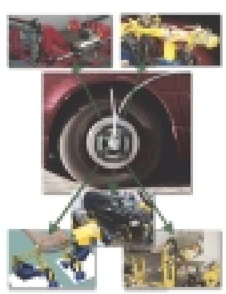
Automated Road Load Data Analysis for Verification of Data Integrity

SAE – FD&E Road Load Data Acquisition Division - October 16 2001

Darragh Murphy - MTS Systems

Presentation Outline



- Background
- "Bad" Data
- Anomaly Detection
- Spike Detection/Correction
- Automated Data Validation Process

Background

- RLDA is one of the most critical steps in product design and evaluation
  - To begin the product evaluation process an understanding of the service environment and accurate load information is required
  - Before accurate load information can be obtained a prototype specimen is typically required
  - Fewer prototypes are available due to cost and shortening development cycles
  - RLDA engineers typically have fewer opportunities to acquire data but are pressed to have faster turn-around of data

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

The need for "good data"

- Accurate road load data is required for the following:
  - Evaluation of product durability, noise and vibration, performance, and handling
  - Refinement of design parameters
  - Fatigue life estimates
  - Source data for modeling
- "Bad" data is typically not caught until late in testing and evaluation exercises when there are multiple factors indicating problems that have no other explanation other than "is the data suspect?"
- Industry leading Engineering Managers estimate between 30 and 50% of engineering time is spent looking for good data or working with bad data

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

RPC Pro Acquire

- Essential attributes of a software data validation application
  - Ability to automatically process large amounts of data
  - Concise presentation of issue list to user
  - Accurate interpretation of and detection of data anomalies
  - Ability to correct data anomalies
  - Minimal detection of false-positives
  - Minimal user interaction and minimal training requirements
  - Performance
  - Supplemental analysis capabilities

Integrated analytical physical virtual Simulation

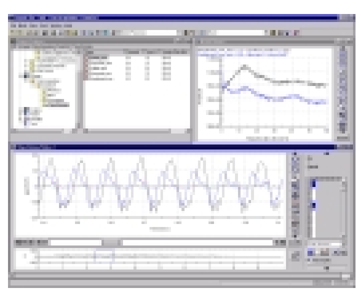
RPC Pro

MTS Proprietary Information

Data Validation - Background

- Data anomaly detection and correction
  - Reference PhD thesis of Tom Hunt, Ford Motor Company – "The Classification and Detection of Anomalies in Digitally Sampled Dynamic Data"
    - This thesis provides extensive reference information regarding data anomaly classification, detection and correction algorithms
  - Modification of and enhancement of Tom Hunt algorithms by MTS R&D group

- Data analysis
  - Use of established RPC signal processing and analysis tools for:
    - Data viewing
    - Signal modification
    - Resampling



Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Presentation Outline

- Background
- “Bad” Data
- Anomaly Detection
- Spike Detection/Correction
- Automated Data Validation Process

Where does “Bad Data” come from?

- Bad data typically comes form one of the following sources:
  - Human error
    - incorrect data acquisition equipment setup
    - Incorrect calibrations, offsets, selection of sample rate
    - incorrect selection and mounting of transducers
    - cabling errors
  - Measurement environment
    - Temperature drift
    - Shocks
    - Noise
    - backlash

What is "Bad Data"?

- Data anomalies can be broken down into two distinct groups:
  - **General anomalies**
    - Anomalies that can be detected by analyzing the data in isolation
    - Simpler anomalies to detect
  - **Specific anomalies**
    - Anomalies that can only be detected by comparing the data with additional information
    - Complicated and difficult to detect
- Real world experience has shown that approximately 80% of dynamic data anomalies are of the general type

RPC Pro

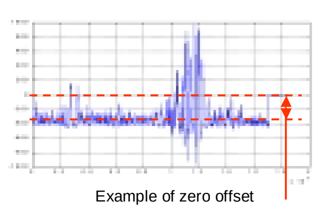
MTS Proprietary Information

General Anomalies

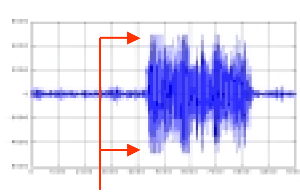
- The following is a common list of General anomalies

  Note: Refer to Tom Hunt's thesis for further details.

