Developing a Set of SAE J1939 Data Fields For Better Data Analysis

ERG:
Michael Sabisch
Doug Jackson
Sandeep Kishan

US EPA:
Carl Fulper

UCR / CE-CERT:
Kent Johnson
Thomas Durbin
Georgios Karavalakis
J. Wayne Miller
Yu Jiang
Mark Villela
Daniel Sandez

Presented at the UCR / CE-CERT 6th Annual PEMS Conference
March 18, 2016

Background and Motivation

- **What's the issue? The challenge of finding the most beneficial (and commonly populated) parameters from among 5000+ SAE J1939DA suspect parameter numbers (SPNs) for a particular study**

- **Opportunity: SAE J1939 offers a wide variety of engine, emission control, and other vehicle operational parameters**, such as:
  - Air and fuel input information; temperatures pressures and flows; engine control and operational conditions, accessory status, engine output power, (absolute) and emissions (NOx and PM), DPF soot accumulation; vehicle speed; ambient conditions, hybrid electrical parameters; drivetrain ratios; payload mass; to name a few.
- **Solution: Develop a "Master List" of fundamental parameters** generally needed for any study. **Supplement with additional study-specific parameters** prior to start of study to avoid overlooking critical parameters and to ensure all necessary data is collected.

Parameter Selection

- **Start with fundamental parameter "Master List"**
- **Consider additional data that is needed for specific study:**
  - Regional emissions or activity study
  - Powertrain and emission control component performance
  - Drive / duty cycle, hot spot quantification
  - GHG
  - Other?
- **Select SAE J1939 parameters based on study needs**- Careful consideration and selection of parameters ***before*** start of study helps ensure consistent data acquisition and collection of parameters needed to conduct analysis
- **Warning – acquisition of too many parameters can slow acquisition rate**

In-Field Data Collection

"Here comes the data!"

- **A PAMS study can result in an enormous amount of data!** For example, 100 parameters / 1 Hz rate / 6 hrs/day, 5 days/week / 3 months = 24 million data fields!
  - Transfer (compressed data) can reach 1 GB / month (or more)
  - Processed data can double or triple in size
- As discussed in an earlier presentation, **cellular is an attractive option**
  - To capture PAMS data "real-time" (daily upload)
  - To monitor in-use performance of test vehicles
  - To monitor in-use performance of PAMS

Data Processing and QC

- **Start data QC process early to identify and correct data issues**
- **Once data is received and processed, an effective method must be used to handle, validate, and analyze the data.**
  - **SAS[1] -** processes data line-by-line, and therefore can accommodate large datasets without any special data handling measures.
  - **R[2]** - generally handles entire data frame in RAM, special consideration required for extremely large datasets
    - Special packages to accommodate large data frames / matrices (i.e., bigmemory, ff, others…)
    - Use of vectorized functions / sub setting data frames / storage within a database (such as MySQL) and access subsets as needed
  - **Others** - Many other types of software / languages available

1 www.sas.com
2 www.r-project.org

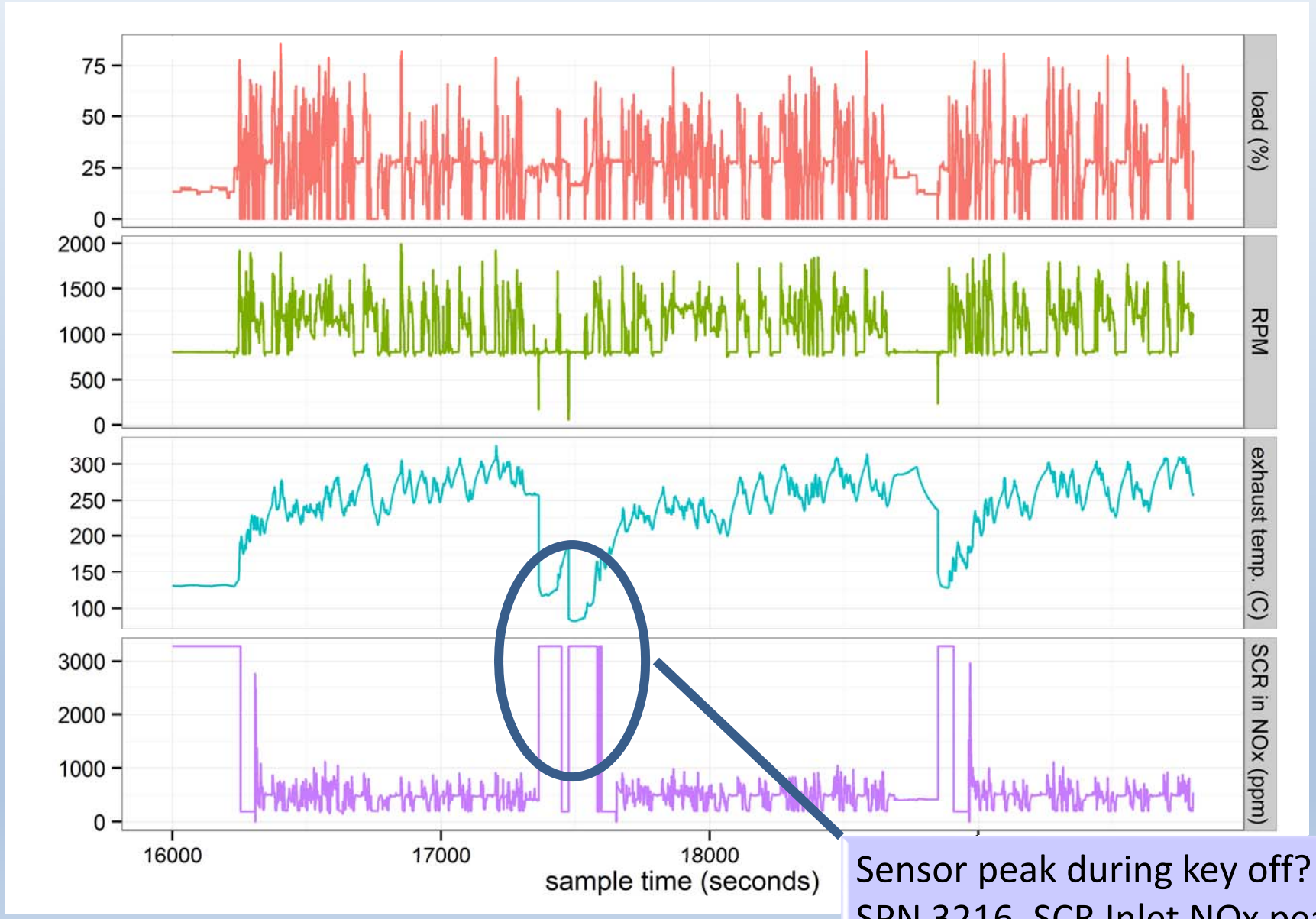
Data Processing and QC Examples

- As will be shown, just as with emissions data, **QC and validation of J1939 data recorded by PAMS is needed to identify data issues**.
- **Same techniques used for PEMS and dyne emissions data can be applied to PAMS data.**
  - Time series analysis, followed by drilling / filtering of data
  - Various scatter plots
  - Data summaries, statistics, identification of outliers, etc.

