

Tackling Fleet Test Data with MATLAB

Will Wilson
Application Engineer

Agenda

- Motivation & Common Understanding
- Case Study: Automotive Fleet Data Analytics
- Key Takeaways

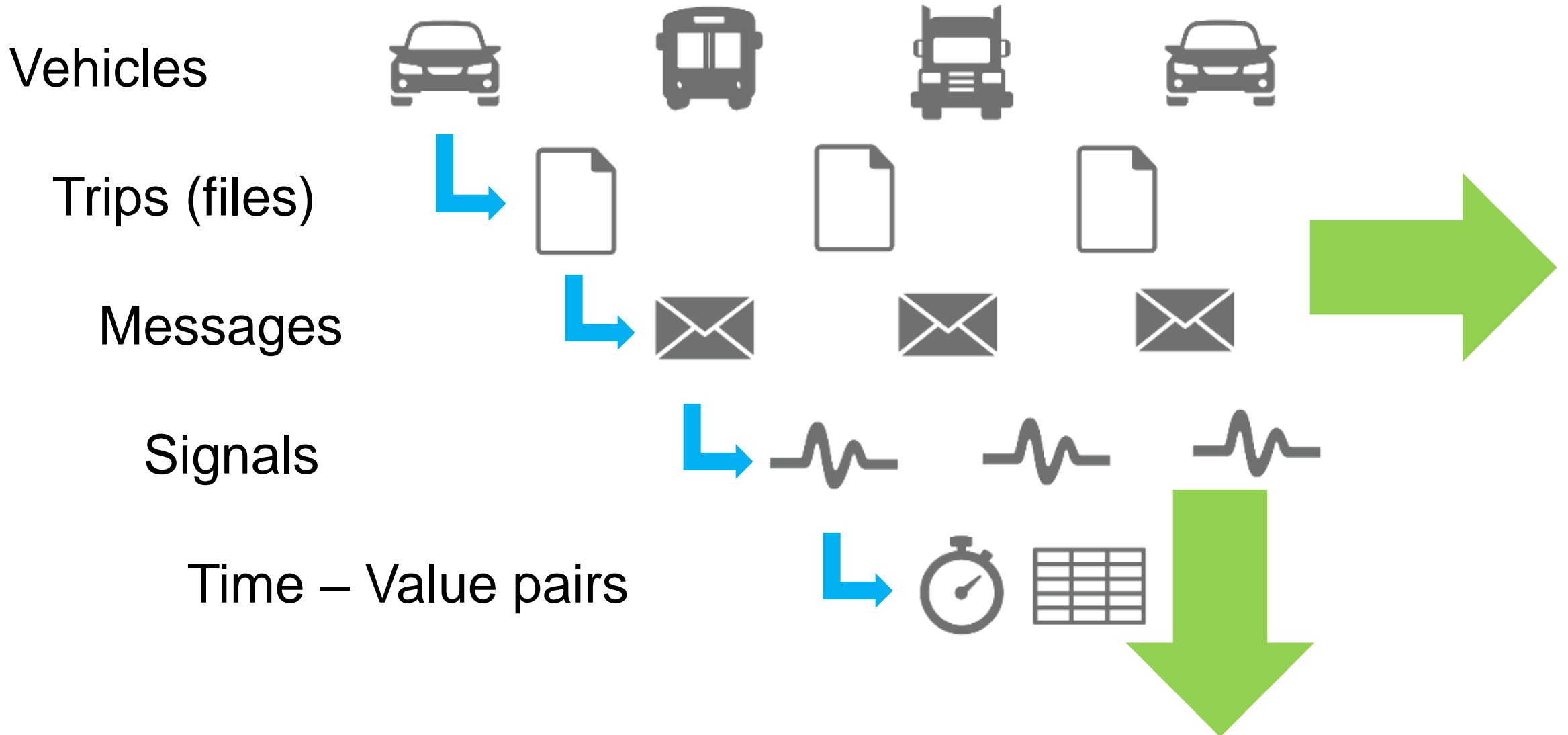
Why Analyze Vehicle Fleet Data?

- Understand real-world behavior (spec vs. actual usage)
- Make decisions with **all** of your data
- Enable Domain Experts to validate component or system behavior ***in situ***
- Identify and open up new market / product opportunities

Challenges with Analyzing Automotive Fleet Data

| Challenge | How does MATLAB Help? |
|--|--|
| Time required to gain insights from our data | Out of the box, MATLAB R2018a is 2x faster than MATLAB R2015a (or older) |
| Too much (Big) data | datastores, tall Arrays, MATLAB MapReduce, Parallel Computing |
| Time Series Data (<i>with different sample rates</i>) | timetable, retime, synchronize |
| Messy / Missing / Incomplete Data | Extensive pre-processing functionality |
| Very Large Individual Files | datastore |
| File Formats | Native MDF support, custom datastores |
| Integrate data-oriented algorithms with classical controls | MATLAB & Simulink in a single platform |

Lesson's Learned About Automotive Vehicle Test Fleets



How Do Customers Apply Analytics on Test Data?

