High frequency data processing: Lessons learned and best practices

Motivation

- Mistakes happen!

- It is easy to ‘compute’ a flux but how does one know it’s right?
  - Prevalence of software packages
  - Overwhelming volumes of data
  - Endless number of corrections

- Things change
  - New science
  - Data network requirements
  - Funding agency/publication stipulations

Step 1: be prepared

Assumption: (Re)processing happens…

…make it easier.

- Organize data
  - Use systematic approach! (file naming, folder structure, data format)
  - Preserve original data sources
  - Don’t ignore ancillary data and metadata
  - Know what was previously done!
    - Data archeology is really hard…
    - Can you reproduce past results?
  - Document, document, document
    - Record metadata (data processing steps)

Software considerations

- Is there technical support available? Documentation?
- Is the code base updated regularly? Are new corrections available?
- Is the code openly available? Is it a ‘black box’?
- Are results reproducible?

- Homebrew code?
  - Has it been evaluated against other codes?
  - Is there a person in charge of maintaining the code?

Examples

Combined experience from AMP site visits, site data reviews, and as site staff.

- Gill ‘w-boost’ issue
- Issues in raw data (x2)
- Gaps in processed data
- Time shifts
- Lag times
- Spectral analysis
- Spectral corrections
- US-ARM

Ex1: Gill ‘w-boost’ issue

First example is a specific case of a general problem: a new finding requires data reprocessing.

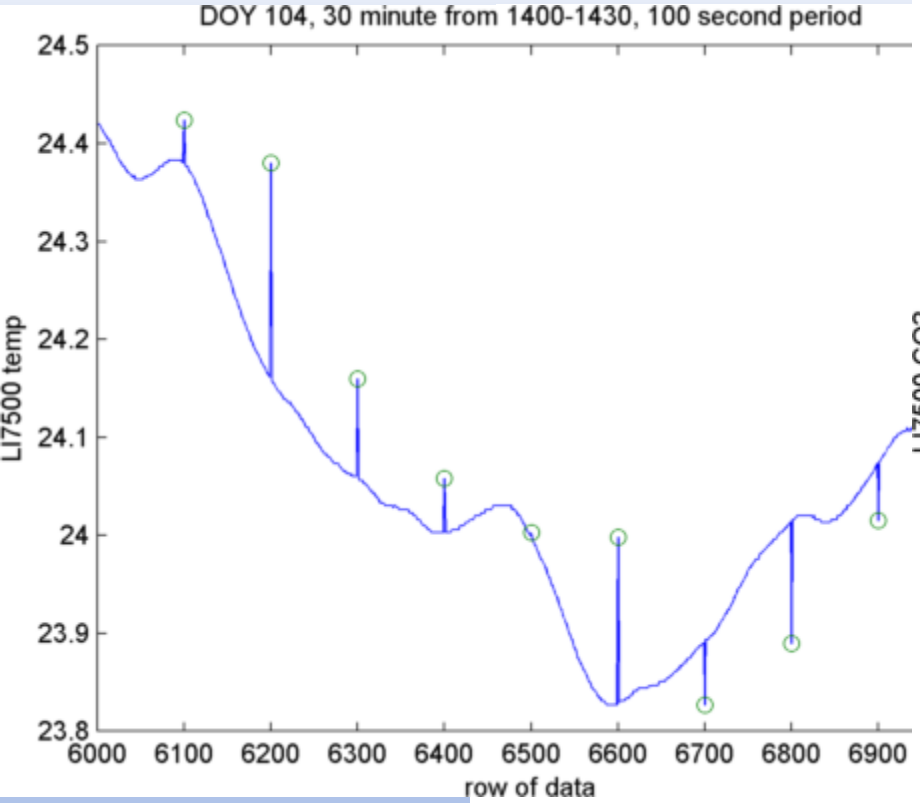
- A firmware problem was found that impacted some Gill sonic anemometers. A calibration factor was missing from the vertical wind (asymmetric).
- Flux impacts averaged 18% (See Billesbach et al., 2019, AFM).

