AE9/AP9/SPM
Radiation Environment Model

Cleaning of Satellite Data Sets

A Brief Overview

October 10, 2012
Christopher Roth
Atmospheric and Environmental Research, Inc. (AER)
www.aer.com

aer
Atmospheric and
Environmental Research

Outline

- ## What is Data Cleaning?

- ## Satellite Data Processing Steps

- ## Data Cleaning Survey Plots

  – Example Filters

- ## Verify Filter Results

- ## Summary

aer

**Atmospheric and
Environmental Research**

What is Data Cleaning?
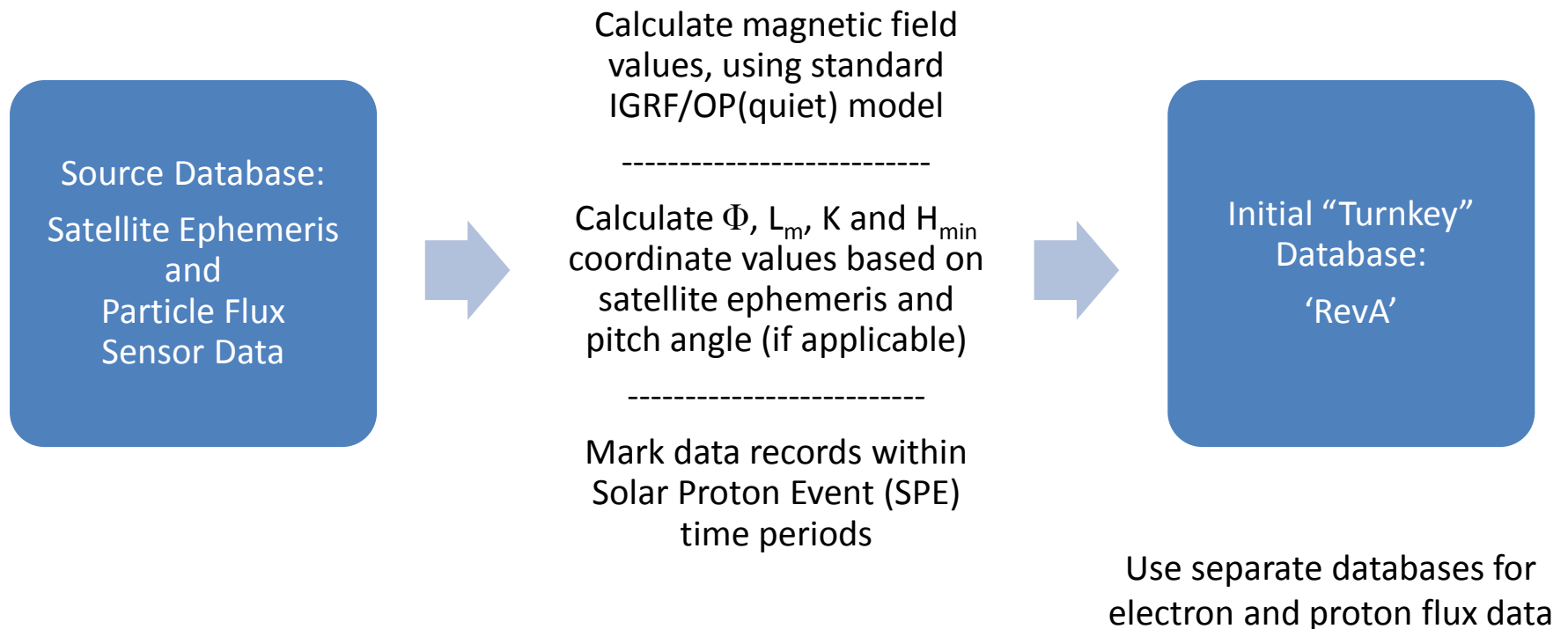
Objective is to identify and remove data that is:

- – Corrupted (sensor errors)
- – Saturated
- – Contaminated
- – Includes significant background
- – Unusual/unexpected

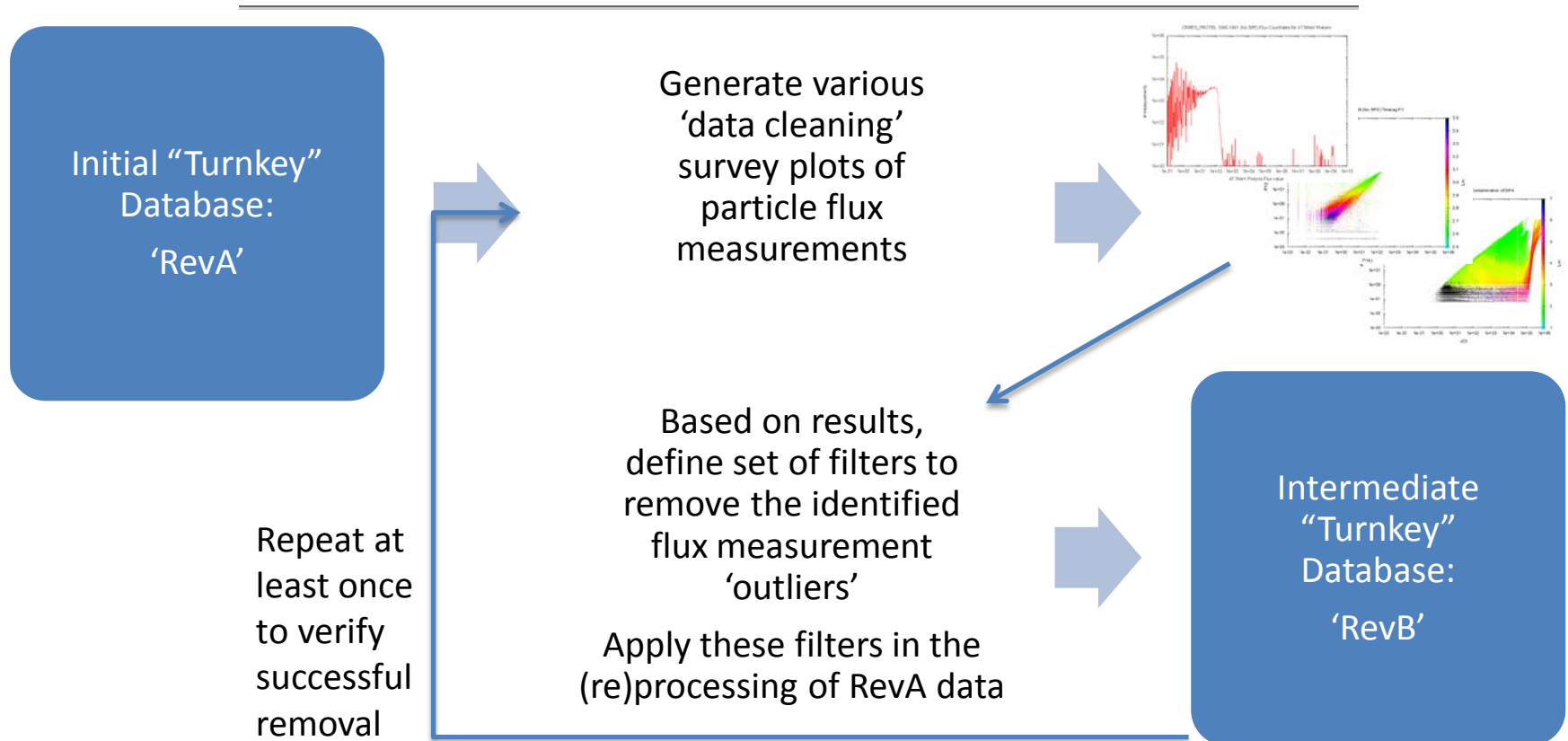This process does not address particle sensor calibration and measurement error bars.

