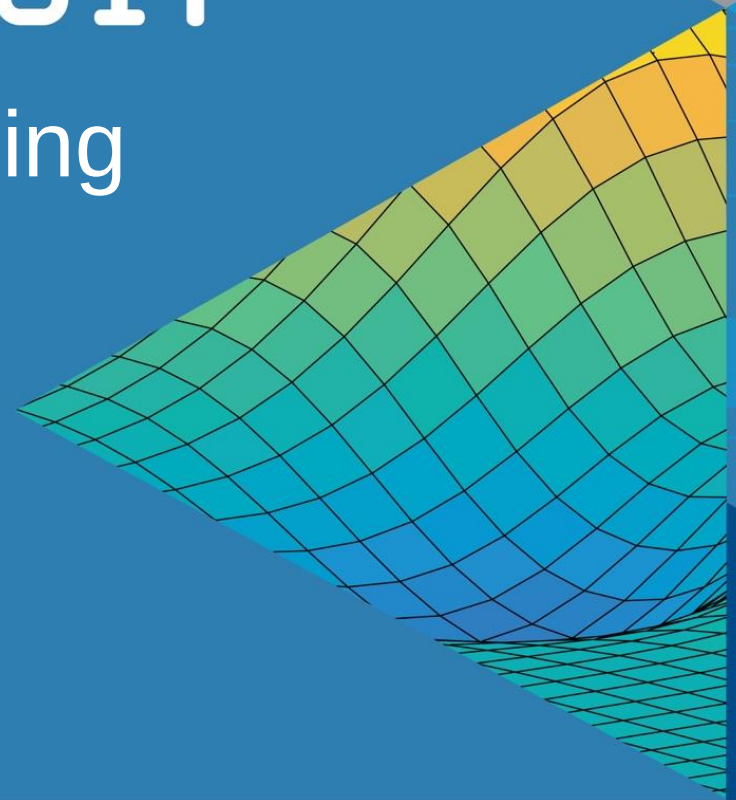
MATLAB EXPO 2017

Big Data and Machine Learning for Predictive Maintenance

Paul Peeling



Agenda

- The Predictive Maintenance Opportunity
- Exploring Big Data
- Machine Learning Approaches
- Deep Learning
- Fault Modelling
- Deploying to the Edge and the Cloud

React or Prevent?

Aaron “tango” Tang on Flicker

Predictive Maintenance software

Sense

Perceive

Decide & Plan

Act

Temperature sensors

Pressure sensors

Vibration sensors

Total of 25 sensors - but which ones were the best predictors?

