

Group discussion

1. What algorithms do you use to gap-fill flux data?

[your_name/group_name]

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; MDS, kNN, ANN

[your_name/group_name]: ...

Taylor Thomas / LI-COR: MDS, Random Forest, ANN

Pushpendra Raghav; MDS;

Peter Isaac/TERN EP: MDS, SOLO neural network (EVI as driver)

a. What are the strengths of these algorithms?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; MDS is fast with ease of access, kNN and ANN for CH₄, N₂O but work also on others

[your_name/group_name]: ...

Taylor Thomas / LI-COR: MDS ease of access to input features

Peter Isaac/TERN EP: MDS is fast but not good for long gaps, SOLO with EVI is good for long gaps but slow

b. What are the major challenges or limitations?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; Highly dynamic agroecosystems, Natural occurring spikes

[your_name/group_name]: ...

Taylor Thomas / LI-COR: Overall performance of MDS on longer timescales

Peter Isaac/TERN EP: Speed so you can do it for a distribution of u* values, noisy input data

2. What variables or predictors do you use to gap-fill flux data?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; As many as reliably collected, biogeochemically making sense

[your_name/group_name]: ...

Taylor Thomas / LI-COR: PPFD, Tair, Humidity

Peter Isaac/TERN EP: Incoming shortwave, air temperature, specific humidity, soil moisture, ground heat flux (sometimes)

a. How do you deal with gaps in those predictor variables? How do the missing drivers affect your gap-filling results?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; Gaps in predictors are being filled as well, if they can not be derived from nearby sensors

[your_name/group_name]: ...

Peter Isaac/TERN EP: gap fill meteorology with automatic weather station data, numerical weather prediction model data (ACCESS), reanalysis data (ERA5), can't fill gaps in fluxes if there are gaps in the driver variables (I think you call these "predictors"?)

- b. What are critical predictors but currently missing?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; Canopy height, Ts rather than Ta, PPFD, SWC

[your_name/group_name]: ...

3. How can we make gap-filling workflows more transparent, reproducible, and scalable for different sites?

[your_name/group_name]: ...

[your_name/group_name]: ...

[your_name/group_name]: ...

Michael Schuppenhauer; EddyPro metadata files and raw data must be shared/accessible so that *how* we got from HF to 30 min and what was filtered there becomes transparent

Web based interface - see MPI MDS GUI becoming defacto standard although not perfect

[your_name/group_name]: ...

Taylor Thomas / LI-COR: Strict data provenance and expression of uncertainty.

Peter Isaac/TERN EP: Open source code, documentation e.g. wiki with block diagrams of high level workflow, lots of examples e.g. evergreen, savannah, wetlands etc.

Feedback is a gift

<https://docs.google.com/forms/d/e/1FAIpQLSfIPGHkiLr6P1Kc6ETotzRVXk2m8Sp1zVbCUQdxT61F7kE98w/viewform?usp=dialog>

Question & Comments during AmeriFlux 25 breakout session

- What is the difference or (dis)advantages between different gap-filling approaches?
- What are the entry points for users unfamiliar with Python?
- What's the recommendation of the data record to apply this?
 - How does it perform in dryland ecosystems with low signals & high variability?

- May not solely depend on the length of record, but whether it's enough to capture the seasonal or desired temporal dynamics
- Predictors: How do you select them & what's the best combination of predictors?
 - Variable downslection
 - Feature importance can be an indicator
 - What does feature importance explain?
 - Resolution of variables
 - GCC from phenocam
 - EVI from MODIS, assuming slow dynamics
 - XGBoost can handle collinearity between the predictors
- What's the best practice to determine the number of iterations for training models? Is it depend on sites, or universal?
- What is the synthetic data and their purpose?
 - How to define an approach to quantify the uncertainty in gap-filling?
 - It reflects the gap influences? Current approach hasn't accounted for random uncertainties & u^* filtering.
- What is the threshold of 25 gC/m²/year? What's the justification for that?
 - Threshold