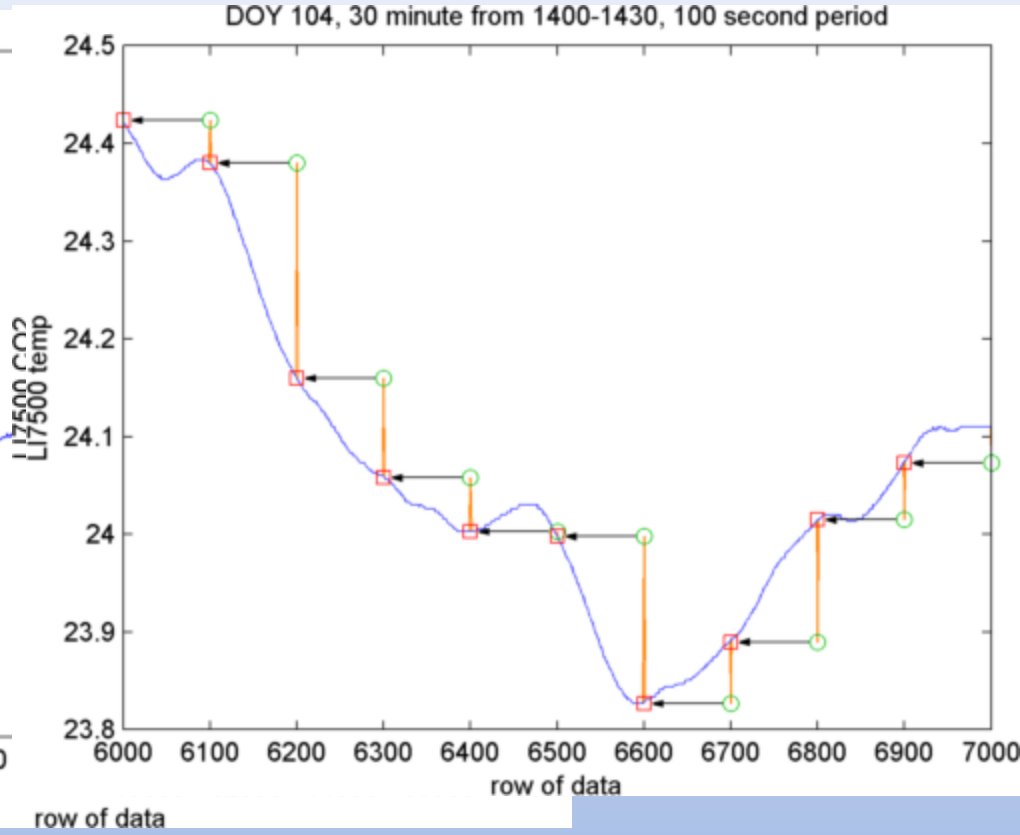
Takeaways:

- Issue was specific to firmware (not just make/model/SN) on instrument (requires good metadata)
- Raw data required to properly correct the issue
- Impacts on other published corrections (Angle of attack)

Ex2a: Issues in raw data

Don’t ignore the raw data! Look at it.



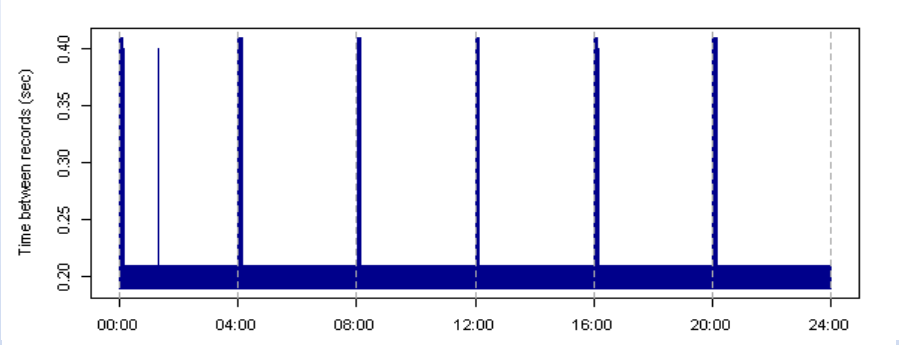


Examine all timescales and all variables recorded

Correcting issues with high frequency data -> reprocessing

Ex2b: Issues in raw data

Don’t ignore the raw data! Look at it.



