AmeriFlux 2016 breakout session report

Time: 2016-09-22, 14:40 - 15:35

Location: AmeriFlux PI meeting, Golden, CO

Title: "Processing best practices and methods including tools and workflows" aka "From field notes to fluxes: Power tools that get the job done"

Chairs: Stefan Metzger, David Durden, Cove Sturtevant, Deb Agarwahl, Gilberto Pastorello, Dario Papale

Participants: 25 – 30

1   Outcomes

- methods, tools and workflows for efficient field operation and data-based science

	tactical: current need, flux networks could support	strategic: availability and leveraging by flux networks
main outcomes	• efficient and transparent data provenance • "from field to final dataset"	• inventory: breakout report • shareability of data and tools across networks • modularity, interoperability • standards for data reporting, representation, and interfaces • cross-network value-added data products
follow-up activities	• mobile tools and file structures in development at AmeriFlux and NEON	• annual meeting of data teams • idea: mutual participation in science steering committees

- work with organizers: make breakout report, presentation, posters available

2   Structure

Moderation: Stefan

- 14:55 - 15:00 (5 min): introduction (moderation: Stefan)
    - exchange on publicly available tools that simplify the mechanics of everyday work

- 15:00 - 15:20 (20 min): 2 breakouts-in breakout: input from the PIs
    - topics

- ■ #1: efficient field operation: proven tools, methods and workflows (Sect. 3, moderation: Cove, Gilberto)
- ■ #2: data-based science: proven tools, methods and workflows (Sect. 4, moderation: Dave, Deb)
  - ○ objectives
    - ■ prioritized list: what tools are being used and recommended for the central flux networks to adopt
    - ■ prioritized list: what tools are missing and are desired, e.g. with support from central flux networks
    - ■ metadata: connections between field operations and data-based science
  - ○ tool for capturing and reporting breakout information: Google Docs

- ● 15:20 - 15:30 (10 min): breakout summaries
  - ○ representatives breakout #1 (Cove, Gilberto): prioritized list
  - ○ representative breakout #2 (Dave, Deb): prioritized list

- ● 15:30 - 15:40 (10 min): outcomes (Sect. 1, moderation: Stefan)

3 Breakout-in-breakout #1: efficient field operation: proven tools, methods and workflows

Moderation: Cove, Gilberto

3.1 What tools are being used and recommended for the central flux networks to adopt

- ● Most/all use paper
  - ○ Waterproof paper with clips, etc. to attach to pants (for unencumbered tower climbing)
  - ○ Transcribed to electronic format back at the lab
- ● Blogging-type software. Make entry, record name. Links to calendar. (access to server necessary)
  - ○ Sean Burns' recent paper lists software in the supplement (Burns et al., 2016)
  - ○ Entered on a local computer. Gets uploaded to a webpage
  - ○ Text-based, but can include images.
  - ○ Open-source
- ● OneNote/EverNote
  - ○ Can install on phone, laptop, office
  - ○ EverNote phone application allows speaking into it
    - ■ Good for dirty applications where writing can be hazardous to phone

- Drawing in the field
  - Blank PDF page. Load in EverNote and draw on it.
- Fulcrum apps
  - Great for customizable applications
  - Offline data entry
  - Limited in the rain unless you have a ruggedized field tablet ($$)
  - Costs $ - $18-25/month per license (one license can be used on several platforms)
- What are the priorities?
  - Tradeoff
    - Structured fields allow standardization and ease of data transfer.
    - Free text field notes have ultimate flexibility. Especially for drawing.
    - Mobile electronic apps take more effort to enter in the field b/c the tablet is cumbersome and fragile, viewing & protecting screen in adverse weather is difficult.
  - Hybrid is best?
    - Field conditions dictate what can be used (rugged computer/tablet, paper)
    - Will never replace the flexibility and lightweight aspect of paper
    - Pictures & drawings are important to include in notes
    - The best solution would is a combo of structured entries and free text notes/drawings

