# Airborne Data Processing and Analysis


David Delene
Department of Atmospheric Sciences
University of North Dakota

# Scientific Programming: Problems and Solution?

- Scientist write program because they understand the problem to be solved; however, they are judged/reward for writing papers, not producing data sets or software.
- Software used to collect, process and analyze measurements and models almost never evaluated.
  - Sometimes by supervises or co-workers
  - No independent source code review. (Class???)
- Mistakes in software result in wrong scientific conclusions!
  - Black Box / White Box Testing

# Accuracy of Scientific Results

- Scientist could only reproduce 6 out of 53 “landmark” articles published by reputable labs in top journals.
  - http://www.nature.com/nature/journal/v483/n7391/full/483531a.html
  - http://www.reuters.com/article/2012/03/28/us-science-cancer-idUSBRE82R12P20120328
- John P. A Ioannidis explains in detail how “It can be proven that most claimed research findings are false.”
  - http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1182327/
- Personnel Experience

# Airborne Data Processing and Analysis (ADPAA) Software Package (http://sourceforge.net/projects/adpaa/)

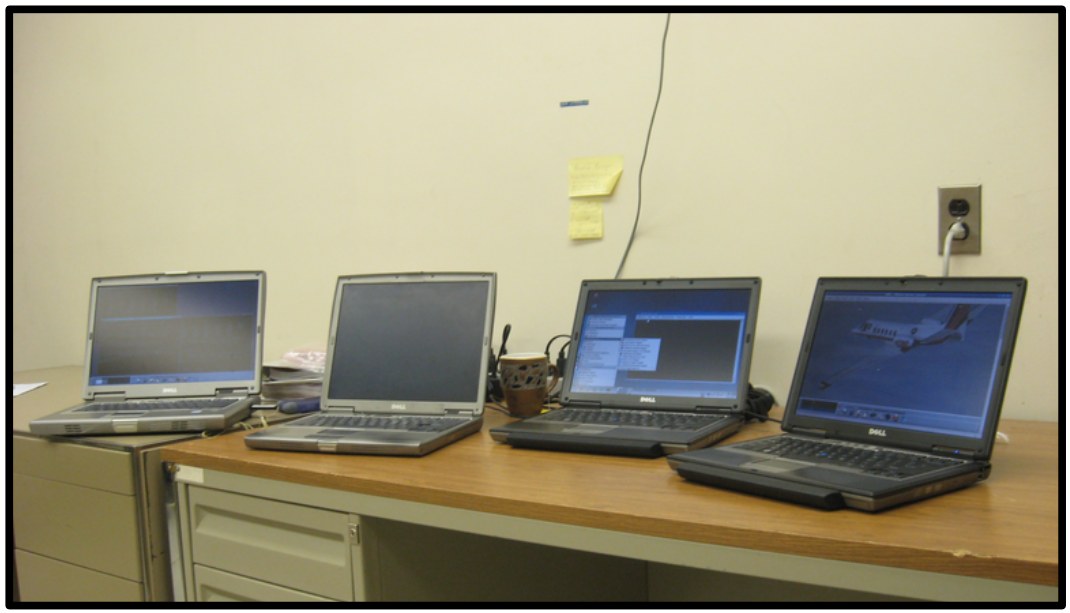
- Independent, Open, and Freely Available
- GNU/GPL v3 Licensed (Only non-commercial use)
- Started in 2007, Version 1618 March 2014
- Approximately 205,051 Lines of Code (IDL, Perl, Bash, csh, FORTRAN, C, Python, etc.)
- Subversion (SVN) Source Code Management System
- Feature Requests, Bug Tracker, Forum and Wiki

Delene, D. J., Airborne Data Processing and Analysis Software Package, Earth Science Informatics, 4(1), 29-44, 2011, URL: http://dx.doi.org/10.1007/s12145-010-0061-4, DOI: 10.1007/s12145-010-0061-4.

