

Instrument Metadata Management

Getting the most out of air
monitoring instrumentation



Mark Stoelting
Bay Area Air Quality Management District
San Francisco, CA



Presented at:
U. S. EPA 2009 National Air Quality Conference
Nashville, TN

November 2-5, 2009

Metadata Management

- What is metadata
- Why is it important
- Where does it come from
- How is it collected
- How can metadata be used efficiently
- How does it improve ambient data quality
- What eventually happens to metadata

What is Metadata

- As defined by Merriam-Webster, metadata is “data that provides information about other data”
- Instrument shelter temperature is metadata
- Instrument metadata provides information about the internal operation of the instrument
 - Real-time clock
 - Temperatures, voltages, pressures, flows, etc.
 - Status conditions
 - Alarms
- Most instruments produce far more metadata than ambient data

Why is Instrument Metadata Important

- Track instrument performance remotely to improve staff efficiency
- Real-time indicator of instrument stability and potential for failure
- Reduces instrument down time and emergency repairs
- Provides information about when and how instrument failures occurred
- Indicates how (or if) compromised data can be corrected
- **Improve overall ambient data quality produced by the instrument**

Where Does Instrument Metadata Come From

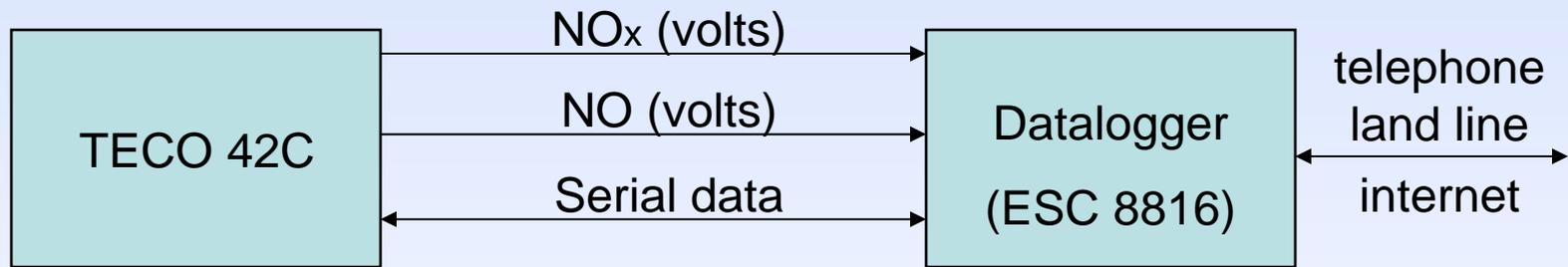
Before the late 1990s, little existed.



- Analog design
- No internal data collection system
- Voltage outputs
- A/D converter
- No serial data capability
- Simple computer functions
- Small storage capacity
- Minimal flexibility

Where Does Instrument Metadata Come From

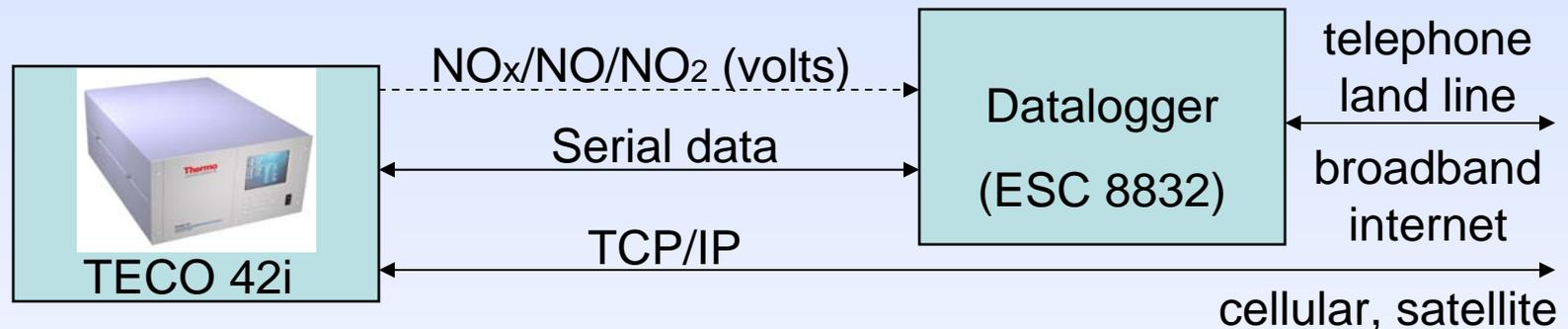
Some available starting in the late 1990s.



- Analog/digital design
- **Simple data acquisition system** for control and measurement
- Limited data storage
- Voltage and serial data output
- Analog and limited digital inputs
- More data channels
- Moderate storage capacity
- Simple programmability

Where Does Instrument Metadata Come From

Extensive metadata available by 2005.