The following examples are preliminary and are based on uncorrected data.

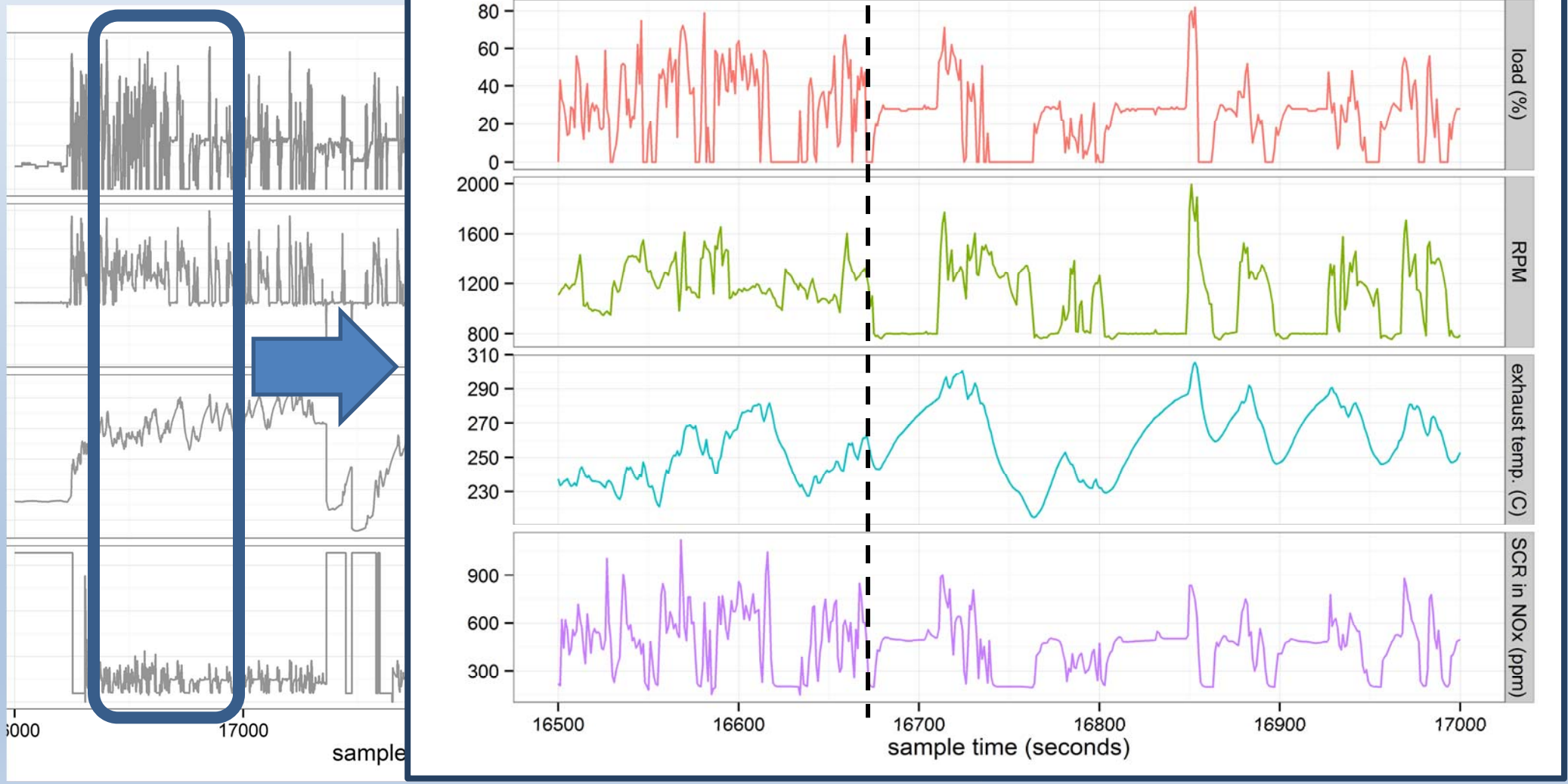
Example Time Series

Time series analysis is critical for reviewing in-network data streams



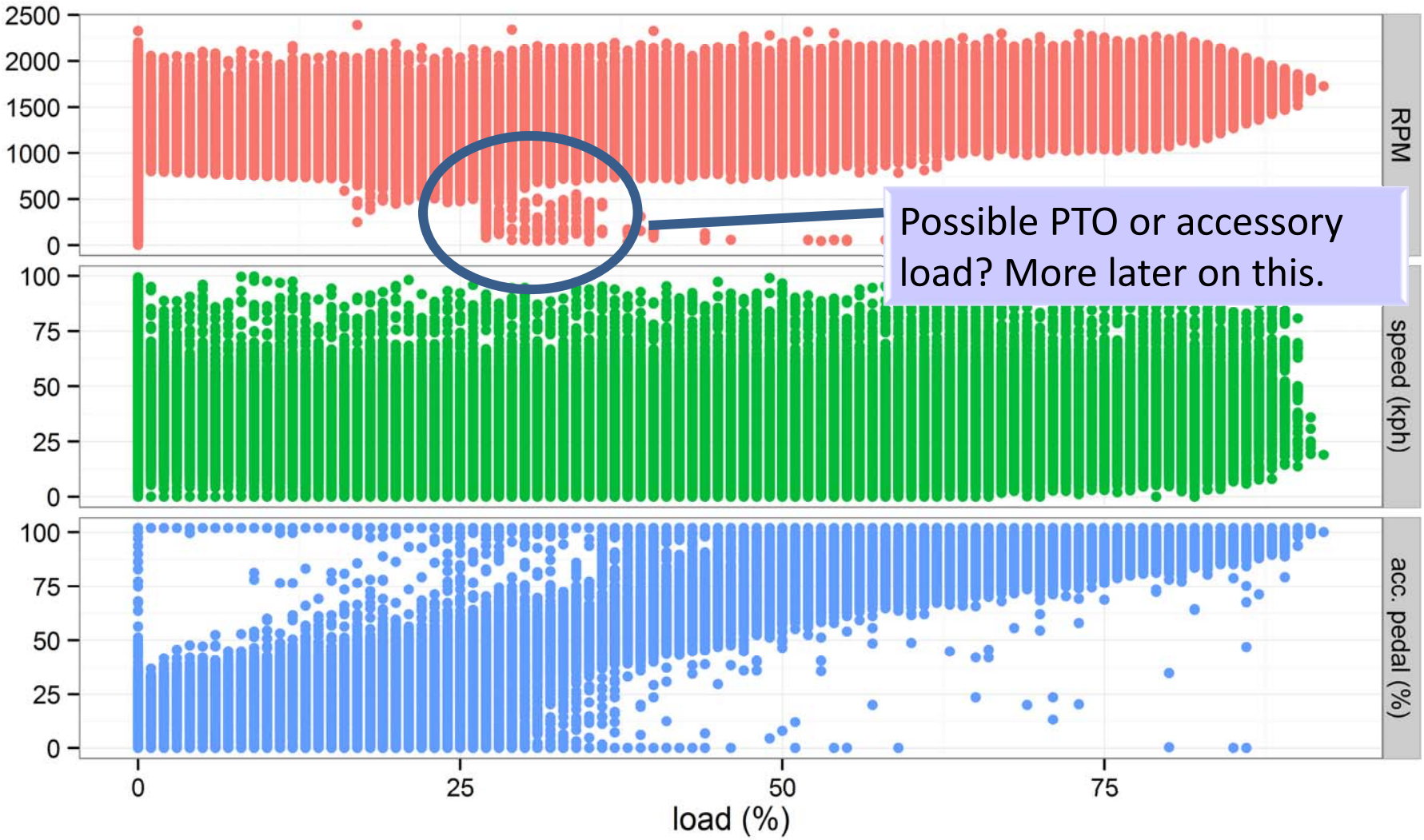
Example Time Series Zoom