  - Zero Drift
  - Zero Offset
  - Zero Shift
  - Spikes
  - Noise
  - Missing Data
  - Intermittent Data
  - Ringing
  - Saturated Data
  - Clipped Data



Example of zero offset



Example of clipped data

RPC Pro

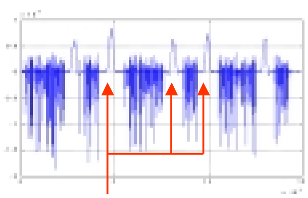
MTS Proprietary Information

Specific Anomalies

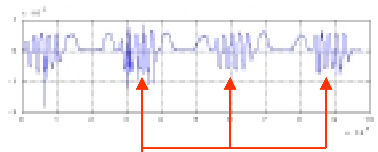
- The following is a common list of Specific anomalies

  Note: Refer to Tom Hunt's thesis for further details.

  - Polarity errors
  - Amplitude errors
  - Data breakup
  - Impulses



Example of data breakup



Example of impulses

MTS Proprietary Information

What is the effect of "Bad Data"?

- In general anomalies cause misleading results when data is further processed, some examples are:
  - Incorrect statistical summaries
  - Incorrect Rainflow histogram data
  - Inaccurate spectral calculations

- This is further compounded when the following potential is considered:
  - Erroneous fatigue life calculations leading to under/over design
  - Erroneous physical simulations leading to incorrect failure mode and location reproduction
  - Inaccurate reproduction of time to failure
  - Erroneous virtual model results

MTS Proprietary Information

Presentation Outline

- Background
- "Bad" Data
- Anomaly Detection
- Spike Detection/Correction
- Automated Data Validation Process

MTS Proprietary Information

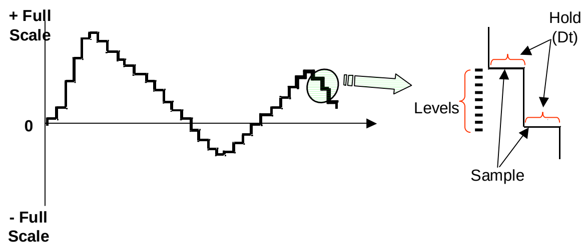
Anomaly Detection

- Detection of anomalies is done in two parts:
  - Initial check of data for the following problems:
    - Under resolved data
    - Under sampled data
    - Saturated channels
    - Missing Data
    - Signal to Noise issues
  - More intensive search for anomalies such as:
    - Spikes *(<- discussion limited to spikes due to time restrictions)*
    - Jumps
    - Data breakup
- Improper detection and correction of anomalies can cause more problems than it fixes!!

MTS Proprietary Information

Anomaly Detection

- Under resolved data
  - This validation calculates the usage of A/D digitization range, when optimized to the range of data, full advantage is taken of the available precision.
  - If the data is using 9 bits or less (512 levels), consider the data under resolved.



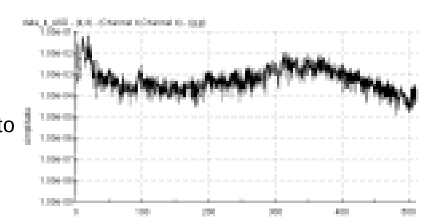
MTS Proprietary Information

Anomaly Detection

- Under sampled data
  - Establish bandwidth of signal (= frequency below which 95% of signal energy found)

Typical spectrum

  - If sample rate to bandwidth ratio is less than the threshold, channel is determined to be under-sampled