# Airborne Data Sets

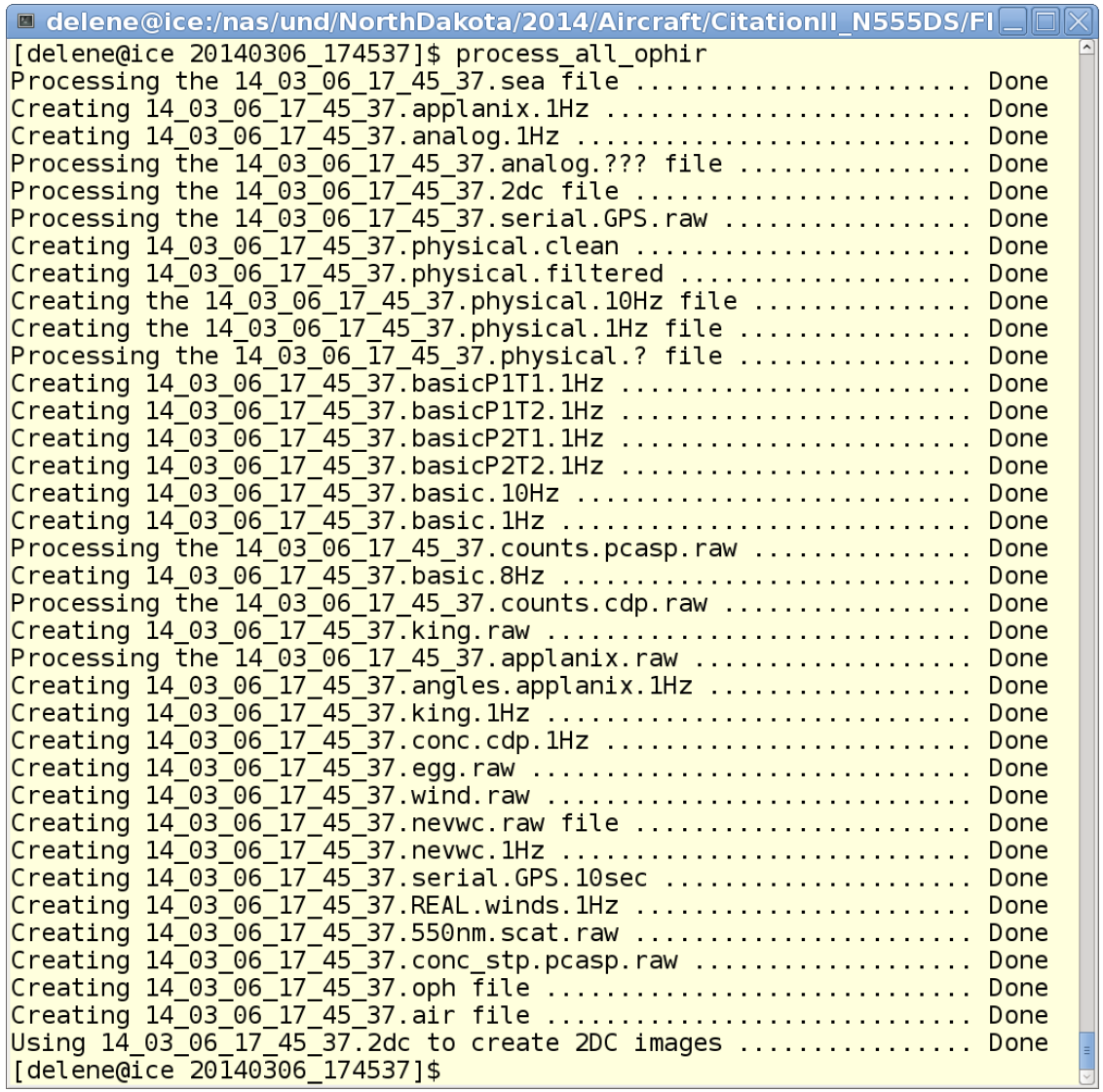
**Quality Control -** The process of conducting tests to check that measurements are being made correctly and accurately.

**Quality Assurance -** The process of reviewing a data set to eliminate measurements that are invalid due to known problems.