3.2 What tools are missing and are desired, e.g. with support from central flux networks

- Electronic drawing tool within mobile app for field note diagrams.
- Seamless integration of paper and electronic notes
- Online/mobile tool with input forms that feed directly into the BADM
  - Standardized set of metadata to collect every time in the field
  - AmeriFlux is already moving toward mobile applications
  - AmeriFlux is creating template with site-specific fields that you can print for each site
- Improved efficiency UI
  - Searchable
  - Flexible
- Resource for finding solutions others have implemented.
  - Central tool repo shared among the networks
  - Blog (e.g. StackExchange) for sharing ideas
    - Searchable "database" for solutions

3.3 Metadata: connections between field operations and data-based science

- The online tool with input forms that feed directly into the BADM tackles the connection between field ops and data processing
  - Searchable
  - AmeriFlux is already moving toward mobile applications

4 Breakout-in-breakout #2: data-based science: proven tools, methods and workflows

Moderation: Dave, Deb

4.1 What tools are being used and recommended for the central flux networks to adopt

- What processing tools are people using to process flux data?
  - EddyPro
    - Several users are using EddyPro (including ICOS and AmeriFlux networks)
  - R-packages
    - eddy4R for raw-data processing to 30-min
      - public software release follows NEON data product rollout
      - base package rollout planned for Dec. 2016
    - REddyProc for higher-level processing starting from 30-min
      - Several people using for partitioning
  - EdiRe (Houson) for raw-data processing to 30-min
    - Flexible workflow is nice
  - EZflux DL raw-data processing to 30-min
    - CR Basic
  - TK3 for raw-data processing to 30-min
  - EddyUH for raw-data processing to 30-min
  - Matlab code for higher-level processing starting from 30-min
    - Package not open at this point (Dario)
- Additional data processing tools
  - LAI tool (Dario)
  - R-package Phenocam ("phenopix")
    - A package for the extraction of chromatic coordinates and analysis of digital images
  - QAQC code (Dario)- to work on half-hourly data
    - Will not need hopefully if everything is QAQC by AmeriFlux

- - - ■ Will be publicly available in the future
    - ○ Biological, Ancillary, Disturbance and Metadata (BADM)
        - ■ A standard input format for accompanying information for AmeriFlux data
    - ○ Docker
        - ■ Allows results to be the same across processing systems
    - ○ GitHub
        - ■ Git is a version control system (VCS) for processing code.
        - ■ GitHub allows cloud-based repository for processing code and enables collaborative development
    - ○ Hierarchical data format (HDF5)
        - ■ Efficient file format that permits metadata allocation

4.2 What tools are missing and are desired, e.g. with support from central flux networks

- Missing high frequency QA/QC in the data analysis
    - ○ Ameriflux looking for a way to consolidate PI provided flags
        - ■ NEON following Smith et al. (2014)
        - ■ R-package eddy4R.qaqc is addressing this issue
- HDF5 Climate and forecasting (CF) metadata conventions not implemented
    - ○ CF conventions currently not implemented for point measurements
    - ○ Atmospheric Radiation Measurement Program is working on this development
    - ○ NEON working on providing metadata in both, EML and CF conventions
- Uncertainty calculation
    - ○ Still a big issue that needs to be addressed
    - ○ NEON working on R-package eddy4R.ucrt to address this issue
- Storage flux calculation
    - ○ Still a big issue that needs to be addressed
    - ○ NEON working on R-package eddy4R.stor to address this issue

4.3 Metadata: connections between field operations and data-based science

- HDF5
    - ○ CF standard conventions in file
    - ○ Ecological Metadata Language (EML), Climate and Forecast (CF conventions)
- BADM standards - BADM Web update - (web services back-end)
    - ○ Easy ingest of BADM files
    - ○ Currently, requires manual formatting for ingest
- Mobile Applications (Fulcrum)
    - ○ Ability to ingest metadata in a machine readable format

5 References

Burns, S. P., Maclean, G. D., Blanken, P. D., Oncley, S. P., Semmer, S. R., and Monson, R. K.: The Niwot Ridge Subalpine Forest US-NR1 AmeriFlux site – Part 1: Data acquisition and site record-keeping, Geosci. Instrum. Method. Data Syst., 5, 451-471, doi:10.5194/gi-5-451-2016, 2016.

Smith, D. E., Metzger, S., and Taylor, J. R.: A transparent and transferable framework for tracking quality information in large datasets, PLoS One, 9, e112249, doi:10.1371/journal.pone.0112249, 2014.