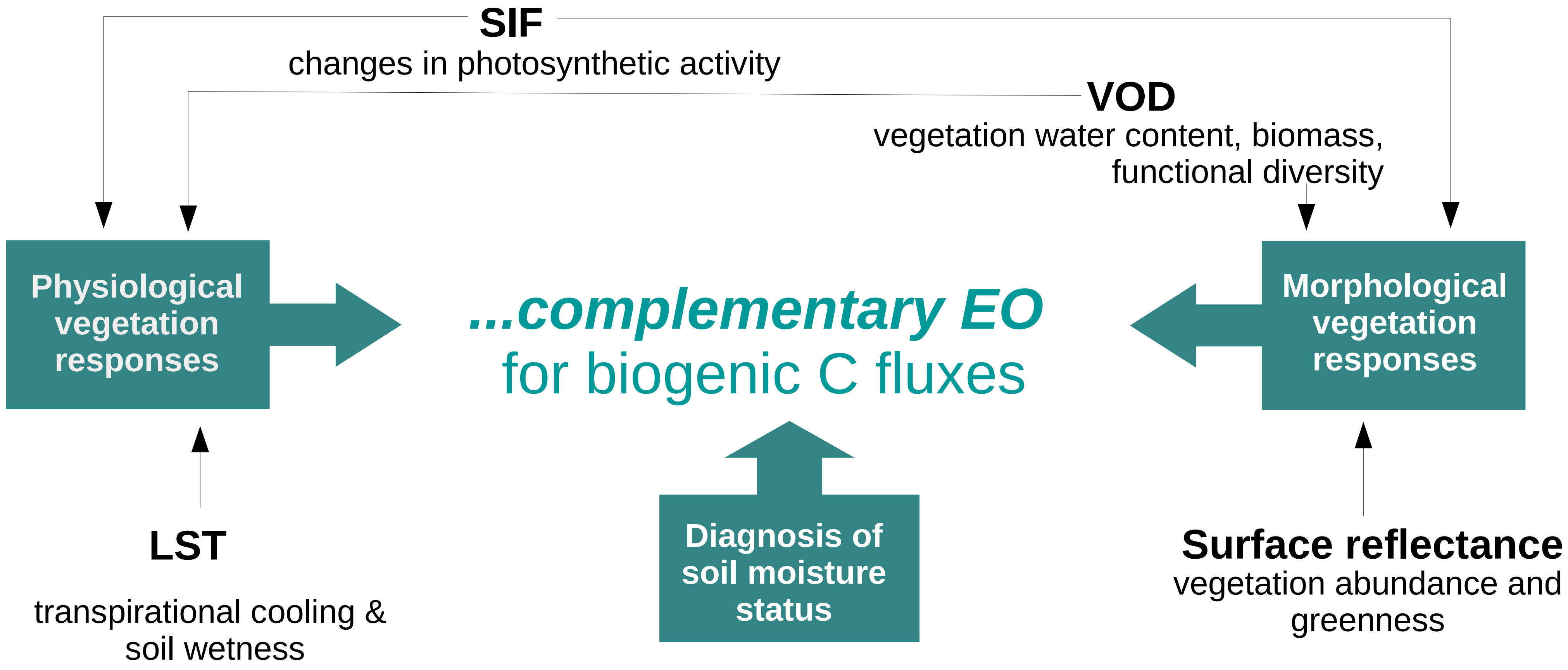
...complementary EO
for biogenic C fluxes

Morphological vegetation responses

Diagnosis of soil moisture status

Surface reflectance
vegetation abundance and greenness

On the role of ...



On the role of ...

SIF
changes in photosynthetic activity

VOD
vegetation water content, biomass, functional diversity

Physiological vegetation responses

...complementary EO
for biogenic C fluxes

Morphological vegetation responses

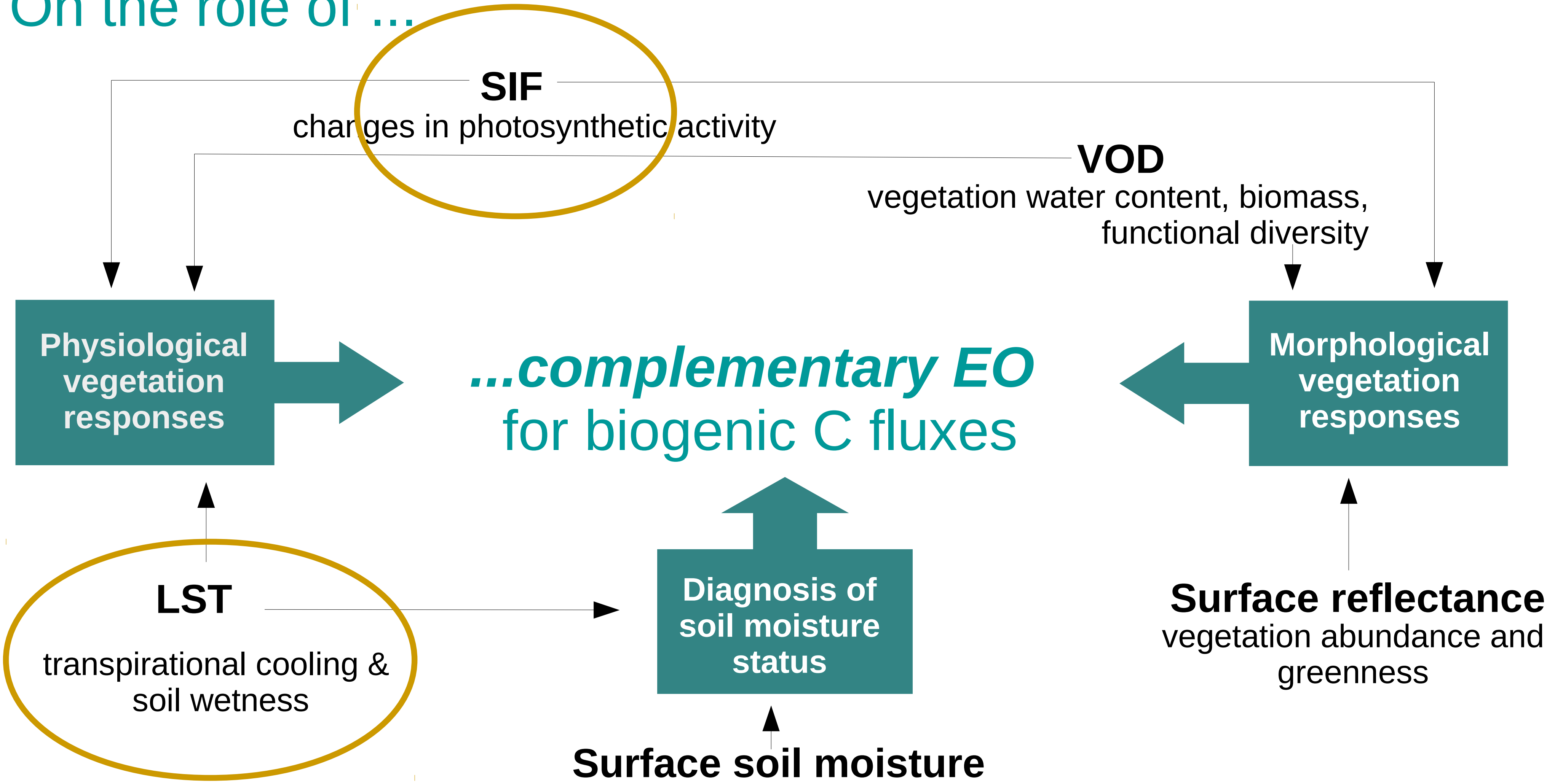
LST
transpirational cooling & soil wetness

Diagnosis of soil moisture status

Surface reflectance
vegetation abundance and greenness

Surface soil moisture

On the role of ...