aer

**Atmospheric and
Environmental Research**

Satellite Data Processing Steps (1)

'RevA' Database Processing

Source Database:

Satellite Ephemeris and Particle Flux Sensor Data

➤

Calculate magnetic field values, using standard IGRF/OP(quiet) model

----

Calculate Φ, L_m, K and H_{min} coordinate values based on satellite ephemeris and pitch angle (if applicable)

Mark data records within Solar Proton Event (SPE) time periods

➤

Initial "Turnkey" Database:

'RevA'

Use separate databases for electron and proton flux data

aer
Atmospheric and Environmental Research

Satellite Data Processing Steps (2)

'RevB' Database Processing

Initial "Turnkey" Database:

'RevA'

Generate various 'data cleaning' survey plots of particle flux measurements

Based on results, define set of filters to remove the identified flux measurement 'outliers'

Apply these filters in the (re)processing of RevA data

Intermediate "Turnkey" Database:

'RevB'

Repeat at least once to verify successful removal

Atmospheric and Environmental Research

Satellite Data Processing Steps (3)

'RevC' Database Processing

Intermediate "Turnkey" Database: 'RevB'

→ Apply channel-specific satellite measurement cross-calibration factors to each flux value →

Final "Turnkey" Database: 'RevC'

Model database construction process
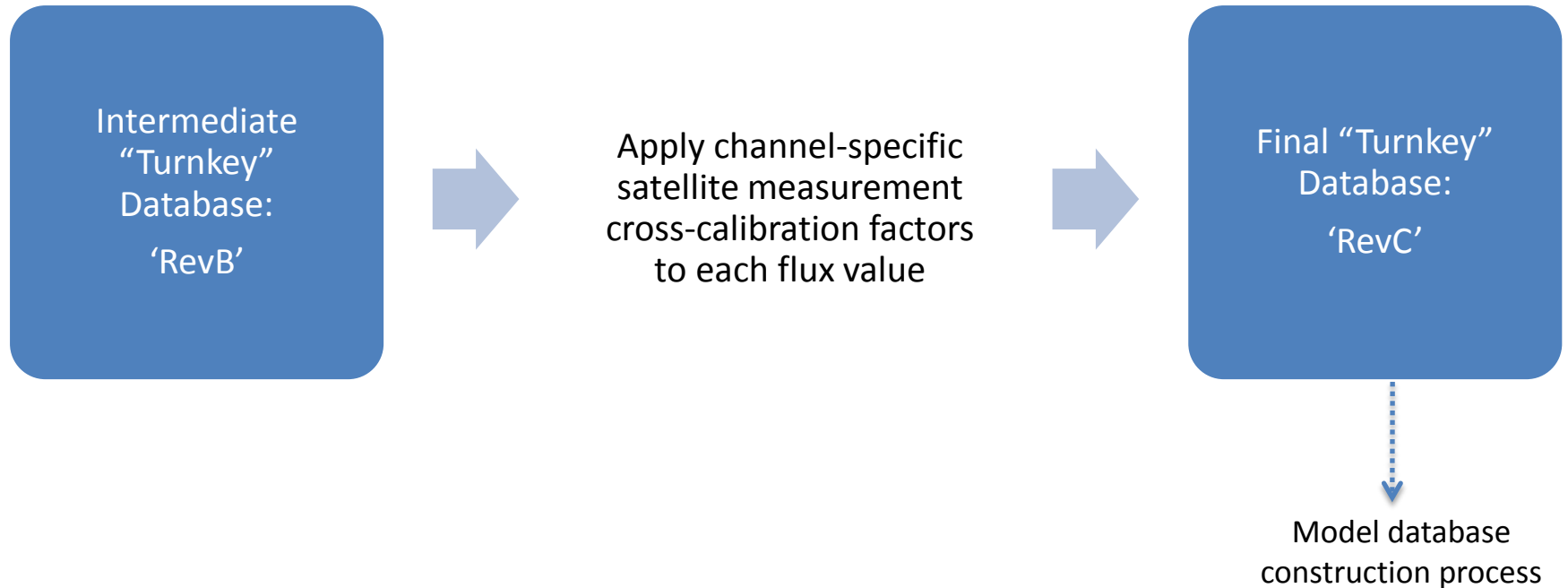
Data Cleaning Survey Plots (1)

- Measurement Frequency Plot
   - A histogram showing the number of occurrences for every possible value of the flux measurements.

