

AGU 2018 NEON SAE Workshop

Contents

Workshop Goals.....	1
Setting.....	1
Introduction (~15 min).....	1
Brief community presentations on utilizing or synergizing with NEON SAE resources (~20 min)	2
Breakout groups (~90 min)	2
Breakout summaries (~15 min).....	6

Workshop Goals

- Explore these NEON SAE resources
- Solicit input to guide development of NEON SAE community resources for 2019
- Test Cyverse cloud compute environment for hosting code tutorials
- Provide a forum to discuss important topics in the SAE community, with the hope to answer the following three questions:
 - What are the goals the breakout groups suggest?
 - What can the community do to achieve goals?
 - How can NEON help achieve goals?

Setting

- Time: Tuesday December 11th from 7:00-9:30 PM ET
- Location: Hilton Garden Inn, 815 14th Street, NW, Washington, DC; Georgetown C
- Hosts: David Durden, Chris Florian, Natchaya Pingingtha-Durden, Cove Sturtevant, Stefan Metzger
- Participants: [50 participants signed up](#), ~ 35-40 participants attended

Introduction (~15 min)

- Introductory presentation (Dave, Stefan, and Chris)
 - History of NEON SAE meeting at AGU
 - Recent developments
 - Previous outcomes driven by AGU NEON SAE meeting
 - Structure of DevOps approach at NEON
 - Goals
 - Discussed expectations
 - Workshop structure
 - Layout of the agenda

Brief community presentations on utilizing or synergizing with NEON SAE resources (~20 min)

- **T1: Introduction to NEON data and usability tools [~5 min]**
 - Overview of NEON data portal, NEON data skills, and NEON EC data (Chris)
- **T2: New budgeting approach reveals source of terrestrial carbon uptake overestimation [~5 min]**
 - Spatially representative fluxes introductory presentation (Anne Griebel)
- **T3: Flux data fusion for ecosystem understanding– flux fusion [~5 min]**
 - Phenocam/flux data fusion introductory presentation (Bijan Seyednasrollah)
- **T4: Developing end-to-end QAQC routines for flux observations – Tovi & Openeddy overview [~ 5 min]**
 - Tovi QAQC introductory presentation (Gerardo Fratini)
 - Openeddy QAQC introductory presentation (Ladislav Sigut)

Breakout groups (~90 min)

- **T1: Getting started with eddy4R hands-on tutorial (Lead by Dave)**
 - Eddy4R vignette was run in the Cyverse Discovery Environment
 - 6 participants in tutorial
 - Notes:
 - The Cyverse environment worked well
 - Participants also interest in NEON EC data exploration vignette for usability tools

What are the goals the breakout groups suggest?	What can the community do to achieve goals?	How can NEON help achieve goals?
Dedicated time devoted to tutorial (prior and during workshop)		Provide early notification and interact with subset of workshop participants interested in the tutorial separately
Bring the community cloud computing resources (CyVerse exploration)	Promote Cyverse as a community code integration platform	NEON could provide a consolidated "How to for CyVerse utilization"
Provide tools for the whole community [pyfluxpro (Python tools)]	Create python tools	Create a python wrapper for eddy4R or create some python tools to interact with NEON EC data (Jupyter notebook), highlight community created tools
Work with NEON data	Determine how they would like to interact with NEON SAE data, provide suggestions	Create usability tools to interact with NEON data (code, vignettes); create a survey to poll how the community wants to interact with HDF5 files
Harmonized data formats		Provide FLUXNET names as metadata in NEON HDF5

Lake buoy fluxes	Begin evaluating fluxes and use IMU to correct fluxes	IMU code publicly available
------------------	---	-----------------------------

- **T2: Scale-aware flux data products and integration with remote-sensing data (Lead by Stefan and Anne)**
 - Discussion centered around several goals
 - provide simple/little processed data products
 - how to use remote sensing and flux tower data together, e.g. to explain the impact of management practices using eddy covariance network footprints together with remote sensing map?
 - flux budgets that are spatio-temporally representative at the scale of remote-sensing/modeling data; impact of such equitable fluxes on energy balance; possibility to apply to gap-filling
 - space-time resolved flux (state variable?) maps around flux towers, to use together e.g. with point-based measurements, remote sensing data etc. for value-added / derived data products

What are the goals the breakout groups suggest?	What can the community do to achieve goals?	How can NEON help achieve goals?
Provide simple/little processed data products	Community can actively participate in exploring (alternative) approaches, carbon budgets	Provide simple and concisely documented "basic" data product, in addition to "expanded" data products
How to use remote sensing and flux tower data together, e.g. to explain the impact of management practices using eddy covariance network footprints together with remote sensing map?	A footprint tool / data product are starting point, but probably difficult for a majority of users to apply -derive end-to-end tool	A software that superimposes footprint map over land surface information (in general, incl. topography, soil types) Could use Google Earth Engine? Create temporal aggregates / quick-looks; e.g. footprint climatology over topography etc. Create useful source area representations / visualizations / products that are useful with non-standard species/aggregated sampling techniques
Flux budgets that are spatio-temporally representative at the scale of remote-sensing/modeling data; impact of such equitable fluxes on energy balance; possibility to apply to gap-filling	Make software (eddy4R.fast) accessible to the community	Adjust to eddy4R terminology, AGPLv3 license