# Data Processing

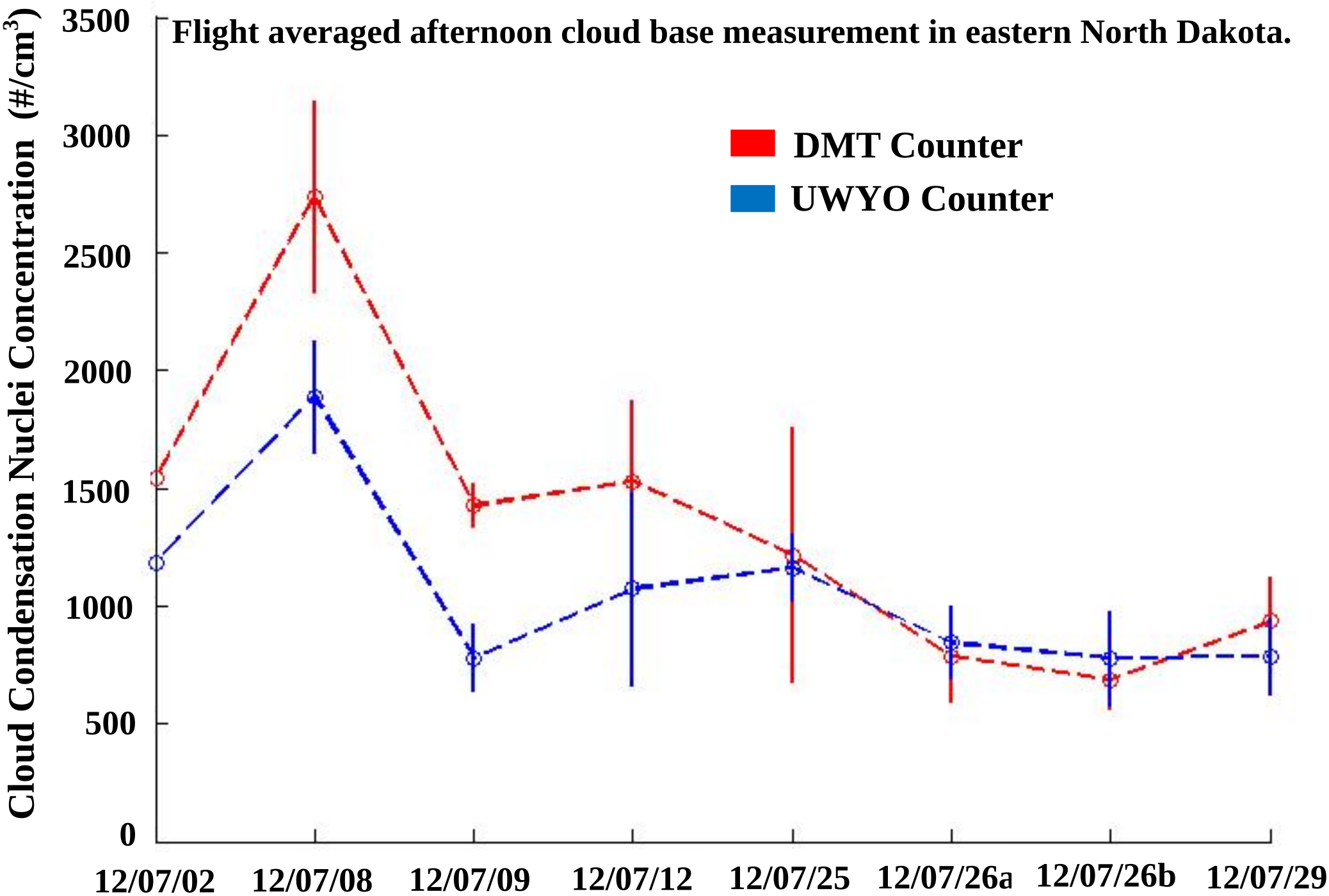
- Data Quality Control
  - Calibration Checks
- Data Missing Values Codes
- Levels of Data Processing
  - Raw Recorded Data
  - Engineering to Physical Units
  - Single Instrument Data Files
  - Combined Instrument Data File
- Data Quality Assurance
  - Scientific Data Review
  - Scripts Search for Unrealistic Values