CRRES_PROTEL 1990-1991 (No SPE) Flux Countrates for 02.5MeV Protons
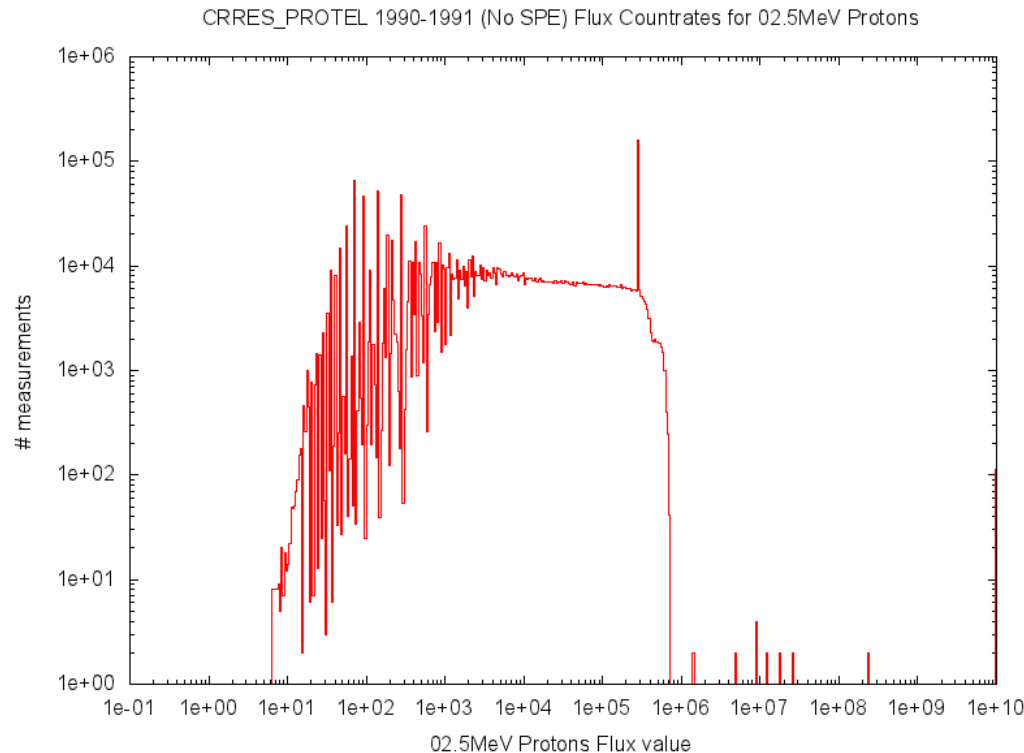
Example Filters (1)

- Measurement Frequency Plot
  - Very easy to define a maximum flux value allowed
 - Other filters will be required to address all problems evident

CRRES_PROTEL 1990-1991 (No SPE) Flux Countrates for 02.5MeV Protons

'stuck' measurement

Filter definition: discard when flux > 7.5×10^5
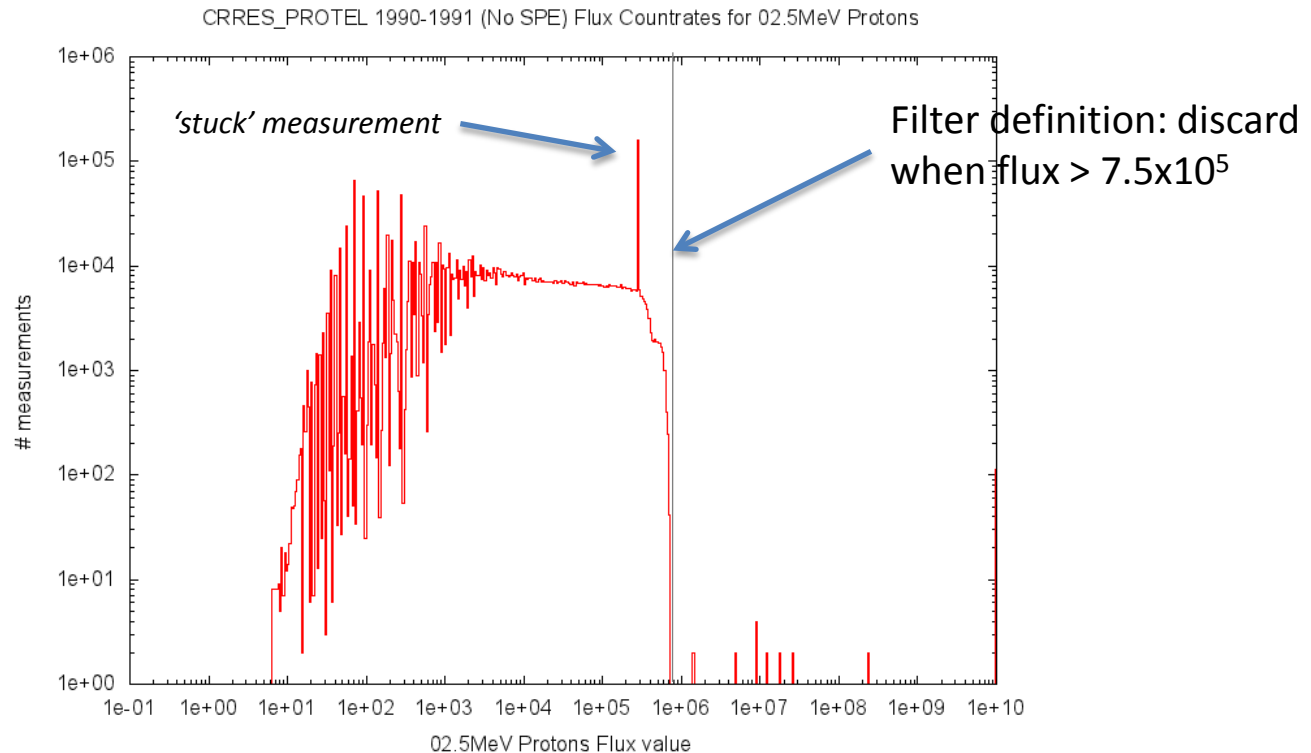
Data Cleaning Survey Plots (2)

- Time-lag Correlation Plot
    - Scatter plot (color-coded) for a single species and energy channel, showing measurements for one time versus measurements for the next time.