- Digital design
- **Powerful data acquisition system** for control and measurement
- Extensive data storage
- **Programmable** serial and TCP/IP **communications**
- Extensive digital data acquisition capability
- Many data channels
- Virtually unlimited storage capacity
- Highly programmable
- **Remote data collection and instrument management**

Where Does Instrument Metadata Come From

In summary, instrument metadata exists because:

- **Inexpensive digital technology** is available for better instrument design
- **Smarter instruments** with data acquisition systems can continuously measure 1-minute metadata
- **More bandwidth** to transmit larger data volume required for metadata
- **Need for better instrument diagnostic information** for improved reliability and higher quality data

Like new cars today, the computer within your NO_x instrument can tell you what is wrong

How is Metadata Collected

- Digital data is provided in 'standard' data records every minute, hour, etc., automatically or on demand
- The TECO 42i 1-minute long-format record is:

```
08:01 10-25-09  flags CC000000 no 32.097 no2 13.138 nox
45.236 intt 30.872 rctt 50.595 pmtt -2.974 convt
324.469 pres 163.245 smplf 0.718 pmtv -729.641 ozonf
0.050*
```
- Only the NO and NO₂ values are ambient data; everything else is instrument metadata
- Additional metadata can be collected by command. For example, the `diag volt mb` command reports the voltages on the instrument motherboard:

```
diag volt mb 24.1 14.9 4.9 3.2 -3.2
```
- Some metadata are 'set' values like `pmtv` while others are responsive like `pres`

How is Metadata Collected

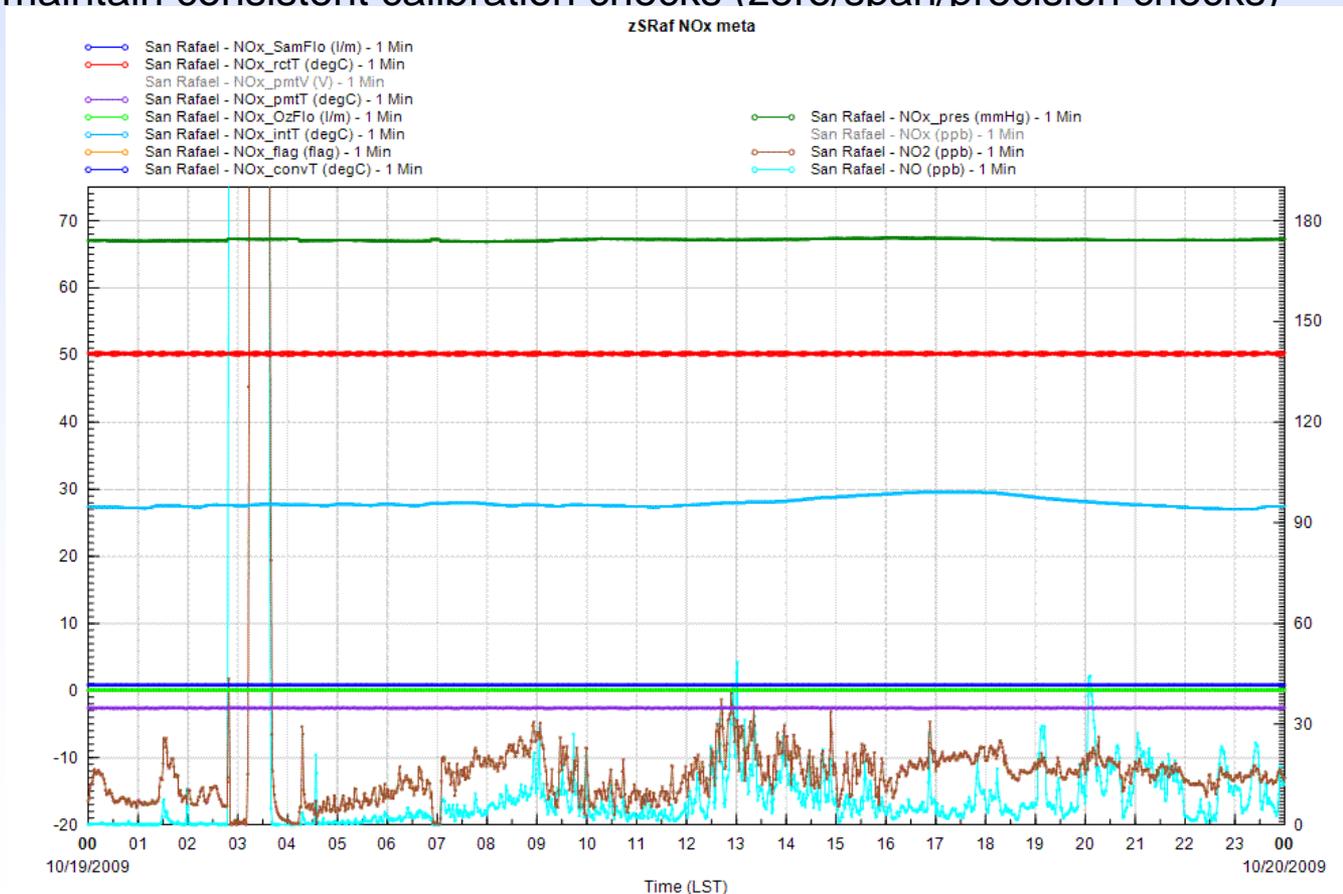
Dataloggers locally store instrument data, insert SiteID and OpCode information, and reformat it for export to a central database. The export format from Bay Area's dataloggers is:

```
<siteID>|NO|10-25-09 08:01|<OpCode>|32.097|
<siteID>|NO2|10-25-09 08:01|<OpCode>|13.138|
<siteID>|NOx|10-25-09 08:01|<OpCode>|45.236|
<siteID>|NOx_flag|10-25-09 08:01|<OpCode>||CC000000
<siteID>|NOx_intt|10-25-09 08:01|<OpCode>|30.872|
<siteID>|NOx_rctt|10-25-09 08:01|<OpCode>|50.595|
<siteID>|NOx_pmtt|10-25-09 08:01|<OpCode>|-2.974|
<siteID>|NOx_convrt|10-25-09 08:01|<OpCode>|324.469|
<siteID>|NOx_pres|10-25-09 08:01|<OpCode>|163.245|
<siteID>|NOx_smplf|10-25-09 08:01|<OpCode>|0.718|
<siteID>|NOx_pmtv|10-25-09 08:01|<OpCode>|-729.641|
<siteID>|NOx_ozonf|10-25-09 08:01|<OpCode>|0.050|
```