"Zoom" windows can reveal interdependencies and be used to identify and investigate data issues and time alignment

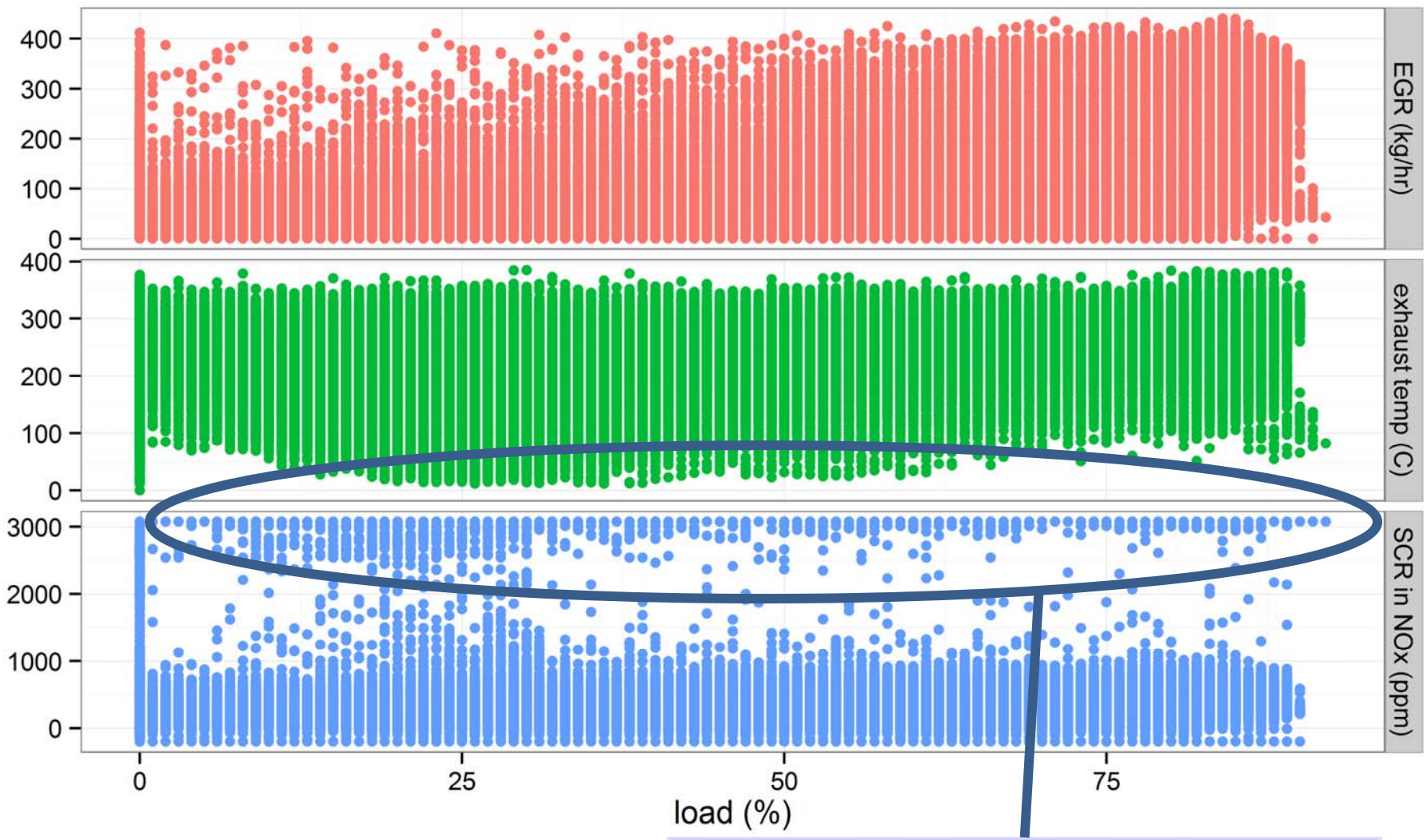


Parameter Profile Review

Scatter plots can give quick visual overview of data and reveal operation ranges and potential problems