ICO 2001-2009 (No SPE) Timelag P1
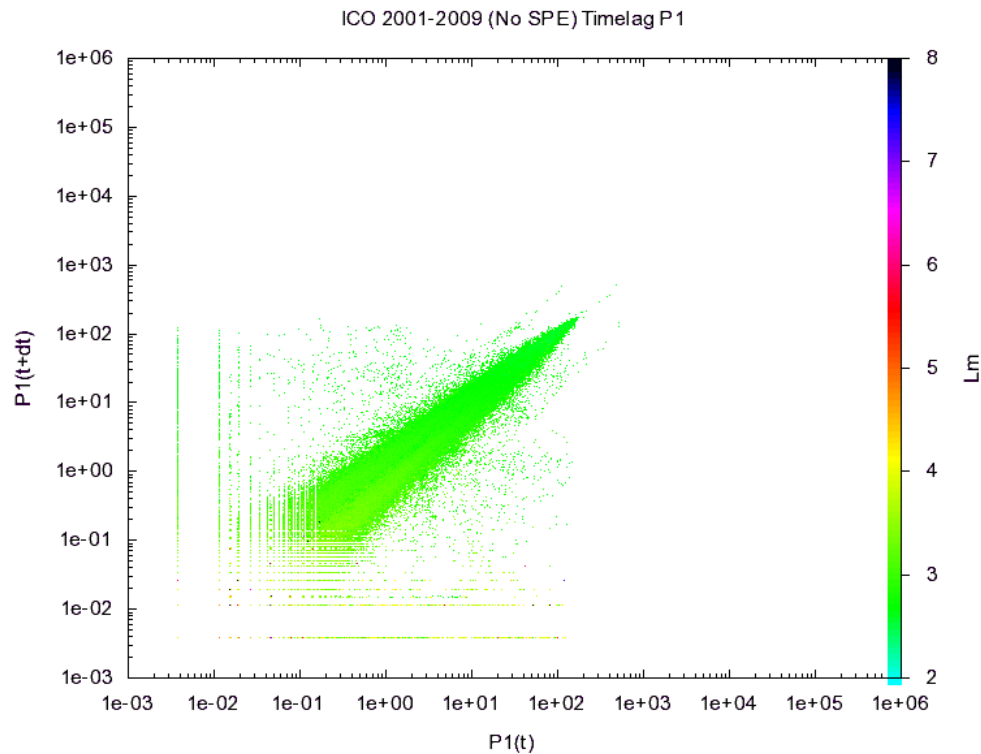
Example Filters (2)

- Time-lag Correlation Plot
    - Define a symmetrical envelope around main body of points, via four lines
    - P1($t+dt$) lower limit : min of { (P1(t)/A)^B , P1(t)/C)^D } where A=4.1 B=1.2 C=2.7 D=2.8
    - P1($t+dt$) upper limit : max of { (P1(t)$^{(1/B)}$)*A, (P1(t)$^{(1/D)}$)*C } for this particular case



ICO 2001-2009 (No SPE) Timelag P1
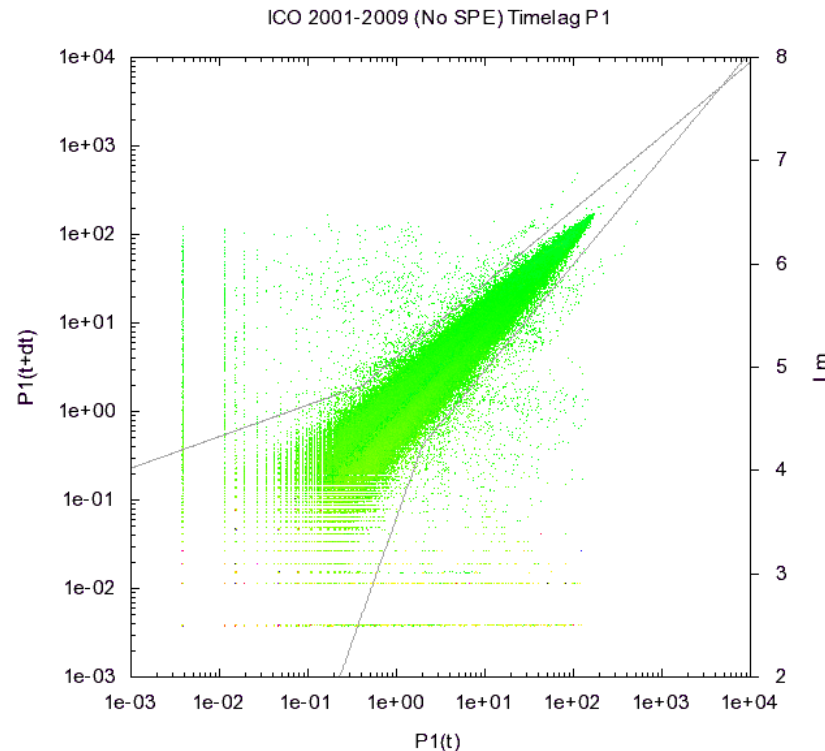
Data Cleaning Survey Plots (3)

- Pitch-lag Correlation Plot (for data sets with multiple look directions per record)
  - Scatter plot (color-coded) for a single species, showing measurements for one pitch angle (look direction) versus measurements for the next pitch angle, within the same data record.


CRRES_PROTEL 1990-1991 (No SPE) Pitchlag Protons at 02.5MeV
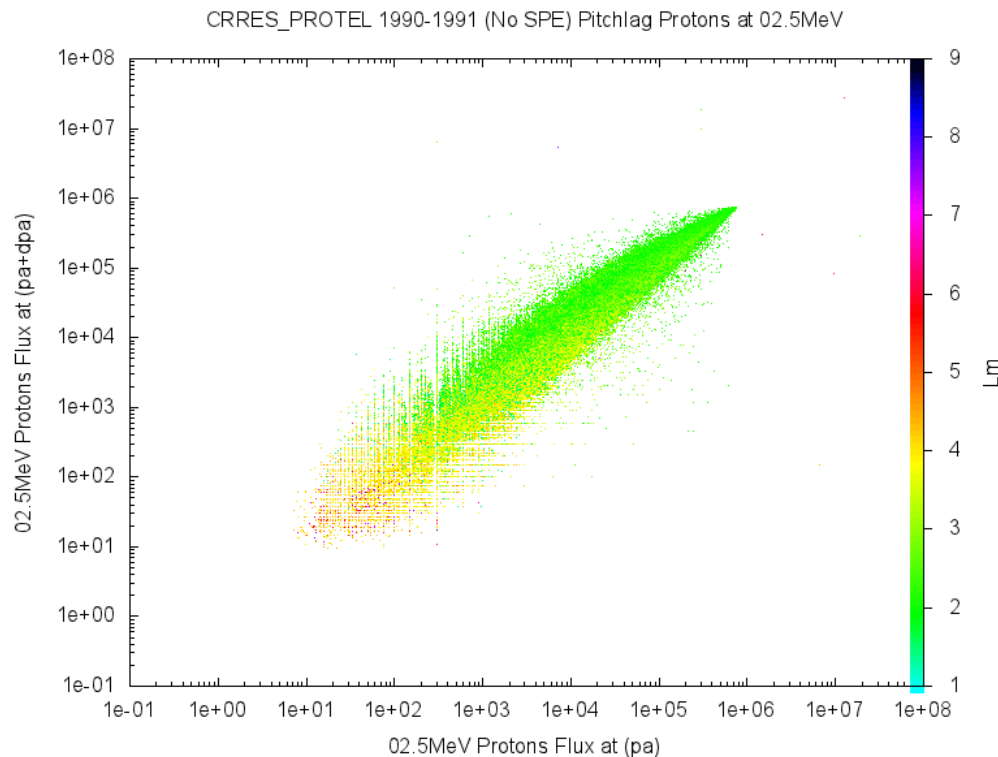
Example Filters (3)