Someday, perhaps all instruments can produce programmable, database-ready information...

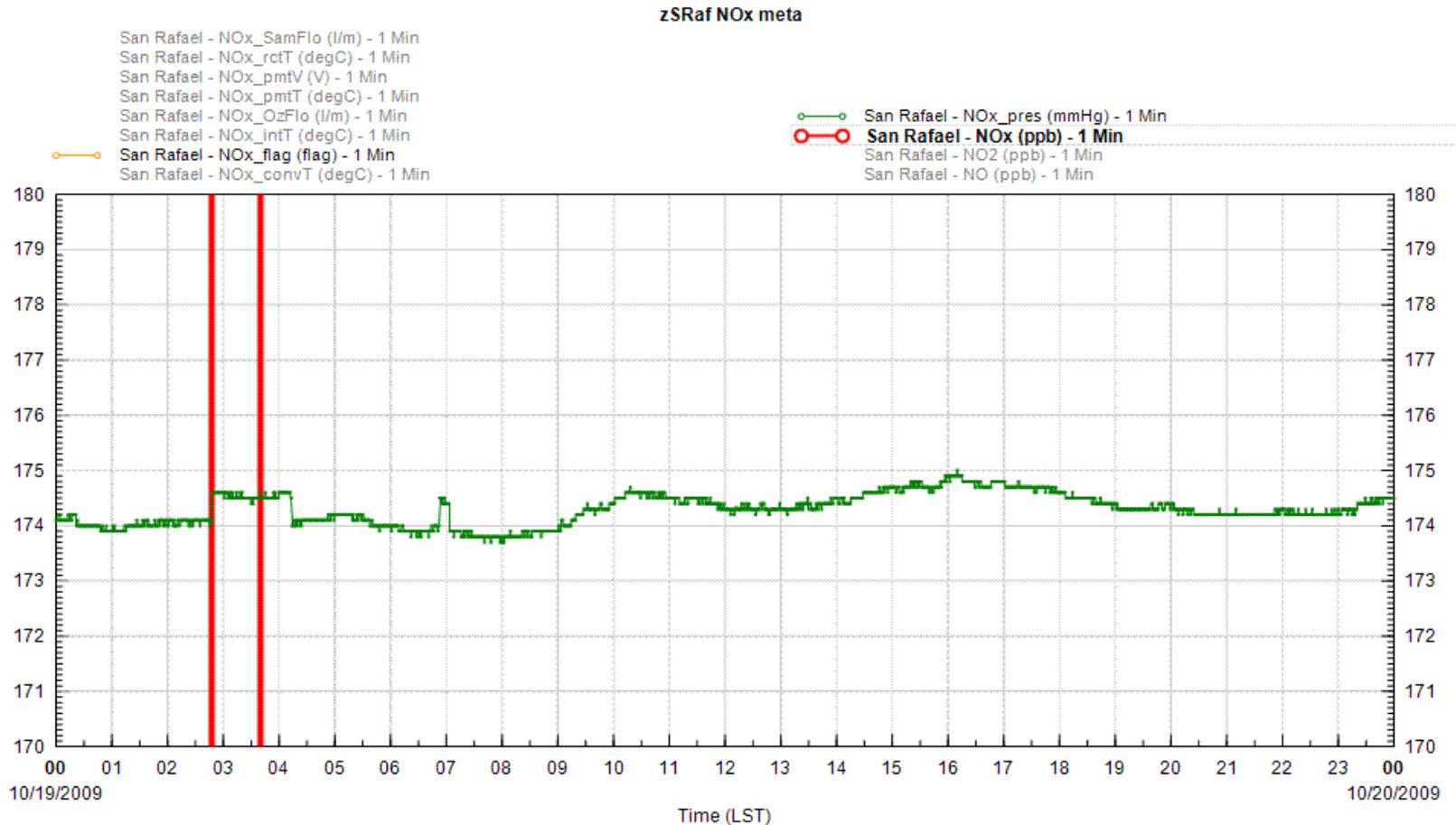
How is Metadata Used

- Metadata is managed in a database along with the ambient data from the same instrument; the one-day graph below shows over 11,000 1-minute values
- Real-time (daily) metadata is difficult to review - the information is buried in small changes
- Good metadata shows an absence of change indicating the instrument is stable and will maintain consistent calibration checks (zero/span/precision checks)