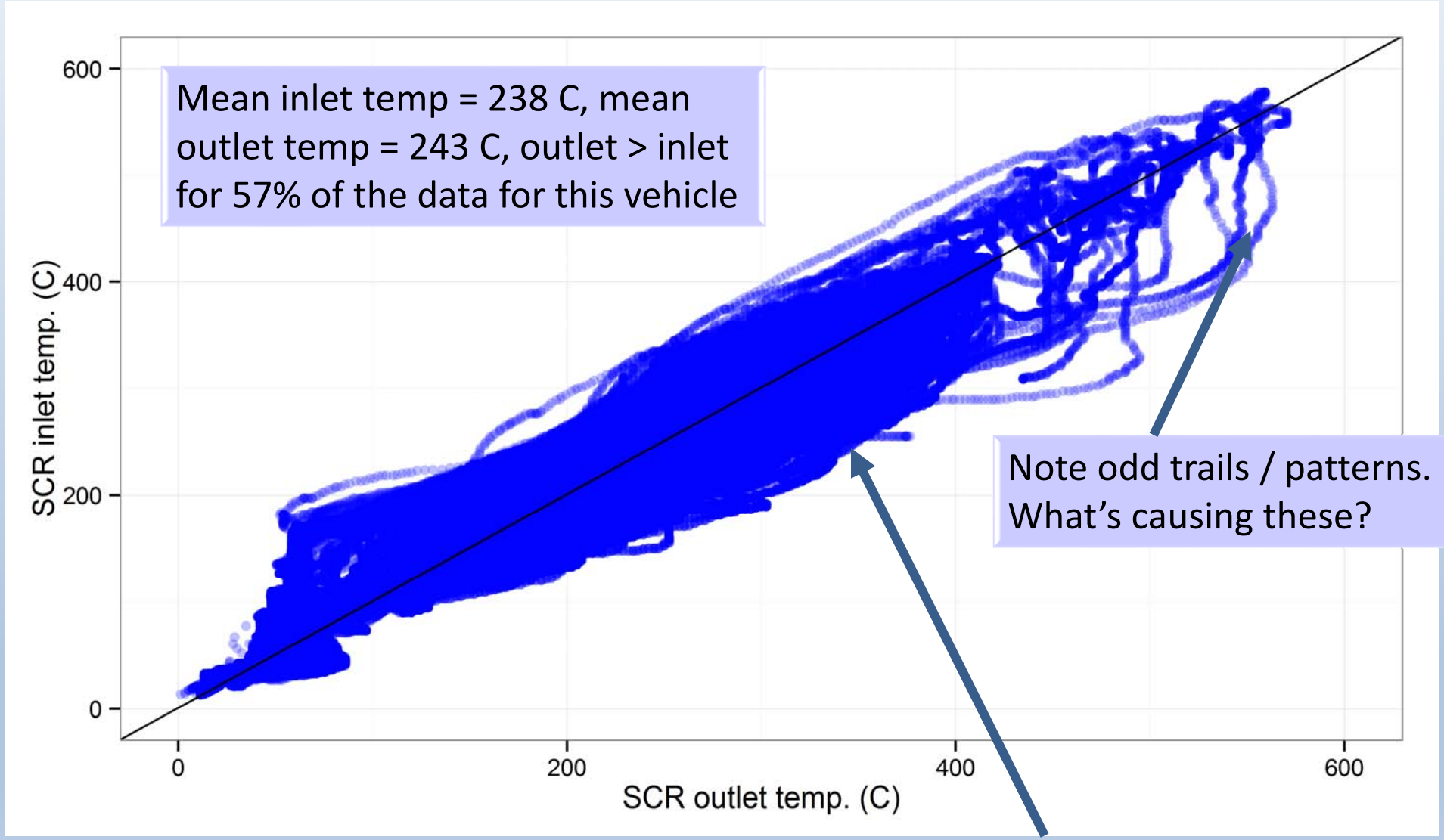


Parameter Profile Review (contd.)



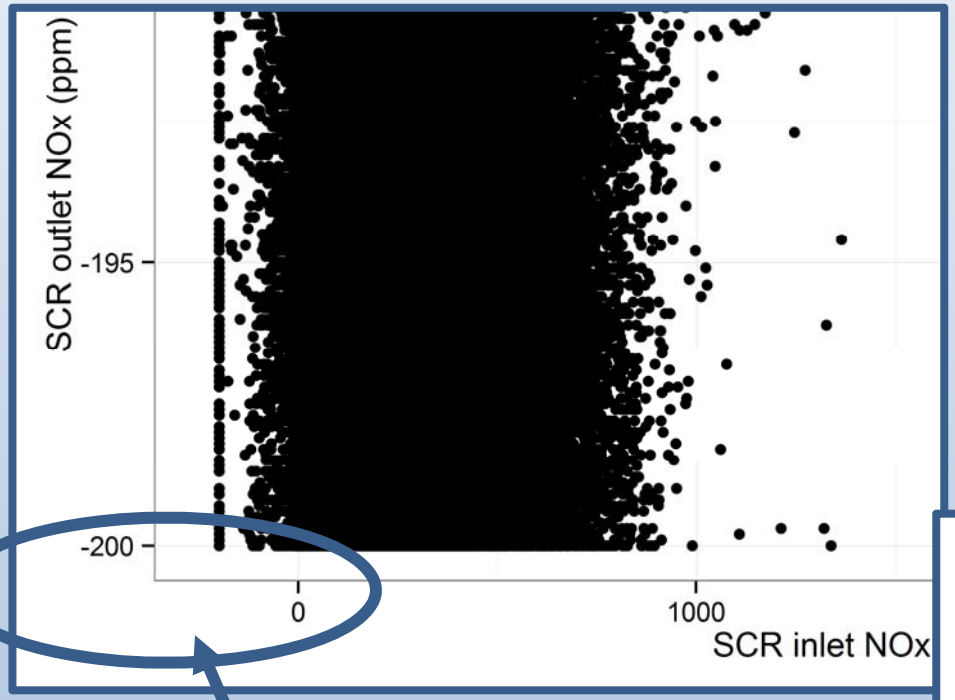
Peaked SCR inlet across all loads suggests inlet NOx reading ***could*** be underestimated at times. We'll revisit.

SCR Inlet vs. Outlet Temperatures



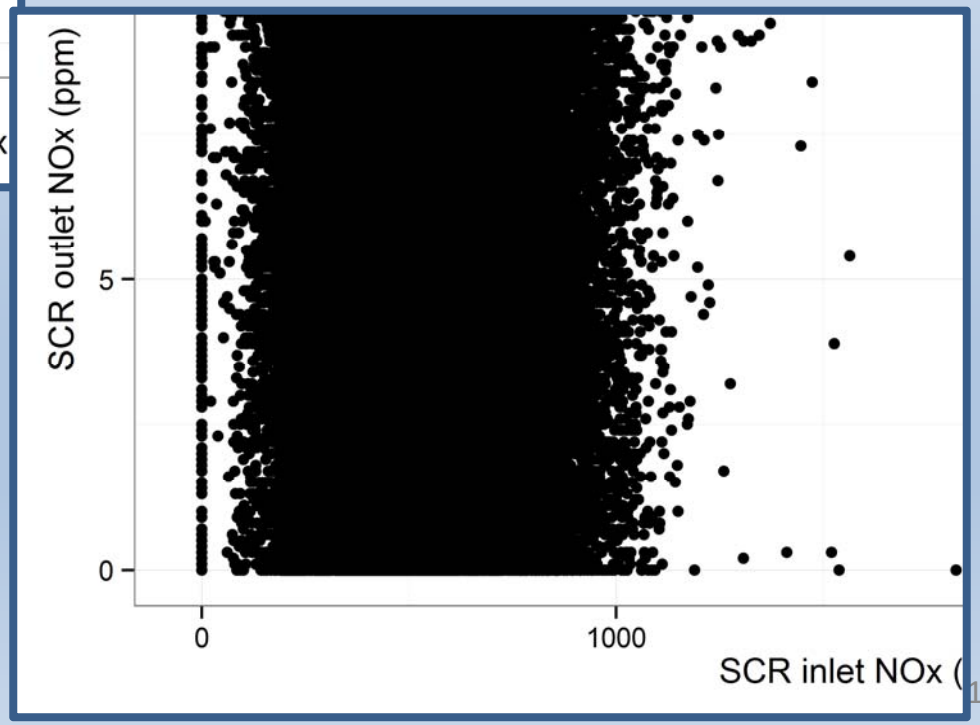
NOx Offset

Review of scale factors, ranges and offsets is necessary to ensure all recorded parameters are valid.