- Pitch-lag Correlation Plot
    - Sometimes no filter is necessary; opinions may vary
    - maximum flux filter shown in plot

CRRES_PROTEL 1990-1991 (No SPE) Pitchlag Protons at 02.5MeV
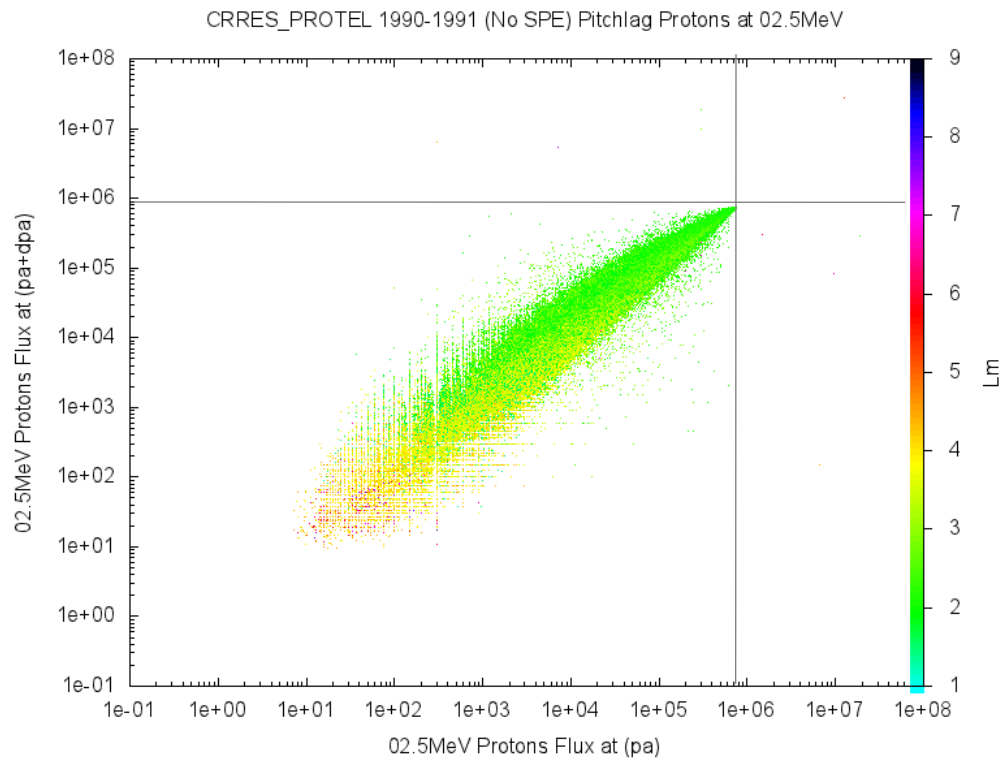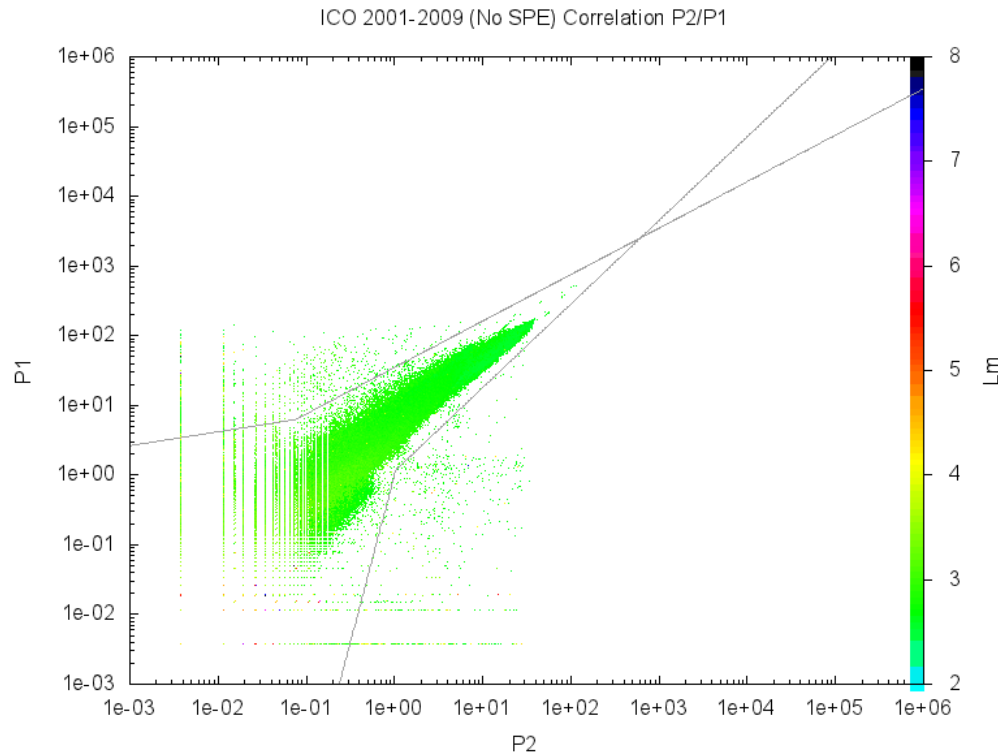
Data Cleaning Survey Plots (4)

- Correlation Plot
  - Scatter plot (color-coded) for a single species, showing measurements for one energy channel versus measurements for the next higher channel.

