ONEFlux Processing Pipeline:
Network-collaborative software development

Gilberto Pastorello, Carlo Trotta, Alessio Ribeca, Dario Papale, Abdelrahman Elbashandy, Alan Barr

Markus Reichstein, Gitta Lasslop, Deb Agarwal, Sebastien Biraud, et al.

Outline

- **Background:** Evolution 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 (as of Sep 7, 2018)

- 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 1600 users (as of Sep 7, 2018)

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

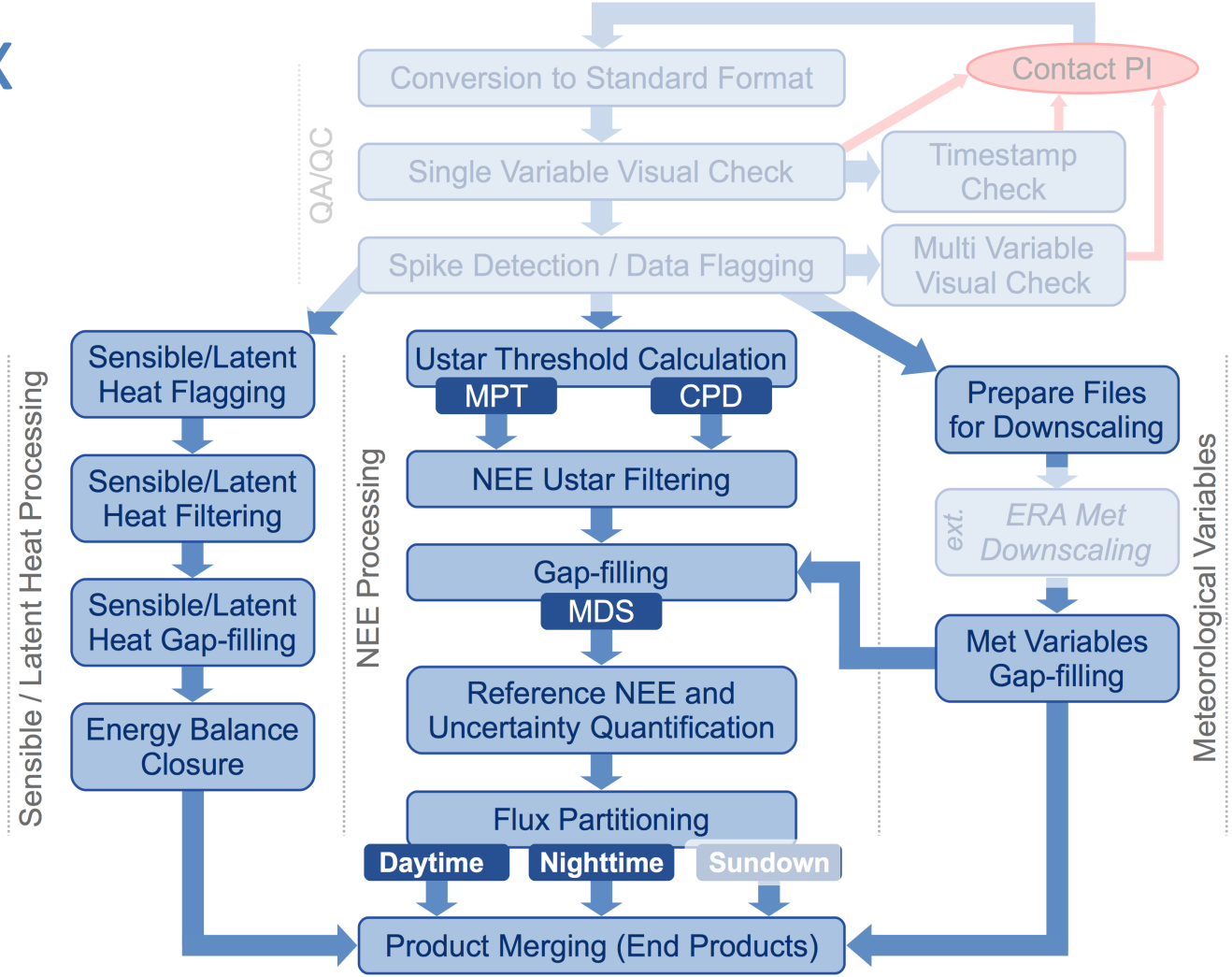
Eddy Covariance Data Processing

- 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?