Low range suggests 200 PPM offset not applied for SPNs 3220 and 3230.

After application of offset.



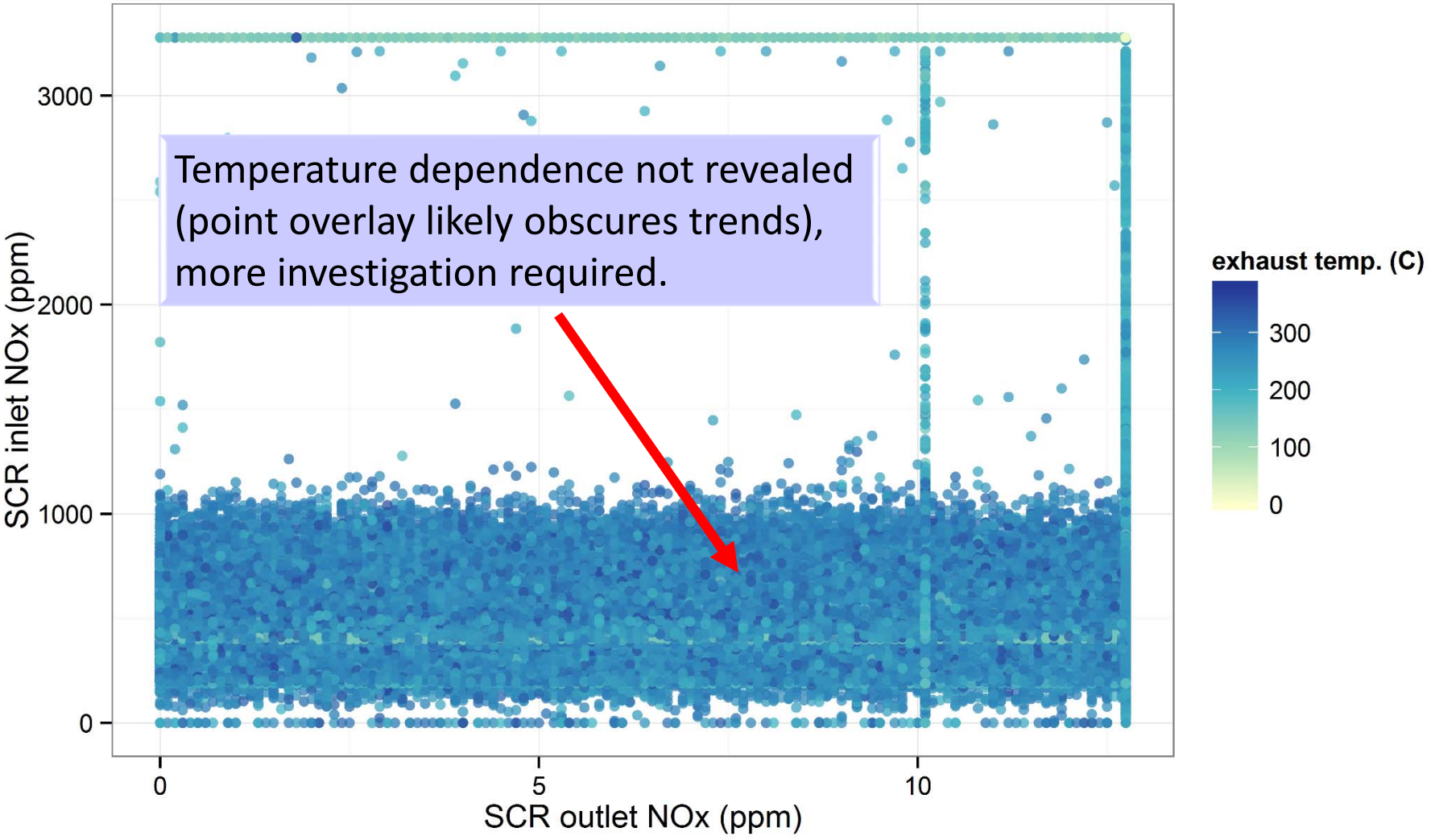
Data Analysis Examples

- **Various parameters may be selected to review system performance as reported by onboard system** – For example,
  - **SCR Performance**: Are exhaust temperatures too low to allow effective SCR performance? How effective is the SCR for the particular use / application?
  - **DPF Performance**: Do frequent short, low-speed drives or extensive PTO operation inhibit DPF regeneration, resulting in excessive soot accumulation?
  - **EGR Performance**: What effect does EGR have on SCR inlet NOx values for the particular use / application?
  - **Others?**

All examples presented are preliminary and are based on uncorrected data.