ICO 2001-2007 (No SPE) Correlation P2/P1
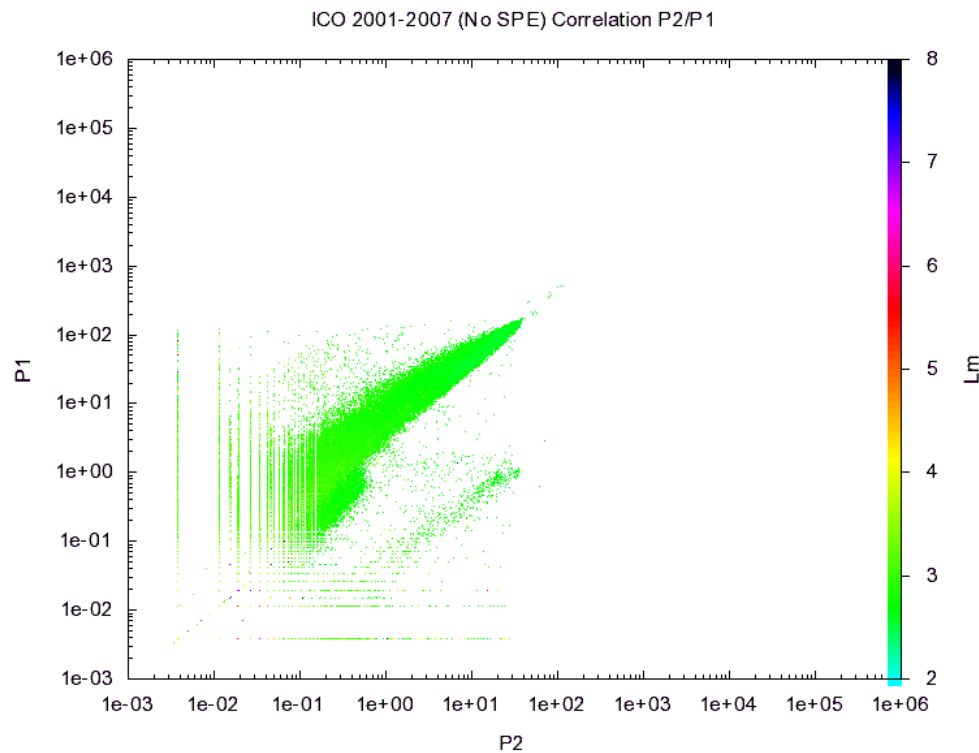
Example Filters (4)

- Correlation Plot
  - Define an envelope around main body of points, via four parameterized lines
  - P1 lower limit= min of { $(P2*A)^B$, $(P2*C)^D$ }   where A=1.111 B=1.2 C=1.0 D=4.8 for this case
  - P1 upper limit: max of { $(P2*D)^E$, $(P2*F)^G$ }   where D=35.0 E=0.666 F=10.5 G=0.2

ICO 2001-2009 (No SPE) Correlation P2/P1
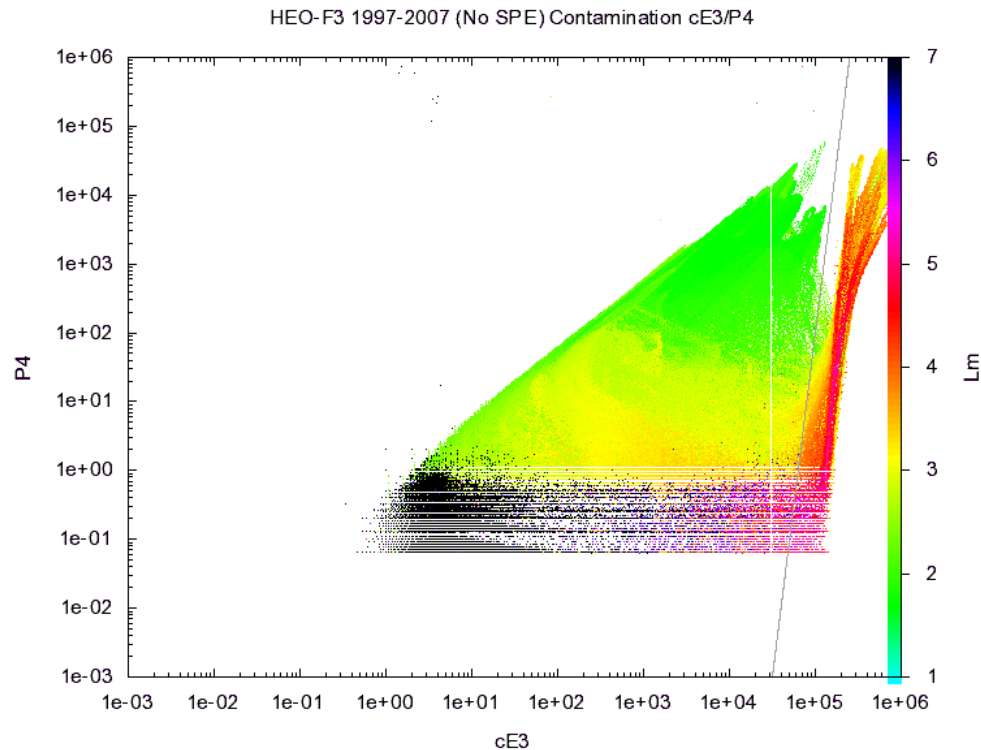
Data Cleaning Survey Plots (5)

- Contamination Plot
  - Scatter plot (color-coded, by date or coordinate) showing measurements for a single energy channel from one species versus measurements for a single energy channel of the other species (ie, electrons versus protons).

HEO-F3 1997-2007 (No SPE) Contamination cE3/P4
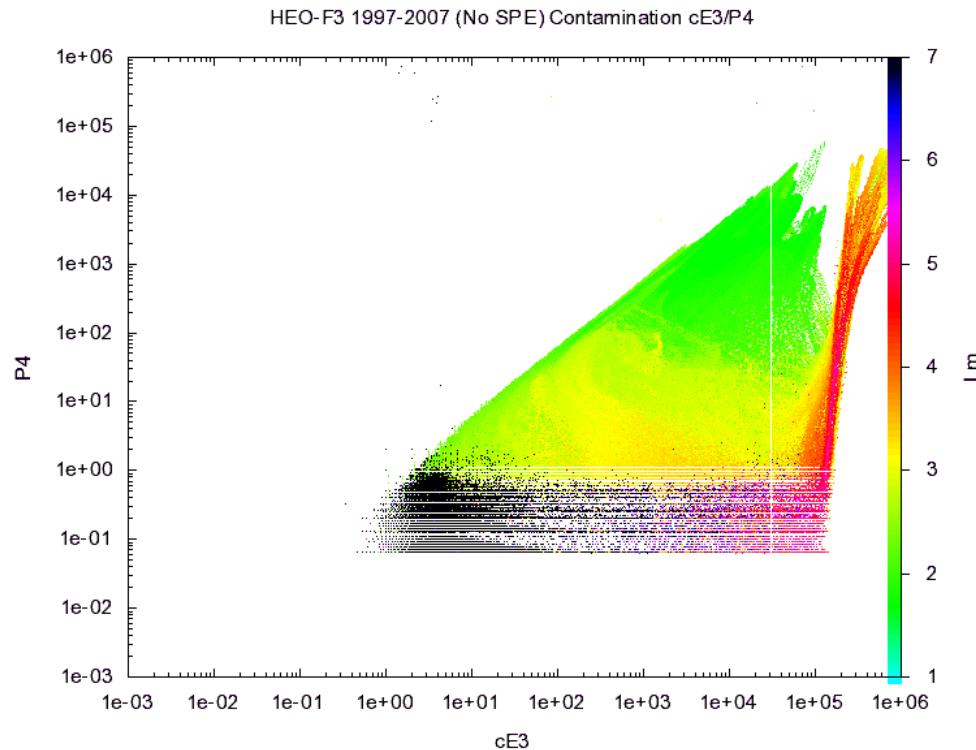
Example Filters (5)