Observations

- @ 141 sites: 2.5mio good quality* samples
  data sets: LaThuile, Fluxnet2015, ICOS Drought2018, warm winter 2020

- predictor variables:
  EO:
  - daily MODIS surface reflectance (MCD43A4), derived vegetation indices, and **LST** (MxD11A1) at the sites (*'FluxnetEO' v2 data set, Walther et al. 2021, Biogeosc Disc*)
  - daily **SIF** from GOME-2 (MetOp-A, Köhler et al. 2016)

* data set QC plus new EC QC (Jung et al. in prep)
** nighttime partitioning

Observations

- @ 141 sites: 2.5mio good quality* samples
  data sets: LaThuile, Fluxnet2015, ICOS Drought2018, warm winter 2020

- predictor variables:
  EO:
  - daily MODIS surface reflectance (MCD43A4), derived vegetation indices, and **LST** (MxD11A1) at the sites (*'FluxnetEO' v2 data set, Walther et al. 2021, Biogeosc Disc*)
  - daily **SIF** from GOME-2 (MetOp-A, Köhler et al. 2016)

  meteorological conditions:
  - hourly actual and potential shortwave incoming radiation, air temperature, vapour pressure deficit

  plant functional type
- target variable: hourly gross primary productivity** (GPP)

* data set QC plus new EC QC (Jung et al. in prep)
** nighttime partitioning

Observations

- @ 141 sites: 2.5mio good quality* samples
  data sets: LaThuile, Fluxnet2015, ICOS Drought2018, warm winter 2020

- predictor variables:
  EO:
  - daily MODIS surface reflectance (MCD43A4), derived vegetation indices, and **LST** (MxD11A1) at the sites (*'FluxnetEO' v2 data set, Walther et al. 2021, Biogeosc Disc*)
  - daily **SIF** from GOME-2 (MetOp-A, Köhler et al. 2016)

  meteorological conditions:
  - hourly actual and potential shortwave incoming radiation, air temperature, vapour pressure deficit

  plant functional type
- target variable: hourly gross primary productivity** (GPP)

Method

- 5-fold cross-validation
- XGBoost as a machine learning model

* data set QC plus new EC QC (Jung et al. in prep)
** nighttime partitioning

How to quantify the `role' of an EO data stream?

- How does the model use the data to make its predictions? How does the value of a predictor variable influence the value of the estimated C flux?

How to quantify the `role' of an EO data stream?

- How does the model use the data to make its predictions? How does the value of a predictor variable influence the value of the estimated C flux?

Quantify via SHAP values φ_j:

$$\hat{Y}_i = \text{baseline value} + \Sigma_j\ \varphi_{j,i}$$

$\hat{Y}$: predicted carbon flux

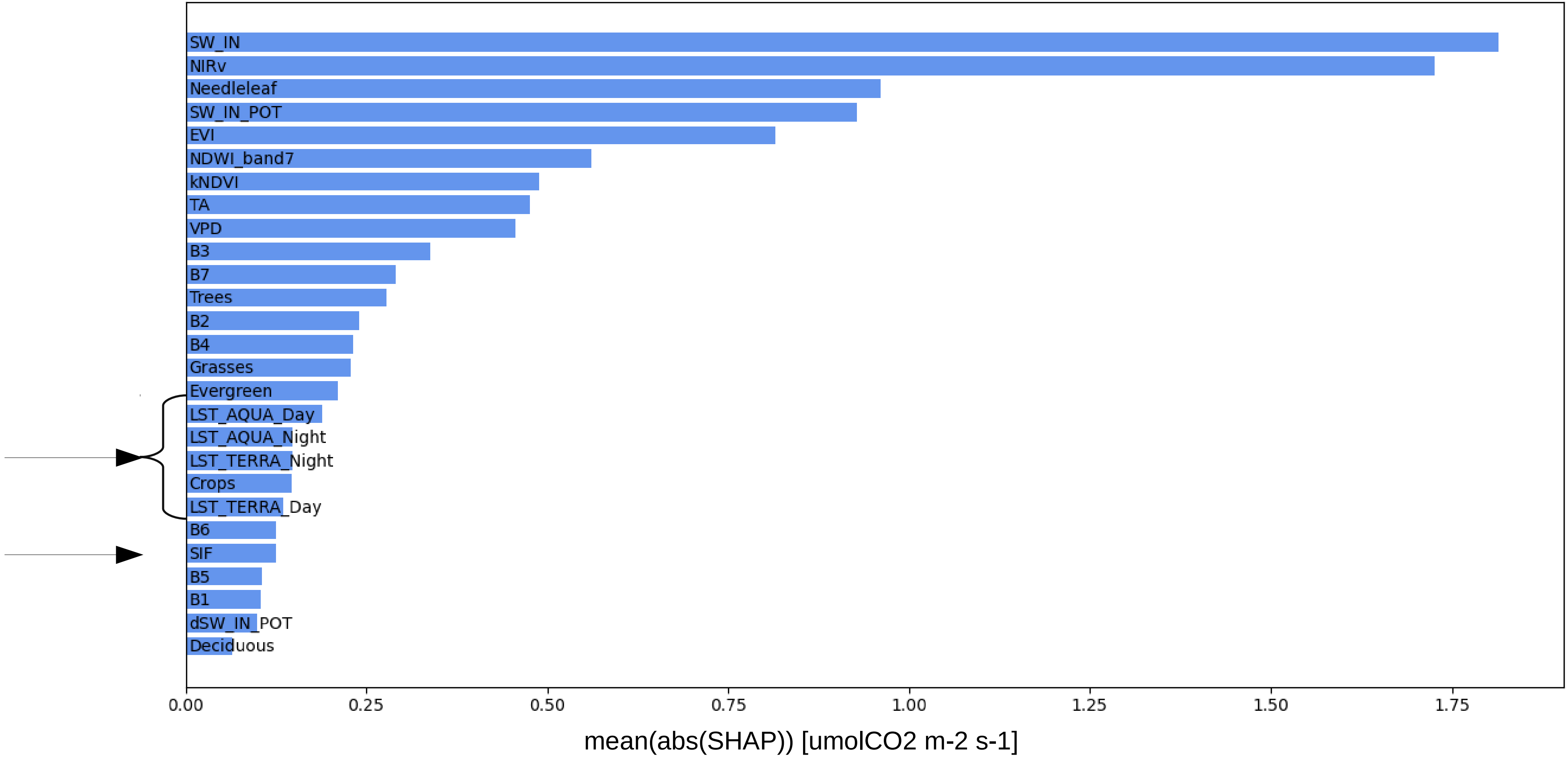
i : sample (one site-hour)