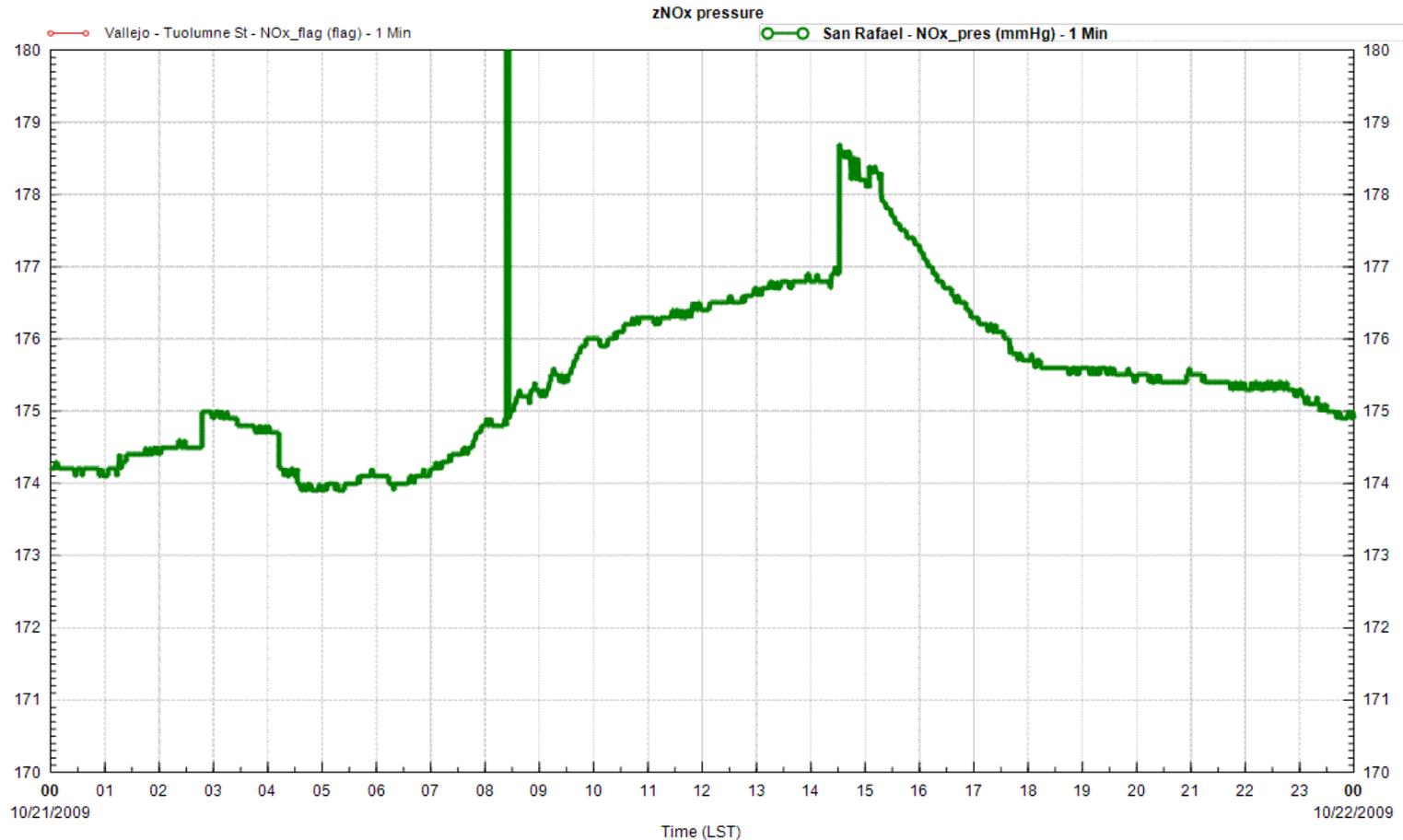
How is Metadata Used

- Expanding the pressure scale reveals a slight pressure change during early morning auto-calibrations
- Auto-cal pressure change of 0.5 mmHg is within normal instrument pressure variation
- Vertical red lines indicate the period of calibrator gas flow



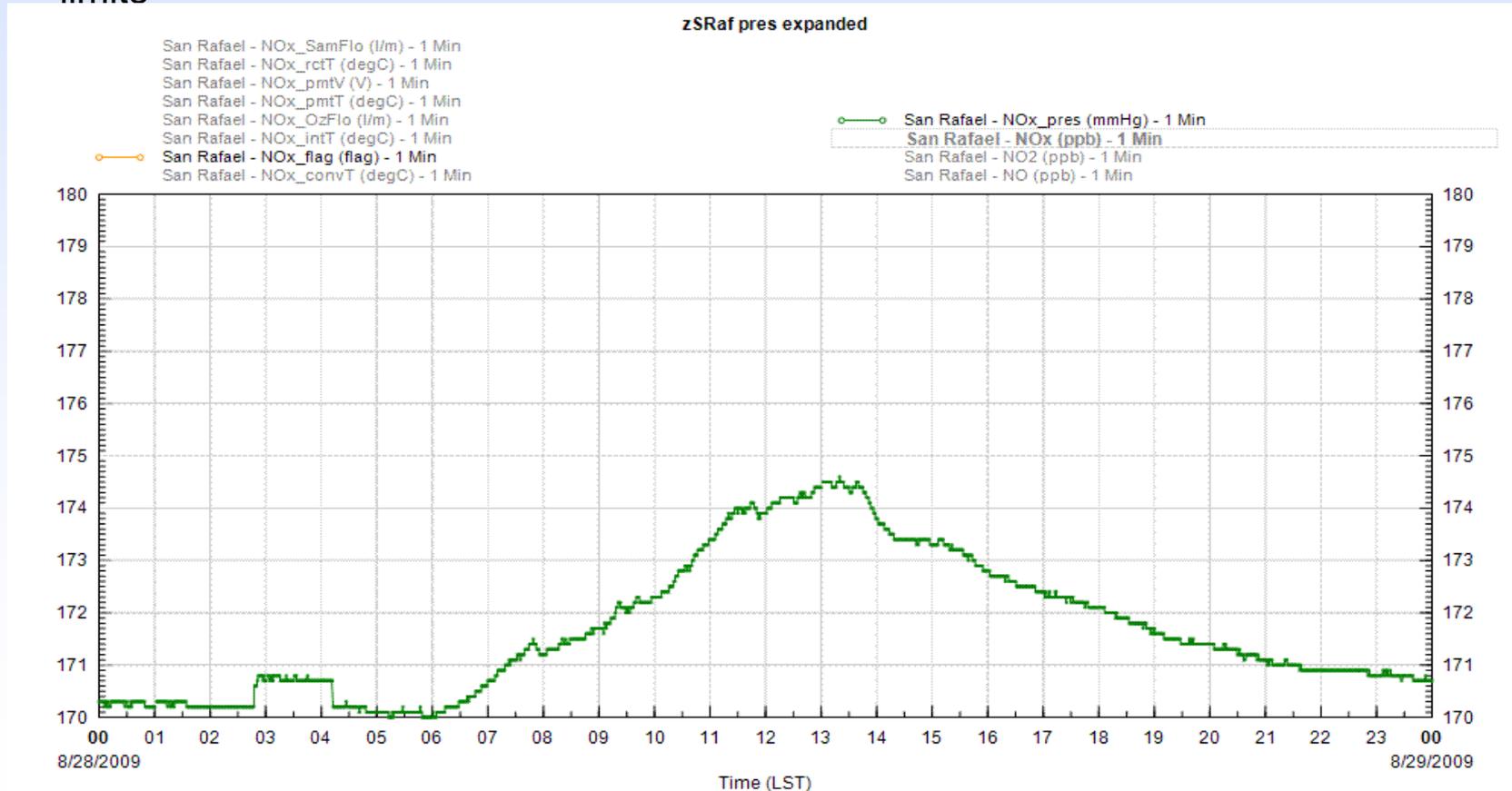
How is Metadata Used

- A different day shows a pressure spike and a small (1.5 mmHg) jump
- Both of these events can be identified by database auto-QC checks and email notices sent to a technician for prompt investigation
- Database auto-QC tests are far superior to human reviewers at identifying small data anomalies within large volumes of instrument metadata