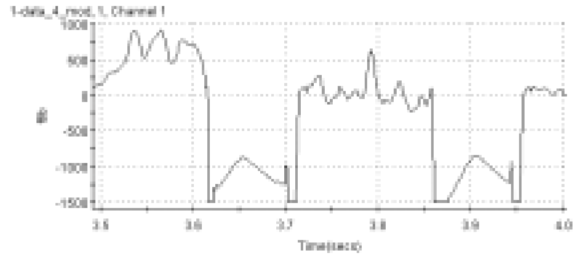


"Under-sampled" spectrum

MTS Proprietary Information

Anomaly Detection

- Saturated channels
  - Examine data for maxima/minima equal to full-scale



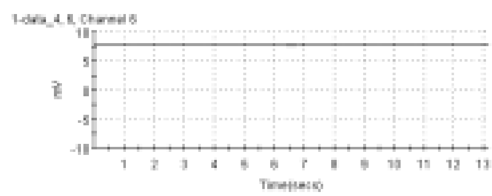
Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Anomaly Detection

- Missing Data
  - All data points found at the “same” value
  - Typically all data equal to zero or plus/minus full scale



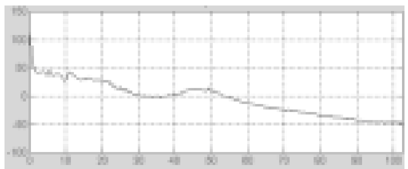
Integrated analytical physical virtual Simulation

RPC Pro

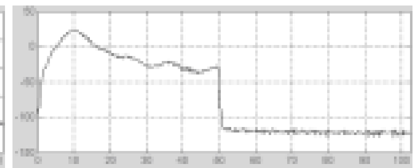
MTS Proprietary Information

Anomaly Detection

- Signal to Noise issues
  - Noise average power is calculated from signal power spectral data
  - Signal power is calculated as (Average Power – Noise Power)
  - Signal to Noise ratio is calculated as: $SNR = 10\log_{10}\left(\frac{P_s}{C}\right)$



SNR ~ 50dB



SNR ~ 60dB

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

MTS

Presentation Outline

- Background
- “Bad” Data
- Anomaly Detection
- Spike Detection/Correction
- Automated Data Validation Process

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

MTS

Spike Detection

- Spike Detection
  - Spikes are high frequency impulses which occur randomly in data and are events which the physical system is not capable of producing.
  - Spikes are difficult to detect
  - While detecting the presence of spikes in data is difficult, determining the correct location of the spike is even more difficult.

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Spike Detection

- Spike Detection – Hunt method
  - Two step approach to enhance the spike and then detect the spike
    - Several enhancement approaches were considered:
      - Gradient based techniques
      - High pass linear filters
      - Auto Regressive models
      - Discrete Cosine Transform
      - Median Filters
      - Matched Filters
  - High Pass Filter enhancer with Crest Factor detector was ultimately selected
  - 20 million data samples were searched to identify approximately 240 spikes

Integrated analytical physical virtual Simulation

RPC Pro

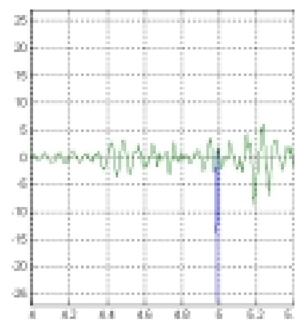
MTS Proprietary Information

Spike Detection

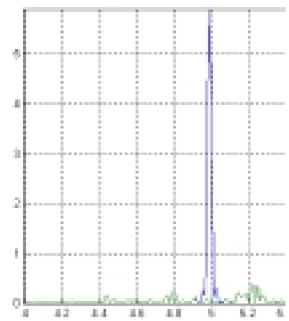
- Spike Detection – Hunt method
  - 1. High Pass Filter enhancer
    - Raw data is filtered to remove low frequencies and enhance spikes
      - A second order butterworth filter is used
      - Filter cut off is set at 80% of nyquist

