ONEFlux Overview

Gilberto Pastorello, Carlo Trotta, Alessio Ribeca, Dario Papale,
Abdelrahman Elbashandy, Alan Barr, Markus Reichstein,
Gitta Lasslop, Deb Agarwal, Sebastien Biraud, et al.

2019 AmeriFlux Data-Tech Workshop

October 29-30, Berkeley, CA

Outline

- **Background:** History of FLUXNET datasets
- **Context:** Eddy covariance data processing
- **What’s included:** Overview of ONEFlux
- **(brief) How-To:** Running ONEFlux
- **Future:** Next steps for network datasets

FLUXNET Datasets

- Marconi 2000: 38 sites / 97 site-years
  - Data contributed by attendees of workshop
  - Access limited to attendees

- LaThuile 2007: 252 sites / 965 site-years
  - Data contributed by many site PIs
  - Start of uniform code base and processing for comparable data products
  - Data access/usage opened to other users, with some restrictions (three tiers)
  - Over 500 users

- FLUXNET2015: 212 sites / 1532 site-years
  - Standardized code base (developed in collaboration among networks)
  - Extensive QA/QC of data (data more comparable and usable, but fewer sites)
  - Data access to anyone, usage has fewer restrictions (two tiers) – tools to keep track of usage
  - Over 2,200 users (as of Sep 10, 2019)

Eddy Covariance Data Processing

- High-frequency data
  - EddyPro (LI-COR),
  - EdiRE (U. Edinburgh),
  - Alteddy (ALTERRA),
  - eddy4R (NEON)
  - Custom codes
- Data products from fluxes
  - USTAR filtering, met/flux gap-filling, flux partitioning, uncertainty estimates, etc.
  - REddyProc (MPI),
  - ONEFlux (AMP-EUDB)
  - Custom codes

Eddy Covariance Data Processing

- High-frequency data
  - EddyPro (LI-COR),
  - EdiRE (U. Edinburgh),
  - Alteddy (ALTERRA),
  - eddy4R (NEON)
  - Custom codes
- Data products from fluxes
  - USTAR filtering, met/flux gap-filling, flux partitioning, uncertainty estimates, etc.
  - REddyProc (MPI),
  - ONEFlux (AMP-EUDB)
  - Custom codes

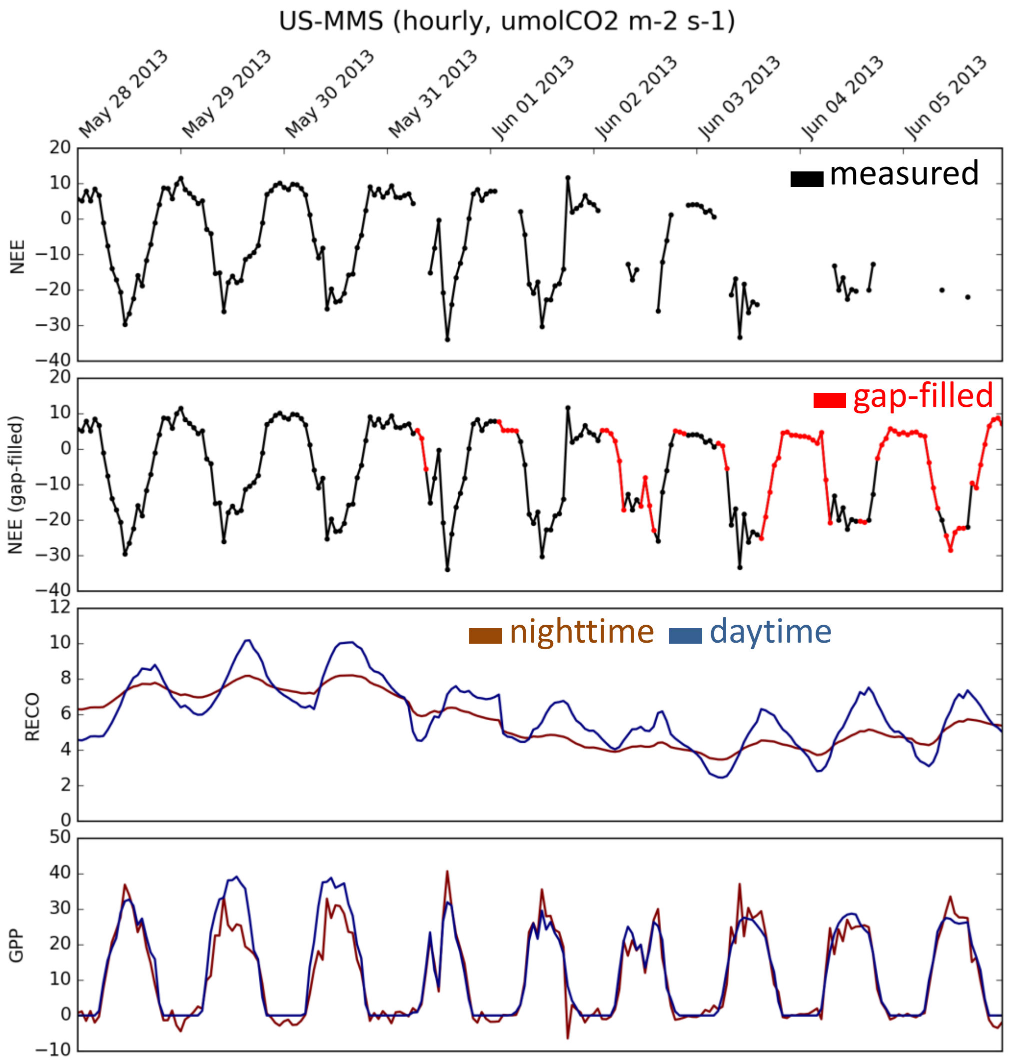
Why yet another software for eddy covariance processing?