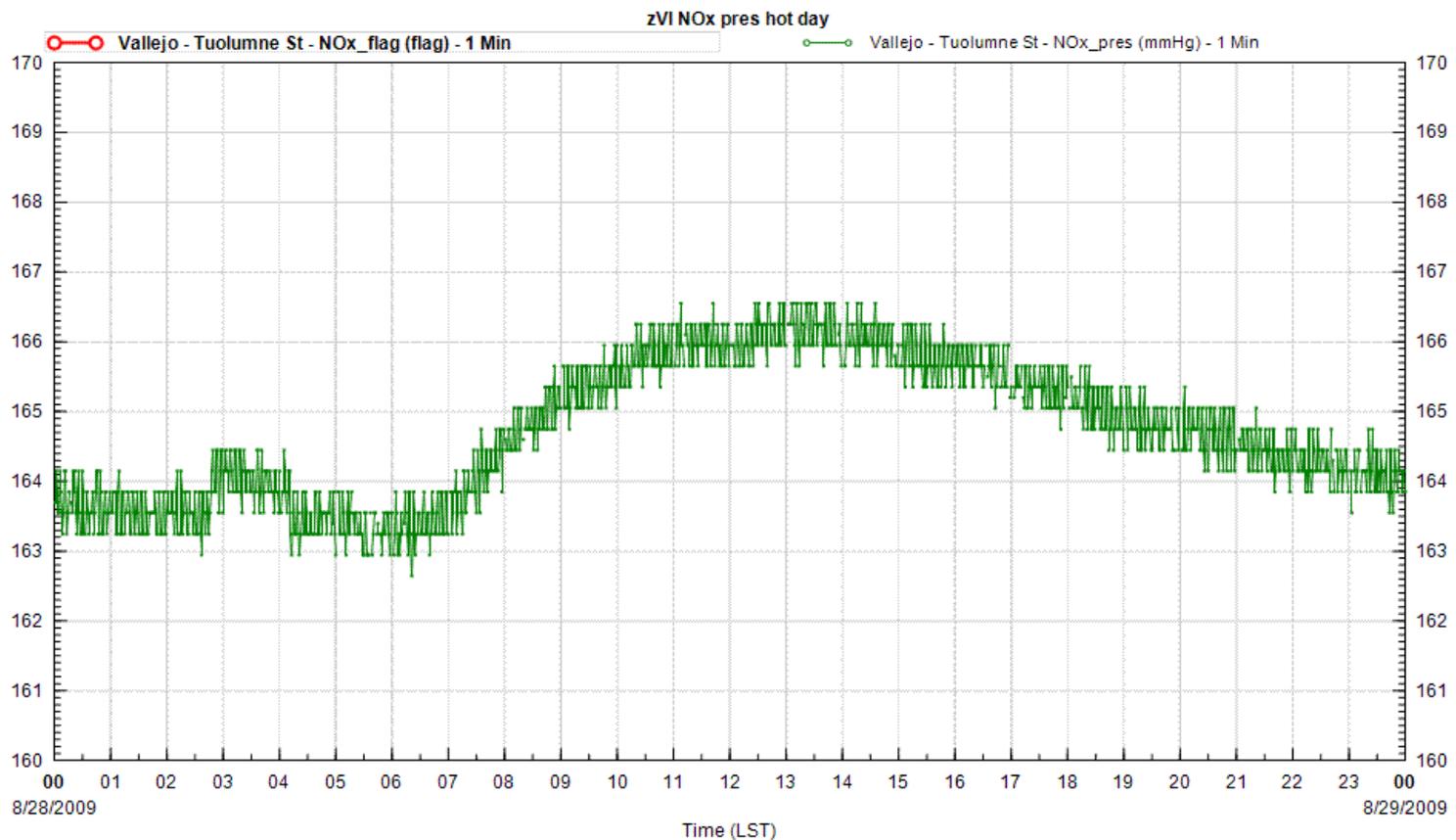
How is Metadata Used

- Depending on individual NO_x sample pumps and the pump location, a hot day can decrease pump efficiency
- Auto-cal NO_x measurements generally show about -1% sensitivity change per 2 mmHg pressure increase
- Midday NO_x measurements are about 2% low, but still within acceptable operating limits