j : predictor variable

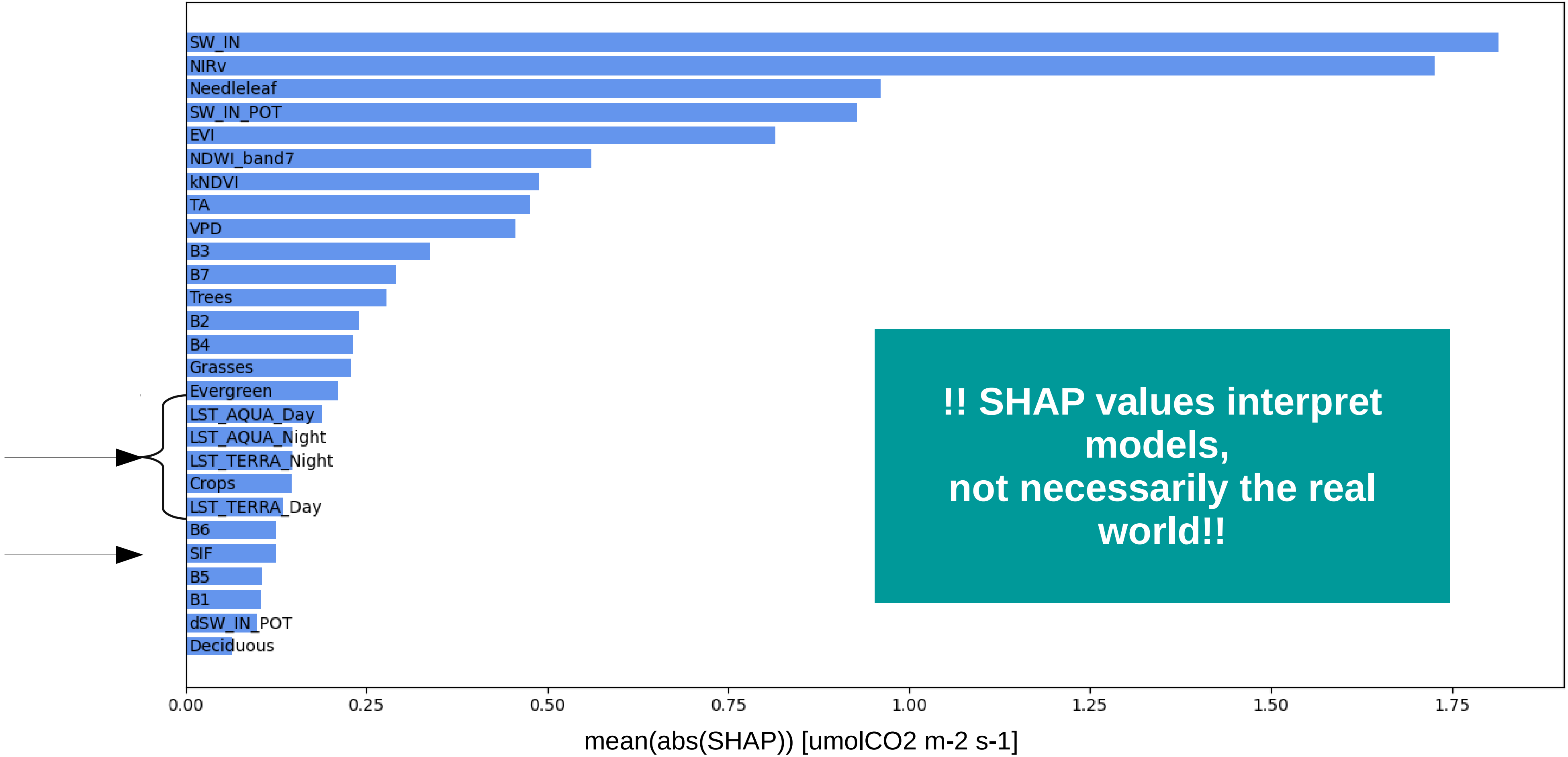
baseline value = const.

φ : shap value

SHAP contributions to GPP predictions



SHAP contributions to GPP predictions



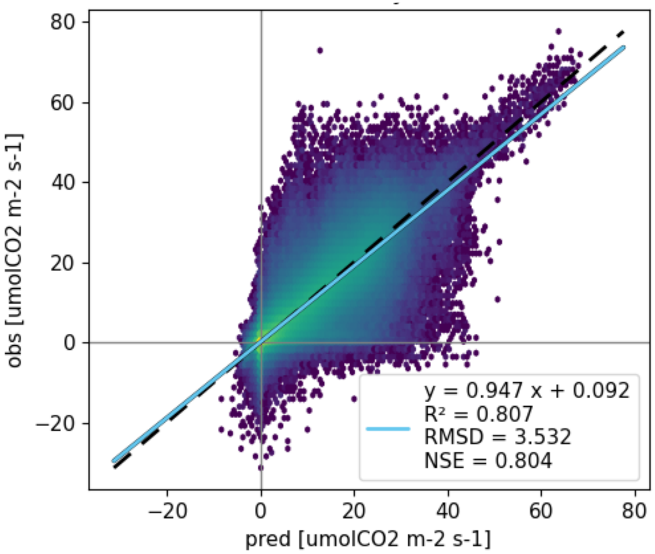
!! SHAP values interpret models,
not necessarily the real world!!

Are the models that we interpret accurate?

Are the models that we interpret accurate?

GPP

hourly

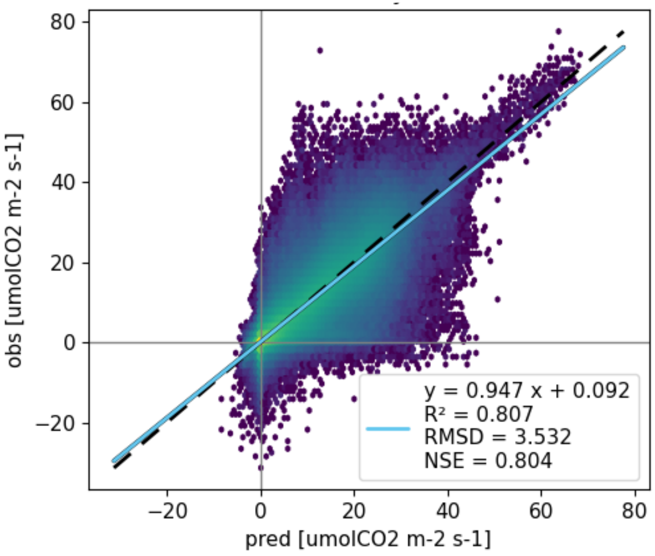


NSE = 0.804

Are the models that we interpret accurate?