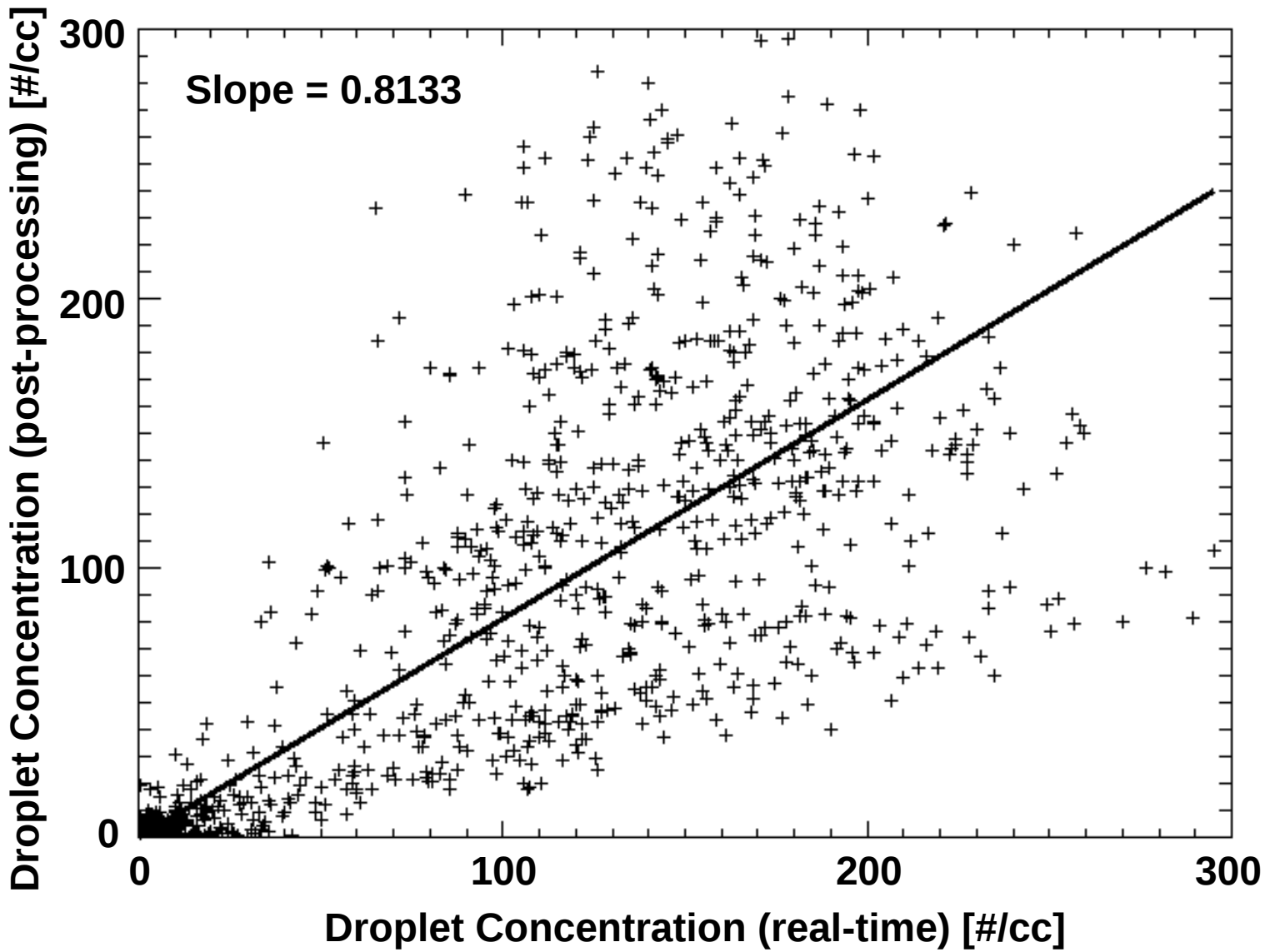
# Comments on Scientific Data

- Quick Visualization of data is very Important.
  - Create a preliminary version of the data using automated processing scripts.
  - Create a final data set after the project is over by applying manual edits to the “raw” data files which replace “bad” data with missing value codes.
- Archive the raw data and any editing files.
- Work with ASCII data as much as possible.
  - Compress ASCII files, if necessary.
- Use a standard data format, which includes Meta data in all data files.