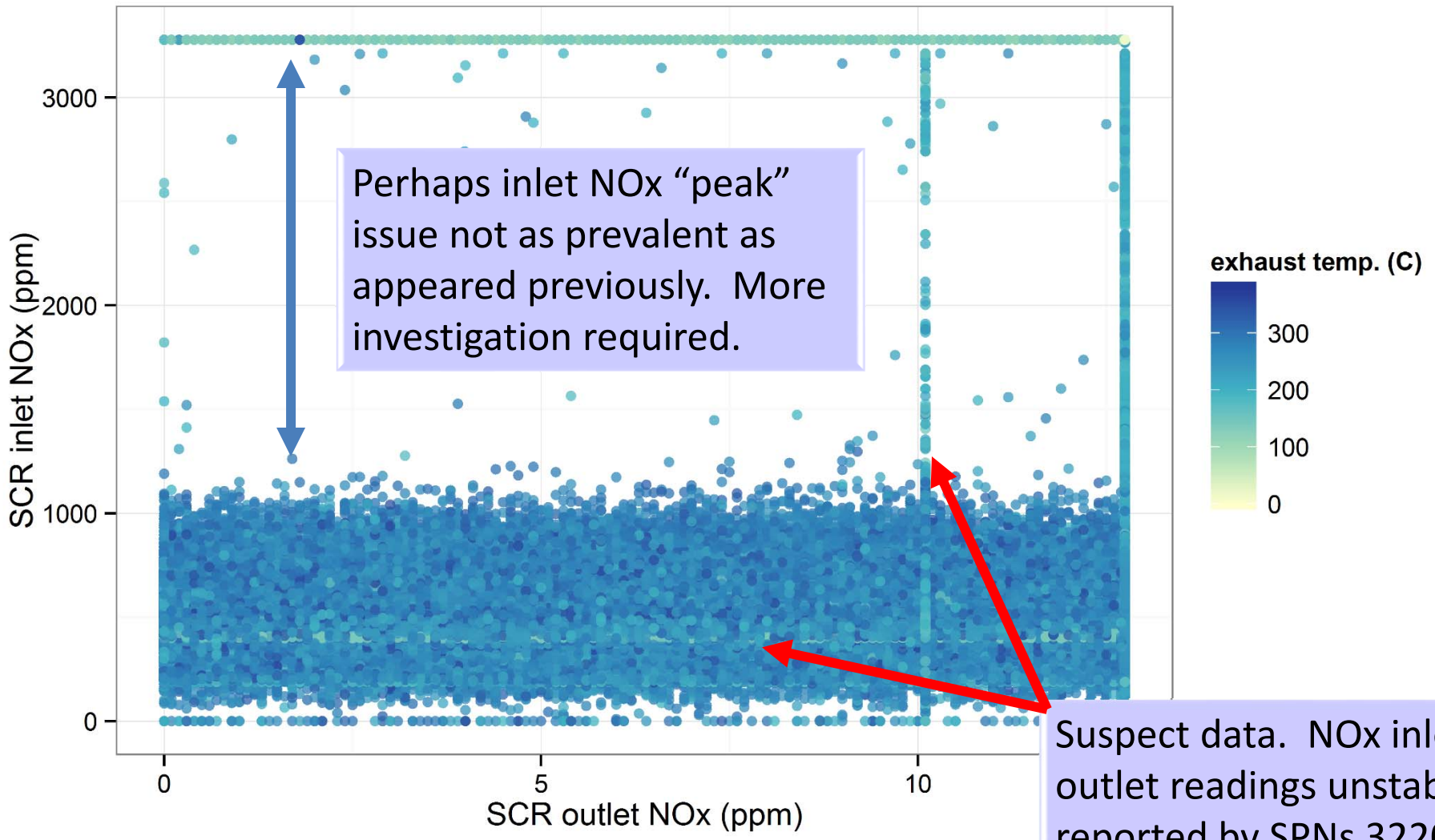
SCR Inlet vs. Outlet NOx

with Exhaust Temperature Indicators



SCR Inlet vs. Outlet NOx (contd.)

Other Observations

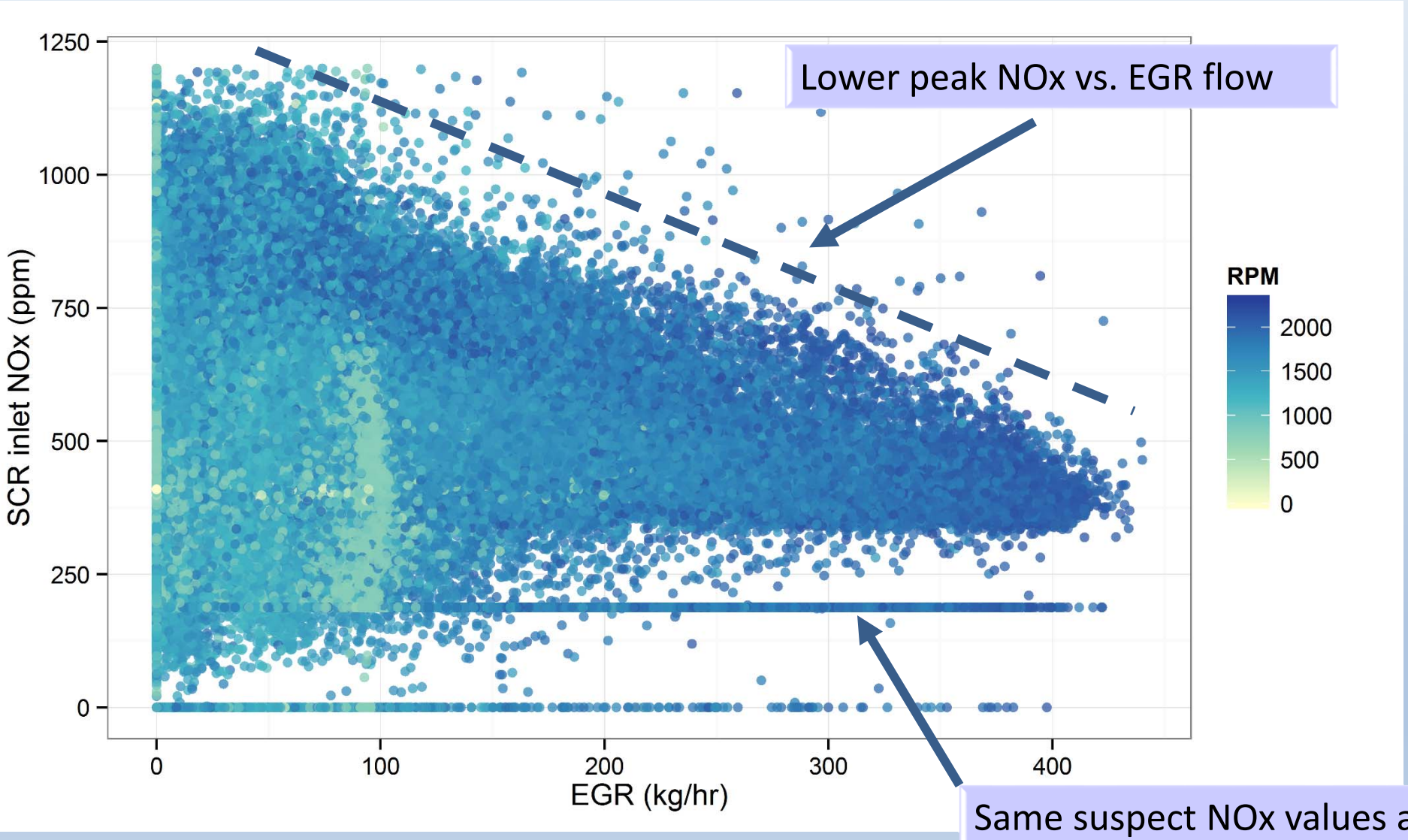


Perhaps inlet NOx "peak" issue not as prevalent as appeared previously. More investigation required.

Suspect data. NOx inlet and outlet readings unstable, as reported by SPNs 3220 (inlet) and 3230 (outlet) indicators of "3".