GPP

hourly

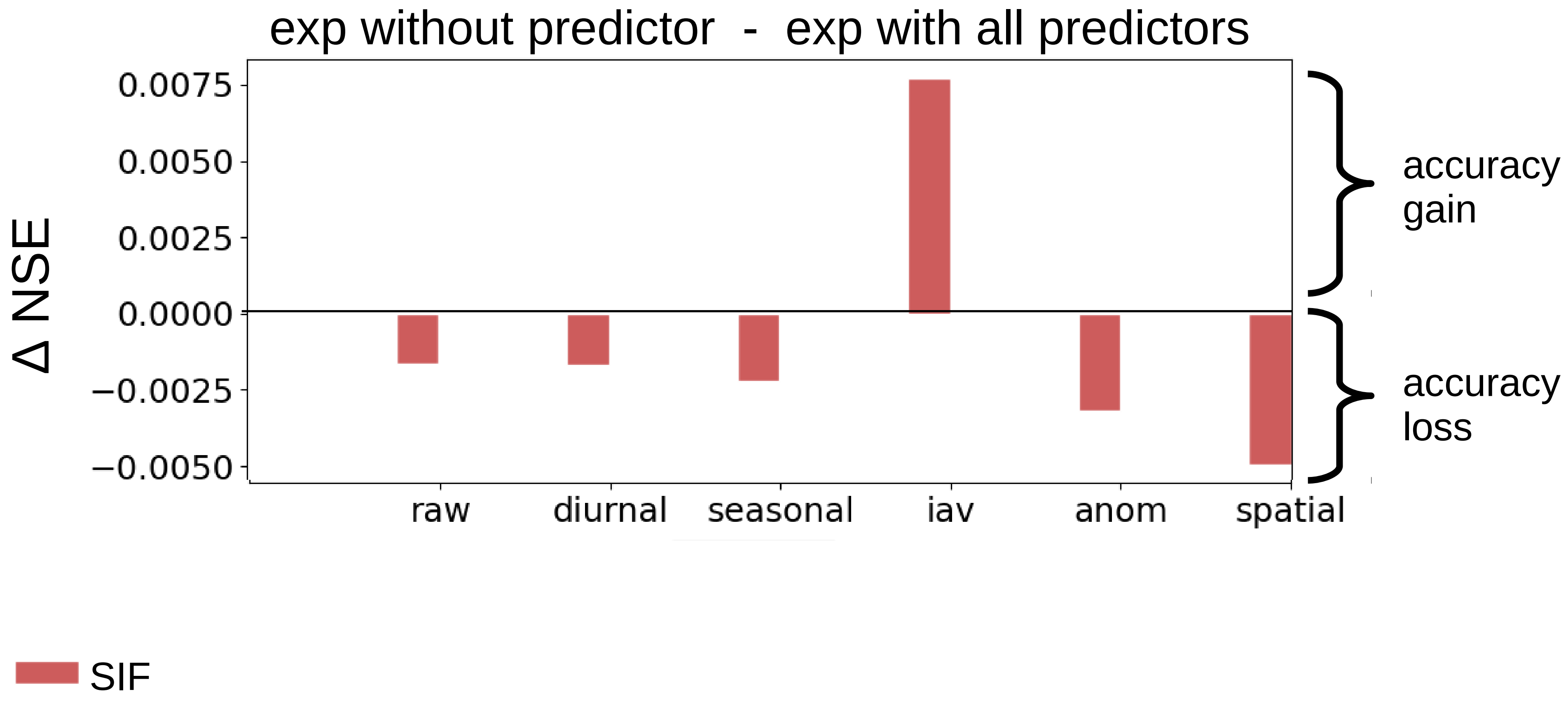


NSE = 0.804

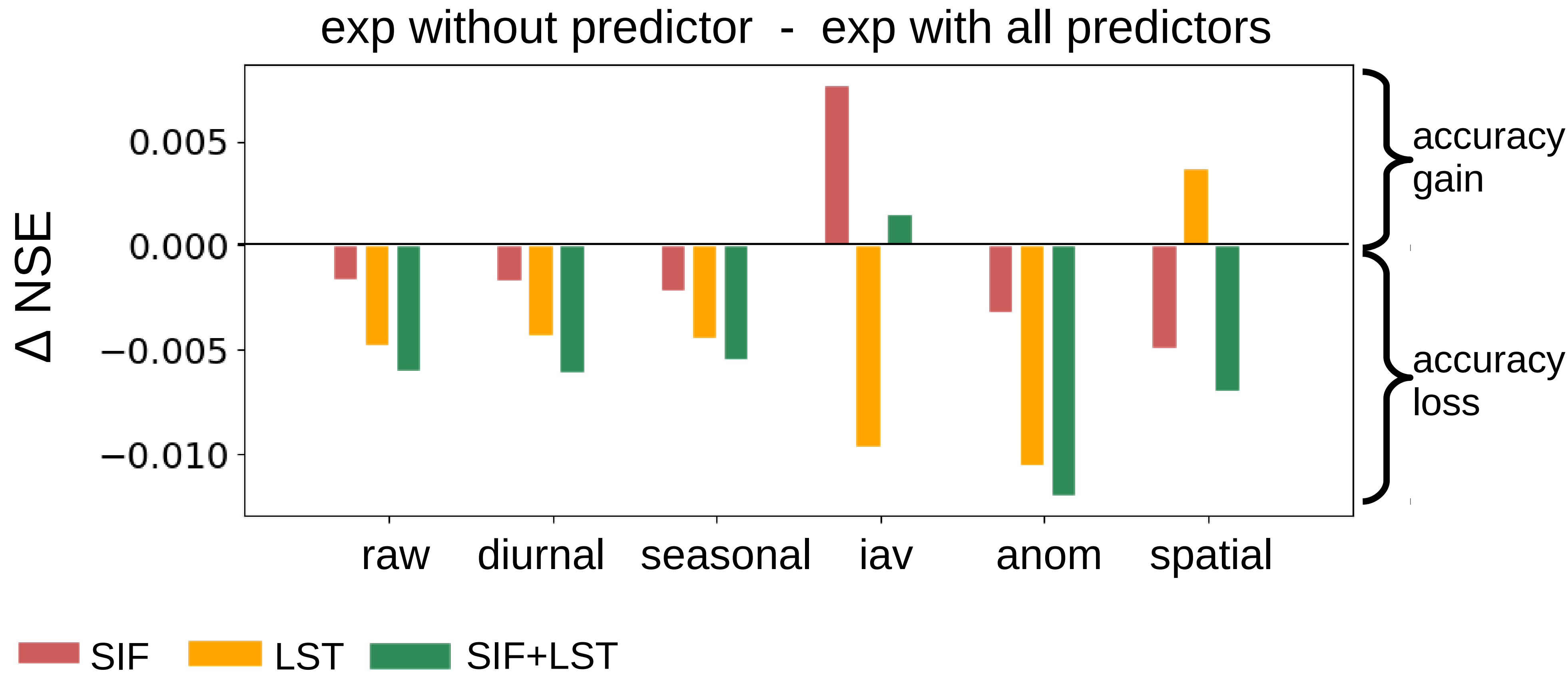
	NSE
diurnal	0.862
seasonal	0.868
spatial	0.752
anomalies	0.35
interannual	0.318

Accuracy loss when excluding EO predictor variable?