- Extensive validation on many site characteristics
- Focus on creating network-level data products (standardized for synthesis studies and cross-site comparisons)
- Not (yet) a software (for users), more of a collection of packages

ONEFlux

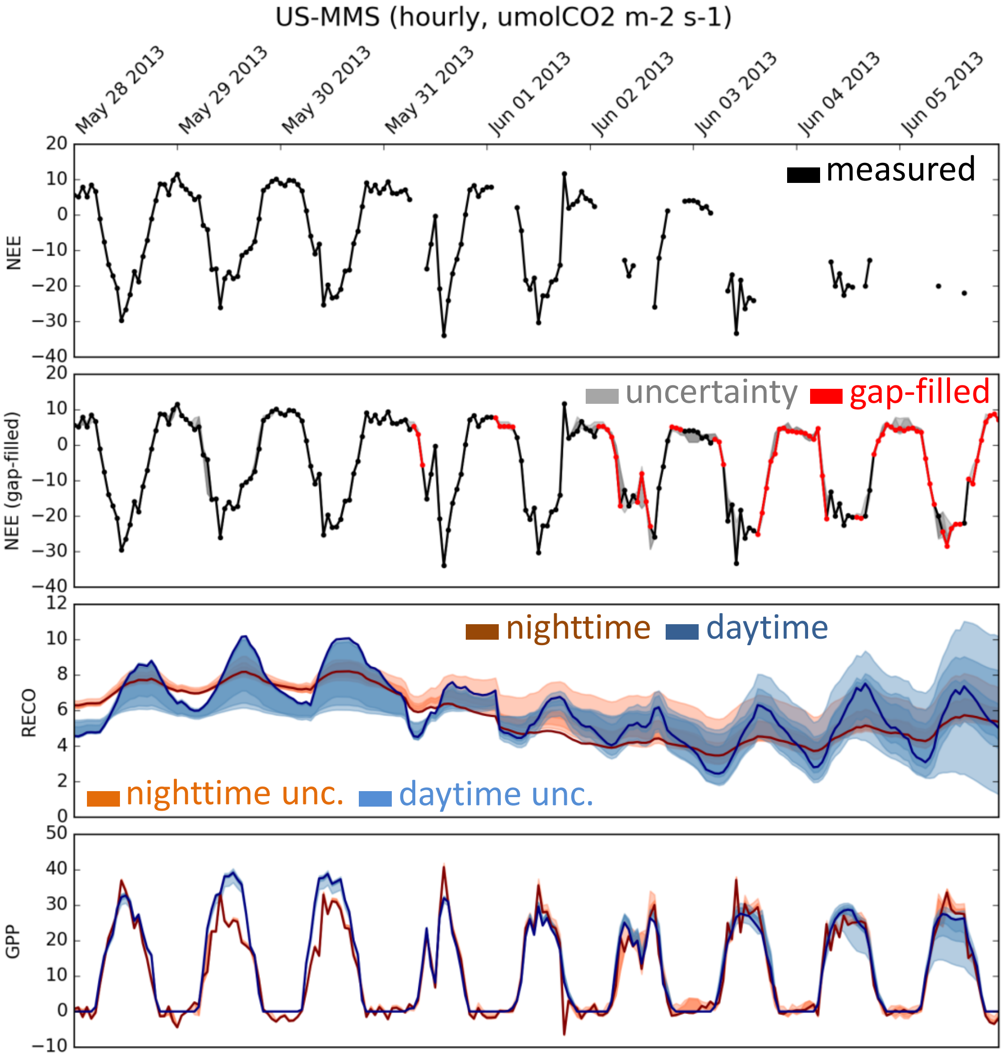
Data Products

- gap-filling
- CO_2 flux partitioning



ONEFlux
Data Products

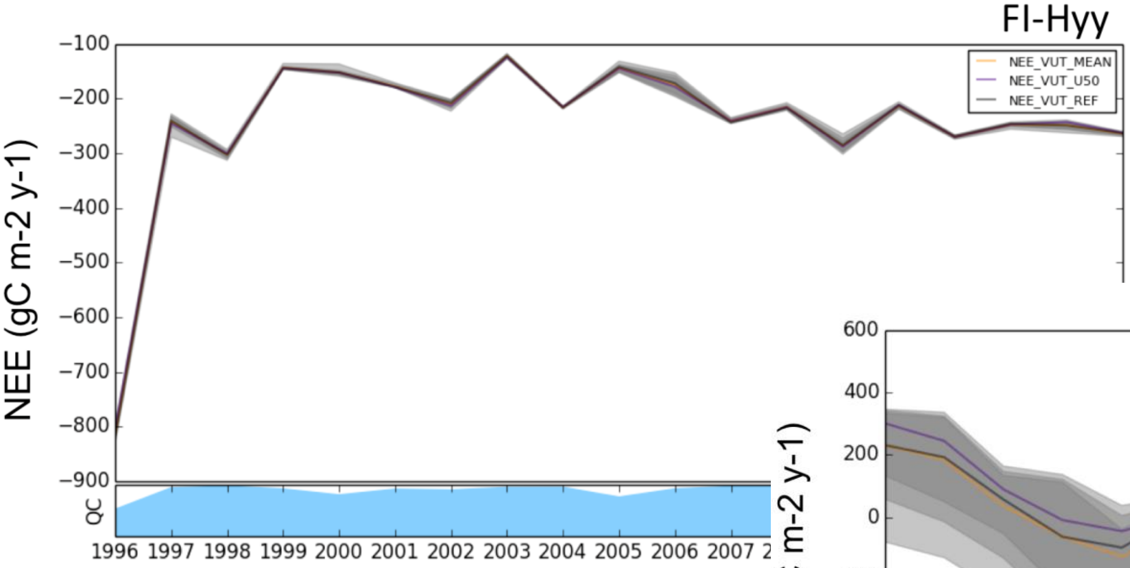
- gap-filling
- CO_2 flux partitioning
- **uncertainty**
  - **USTAR**
  - **"Model"**