Green = data without spike

Blue = data with spike



“Raw” data



Enhanced data

MTS Proprietary Information

Spike Detection

- Spike Detection – Hunt method
  - 2. Crest Factor detector
    - The detection algorithm uses a 1000 point window
      - A “local” RMS and crest factor is calculated for the window
        - » A recursive algorithm is used to calculate the RMS at any given sample, k

$$x_{rms}(k) = \sqrt{\frac{(k-1)x_{rms}^2(k-1) + x^2(k)}{k}}$$

      - Presence of a spike is determined by a crest factor threshold (nominally 15 to 40)
        - » The Crest Factor of a given set of points defined as:

$$\boldsymbol{y} = [y(1), y(2), ..., y(N)]^T$$

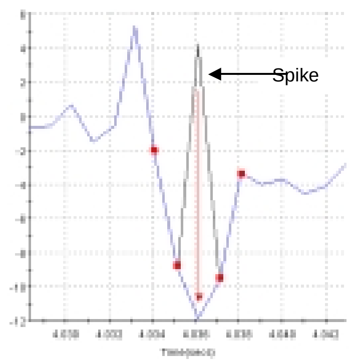
        - » is determined as:

$$CF = \frac{\max(|\boldsymbol{y}|)}{\mathrm{RMS}(|\boldsymbol{y}|)} = \frac{U}{V}$$

MTS Proprietary Information

Spike Correction

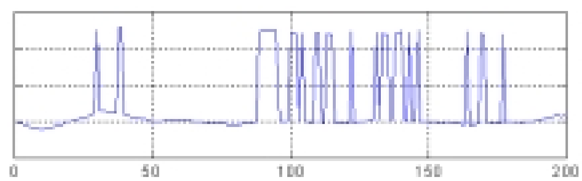
- Spike Correction
  - A spike can never be truly corrected since the real value of the data point is not known. The spike correction technique attempts to replace the spike data points with more appropriate values.
  - A 5 point cubic spline interpolation technique is used, adjusting the center "spike value"



MTS Proprietary Information

Dealing with Spikes

- Spike Detection/Correction Issues
  - If the correct location of a spike is not known, correction of that spike may introduce additional spikes
  - Groups of spikes are more difficult to correctly detect and locate



Extreme example of spike grouping

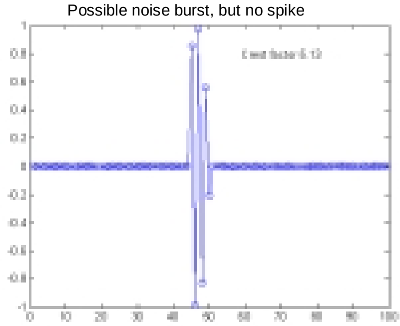
- 5 point spline correction no longer appropriate

MTS Proprietary Information

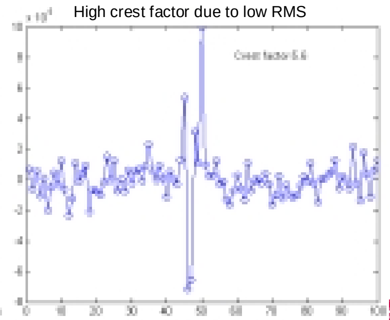
Spike Detection

- Spike Detection – MTS enhancements
  - Utilize basic spike detection algorithm with lower threshold
    - Minimize potential for missing spikes
  - New method for separating spikes form real data
    - Crest factor alone will tend to overestimate the number of spikes

Possible noise burst, but no spike



High crest factor due to low RMS



Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

MTS

Spike Detection

- Two topics were discussed – notes not available:
  - Spike Detection – MTS enhancements (patent pending)
    - Cluster analysis used to pin-point spike location
  - A new MTS spike removal algorithim was also presented