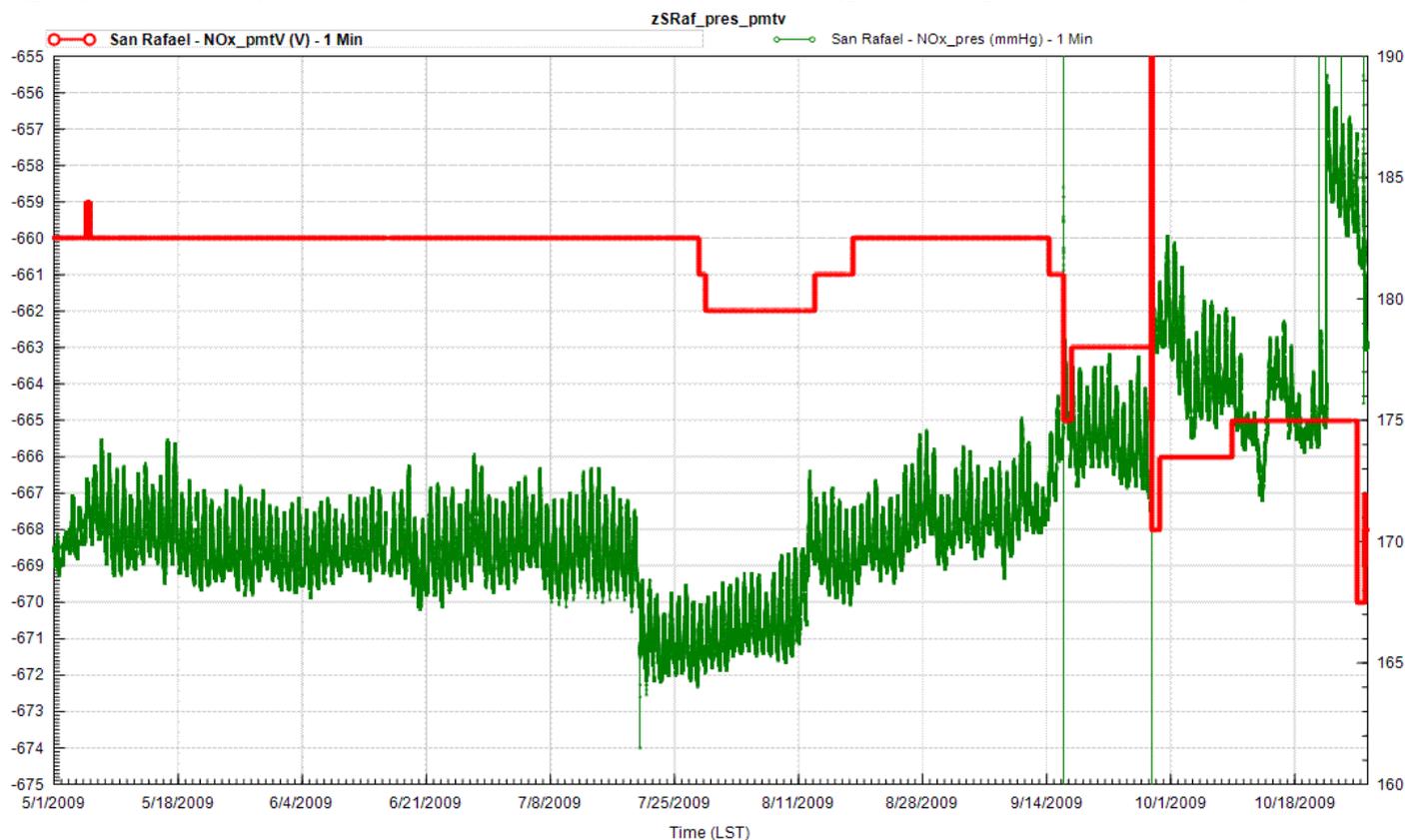
How is Metadata Used

- TECO 42C NOx instruments have a lower resolution data acquisition system than the newer 42i models as shown in the graphic below
- Internal (metadata) measurement precision affects instrument performance, so 42i models generally perform better than the older 42C models