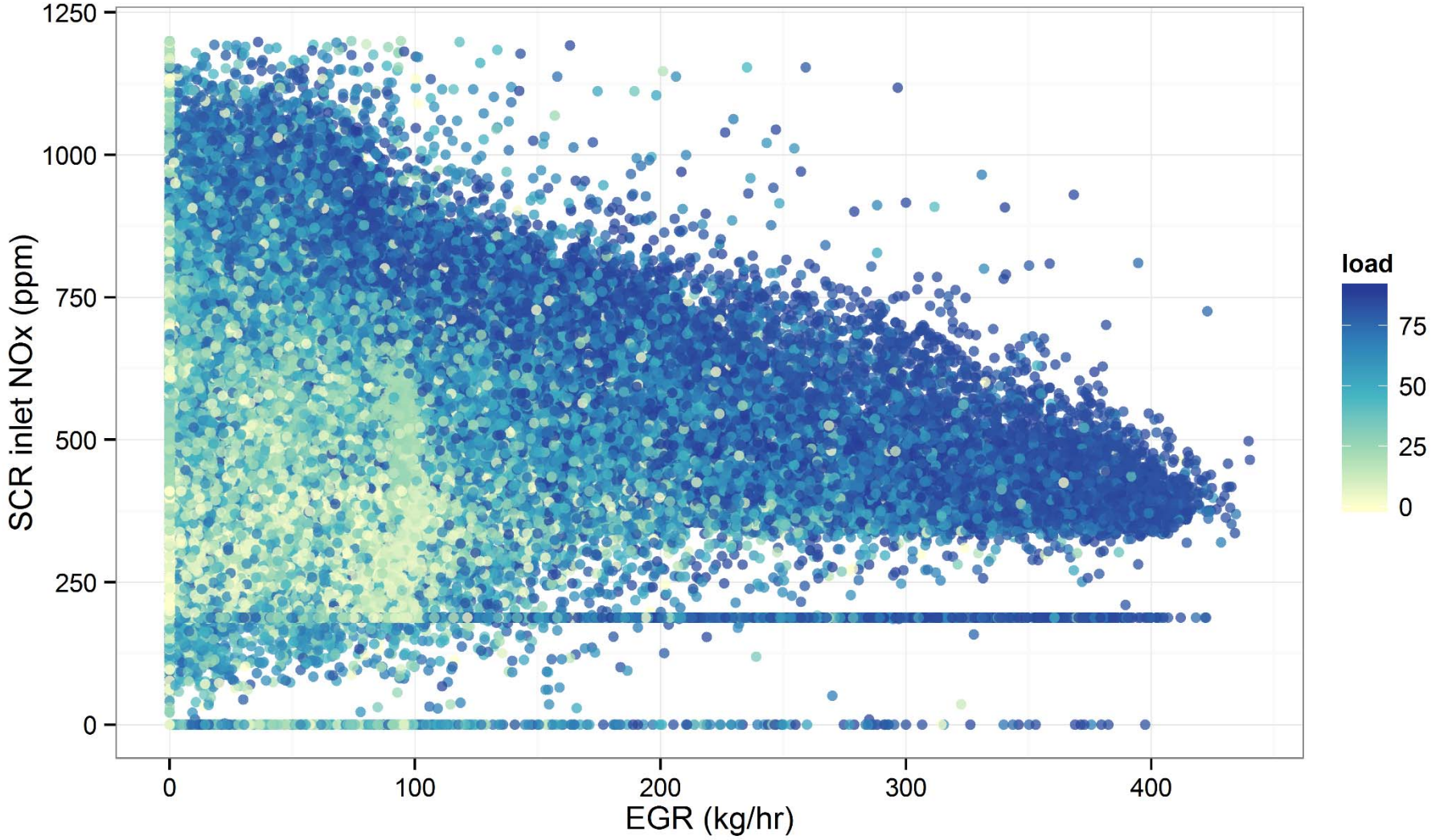
EGR Influence on NOx ("by RPM")

Some possible ERG influence can be seen on peak NOx readings



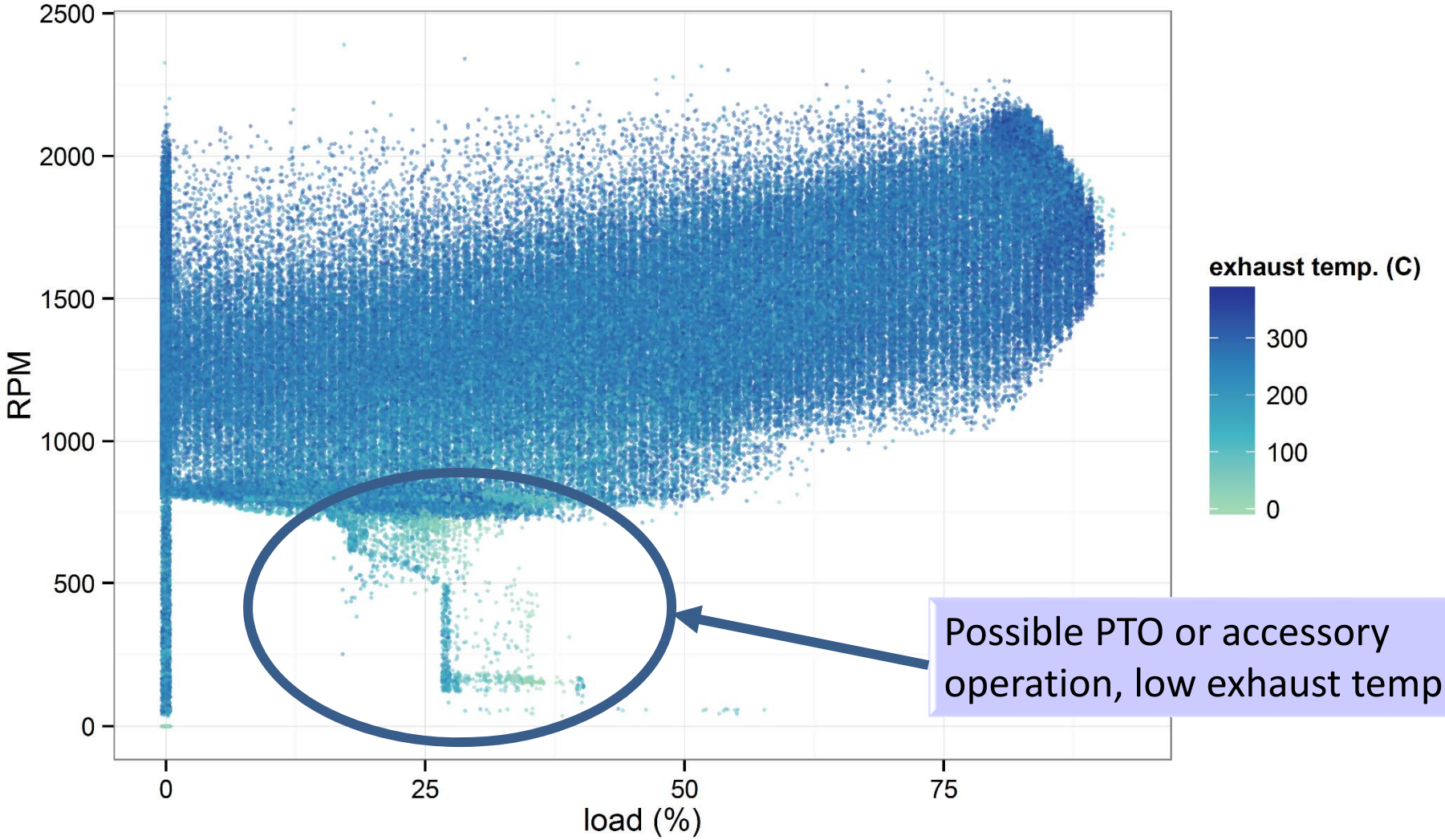
EGR Influence on NOx ("by Load")