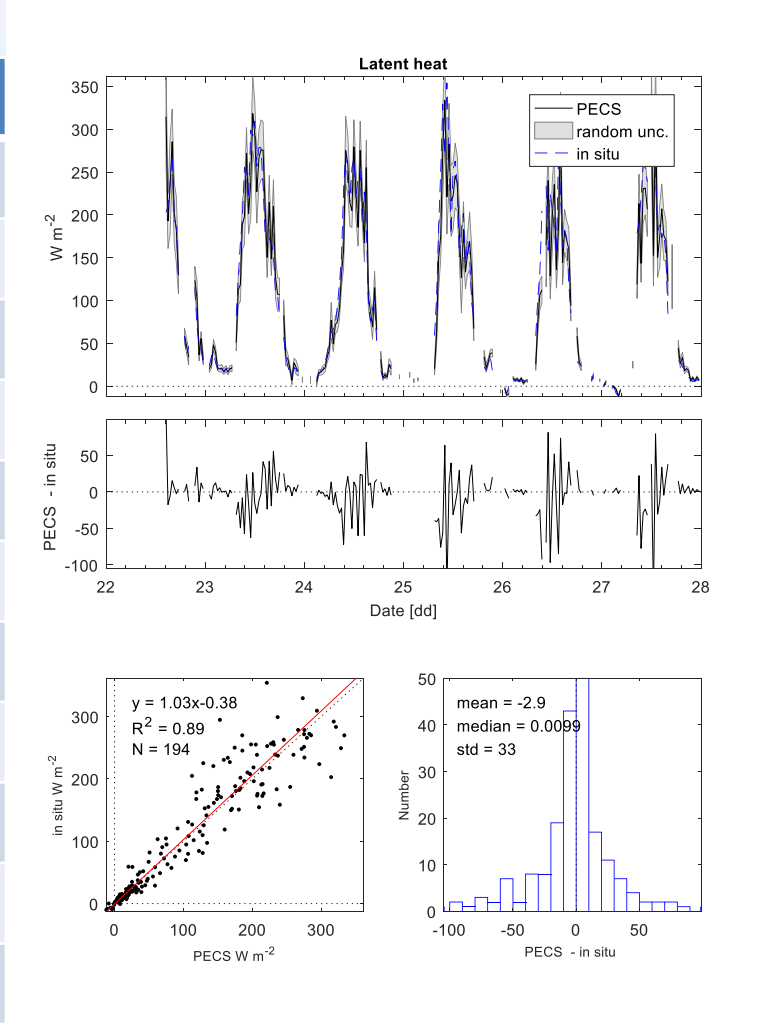
Missing records at the start of each logger file.

Know your software:
EddyPro does not detect timestamps within high frequency data (uses filename). Therefore, these gaps effectively impose a time shift in the processed data.