GPP

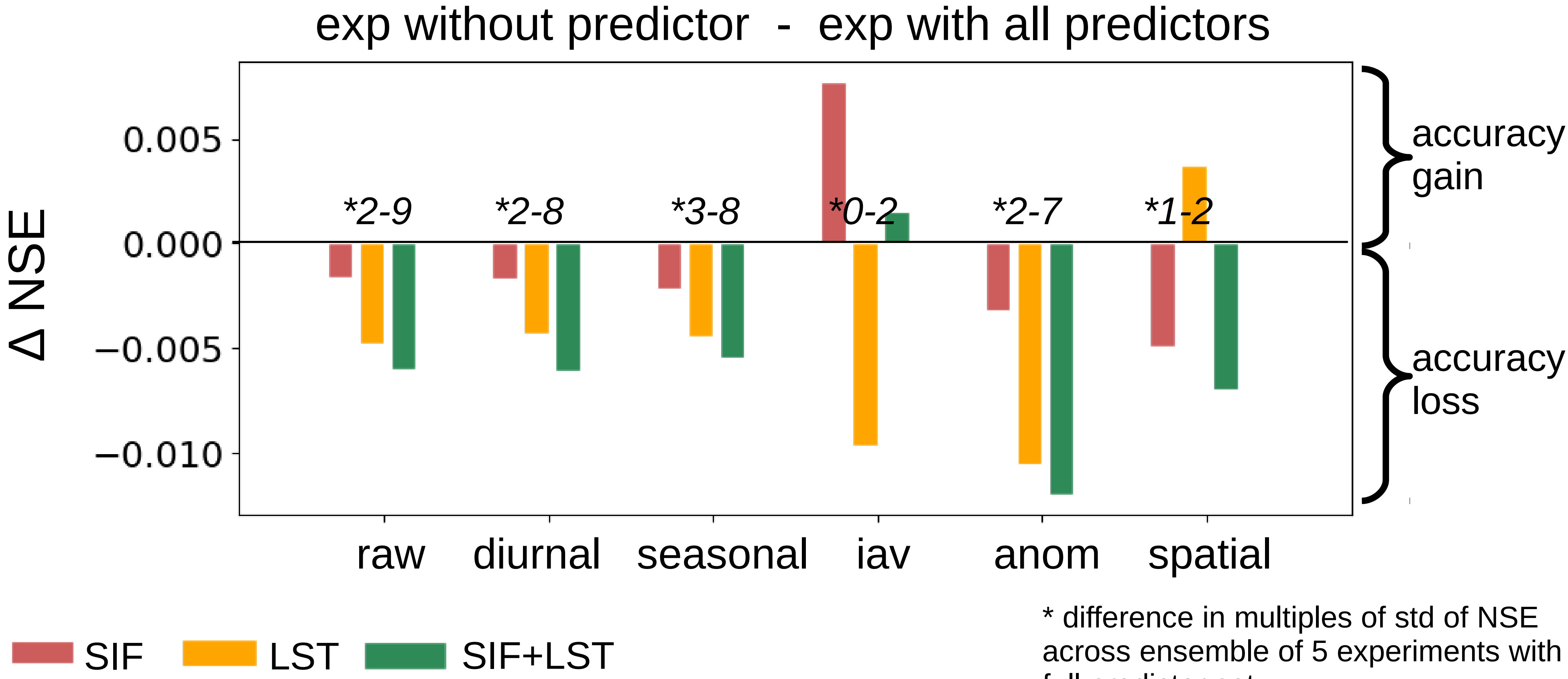


Accuracy loss when excluding EO predictor variable?
GPP



Accuracy loss when excluding EO predictor variable?

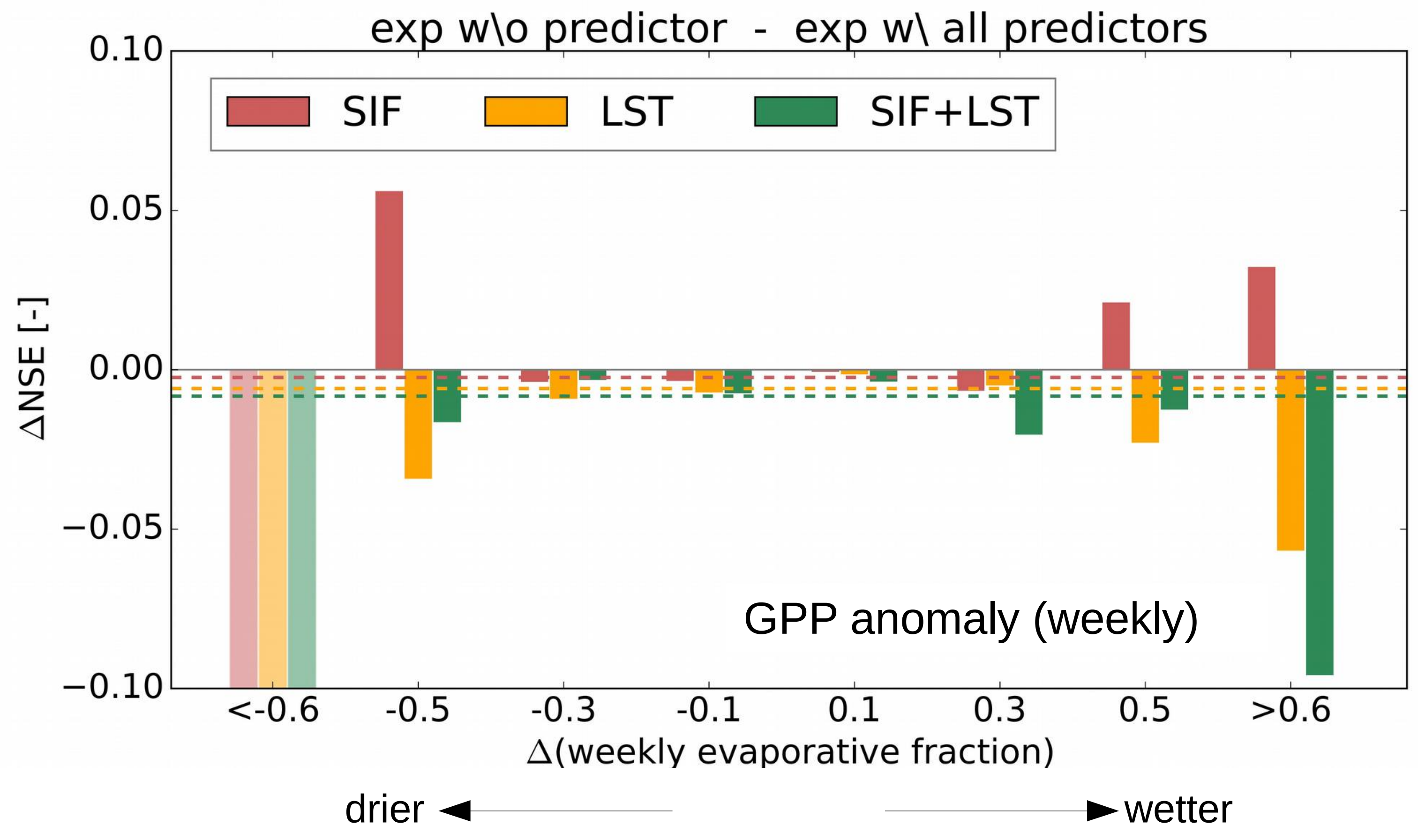
GPP



* difference in multiples of std of NSE across ensemble of 5 experiments with full predictor set

Do SIF and LST increase the accuracy of water effects?

Do SIF and LST increase the accuracy of water effects?

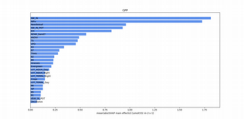


Take away: it all depends...

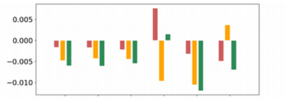
-

Take away: it all depends...

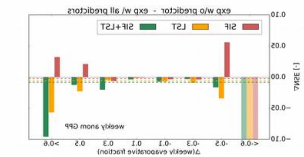
- Shap contribution of SIF and LST to predicted GPP values comparatively low



- contribution of LST and SIF to GPP accuracy depends on scale, synergistic effects (flux, sampling)
  → effect of LST > SIF
  → ΔNSE(GPP) strongest for hourly to seasonal time scales, anomalies

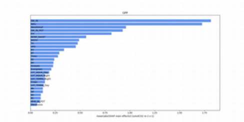


- EO contribution (pos & neg) to ΔNSE of flux anomalies increases with magnitude of moisture anomalies

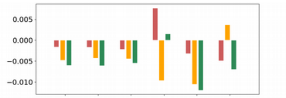


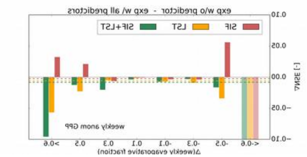
Take away: it all depends...

