The Future of Materials Quality 'Big Data'

Dr. Andrea C. Oehler
Fujimi Corporation
October 21, 2021
What is Big Data?

In 2001, Doug Laney summarized the challenges of Big Data management as having three distinct dimensions, introducing the 3V’s of Big Data:

- **Big Data is high-Volume, high-Velocity and/or high-Variety** information assets that demand cost effective, innovative forms of information processing that enable enhanced insight, decision making and process automation.

- **Veracity**, integrity or “truth” of the data or analysis

- **Value**, big data or its analysis used as an asset
### 3V’s

#### Volume
- **Internal**
  - Manufacturing in process
- **External (private)**
  - Raw Material
  - Customer performance feedback
  - Local Warehouse
- **External (public)**
  - Weather
  - Water Quality

#### Variety
- **Structured**
  - Product Analysis
- **Semi-Structured**
  - Raw Material Lots
  - Customer Qualitative Feedback – Lists of Good/Bad
- **Unstructured**
  - Photos, pictures, sensory characterization
  - Operator observations, batch notes

#### Velocity
- **Very fast**
  - in process censors
- **Fast**
  - Wafer/customer functional performance
- **Slow**
  - Batch Card entries
- **Very slow**
  - Per manufacturing run

---

**CMP Manufacturing Big Data: High-Variety at a range of volume and velocity**
Qualitative Particle Characteristics are Unstructured Data
Innovative Forms of Information Processing

Cell Plot using OOC Characteristics
Innovative Form of Processing

Response screening is a way to look for relationships and avoid p-hacking – identifying false relationships by chance – through FDR p-value – False Discovery Rate p-value.

Response Screening to Avoid p-Hacking
PCS - Continuous Improvement through Variation Reduction
• DOE’s to identify critical to control parameters
• Metrology variation reduction and improved precision
• Trend Reviews and Investigation

Big Data – Continuous Improvement through Collaboration
• High Variety – Process Settings, Product Monitors, Product Performance
• High Volume – Internal and External sources, Supplier through Customer
• Innovative Forms of Information Processing

Share Big Data to Drive Continuous Improvement
Monitor Gamma

Levey Jennings chart of Monitor Gamma
Continuous Improvement Phase

<table>
<thead>
<tr>
<th>Start</th>
<th>PCS 1</th>
<th>PCS 2</th>
<th>PCS 3</th>
</tr>
</thead>
</table>

PCS Program Implementation and Continuous Improvement Successful
Levey Jennings chart of Monitor Alpha

Continuous Improvement Phase

Start | PCS | Big Data Share

Monitor Alpha

Continuous Improvement Not Successful
Focus on Veracity

Prepare data for Sharing
- Standardizing data \( (x - x_{\text{min}})/(x_{\text{max}} - x_{\text{min}}) \) or Normalize data \( (x - \mu)/\sigma \)
- Summarize Data by Lot
- Characterize unstructured data

Remove Noise
- Understand sources of variation such as metrology, applications, environmental conditions
- Identify gaps in data, missing data
- Identify covariates
- Identify inputs versus output

Statistical Techniques
- Scatter Plot Matrix
- Response Screening for False Discovery Rate

Validate Models
- Apply expertise and experience
- Validate causation versus correlation

“The signal is the truth. The noise is what distracts us from the truth” – Nate Silver
Scatter Plot Matrix

Customer Response
Including: \( C_\alpha, C_\beta \)

Product Monitors
Including: \( M_\alpha, M_\beta, M_\gamma, M_\delta \)

Supplier Settings and Monitors
Including: \( S_\alpha, S_\beta, S_\gamma, S_\delta \)
Veracity Check

Response Screening

\[C_{\alpha}, C_{\beta}, M_{\alpha}, M_{\beta}, M_{\gamma}, M_{\delta}\]
<table>
<thead>
<tr>
<th></th>
<th>East</th>
<th>North</th>
<th>South</th>
<th>West</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Customer Beta</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Customer Alpha</strong></td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Monitor Alpha</strong></td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Smooth(Customer Alpha)</strong></td>
<td><img src="image9" alt="Graph" /></td>
<td><img src="image10" alt="Graph" /></td>
<td><img src="image11" alt="Graph" /></td>
<td><img src="image12" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Smooth(Monitor Alpha)</strong></td>
<td><img src="image13" alt="Graph" /></td>
<td><img src="image14" alt="Graph" /></td>
<td><img src="image15" alt="Graph" /></td>
<td><img src="image16" alt="Graph" /></td>
</tr>
<tr>
<td><strong>Smooth(Setting Gamma)</strong></td>
<td><img src="image17" alt="Graph" /></td>
<td><img src="image18" alt="Graph" /></td>
<td><img src="image19" alt="Graph" /></td>
<td><img src="image20" alt="Graph" /></td>
</tr>
</tbody>
</table>

**Setting Gamma related to Customer Alpha**
Restricting range of Setting Gamma, reduces variation of Monitor Alpha
Is Big Data the Future of Quality

- High-Variety of data
- Innovative forms of information processing
- Expertise to prepare data and develop models for testing
- Collaboration with Supplier, Manufacturing and Customer