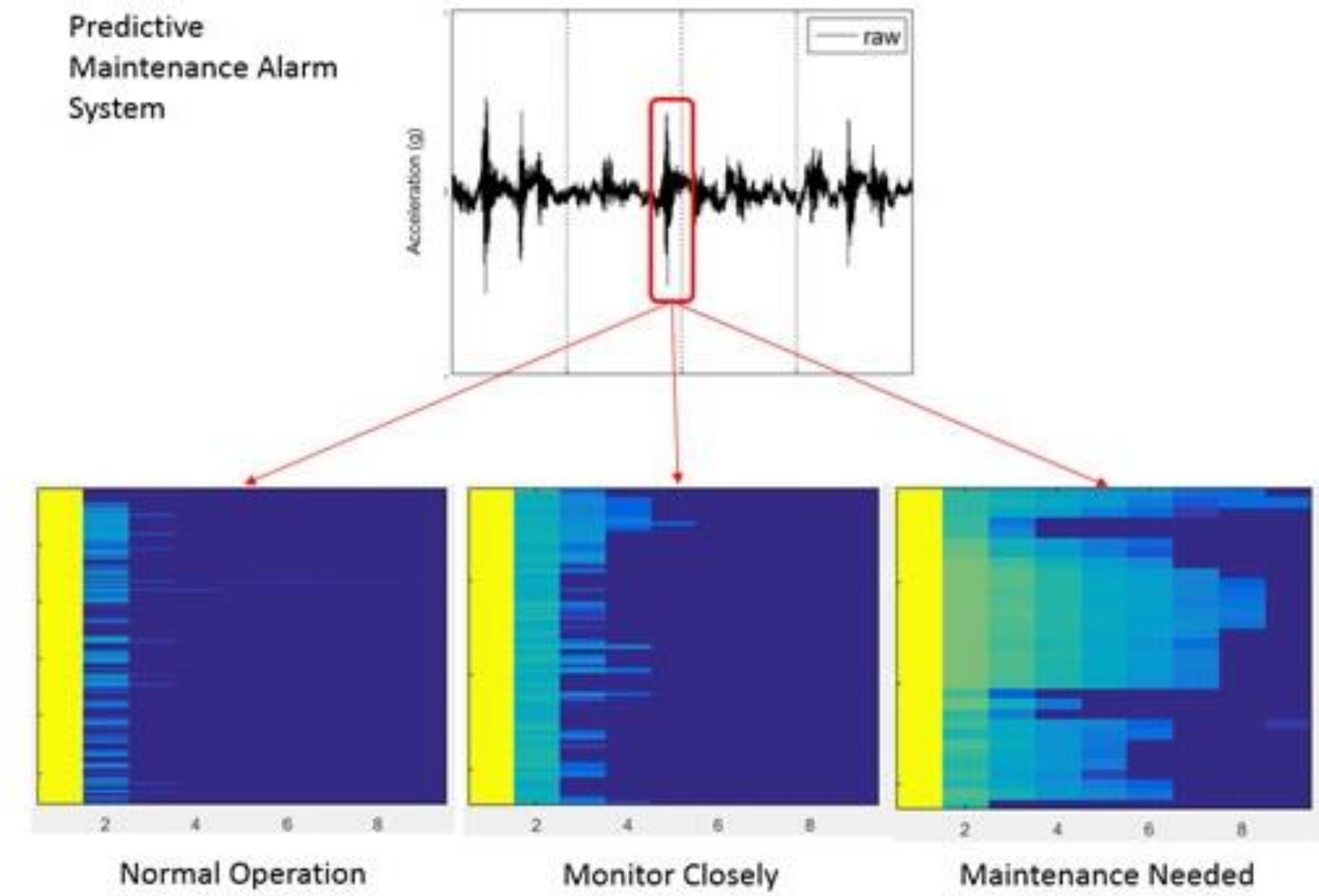
Predictive Maintenance software

Sense

Perceive

Decide & Plan

Act

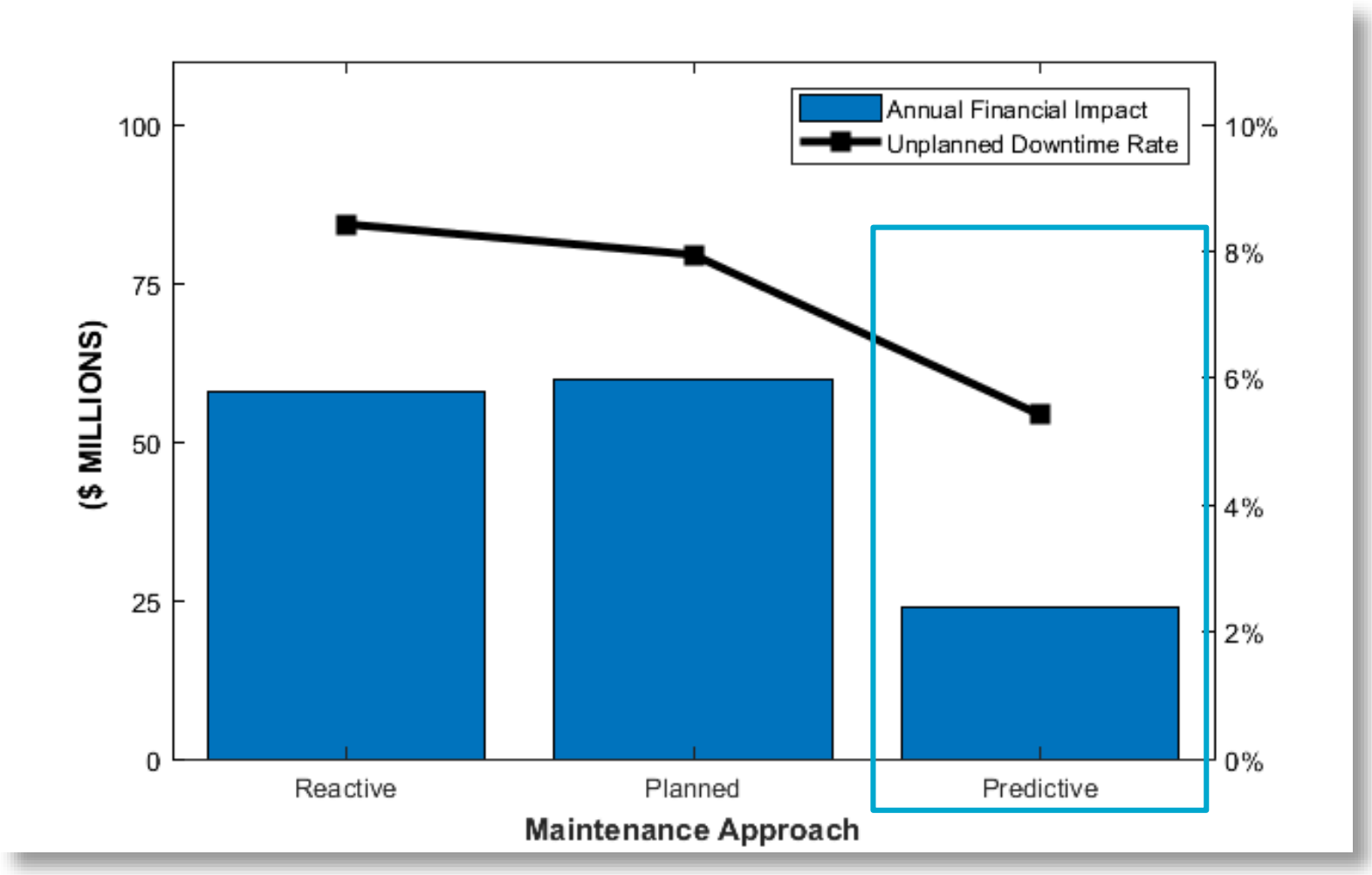


What do we mean by Predictive Maintenance?