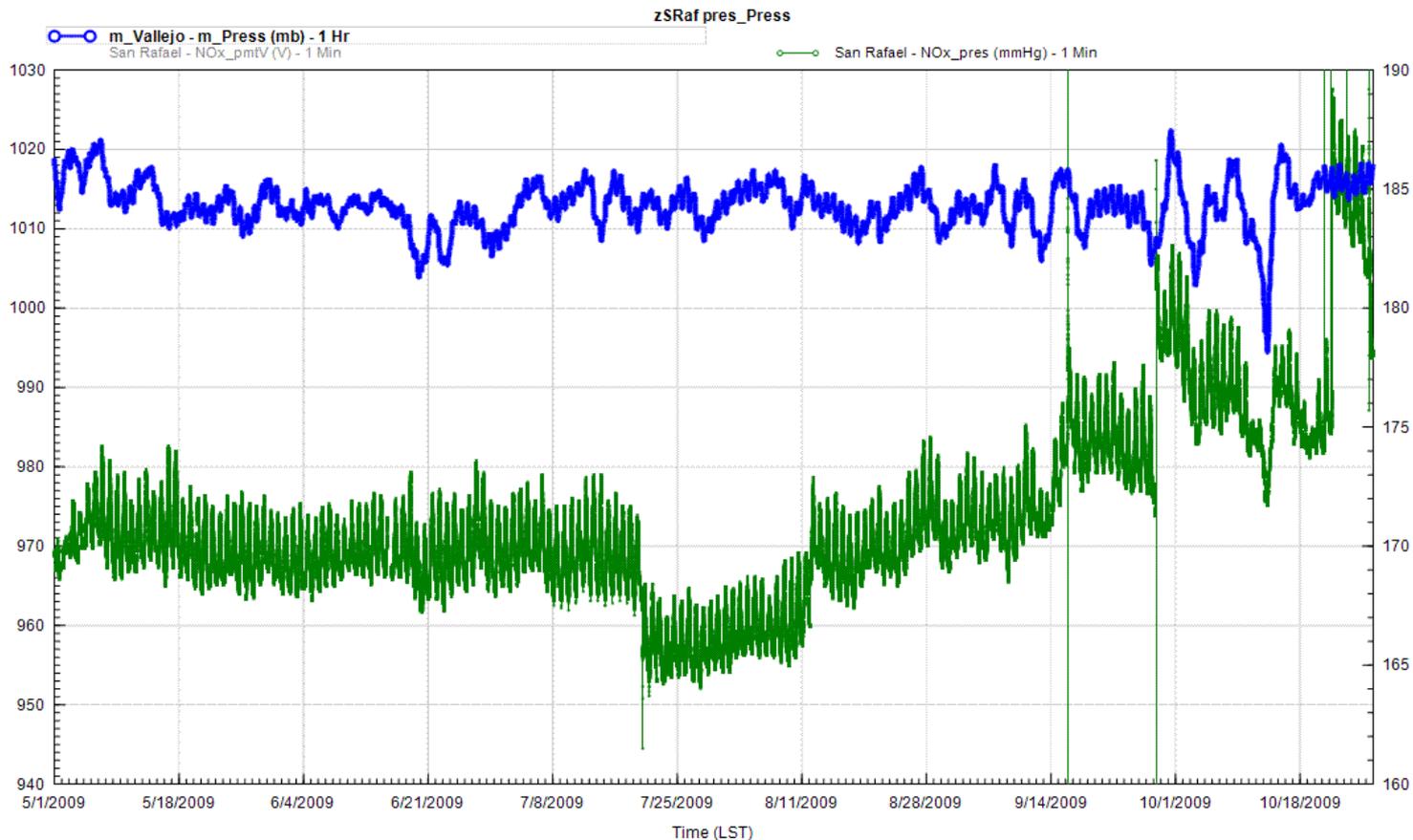
How is Metadata Used

- Compressing the NO_x pressure metadata on a 6-month time scale shows long-term instrument pressure stability
- Pressure discontinuities indicate instrument problems, sample pump changes or other maintenance
- Both pressure and PMT voltage increase instrument sensitivity toward the bottom of the graph so as pressure increases, PMT voltage is adjusted to compensate



How is Metadata Used

- NOx pressure is affected by ambient atmospheric pressure
- A distinctive change in pattern occurs about October 13 during passage of an unseasonably strong storm
- Plotting ambient pressure shows a strong correlation to the NOx pressure and explains the pattern change



How can Metadata Improve Ambient Data Quality

- More consistent instrument calibration checks
- Improved data confidence levels
- Better ambient data completeness
- More robust data correction procedures
- A more defensible data set overall

What Eventually Happens to Instrument Metadata

- After all data reviews are complete and the ambient data (NO and NO₂ for our examples) are final, metadata is no longer needed
- After final ambient data is submitted to AQS, instrument metadata may be purged from the database
- One record/day is retained from purged metadata so that long-term instrument trends may be analyzed
- If necessary, metadata may be selectively reinserted into the database from backups

In Summary

- Instrument metadata is a relatively new tool available to improve ambient data quality and provide efficient remote instrument management
- Metadata management requires a capable datalogger and adequate communication bandwidth
- Good metadata is very stable; auto-QC tools are extremely useful to detect metadata changes and alert technical staff
- Start metadata analysis with a parameter that gives the most ‘bang for the buck’ – something easy to interpret the provides good ambient data quality improvement
- Analyze more metedata as resources permit
- When a failure occurs, do a forensic analysis of the metadata learn what happened and how it might be prevented in the future
- Let metadata tell a story about the ambient data

Contact Information

Mark Stoelting

Bay Area Air Quality Management District

939 Ellis St.

San Francisco, CA 94109

mstoelting@baaqmd.gov

415-749-4619