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

ONEFlux Pipeline



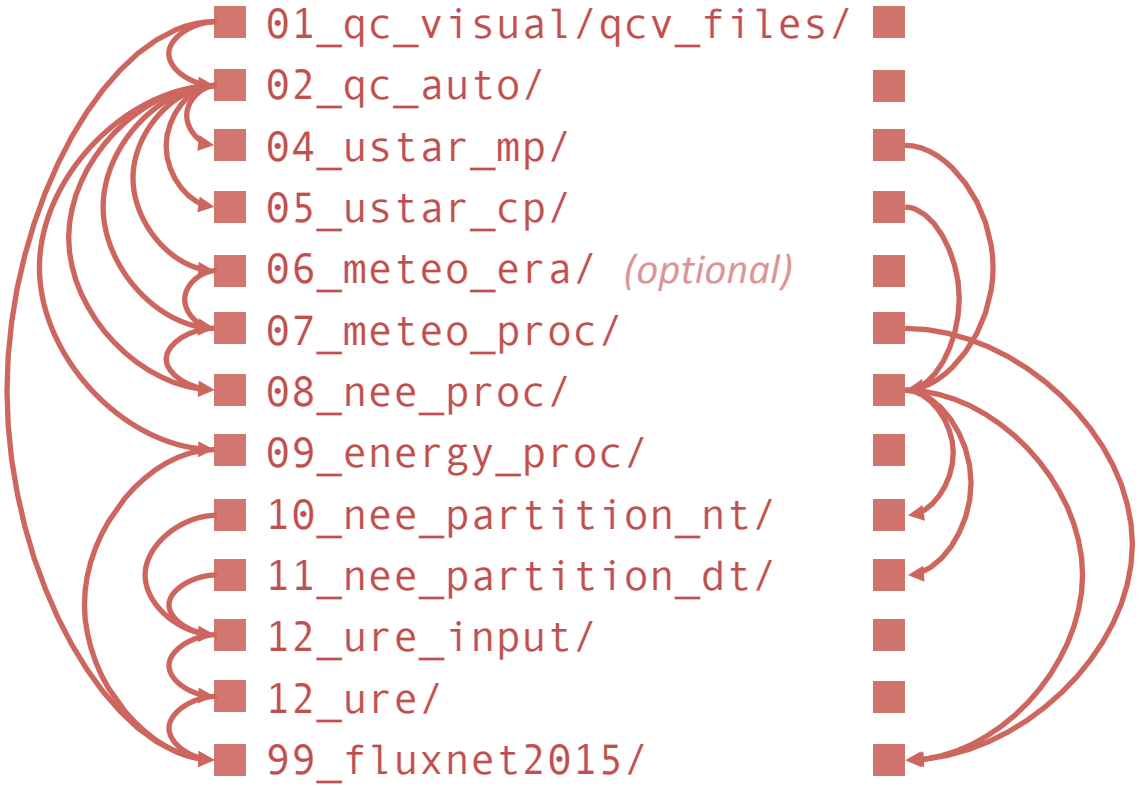
Processing Steps

```
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/

Processing Steps: Dependencies



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
```