Vehicle data, driver profiles



Cold Storage

Historic data:

- **Batch processing**
- Large data on cluster
- Explore long term trends
- Build models



Hot Storage

Streaming data:

- **Near real-time**
- Test and implement model for new data
- Stream processing



Thinking Back to MAC 2017

Big Data, Data Analytics, and Machine and Deep Learning Infrastructure at Caterpillar

Larry Mianzo, Caterpillar

Caterpillar, in collaboration with MathWorks, developed a big data infrastructure, with a web front end to leverage external labelers, a database for searching and retrieving labeled ground-truth, and an engineer interface. This interface for machine learning, visualization, and code generation enables function developers to use the labeled ground-truth for training, validating, and deploying classifiers.

By automating the task of labeling field data, the system reduces the need for human intervention. It directs engineers to focus their data collection effort on the conditions of critical needs. The infrastructure is also scalable in the number of users, the amount of data, and the amount of available compute power.

Advantages of using MATLAB:

- Speed up design iterations
- Automate labor-intensive tasks such as ground-truth labeling and comparison of classifier performance
- Quickly go from collected data to an improved classifier running on the machine using machine learning apps and code generation

“ We can access machine learning capabilities with **a few lines of MATLAB code**. Then, using code generation, engineers can **deploy their trained classifier** into the machine without manual intervention or delays in the process. ”

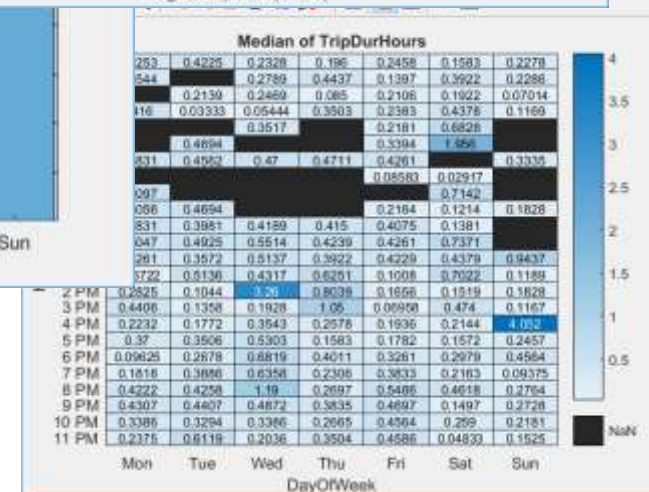
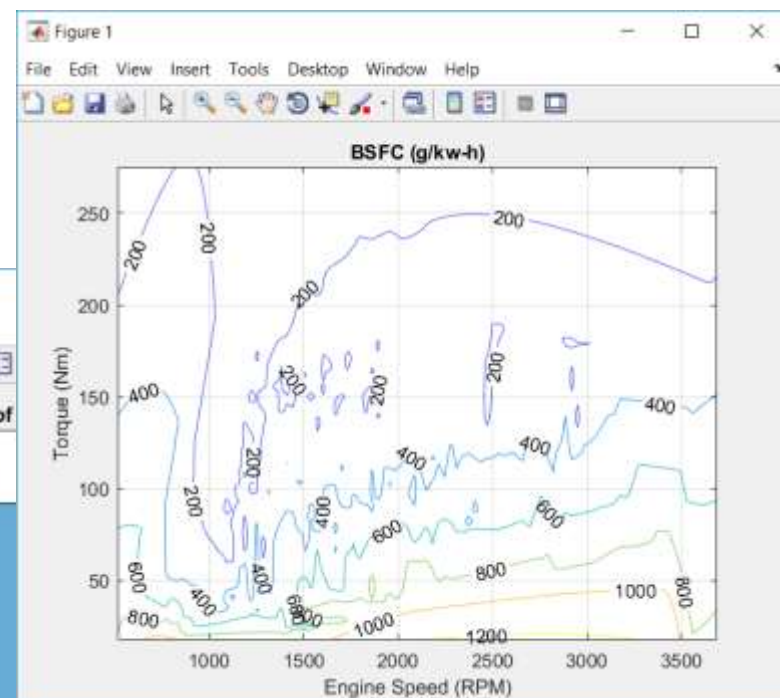


Agenda

- Motivation & Common Understanding
- **Case Study: Automotive Fleet Data Analytics**
- Key Takeaways