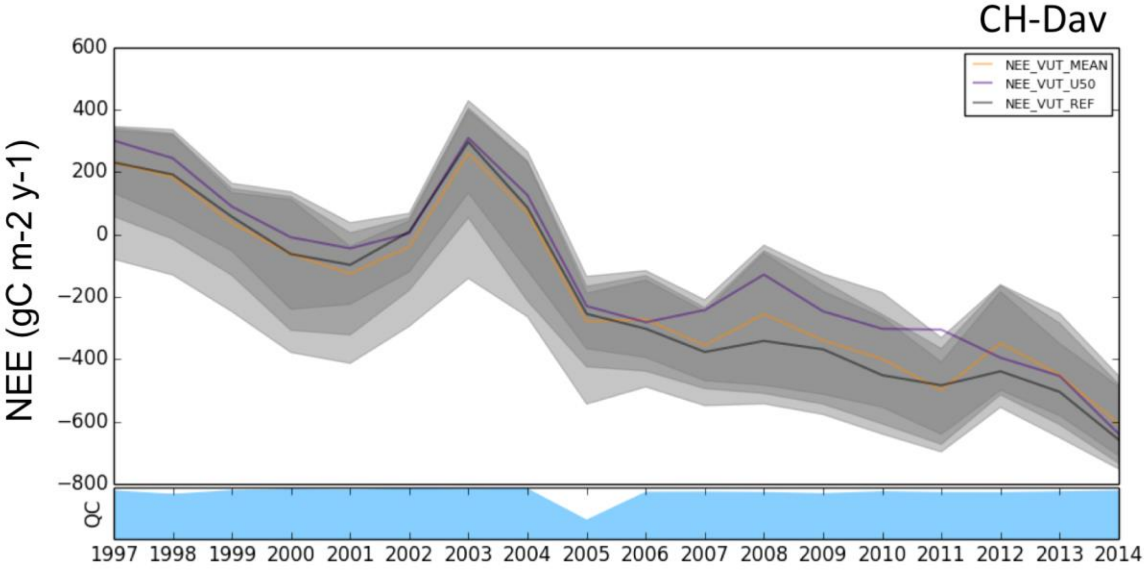


ONEFlux

Data Products

Uncertainty from USTAR filtering



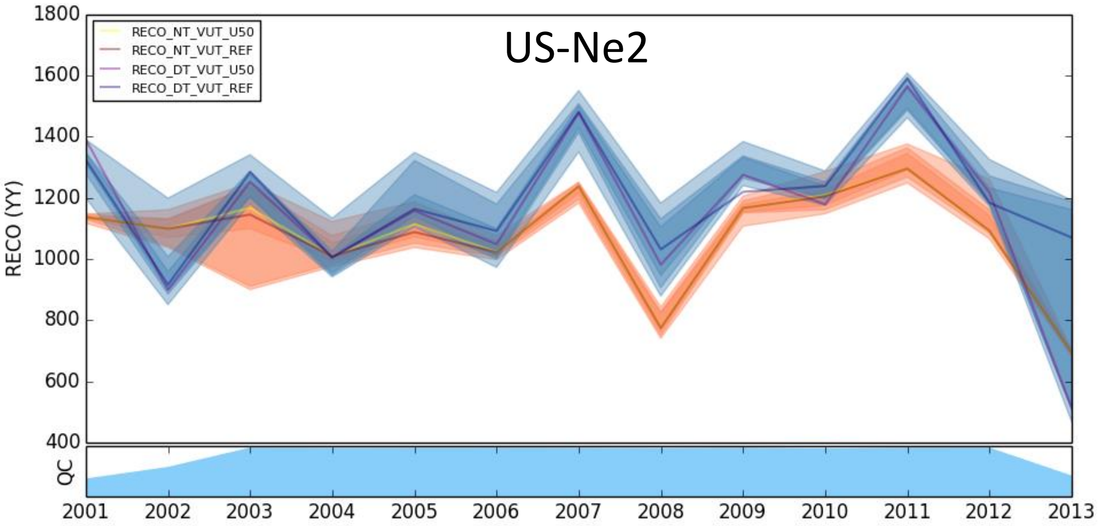


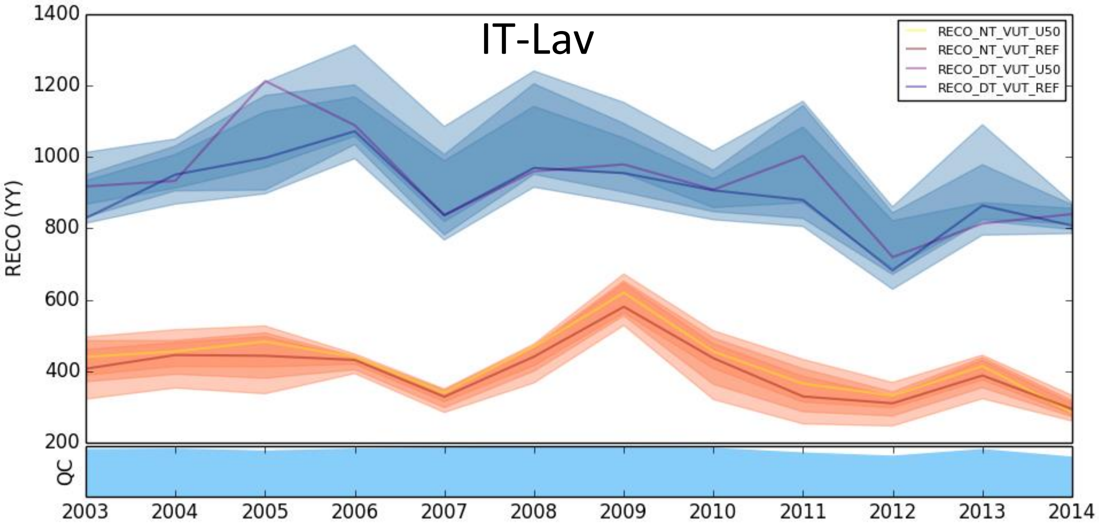
NEE

ONEFlux

Data Products

RECO





Uncertainty from partitioning methods helps to evaluate site consistency

Processing Steps

SITEDIR

- 01_qc_visual/qcv_files/
- 02_qc_auto/
- *
- 04_ustar_mp/
- 05_ustar_cp/
- 06_meteo_era/ *(optional)*
- 07_meteo_proc/
- 08_nee_proc/
- 09_energy_proc/
- 10_nee_partition_nt/
- 11_nee_partition_dt/
- 12_ure_input/
- 12_ure/
- * * *
- 99_fluxnet2015/