- Contamination Plot
  - Define a limit for the expected valid values as a function of the other:

  if ($\log_{10}(cE3) > \log_{10}(P4)/10.0 + 4.80$) P4 = invalid

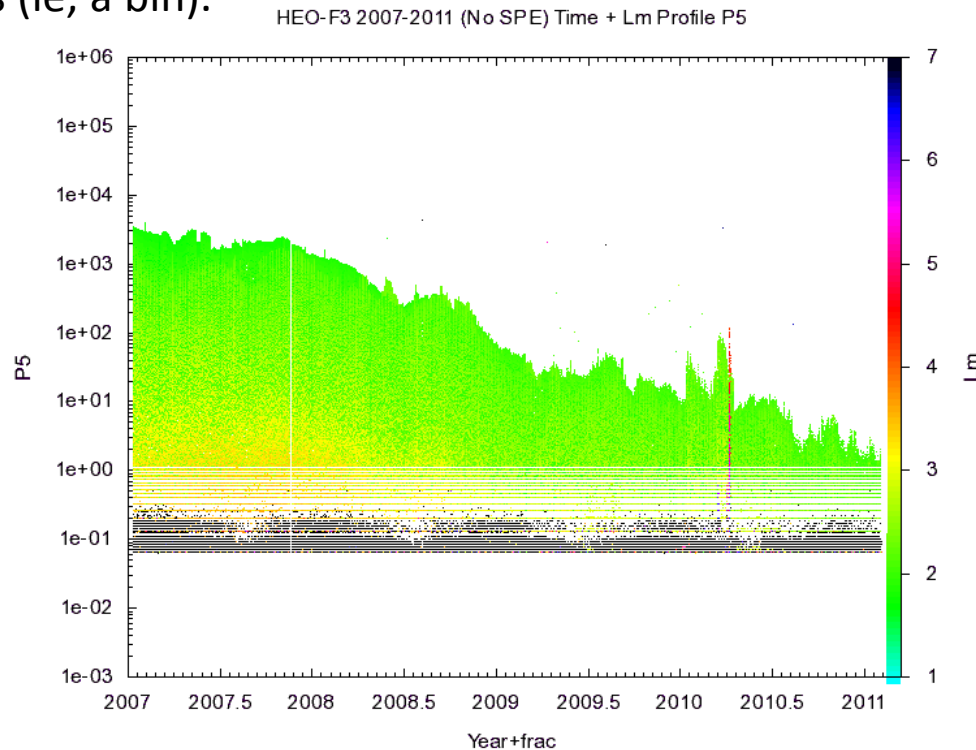HEO-F3 1997-2007 (No SPE) Contamination cE3/P4

Data Cleaning Survey Plots (6)
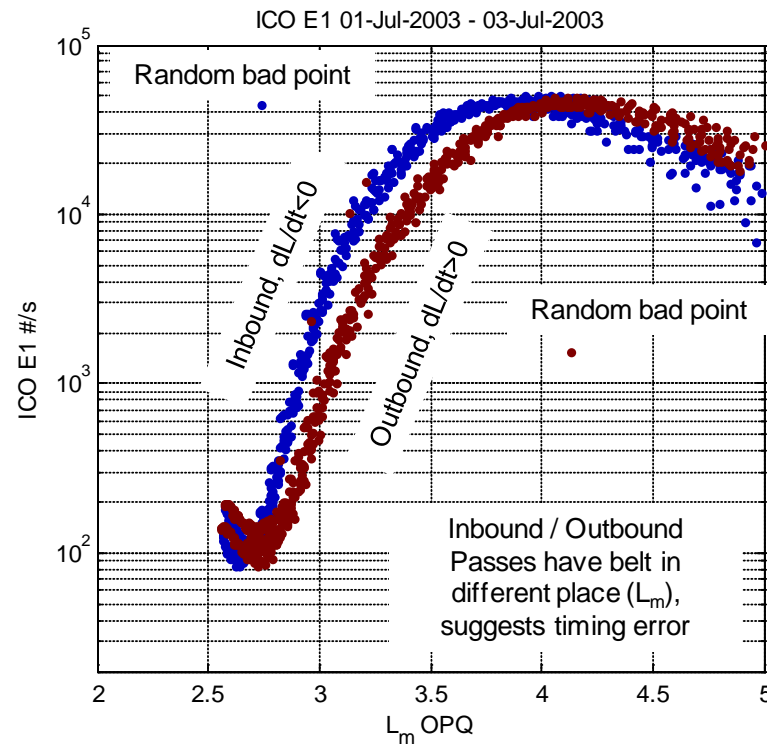
- Time/Bin Plot
  - Line plot for a single species and energy channel, showing measurements versus a long period of time. Color values based on an appropriate coordinate, such as Lm or pitch-angle; or only plot a small range of those coordinates (ie, a bin).

HEO-F3 2007-2011 (No SPE) Time + Lm Profile P5

Data Cleaning Survey Plots (7)

- Pass Plot
 - Line plot plot for a single species and energy channel, showing measurements versus a meaningful positional coordinate (ie, Lm or K), usually for a single day time period or shorter.