Ex3: Gaps in processed data

Missing

R1
R2
R3
R4
R5
R6
R7
R8
R9
…
R864000



C4

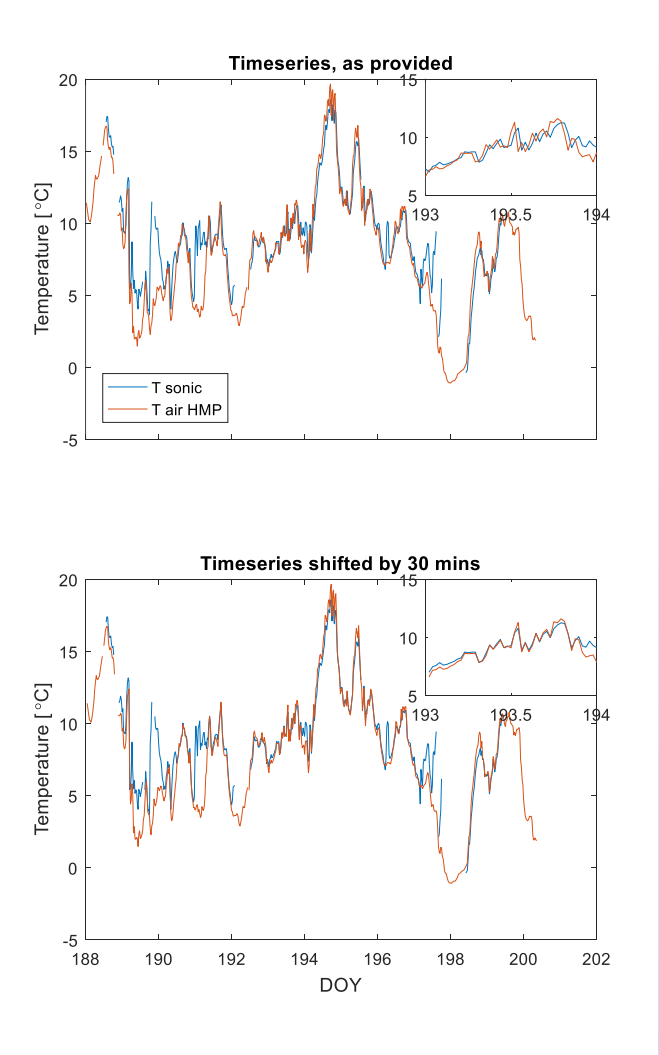
energy flux [W m-2]

Processing errors can be hard to detect!

Don’t ignore the ‘unusual’ data

Ex4: Time shifts

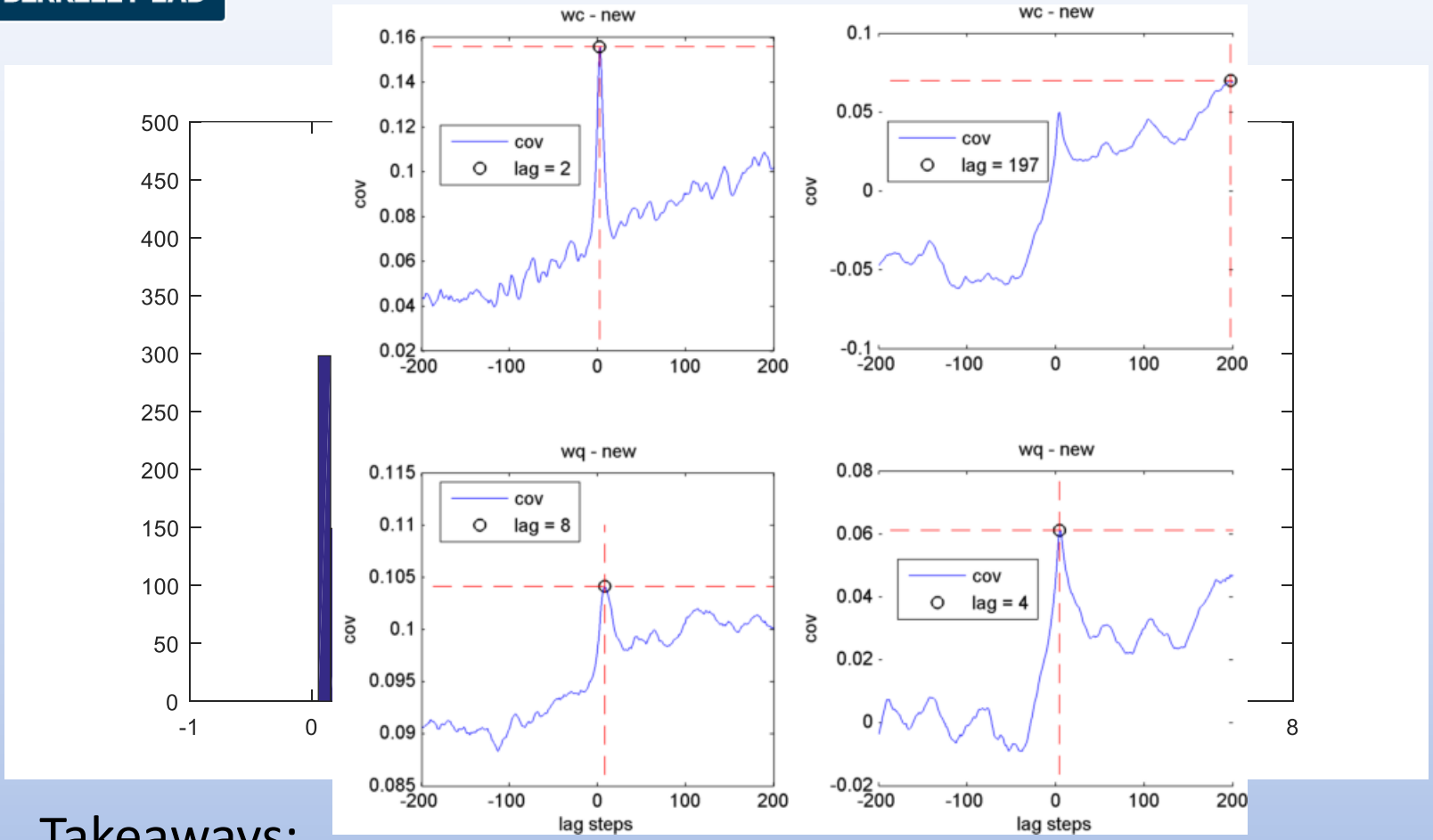
Very common (and easy to do), especially when merging files.