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

MTS

Presentation Outline

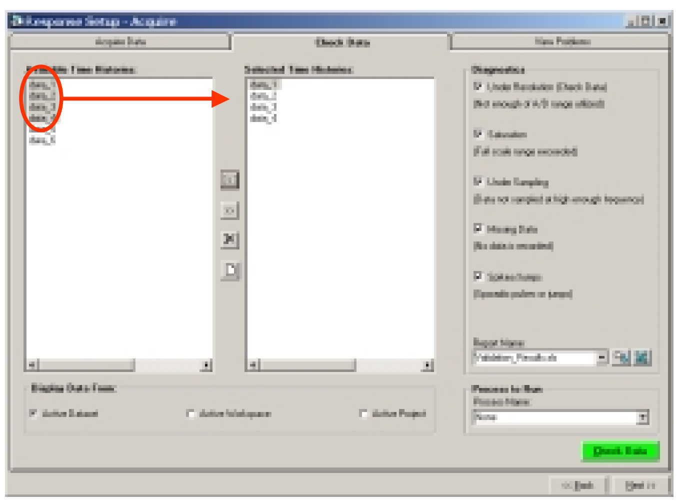
- Background
- “Bad” Data
- Anomaly Detection
- Spike Detection/Correction
- Automated Data Validation Process

Integrated analytical physical virtual Simulation
RPC Pro

MTS Proprietary Information

Automated Data Validation Process

- 1. Identify the “Raw” dataset

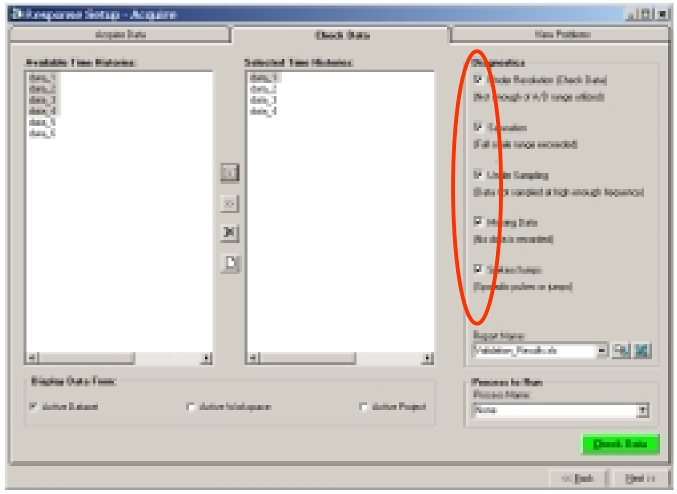


Integrated analytical physical virtual Simulation
RPC Pro

MTS Proprietary Information

Automated Data Validation Process

- 2. Select the appropriate anomaly detections



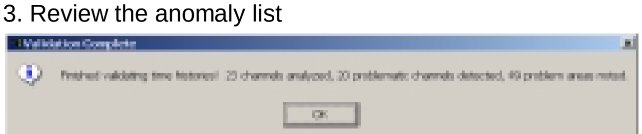
Integrated analytical physical virtual Simulation

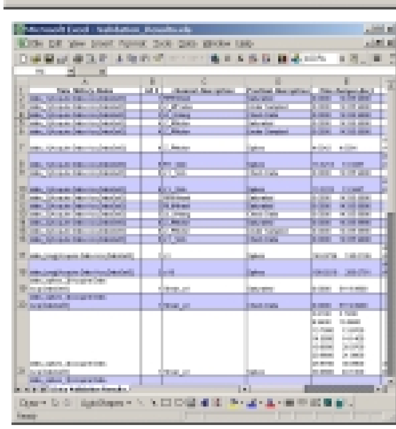
RPC Pro

MTS Proprietary Information

Automated Data Validation Process

- 3. Review the anomaly list





Anomaly list:

- **Automatically generated**
- **Entry tracked by defect, identifying problem and location of defect in the Time History**

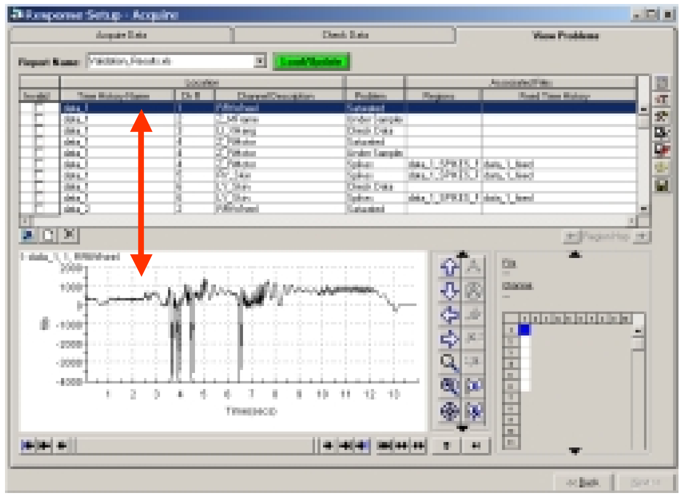
Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Automated Data Validation Process