# Instrument Comparison Example

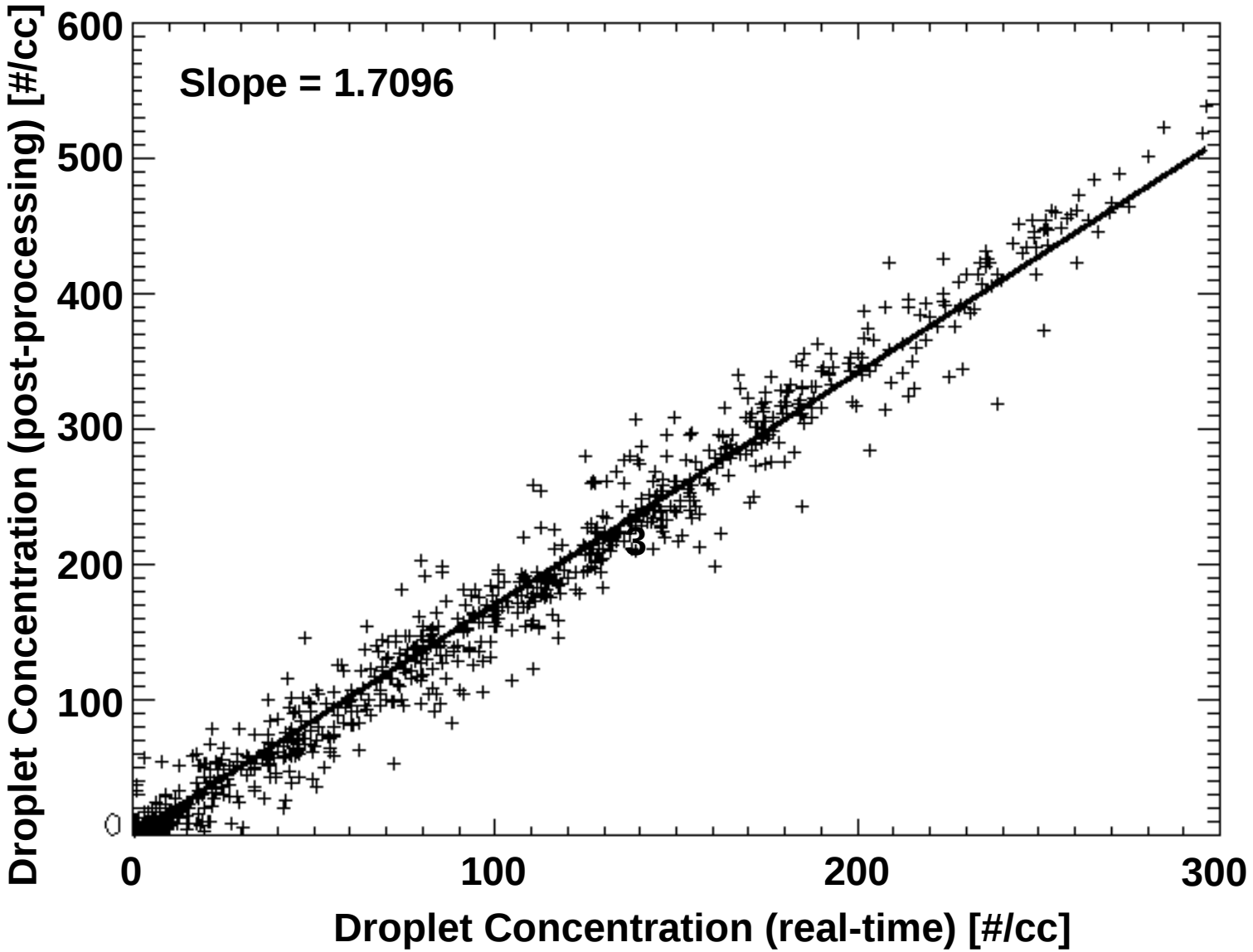


# Comparison of Software Processing Methods




**Comparison of the M300 real-time data processing method (x-axis) and Airborne Data Processing and Analysis (ADPAA) post-processing method for the Forward Scattering Spectrometer Probe. All 1 Hz average data from the second flight on January 10, 2008 are included.**

# Comparison of Software Processing Methods




**Comparison of the M300 real-time data processing method (x-axis) and post-processing method (y-axis) after fixing bead fraction problem. All 1 Hz average data from the second flight on January 10, 2008 are included. Processing includes beam fraction correction but not coincidence and dead time corrections.**

# Scientific Processing Summary

- Different software methods (codes) can disagree and any disagreement needs to be resolved.
  - Airborne Data Processing Workshop (Boston, July 5 & 6 2014)
- Well calibrated instruments and validated software is critical for the scientific progress.
- Peer reviews of papers should require not only open data sets but open source software.
- If a paper's major conclusions are shown to be wrong, the papers should be retracted.
  - http://retractionwatch.com/2012/09/25/if-a-papers-major-conclusions-are-shown-to-be-wrong-we-will-retract-the-paper-plos/

# Proposed Solution

- Private company for scientific data processing.
  - Investment:  Develop necessary tools (software).
  - Reward:  Could pay software developers at market rate.
- Focus
  - Aircraft Data Sets (Instrument development companies are not software development companies)
  - Time Series Measurements (UAS, Ground Stations)
- Model
  - Open source software, open data sets.
  - Use best tools (software) available.
  - Support instruments from all companies.

# Revenue Sources

- Creation of “Analysis Ready” data sets for field project.
  - Groups that can't support software development personnel would support the processing of data so they can conduct there analysis.
- Yearly Maintenance contracts
  - Continue support and development
- Open Source Existing Code
  - Refactoring software into open repositories.

# Obstacles

- Scientist and project manages don't like paying for software.
  - Buy instruments, so software is similar.
- Development of client list.
  - Work with instrument development companies.