Similar trends can be seen when viewed by load



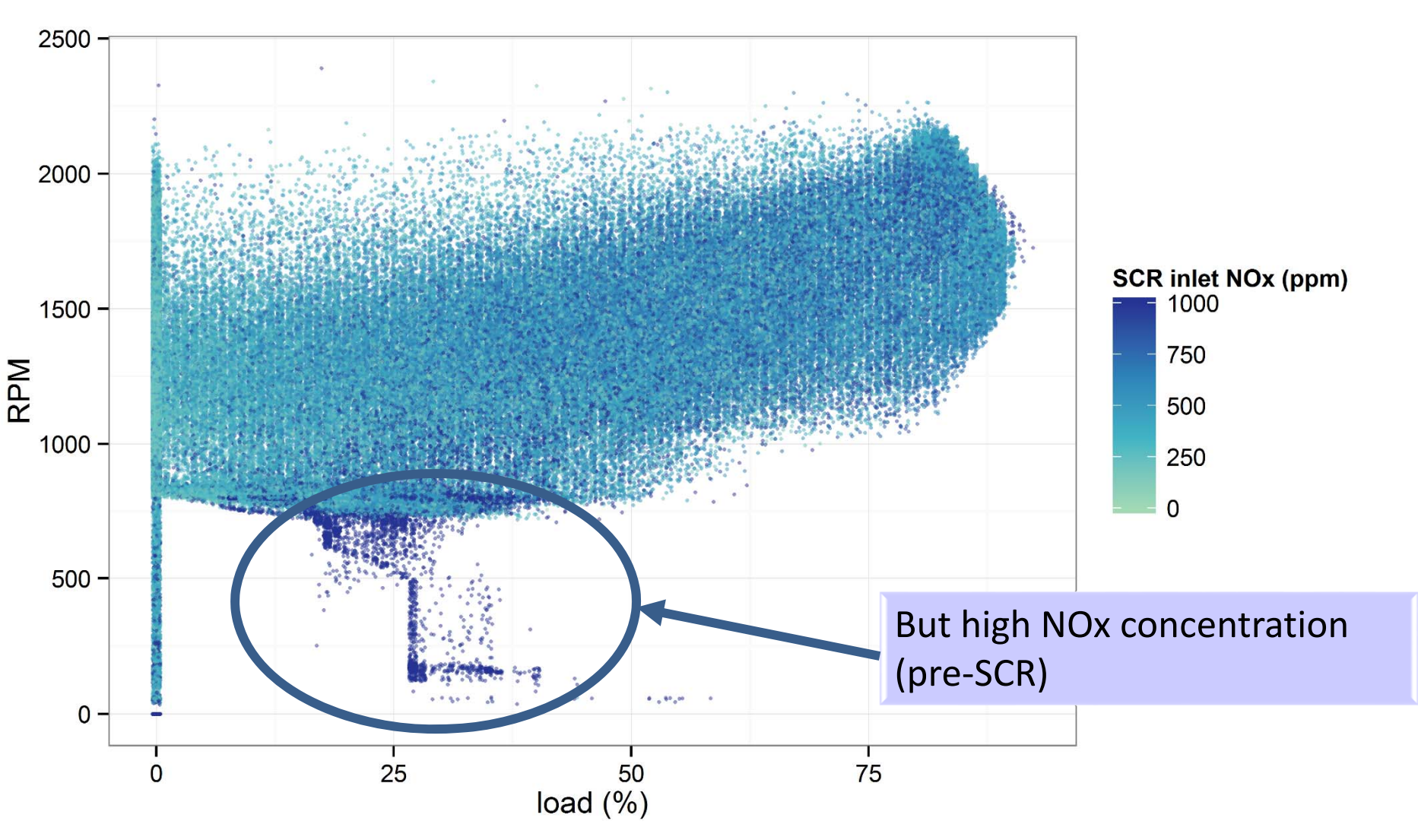
RPM vs. Load (with exhaust temp)

Plotting against various parameters can reveal interesting trends



RPM vs. Load (with SCR inlet NOx)

Plotting against various parameters can reveal interesting trends



Data Validation and Compilation

- **Follows data QC and analysis**
- **Data validation:** Consists of verification and conversion, as needed, of field lengths and types (integer and decimal numeric, characters, mixed, dates, nulls, etc.) for input into database,
- **Data compilation:** consists of populating tables with metadata (establishment, truck, and test information), and creating repository database with all test program data and appropriate data table relationships and documentation
- **Documentation:** To be accompanied by complete database schema, data dictionary and other necessary documentation

Summary and Recommendations

- Develop a “Master List” of fundamental SAE J1939 DA parameters, and spend time before start of study supplementing the “Master List” with study-specific parameters based on study goals
- If possible, capture and analyze data “real-time” to identify and correct problems as soon as possible
- Activity study data accumulation can be enormous, be prepared when designing methods to store, transfer, and analyze the data
- For various reasons, even on-board system data can require QC. Spend the necessary time to QC and assess your data
- Provide a data dictionary and database schema with final database

Acknowledgements

ERG gratefully acknowledges the following study partners and presentation co-authors :

Kent Johnson, Thomas Durbin, Georgios Karavalakis, J. Wayne Miller, Yu Jiang, Mark Villela and Daniel Sandez (**UCR / CE-CERT)** for conducting the activity study from which this data was acquired, for project collaboration and for providing technical support

Carl Fulper **(US EPA)** for providing ERG support funding, equipment support, and technical study support

Doug Jackson and Sandeep Kishan **(ERG)** for providing technical study support

Thank you for your time!

michael.sabisch@erg.com