- 01: input (not-quite FP-Standard)
- 02: flagging/filtering
- 04: USTAR threshold method 1 (unc.)
- 05: USTAR threshold method 2 (unc.)
- 06: downscaling for met gapfilling
- 07: met gapfilling (MDS + downscaling)
- 08: NEE filter/storage/unc./ref/gapfilling
- 09: H & LE corr. factors/gapfilling/unc.
- 10: NEE partitioning (nighttime method)
- 11: NEE partitioning (daytime method)
- 12 (i): create input for 12
- 12: unc. from multiple USTAR thresholds
- 99: merge into FP-Standard compliant output

http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/

ONEFlux Required Variables

- *$\mathbf{CO_2}$ (µmolCO_2 mol-1): Carbon Dioxide (CO_2) mole fraction in wet air
- ***FC** (µmolCO_2 m-2 s-1): Carbon Dioxide (CO_2) turbulent flux, without storage correction
- ***SC** (µmolCO_2 m-2 s-1): Carbon Dioxide (CO_2) storage flux, optional if tower shorter than 3m
- **NEE** (µmolCO_2 m-2 s-1): net ecosystem exchange, used for validation only of NEE computed from FC and SC
- ***H** (W m-2): sensible heat turbulent flux, without storage correction
- ***LE** (W m-2): latent heat turbulent flux, without storage correction
- ***WS** (m s-1): wind speed
- ***USTAR** (m s-1): friction velocity
- ***TA** (degrees C): air temperature
- ***RH** (%): relative humidity (range 0-100)
- ***PA** (kPa): atmospheric pressure
- **G** (W m-2): soil heat flux, not mandatory, but no energy balance calculations if missing
- **NETRAD** (W m-2): net radiation, not mandatory, but no energy balance calculations if missing
- ***SW_IN** (W m-2): potential incoming shortwave radiation; if missing, can be calculated from PPFD_IN, but introduces seasonal error
- **SW_IN_POT** (W m-2): potential incoming shortwave radiation (top of atmosphere theoretical maximum radiation), calculated based on the site coordinates
- **PPFD_IN** (µmolPhoton m-2 s-1): incoming photosynthetic photon flux density
- **P** (mm): precipitation
- **LW_IN** (W m-2): incoming longwave radiation
- **VPD** (hPa): vapor pressure deficit, used for validation of VPD computed from TA and RH
- **SWC** (%): soil water content (volumetric), range 0-100
- **TS** (degrees C): soil temperature

Running ONEFlux

```
usage: runoneflux.py [-h] [--perc [PERC [PERC ...]]]
                     [--prod [PROD [PROD ...]]] [-l LOGFILE] [--force-py]
                     [--mcr MCR_DIRECTORY] [--ts TIMESTAMP] [--recint {hh,hr}]
                     COMMAND DATA-DIR SITE-ID SITE-DIR FIRST-YEAR LAST-YEAR

positional arguments:
  COMMAND               ONEFlux command to be run [all, partition_nt, partition_dt]
  DATA-DIR              Absolute path to general data directory
  SITE-ID               Site Flux ID in the form CC-XXX
  SITE-DIR              Relative path to site data directory (within data-dir)
  FIRST-YEAR            First year of data to be processed
  LAST-YEAR             Last year of data to be processed

optional arguments:
  -h, --help            show this help message and exit
  --perc [PERC [PERC ...]]
                        List of percentiles to be processed
  --prod [PROD [PROD ...]]
                        List of products to be processed
  -l LOGFILE, --logfile LOGFILE
                        Logging file path
  --force-py            Force execution of PY partitioning (saves original
                        output, generates new)
  --mcr MCR_DIRECTORY   Path to MCR directory
  --recint {hh,hr}      Record interval for site
```

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

$ python runoneflux.py **all** \
"../datadir/" US-ARc "US-ARc_sample_input" \
2005 2006 -l fluxnet_pipeline_US-ARc.log \
--mcr ~/bin/matlab/v94/ --recint hh

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory (within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

```
$ python runoneflux.py all                            \
    "../datadir/" US-ARc "US-ARc_sample_input"      \
    2005 2006 -l fluxnet_pipeline_US-ARc.log       \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