ICO E1 01-Jul-2003 - 03-Jul-2003

Random bad point

Inbound, dL/dt<0

Outbound, dL/dt>0

Random bad point

Inbound / Outbound Passes have belt in different place (L_m), suggests timing error

ICO E1 #/s

L_m OPQ

aer

Atmospheric and Environmental Research
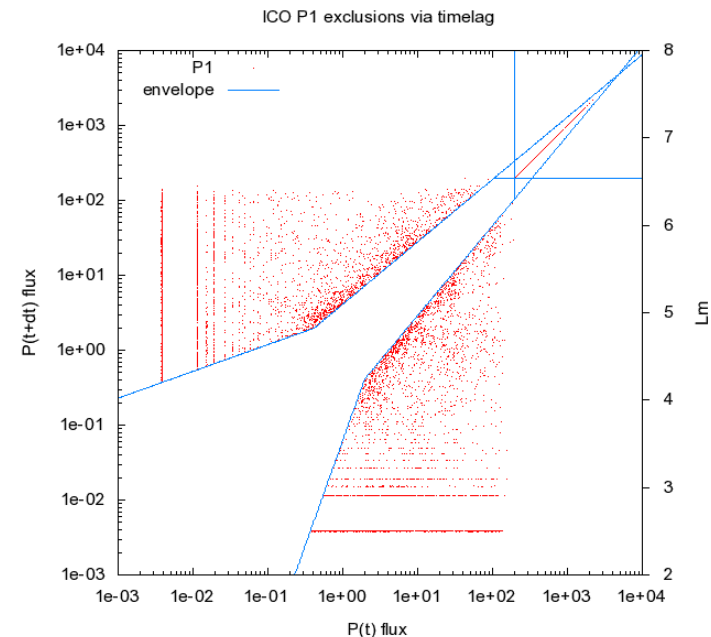
Data Cleaning Survey Plots (8)

- 'Other' Plot
  - Depending on the satellite and/or sensor characteristics, other types of plots may reveal other flux measurement anomalies.
  - Detailed knowledge of sensor is helpful in determining aspects to examine.
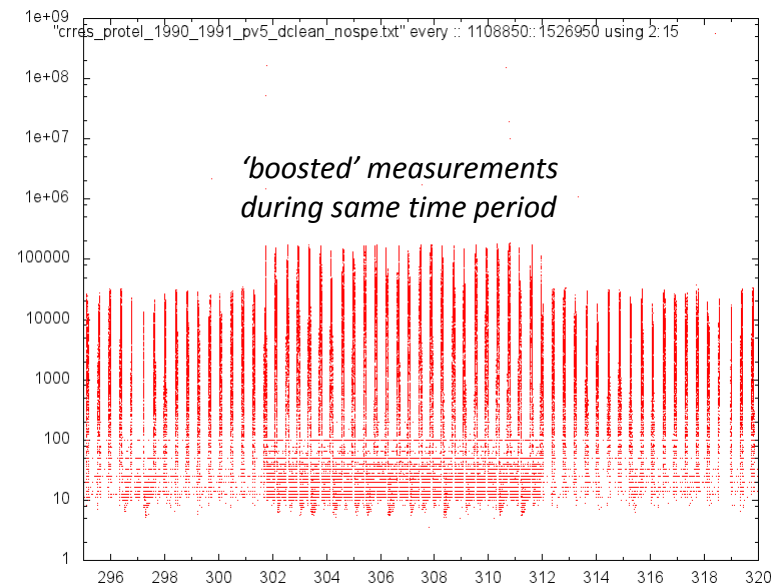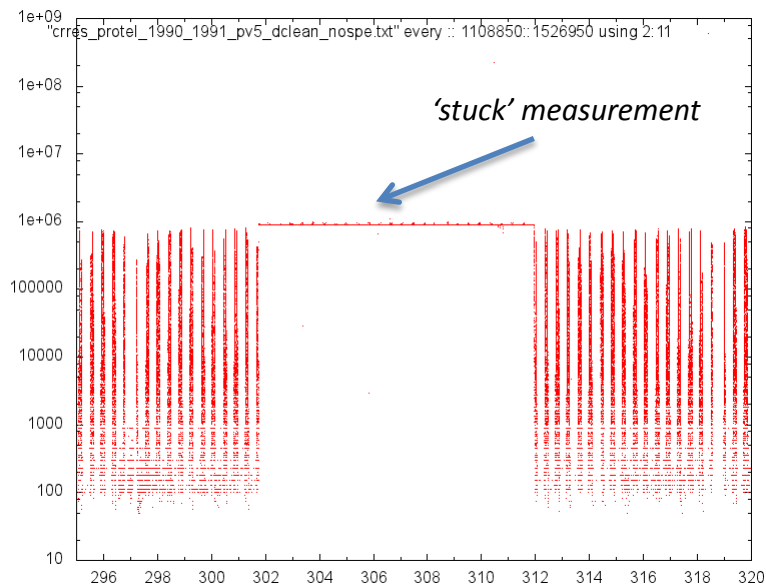
Any number of other types of filters may be defined,
such as the removal of flux data for specific channel and time period, etc.

Verify Filter Results

- After filters defined and database reprocessed, regenerate data cleaning plots
    - Verify that the filters eliminated the intended data values from the database
    - Verify that only the intended data values were removed

HEO-F3 1997-2007 (No SPE) Contamination cE3/P4

ICO P1 exclusions via timelag
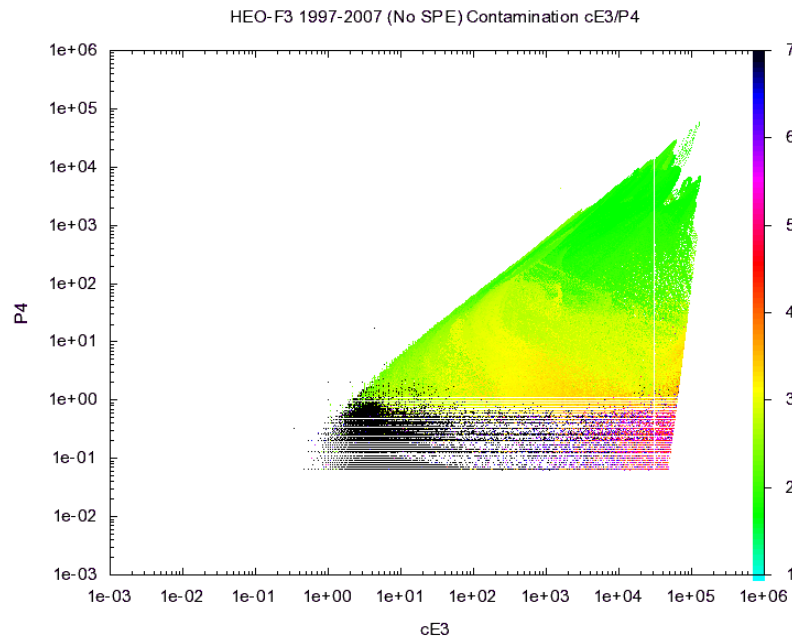
Summary

- Data cleaning removes the abnormal flux values from the sensor data set

- Abnormal flux values are identified through the analysis of various types of survey plots

- Definition of data filters for the reprocessing is a mix of science and art

- Improves overall sensor results & statistics

Questions?

For more information, see **Data Cleaning Guidelines for AE-9/AP-9 Data Sets**
TOR-2012(1237)-4 (DRAFT) Paul O'Brien, The Aerospace Corporation