- 4. Graphically inspect anomalies



– The graphical viewer shown the currently selected anomaly

Integrated analytical physical virtual Simulation

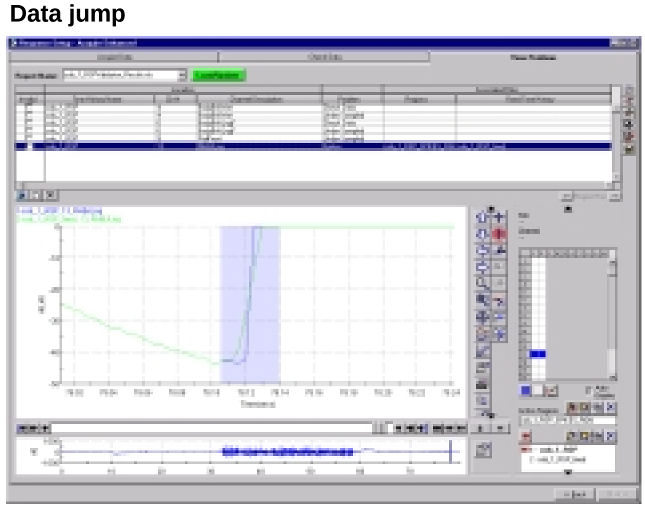
RPC Pro

MTS Proprietary Information

Anomaly Visualization

- Many data anomalies can be observed in the time domain:

Data jump



Integrated analytical physical virtual Simulation

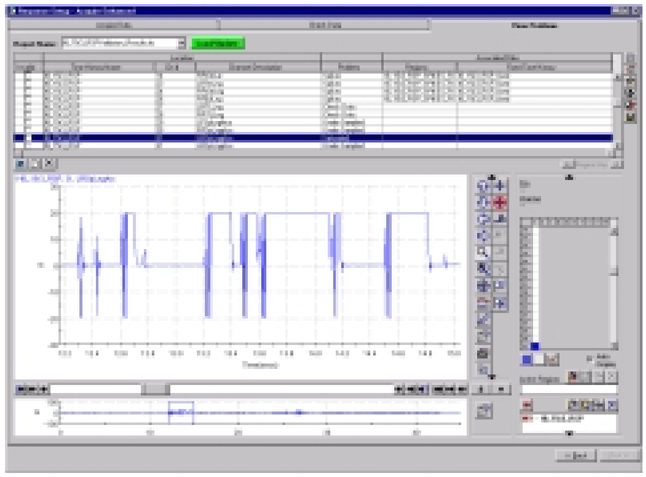
RPC Pro

MTS Proprietary Information

Anomaly Visualization

- Many data anomalies can be observed in the time domain:

Saturated data



Integrated analytical physical virtual Simulation

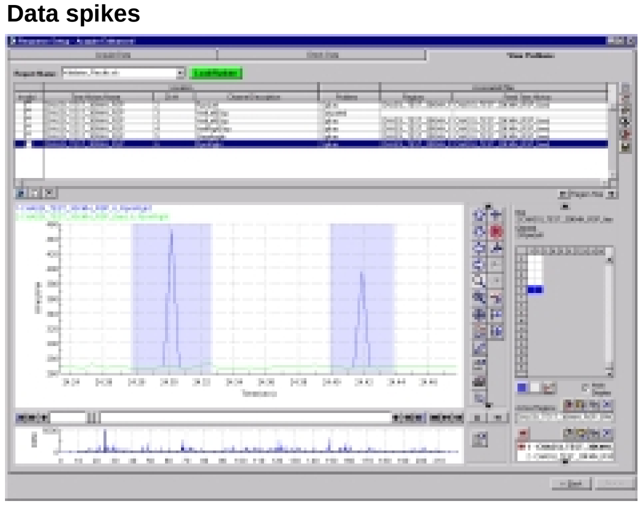
RPC Pro

MTS Proprietary Information

Anomaly Visualization

- Many data anomalies can be observed in the time domain:

Data spikes



Integrated analytical physical virtual Simulation

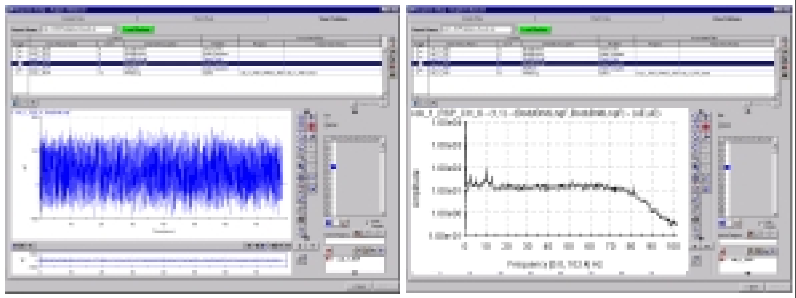
RPC Pro

MTS Proprietary Information

Anomaly Visualization

- However others are better observed in the frequency domain:

Under-sampled data

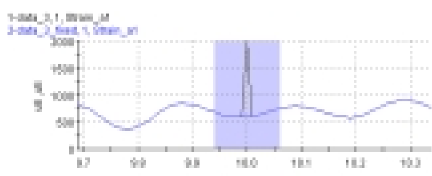


RPC Pro

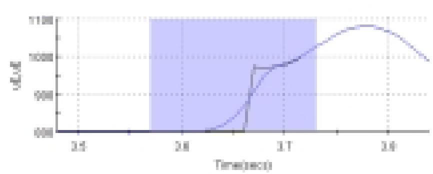
MTS Proprietary Information

Automated Data Validation Process

- 5. Verify corrected data

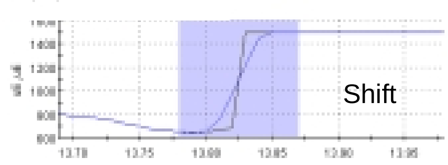


Spike



Jump

– Original Data is shown in Black, corrected in Blue



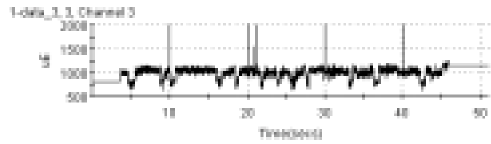
Shift

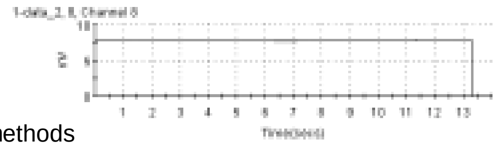
RPC Pro

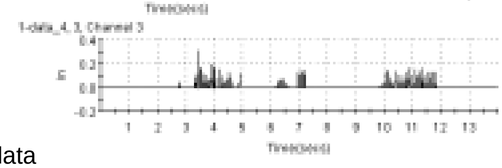
MTS Proprietary Information

Automated Data Validation Process

- 6. Evaluate uncorrected anomalies







Alternative methods must be used to deal with these type of anomalies, such as:

- removal / labeling of data
- manual data modification
- smoothing / tapering

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Summary

- Getting good data early is the key to the success of the development process
- Data validation must be automatic and present a minimal amount of information to the user
- Automated Data Validation allows engineers to:
  - Minimize loss of development time
  - Minimize cost due to errors found late in the cycle
  - Minimize warranty issues

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

Questions?

Automated Road Load Data Analysis for Verification of Data Integrity

Integrated analytical physical virtual Simulation

RPC Pro

MTS Proprietary Information

MTS