- Shap contribution of SIF and LST to predicted GPP values comparatively low



- contribution of LST and SIF to GPP accuracy depends on scale, synergistic effects (flux, sampling)
  → effect of LST > SIF
  → ΔNSE(GPP) strongest for hourly to seasonal time scales, anomalies





- EO contribution (pos & neg) to ΔNSE of flux anomalies increases with magnitude of moisture anomalies

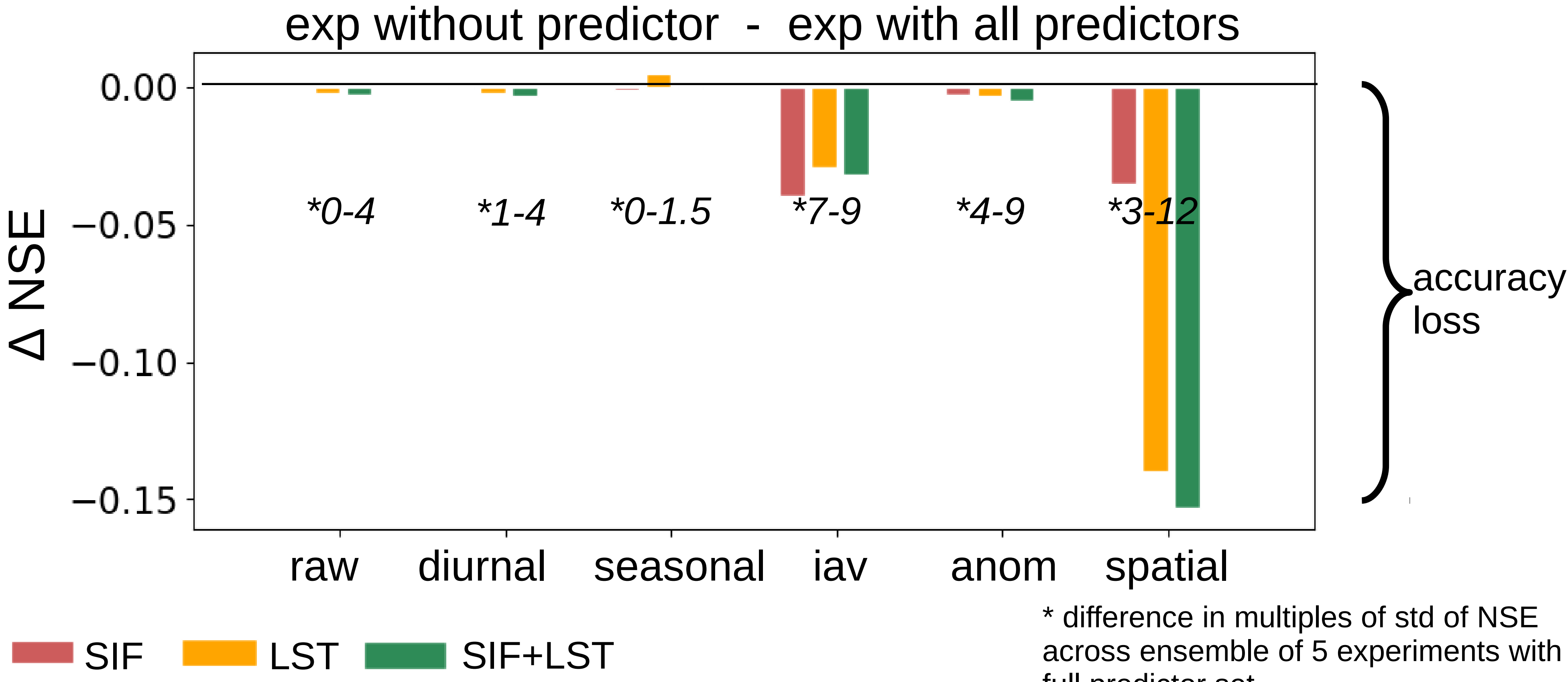
On the to-do list:

- detailed analysis of where and when do the EO improve predicted site flux accuracy
- role of acquisition and retrieval properties
- more EO predictors (VOD, soil moisture)
- production of global data sets and their analysis

Thanks!

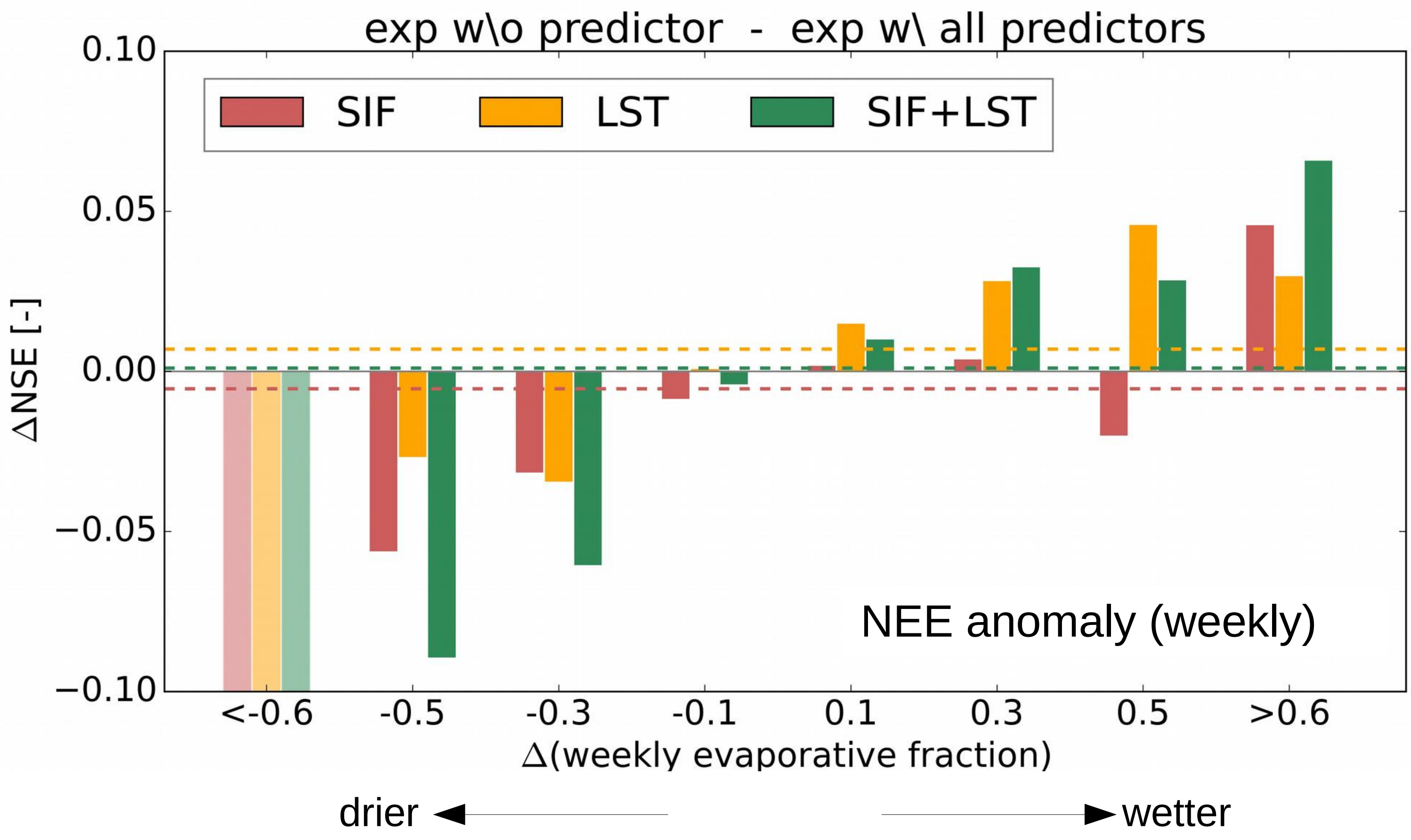
Accuracy loss when excluding EO predictor variable?