- Know whether timestamp denotes beginning or end of period (or middle).
- Sometimes shifts are indication of larger problems (e.g., file management).

Takeaway: always check for temporal consistency within a file(s).

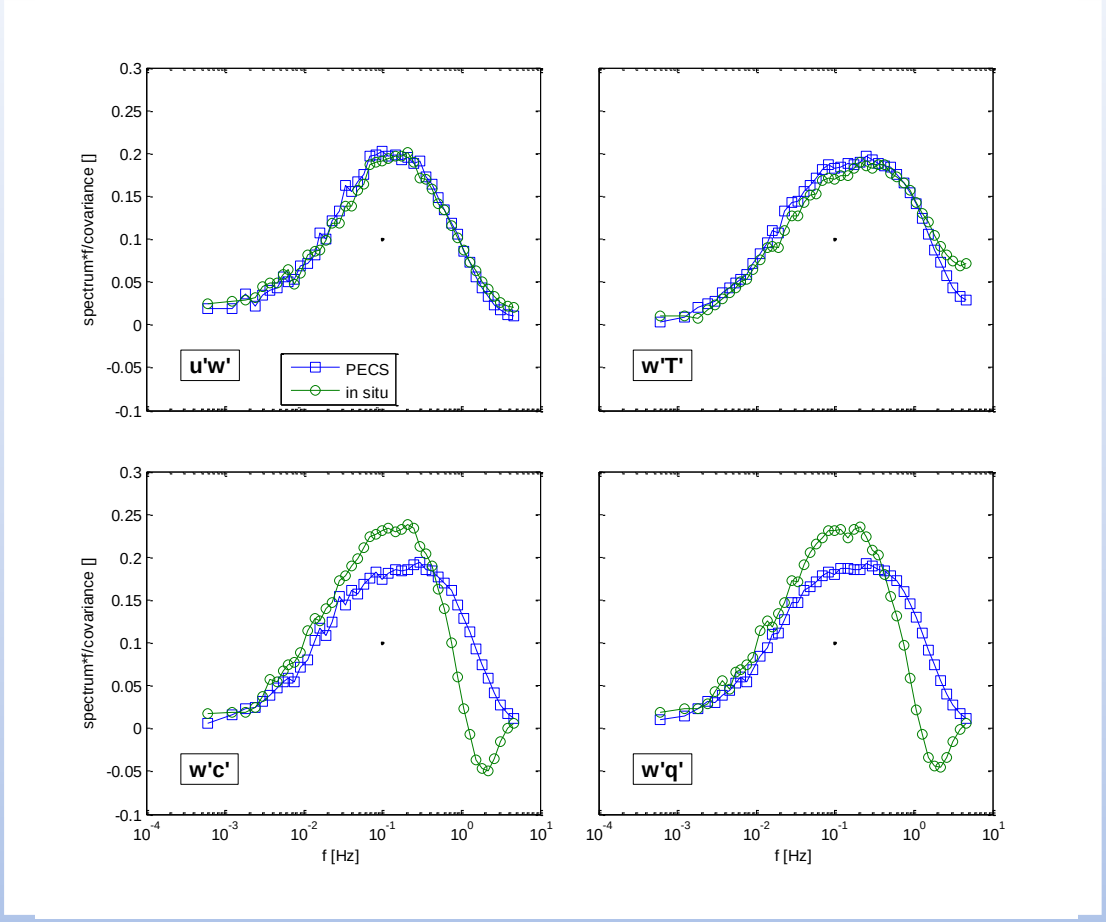