Case Study: MathWorks Automotive Fleet

- 1300 trip log files
- 21 unique vehicles
- Approx 40 unique signals
- Data collected over 1.5 years



Case Study: MathWorks Vehicle Fleet

Challenge

- Given a large set of vehicle fleet data:
 - Efficiently explore all of the available data to deepen our understanding
 - Develop and implement “event detection” functionality
 - Develop a Machine Learning model to classify driving behavior
 - Scale up to run on a Spark enabled Hadoop cluster

Solution

- Utilize the MATLAB platform and `tall` arrays to develop analytics on the desktop and then scale up to the Hadoop cluster

Vehicle Fleet Analytics – Live Script Example

AutomotiveFleetDataAnalyticsWithMATLAB.mlx

Automotive Fleet Data Analytics with MATLAB

Built using MATLAB R2018a
Copyright 2018, The MathWorks, Inc.

Data Analytics Workflow

1 Access and Explore Data

- Files
- Databases
- Sensors

2 Preprocess Data

- Working with Messy Data
- Data Reduction/Transformation
- Feature Extraction

3 Develop Predictive Model

- Model Creation e.g. Machine Learning
- Parameter Optimization
- Model Validation

4 Scale Up

- Big Data Models and Architectures (Spark, Hadoop)
- Clusters and Hardware (EMR, AWS)
- Cloud (Microsoft Azure, Amazon)

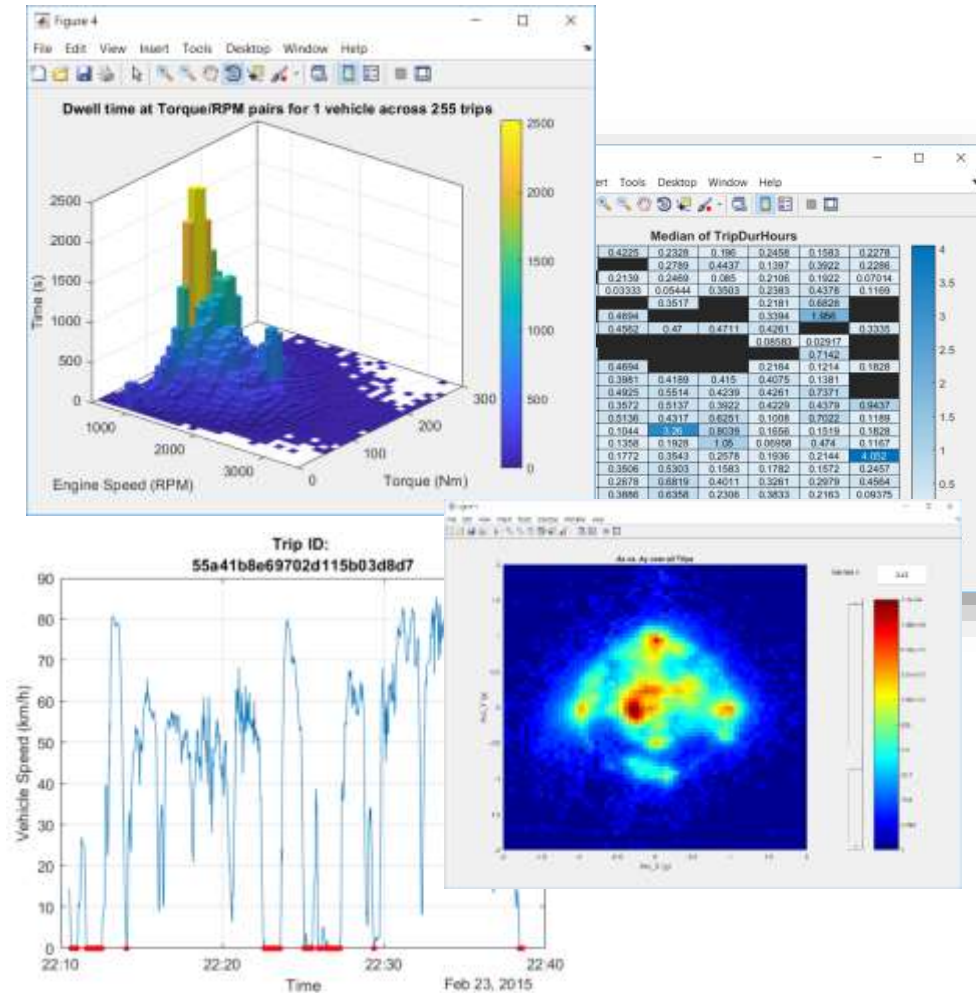
5 Integrate with Production Systems

- Desktop Apps
- Enterprise Scale Systems (MATLAB C/C++ Runtime)
- Embedded Devices and Hardware