```
$ python runoneflux.py all                                 \
    "../datadir/" US-ARc "US-ARc_sample_input"     \
    2005 2006 -l fluxnet_pipeline_US-ARc.log       \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

$ python runoneflux.py **all** \
"../datadir/" US-ARc "US-ARc_sample_input" \
2005 2006 -l fluxnet_pipeline_US-ARc.log \
--mcr ~/bin/matlab/v94/ --recint hh

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

```
$ python runoneflux.py all                                    \
    "../datadir/" US-ARc "US-ARc_sample_input"     \
    2005 2006 -l fluxnet_pipeline_US-ARc.log       \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

```
$ python runoneflux.py all                              \
    "../datadir/" US-ARc "US-ARc_sample_input"     \
    2005 2006 -l fluxnet_pipeline_US-ARc.log        \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Running ONEFlux

COMMAND   ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR   Absolute path to general data directory SITE-ID   Site Flux ID in the form CC-XXX
SITE-DIR   Relative path to site data directory
(within data-dir)
FIRST-YEAR   First year of data to be processed
LAST-YEAR   Last year of data to be processed

Example:

```
$ python runoneflux.py all                              \
    "../datadir/" US-ARc "US-ARc_sample_input"      \
    2005 2006 -l fluxnet_pipeline_US-ARc.log        \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Running ONEFlux

COMMAND ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR Absolute path to general data directory SITE-ID Site Flux ID in the form CC-XXX
SITE-DIR Relative path to site data directory
(within data-dir)
FIRST-YEAR First year of data to be processed
LAST-YEAR Last year of data to be processed

Example:

```
$ python runoneflux.py all \
    "../datadir/" US-ARc "US-ARc_sample_input" \
    2005 2006 -l fluxnet_pipeline_US-ARc.log \
    --mcr ~/bin/matlab/v94/ --recint hh
```

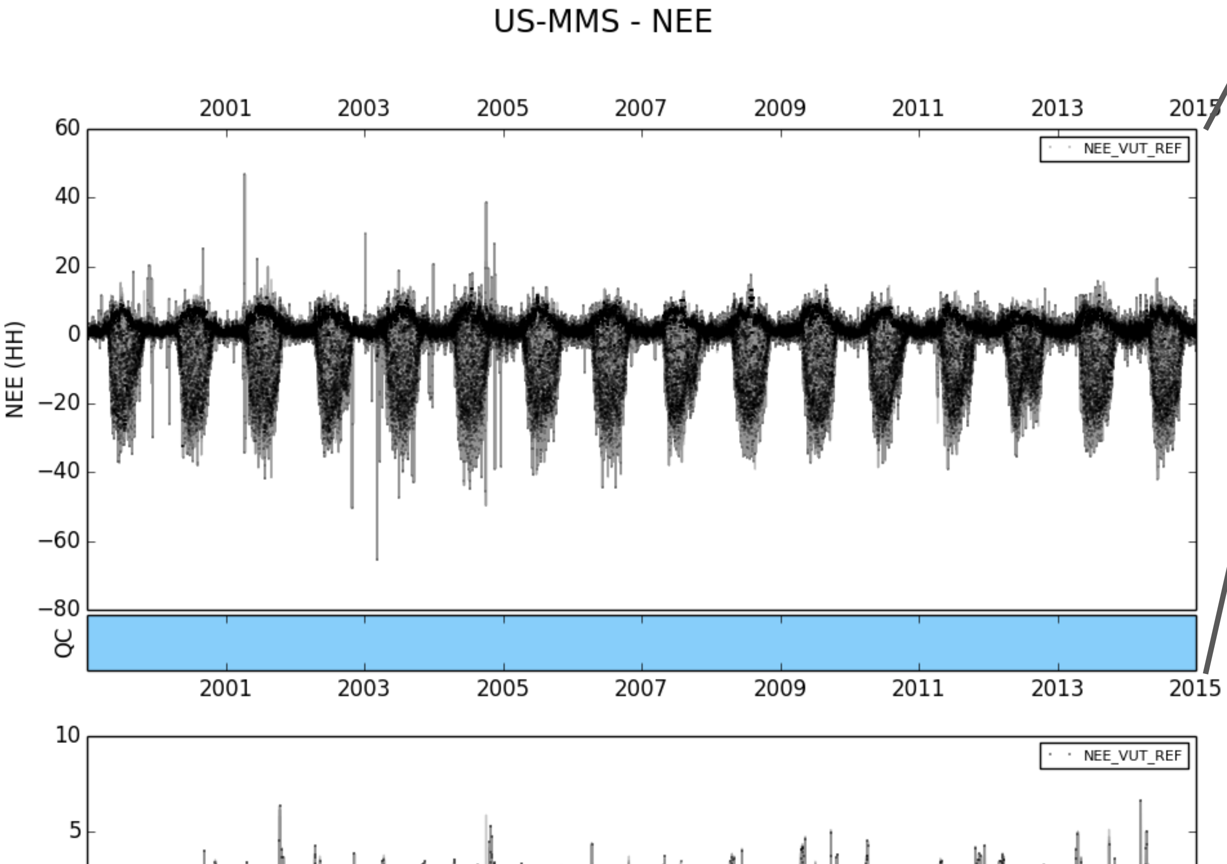
Running ONEFlux

COMMAND   ONEFlux command to be run
[all, partition_nt, partition_dt]
DATA-DIR   Absolute path to general data directory SITE-ID   Site Flux ID in the form CC-XXX
SITE-DIR   Relative path to site data directory
(within data-dir)
FIRST-YEAR   First year of data to be processed
LAST-YEAR   Last year of data to be processed

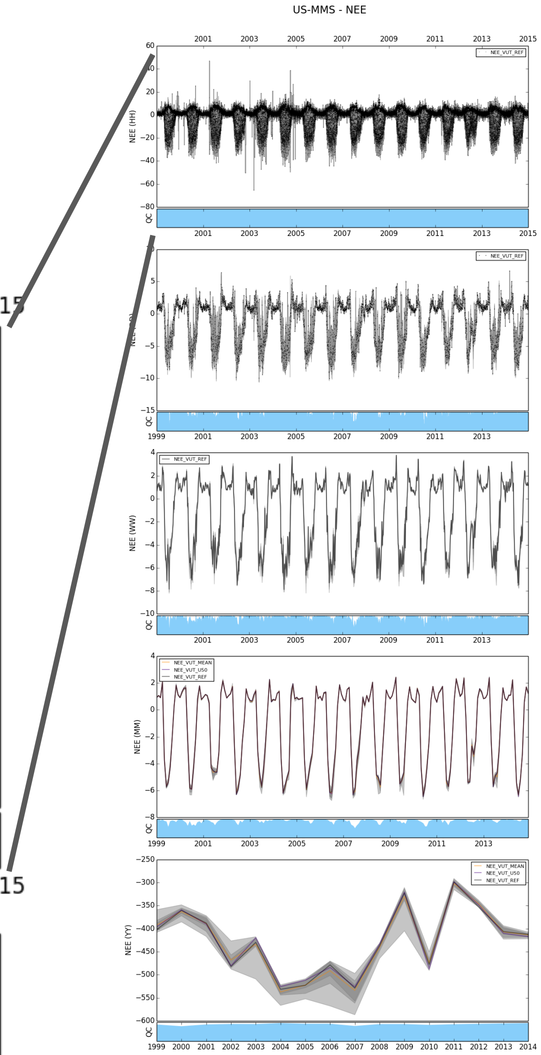
Example:

```
$ python runoneflux.py all                          \
    "../datadir/" US-ARc "US-ARc_sample_input"      \
    2005 2006 -l fluxnet_pipeline_US-ARc.log        \
    --mcr ~/bin/matlab/v94/ --recint hh
```

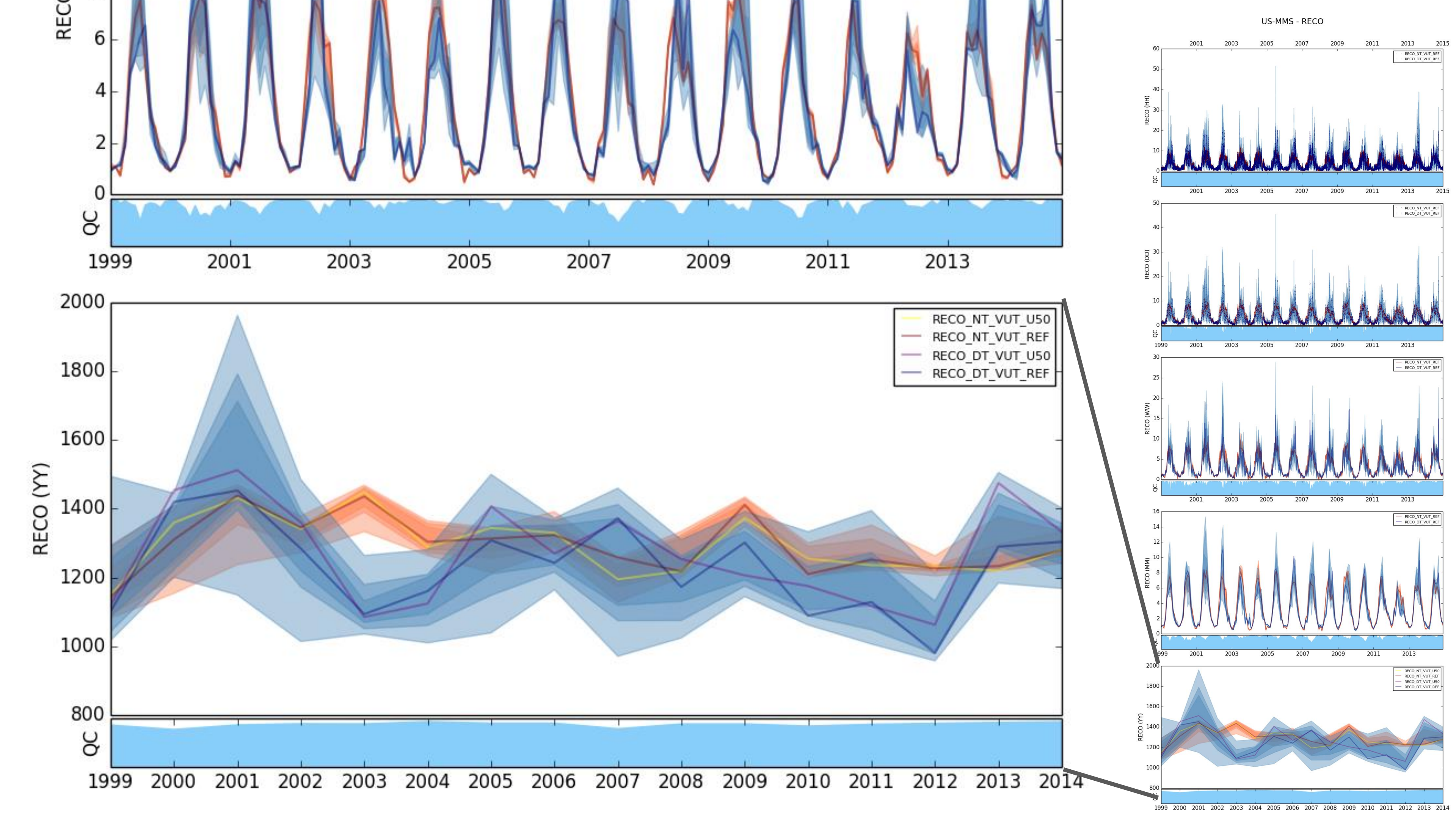
Results



US-MMS - NEE
NEE_VUT_REF
NEE (HH)
QC
NEE_VUT_REF



US-MMS - NEE
NEE (HH)
NEE (WW)
NEE (MM)
NEE (YY)
QC


US-MMS - RECO
RECO_NT_VUT_U50
RECO_NT_VUT_REF
RECO_DT_VUT_U50
RECO_DT_VUT_REF
RECO (YY)
RECO (HH)
RECO (DD)
RECO (WW)
RECO (MM)
QC

From individual to collaborative development

- Shared code repository
  - all codes in common structure
  - around 70k lines of code (v0.3.0-beta)
- Shared documentation
  - online documentation creation
- Execution and testing environments
  - replicated across partners
  - from supervised & manual execution (days to weeks to run a site)
  - to automated pipeline (average 4h per site-year)
- Validation and testing protocols
  - around 350k lines of code
  - 280k lines for validation, testing, and comparisons
  - reproducibility of results key; for each code change: matching data products or full trace of differences in data

ONEFlux coding team

Next Steps: Datasets

- Regional networks: continuous processing
  - AmeriFlux and European Cluster DB (ICOS), plus OzFlux & LBA
  - Execution of same code in multiple institutions, but exactly same data product
  - Can be used as “proxy” for global product (augmenting FLUXNET2015, fully compatible)
- Global product: approach TBD

Next Steps: Codebase

- First priority: new data products
- Improvements to code:
  - bug fixes, performance, use of all steps, input format… (i.e., general “caveat-reduction”)
  - “new” features / algorithm improvements
  - usability for end users (?)
- Development process: moving to full-fledged open-source-style code development

Questions? Comments? Suggestions?

contact:
ameriflux-support@lbl.gov
fluxdata-support@fluxdata.org
gzpastorello@lbl.gov

code:
https://github.com/ameriflux/oneflux