Ex5: Lag issues



Takeaways:

- If lags are incorrect, fluxes will be too.
- Check for consistency across data record.
- Review lag times: are they realistic? Could they be optimized?

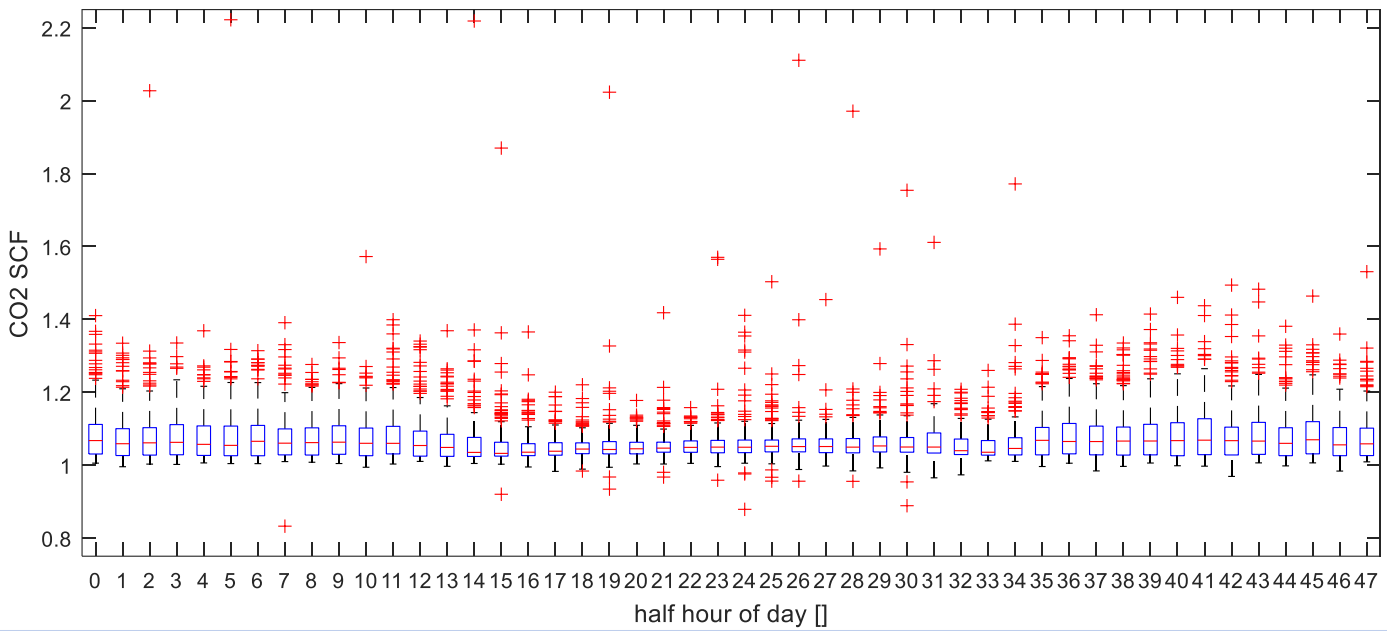
Ex6: Spectral analysis



Takeaways:

- Useful for detecting issues with instruments
- Identifying signal aliasing issues
- Ogives can evaluate appropriateness of averaging window

Ex7: Spectral corrections

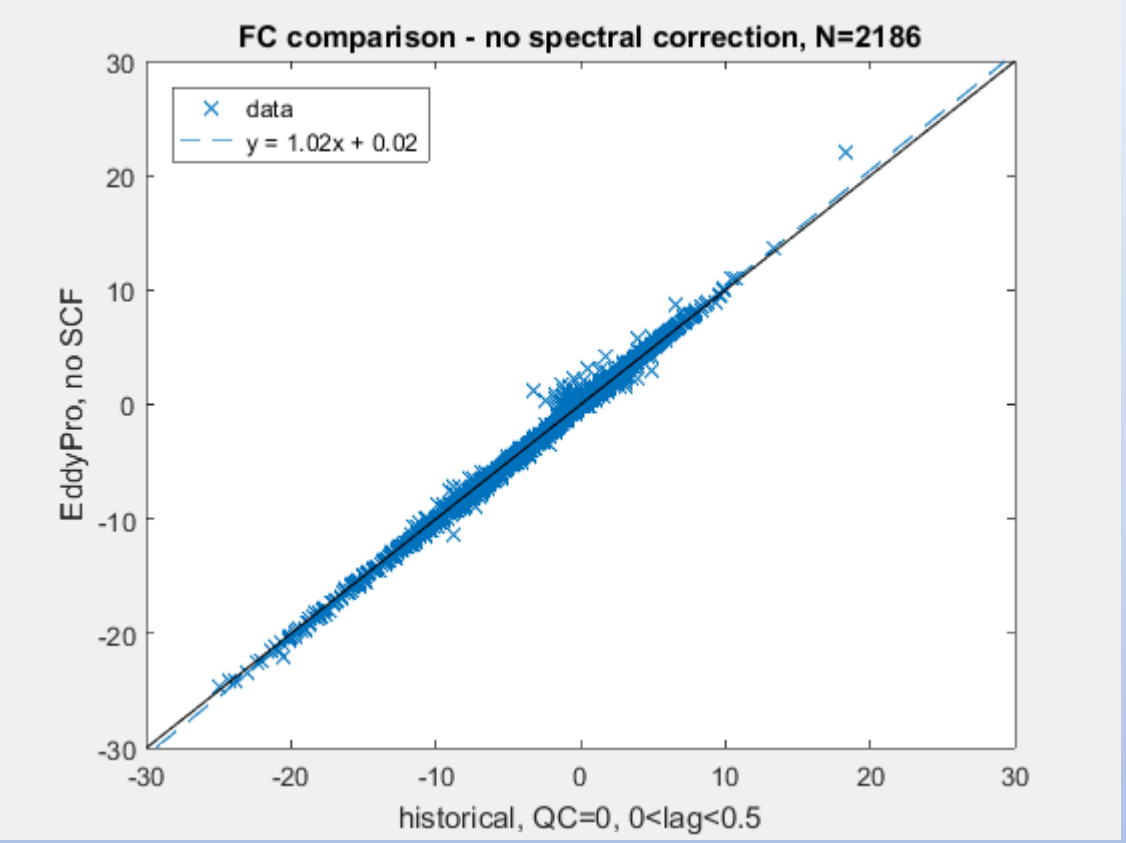


Takeaways:

- Check for consistency across data record.
- Review correction magnitudes: are they realistic? Could they be optimized?

Ex8: US-ARM

Updating a system (instruments and data processing).



Understanding the historical data processing was first step and emphasized the need for reprocessing (2002-15)

Conclusion

FACT: Reprocessing happens…

…make it easier.

- Organize data
  - Use systematic approach! (file naming, folder structure, data format)
  - Preserve original data sources
  - Don’t ignore ancillary data and metadata
  - Know what was previously done!
    - Data archeology is really hard…
    - Can you reproduce past results?
  - Document, document, document
    - Record metadata (data processing steps)

Thank you!

Questions / comments?

Email:
ameriflux-tech@lbl.gov