NEE

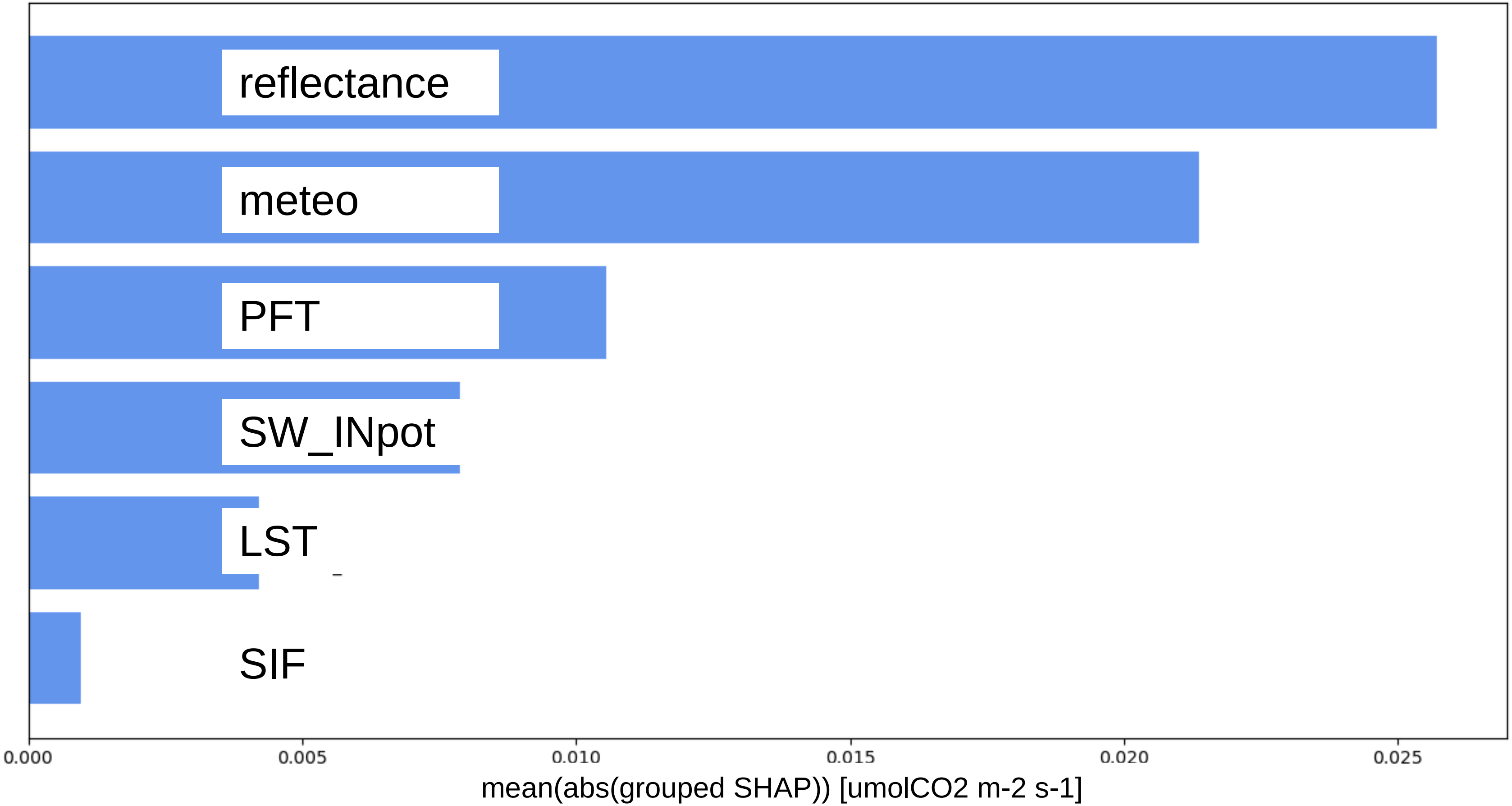


* difference in multiples of std of NSE across ensemble of 5 experiments with full predictor set

Do SIF and LST increase the accuracy of water effects?

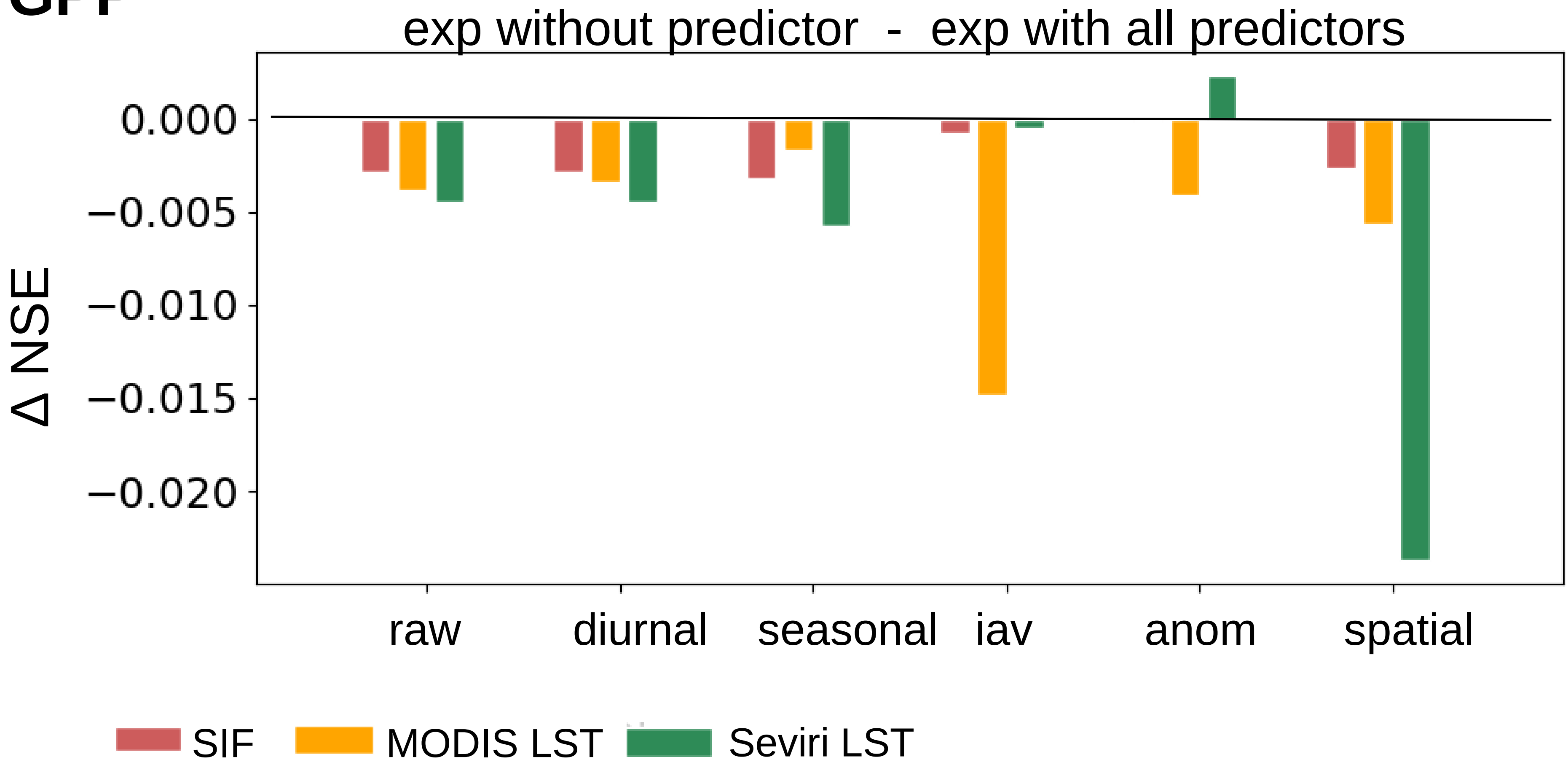


SHAP contributions to GPP predictions



Acquisition properties: GEO and LEO

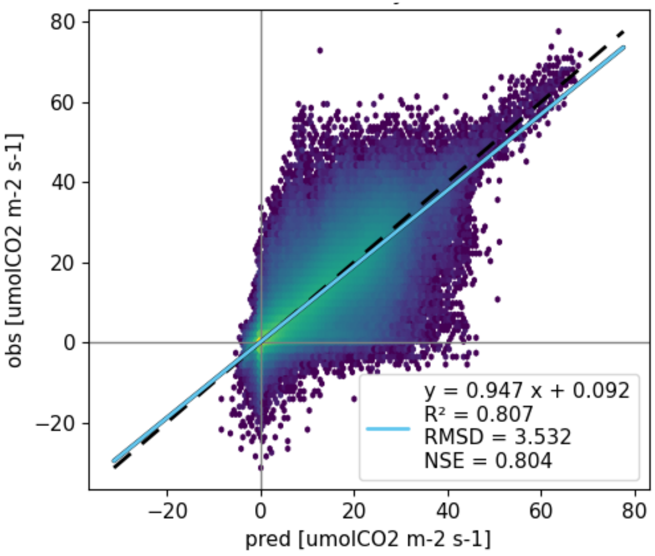
GPP



Are the models that we interpret accurate?