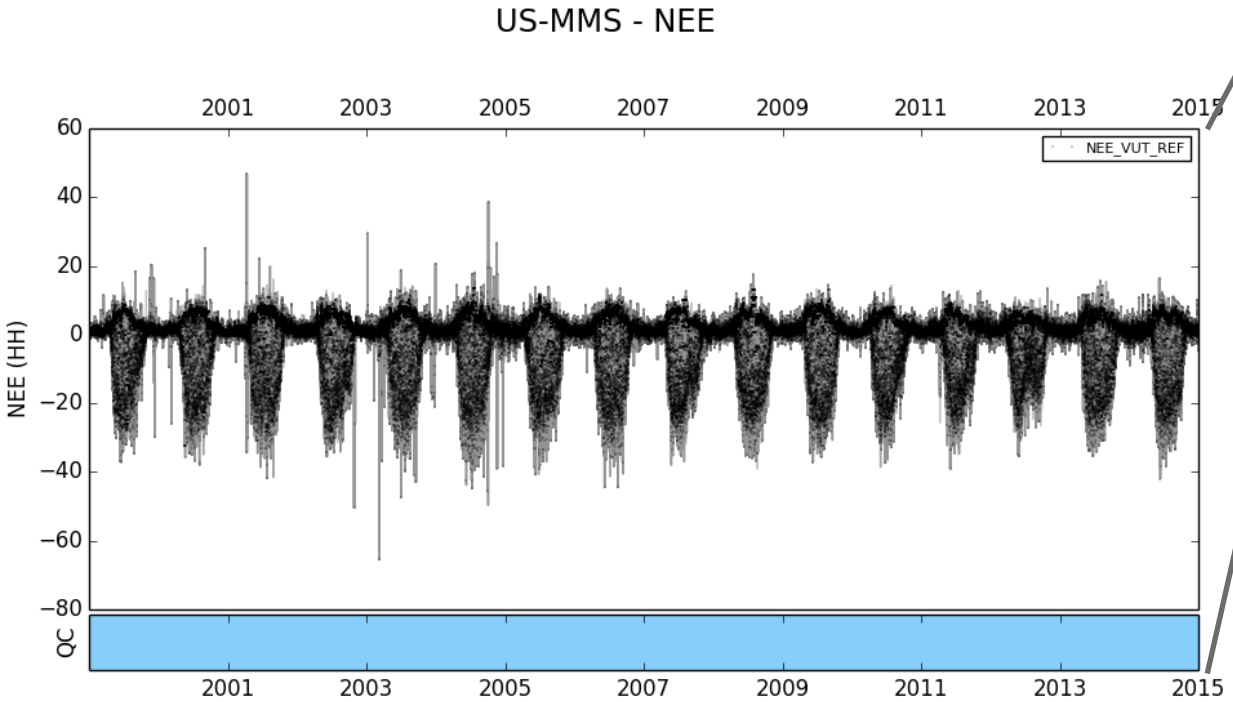
Running ONEFlux

```
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
```

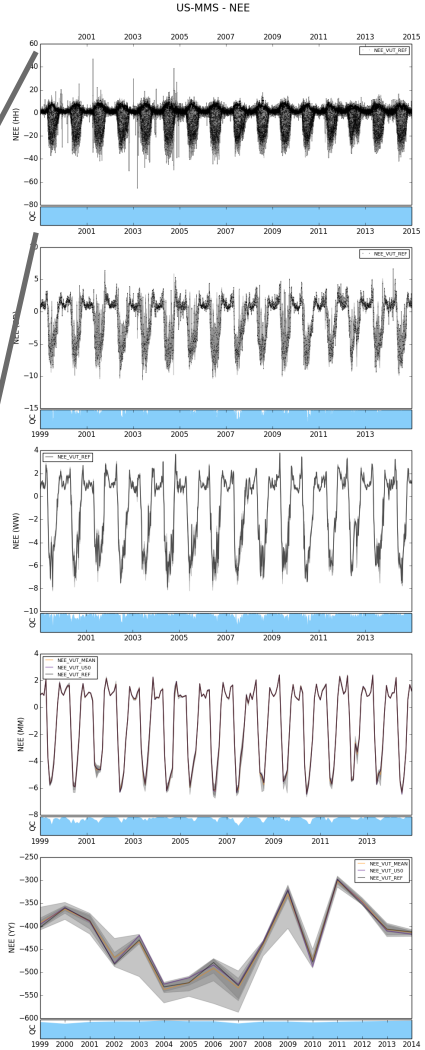
Example:

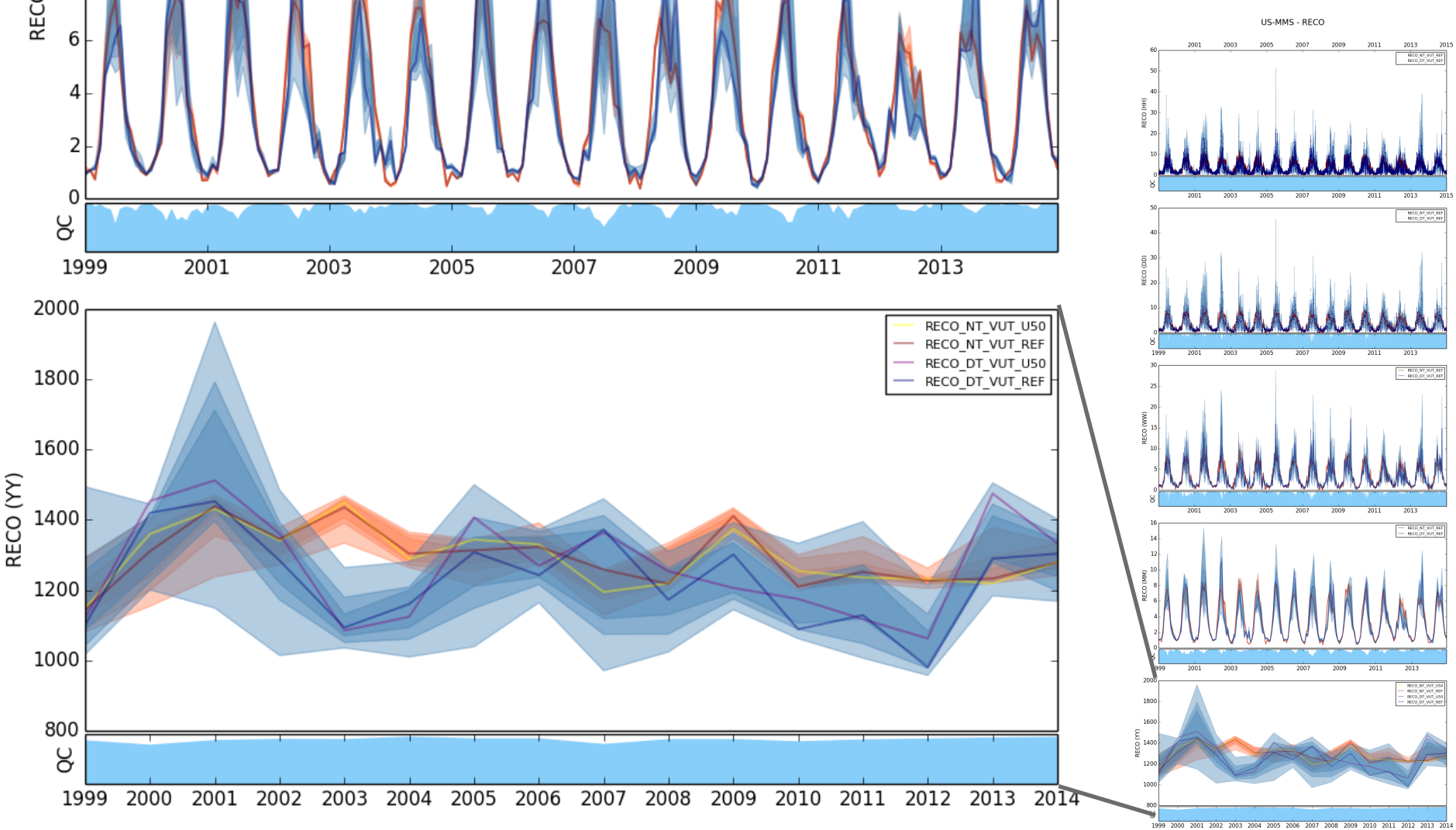
```
$ python runoneflux.py all                          \
    "../datadir/" US-ARc "US-ARc_sample_input"      \
    2005 2006 -l fluxnet_pipeline_US-ARc.log        \
    --mcr ~/bin/matlab/v94/ --recint hh
```

Results


US-MMS - NEE
NEE_VUT_REF
NEE (HH)
QC
2001
2003
2005
2007
2009
2011
2013
2015

NEE_VUT_REF


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


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

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

AmeriFlux

Gilberto Pastorello

Abdelrahman Elbashandy

ICOS

European Fluxes Database Cluster

Carlo Trotta

Alessio Ribeca

Dario Papale

Environment and Climate Change Canada

Alan Barr