10

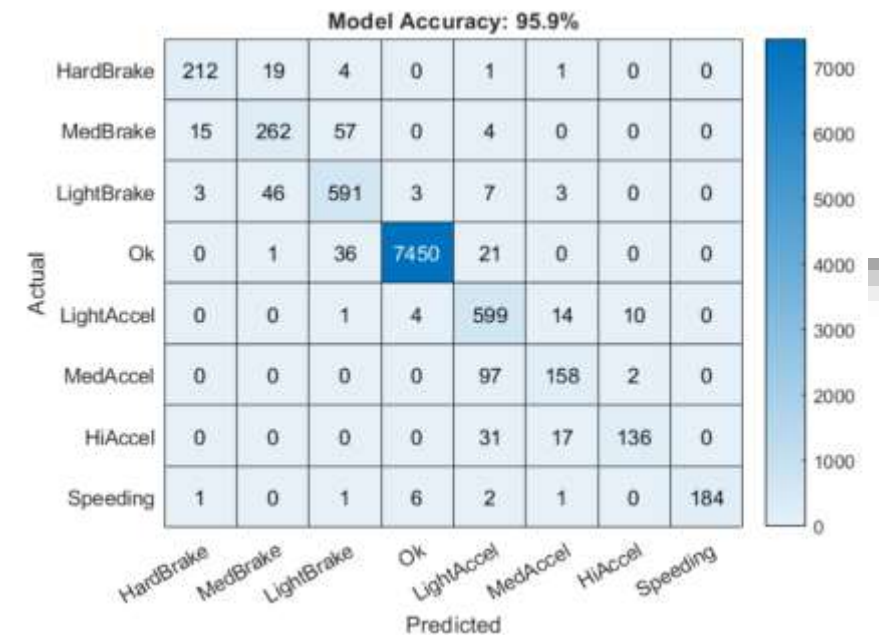
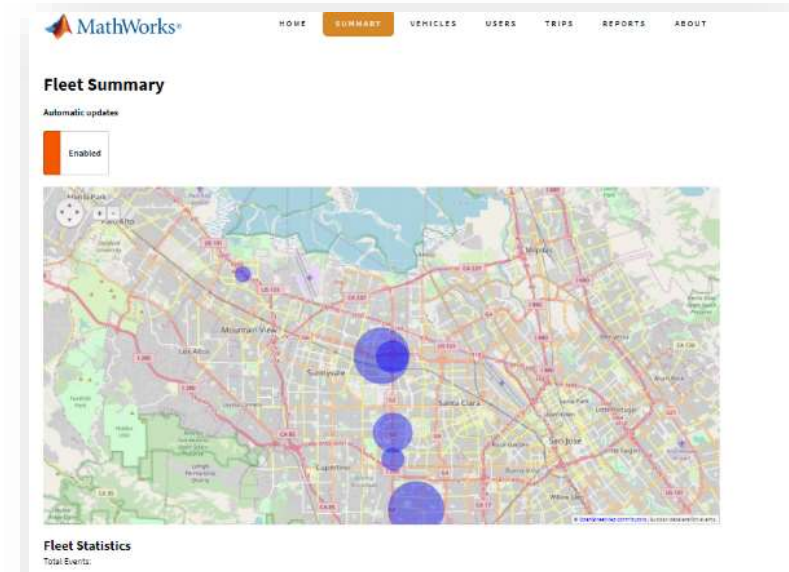
Case Study Results – Exploration & Discovery

- Worked with all of our data with `cell` arrays
- Extracted performance metrics & created summary visualizations
- Applied an "Event Detection" function to all of our data to find specific information



Case Study Results – Model & Predict

- Developed a Machine Learning Model to predict driving behavior
- Interactively explored classification algorithms with the Classification Learner App
- Used the Machine Learning Model to test it against new data



Case Study Recap – What We Did Today

- Access out of memory data
- Work with subsets of your data
- Develop functions for event detection and calculation
- Apply functions to all of your data
- Aggregate, summarize, & visualize
- `datastore & tall`
- `findgroups, splitapply, cellfun`
- Normal MATLAB code
- `cellfun`
- `table, histogram, heatmap, boxplot, binScatterPlot`

Agenda

- Motivation & Common Understanding
- Case Study: Automotive Fleet Data Analytics
- **Key Takeaways**

Key Takeaways / Call To Action

- Working with fleet test data is hard. MATLAB enables **Domain Experts** to implement efficient **Data Analytics** ranging from simple exploratory work to full featured **Machine Learning** Models.
- Learn something new (free online training)
 - www.mathworks.com
 - MATLAB Onramp (65k people in the last year)
 - MATLAB for Deep Learning Onramp (4.5k people in the last year)
- Ask me about Data Analytics: will.wilson@mathworks.com



© 2018 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.