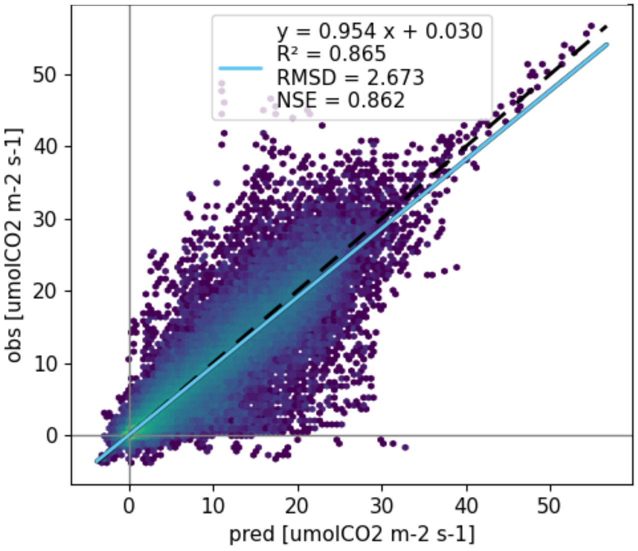
GPP

hourly



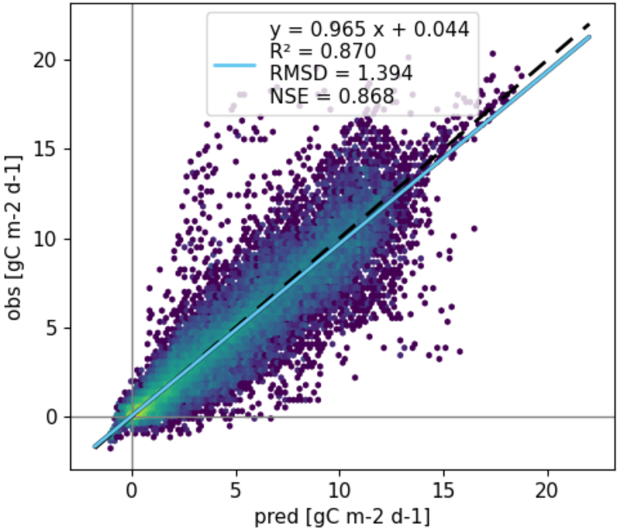
NSE = 0.804

monthly diurnal cycle



NSE = 0.862

daily seasonality

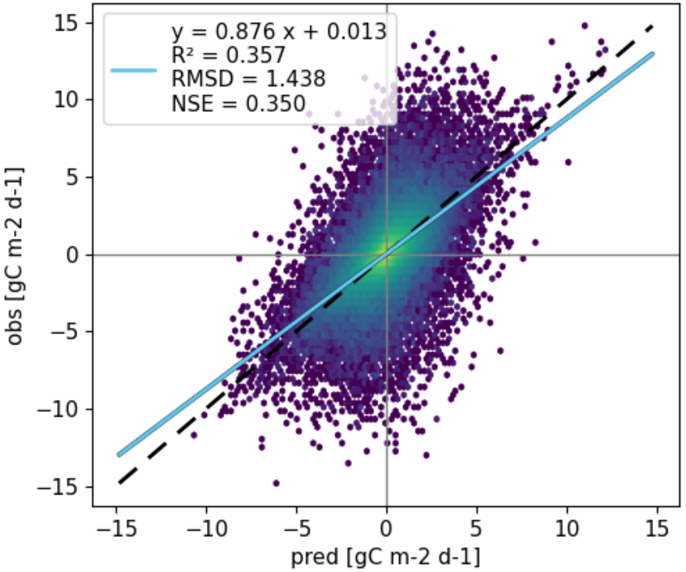


NSE = 0.868

Are the models that we interpret accurate?

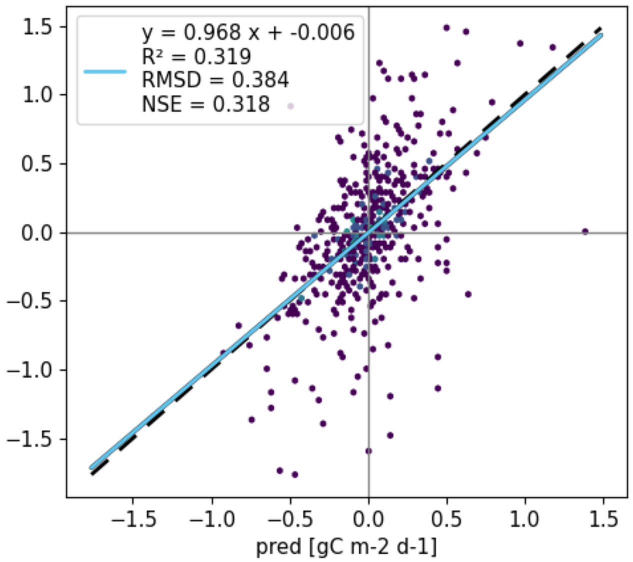
GPP

daily anomalies



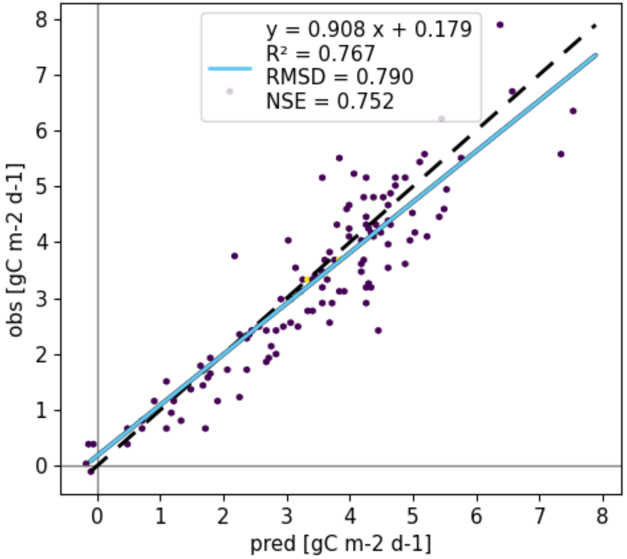
NSE = 0.350

interannual



NSE = 0.318

spatial



NSE = 0.752

But do the models, that we interpret here, actually make sense? How do they reproduce the observations?
Scatterplots and NSE of full model for all time scales

Time series plots for single sites, for drought, seasonal and 2018 Xin paper

How do the EO data sets contribute to NSE?

They help for accurate fluxes, but to what extent to acquistion properties translate to the flux accuracy?

Open questions:
NEE, spatial mismatch, retrieval effect, more data sets