Beyond Excel: Enhancing Your Data Analysis with MATLAB

David Willingham
Senior Application Engineer – Data Analytics
“Data is the sword of the 21st century, those who wield it the samurai.”

- Big data — how to create it, manipulate it, and put it to good use.
- “If you want to work at Google, make sure you can use MATLAB.”

Google’s Former SVP - Jonathan Rosenberg
What Analytics Challenges are you facing?

1. Analytics that increasingly require both business and engineering data
   - Engineering, Scientific, and Field
   - Business and Transactional

2. Developing embedded systems which have increasing analytic content
   - Smart Connected Systems

3. Deploying applications that run on both traditional IT and embedded platforms

4. Enable Domain Experts to be Data Scientists
Data Analytics Workflow

Access
- Files
- Software
- Hardware
- Code & Applications

Explore & Discover
- Data Analysis & Modeling

Share
- Reporting and Documentation
- Outputs for Design
- Deployment

Automate
Frontier Advisors Develops Web-Based Platform for Portfolio Analytics

Challenge
Provide clients with an industry-first web platform for portfolio modeling and analytics

Solution
Use MATLAB to develop and test analytics modules, and use MATLAB Compiler SDK to deploy them into a production .NET environment

Results
- Quantitative development decoupled from interface development
- Stable, responsive system deployed
- Rapid delivery of new features enabled

“MATLAB and MATLAB Compiler SDK enabled us to rapidly deliver a sophisticated portfolio analytics web application with confidence that it will return accurate results extremely quickly, ensuring a highly usable and stable platform for our clients.”

Lee Eriera
Frontier Advisors

Link to user story
Today’s Objectives

- Introduce you to data analysis with MATLAB
- Show how you can overcome common data analysis challenges with MATLAB
- Demonstrate multiple ways of sharing your analysis and results with others
Common Data Analysis Challenges using Excel

- Complex calculations
- Messy Data
- Speed of Execution
- Automation
- Batch Processing
- Report Generation
- Deployment
Demo: Solar Radiation Estimation
Introduction to Data Analysis with MATLAB

- **Goal:**
  - Estimate daily mean global solar radiation given low cost and easily obtained measurements

- **Approach:**
  - Process historical measurements
  - Develop predictive model
  - Document analysis in a report
  - Apply analysis on multiple files
Modeling Global Solar Radiation

\[ R_s = a (1 + bH)(1 - e^{-c \Delta T^n}) \]

Solar Ratio \((R_s) = \frac{\text{Global solar radiation}}{\text{Extraterrestrial solar radiation}}\)

Daily Temperature Difference \((\Delta T) = T_{\text{DailyMax}} - T_{\text{DailyMin}}\)

\(H\) is Relative Humidity

\(a, b, c, n\) are the model coefficients
Demo Summary
Solar Radiation Estimation

Access
- Files
- Software
  - Code & Applications
- Hardware

Explore & Discover
- Data Analysis & Modeling
- Algorithm Development
  - For k=1:max
  - x = fft(dat)
  - y = 20*log1

Application Development

Share
- Reporting and Documentation
- Outputs for Design
  - MATLAB
  - Excel
  - .NET
  - C/C++
  - Java
  - .dll

Deployment

Automate

Products Used
- MATLAB
- Curve Fitting Toolbox
Sharing Results from MATLAB

- Automatically generate reports
- Create and package applications
- Deploy to other environments
Using MATLAB with Excel

- Passing data between MATLAB and Excel
  - MATLAB

- Accessing MATLAB from an Excel spreadsheet
  - MATLAB
  - Spreadsheet Link EX

- Deploying MATLAB as an Excel add-in
  - MATLAB
  - MATLAB Compiler
MATLAB Application Deployment

- Share MATLAB programs with people who do not have MATLAB
  - Royalty-free distribution
  - Encryption to protect your intellectual property

- Create both standalone applications and components for integration

- Deploy to desktop, web, and enterprise applications
Sharing Standalone Applications

1. **Application Author**
   - **MATLAB**
     - **Toolboxes**

2. **MATLAB Compiler**
   - Standalone Application
   - Excel Add-in
   - Hadoop

3. **End User**
   - **MATLAB Runtime**
Common Data Analysis Challenges

- Handling complex and messy data
- Modeling with many predictors
Demo: Preparing Late Plane Data
Handling Complex and Messy Data

- **Goal:**
  - Prepare late plane data for further analysis

- **Approach:**
  - Load mixed data from files
  - Filter data and replace missing data
  - Merge observations from different time intervals into a single data set
Accessing Data from MATLAB

### Access

- **Files**
  - Excel, text, or binary
  - Audio and video, image
  - Scientific formats and XML

- **Web Services**
  - JSON, CSV, and image data

- **Applications and languages**
  - C/C++, Java, FORTRAN
  - COM, .NET, shared libraries
  - Databases *(Database Toolbox)*

- **Measurement hardware**
  - Data acquisition hardware *(Data Acquisition Toolbox)*
  - Stand-alone instruments and devices *(Instrument Control Toolbox)*
Common Data Analysis Challenges

- Handling complex and messy data
- Modeling with many predictors
Machine Learning

Machine learning uses **data** and produces a **program** to perform a **task**

**Standard Approach**

Hand Written Program

- If \( X_{\text{acc}} > 0.5 \) then “SITTING”
- If \( Y_{\text{acc}} < 4 \) and \( Z_{\text{acc}} > 5 \) then “STANDING”

<table>
<thead>
<tr>
<th>Formula or Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_{\text{activity}} = \beta_1 X_{\text{acc}} + \beta_2 Y_{\text{acc}} + \beta_3 Z_{\text{acc}} + \ldots )</td>
</tr>
</tbody>
</table>

**Machine Learning Approach**

**model**: Inputs \( \rightarrow \) Outputs

<table>
<thead>
<tr>
<th>Formula or Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{model} = \langle \text{Machine Learning Algorithm} \rangle (\text{sensor_data, activity}) )</td>
</tr>
</tbody>
</table>
Demo: Machine Learning Using Mobile Phone Data

Data:
- 3-axial Accelerometer data
- 3-axial Gyroscope data
Demo: Portfolio Optimisation
Computing the Efficient Frontier

- Goal: Compute an Efficient Frontier in:
  - Excel Solver
  - MATLAB

- Compare the 2 approaches in:
  - Performance
  - Automation of Workflow
Workflow Portfolio Optimization

- Convert prices to returns.
- Expected Returns.
- Covariance matrices
- Calculate Efficient Frontier
  - Optimize to Maximise the return
  - Optimize to Minimise the risk
  - Optimize multiple times between Min Risk & Max Return
MATLAB Central

- Community for MATLAB and Simulink users
  - Over 70k daily visits

- File Exchange
  - Access more than 10k free files including functions, apps, examples, and models

- MATLAB Answers
  - Ask programming questions or search
  - 18k+ community-answered Questions

- Blogs
  - Read commentary from engineers who design, build, and support MATLAB and Simulink
Expand Your Analysis Capabilities

- **Machine learning**
  *(Statistics and Machine Learning Toolbox, Neural Networks Toolbox)*
  - “Learn” from your data without assuming an equation as a model
  - www.mathworks.com/machine-learning

- **Parallel programming**
  *(Parallel Computing Toolbox)*
  - Speed up your analysis using multicore computers, GPUs, and computer clusters
Today’s Objectives

- Introduce you to data analysis with MATLAB
- Show how you can overcome common data analysis challenges with MATLAB
- Demonstrate multiple ways of sharing your analysis and results with others