Space-time resolved flux (state variable?) maps around flux towers, to use together e.g. with point-based measurements, remote sensing data etc. for value-added / derived data products	Make flux maps, and subsequently software (eddy4R.erf) accessible to the community	Determine a process to leverage network-level data product development that allows clarifying patent pending, code hardening, making it more efficient; code parameter sensitivity studies (spatial/temporal resolution); automated driver selection; team up between NEON and AmeriFlux (Housen POC)
--	--	---

- **T3: Fusing flux data with other data products (Lead by Chris and Natchaya)**
 - Discussion centered around mostly on how NEON could help facilitate data fusion, such as
 - Improvements to discoverability of NEON API information
 - Usability tools for NEON EC HDF5 data
 - Flux partitioned data products
 - Working toward easy integration with outside data products
 - OCO3 satellite NEON site cutouts
 - Phenocam derived products (greenness)

What are the goals the breakout groups suggest?	What can the community do to achieve goals?	How can NEON help achieve goals?
Information about NEON API is difficult to find		
Additional data that can be derived from phenocam images		Algorithm to extract snow depth (read the numbers above the snow, we might already have this)
Difficult to access one specific data product in the h5 file		NEON provides usability documentation for HDF5 files
Keep soil CO2 data as an add on to keep things as separate variables, so people don't compare things that aren't comparable		Keep documentation clear, and not get in the way of AmeriFlux intercomparison
Flux partitioning	Flux partitioning	Implement in future already exists as feature "Annual flux data product: Calculate annual NEE, GPP, and Re"
Validation of neon data (intercomparison of NEON with other networks)		Implement drift correction to improve cross comparison between
OCO3 satellite will focus on NEON sites (launch in March), SIF is important to integrate with the remote sensing community		interesting synthesis with flux data

eddy4R improvements		metScanR, add functionality to work for EC. Would be nice to display co-located sites from other networks on the neon website Make sure updates to naming conventions don't break anyone's code (add new variables to the end of the data frame) Add plotting function to display key data to eddy4R, ideally one that can compare multiple sites
---------------------	--	--

- **T4: QAQC routines for EC and Met (Lead by Cove)**
 - Discussions centered around addressing the following:
 - What are the tendencies of a QC test to remove certain kinds of data. Where does it do well? Where does it fail? What are the systematic problems with specific tests at certain site types.
 - Can we develop a scheme for all sites, or how do sites need to be characterized to know what tests are applicable (e.g. forests, tall tower, etc?). How about developing a decision tree for what tests to employ, given the measurement and site type.
 - Define a metric that characterizes the efficiency of QC tests. Efficiency is defined as the balance between the amount of removed data compared to how much that loss of data increases uncertainty.
 - What variables do we need to be reporting? e.g. wbar residual for planar fit. Need to balance amount of variables reported with needs for QC. Don't want data deluge.
 - The group thought that the community could perform the analysis and NEON could help by providing the standardized data to perform the analysis

What are the goals the breakout groups suggest?	What can the community do to achieve goals?	How can NEON help achieve goals?
Evaluate the question: What are the tendencies of a QC test to remove certain kinds of data. Where does it do well? Where does it fail? What are the systematic problems with specific tests at certain site types.	Tons of data out there. Do analysis.	Standardized data for testing - applicable for all of these
Evaluate the question: Can we develop a scheme for all sites, or how do sites need to be characterized to know what tests are applicable (e.g. forests, tall tower, etc?).	Tons of data out there. Do analysis.	Standardized data for testing - applicable for all of these. Fits into site-specific parameterization

How about developing a decision tree for what tests to employ, given the measurement and site type.		feature, this could be used as a way to communicate and visualize
Define a metric that characterizes the efficiency of QC tests. Efficiency is defined as the balance between the amount of removed data compared to how much that loss of data increases uncertainty.	Tons of data out there. Do analysis.	Standardized data for testing - applicable for all of these
What variables do we need to be reporting? e.g. wbar residual for planar fit. Need to balance amount of variables reported with needs for QC. Don't want data deluge.	Tons of data out there. Do analysis.	Standardized data for testing - applicable for all of these. Can add to expanded file to avoid data deluge in the basic file. TWG can be used to advise what is important to add

Breakout summaries (~15 min)

- We ran out of time to come back as a group for breakout summaries
- However, consolidated notes from breakouts were taken and are presented in above tables.
- The groups gathered input on the following questions:
 - What are the goals the breakout groups suggest?
 - What can the community do to achieve goals?
 - How can NEON help achieve goals?
- Wrap-up
 - The insights from the meeting were incorporated into the NEON SAE backlog
 - These suggestions will be prioritized along with other user stories by the NEON SAE Technical Working Group (TWG)
 - These suggestions and input ultimately steer NEON SAE developments