- Monitor equipment to avoid future failure.
- Schedule maintenance when it's needed.
- Identify the root cause of issues.
- How?
  - Predictive models and sensor data.
  - Deploying to the equipment and cloud.

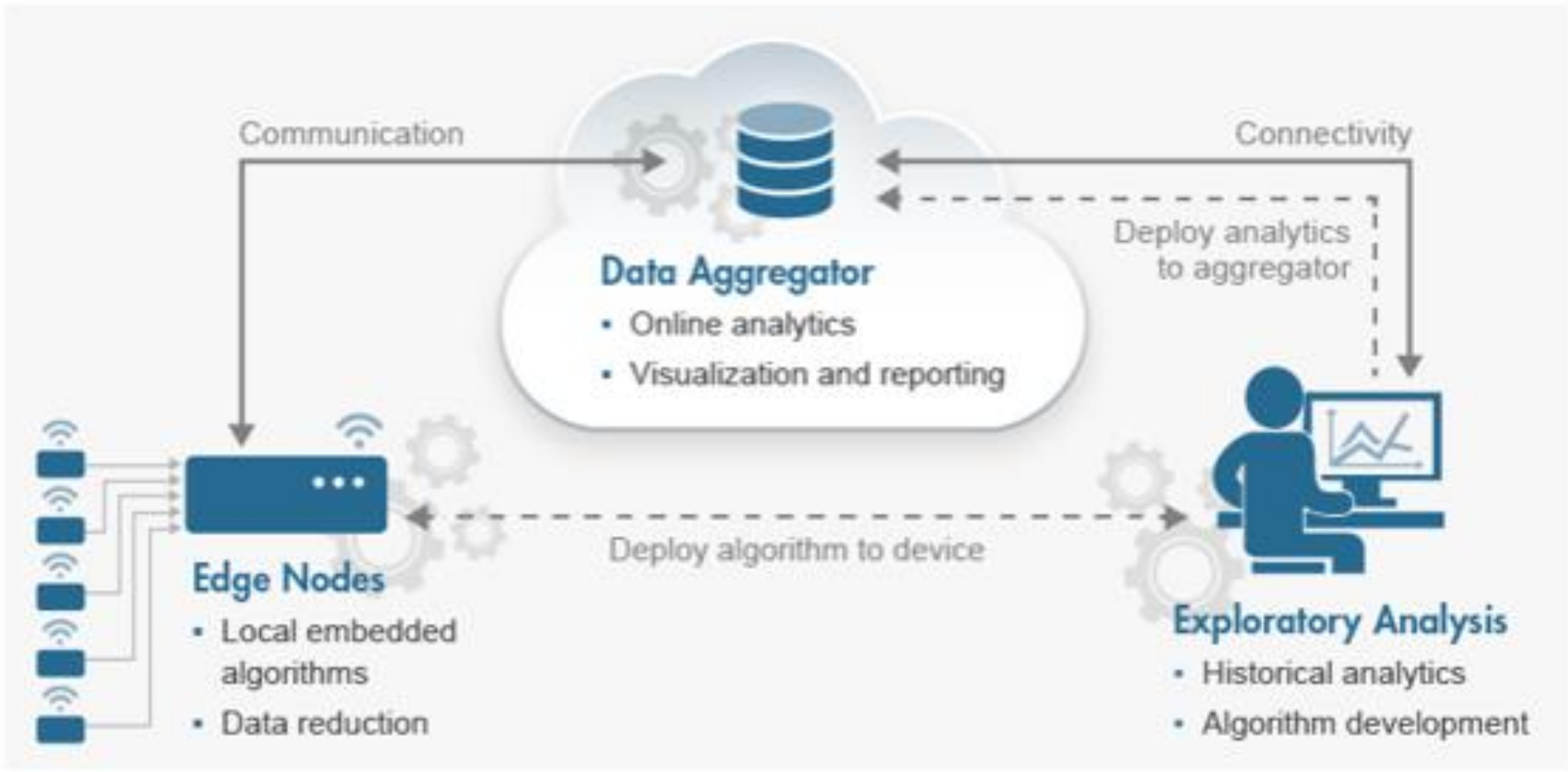
Why is Predictive Maintenance Important?

Why is Predictive Maintenance Important?



Source: GE Oil & Gas

Deploying Predictive Maintenance Algorithms



Aside: What if … ?

- I’m not in the business of Predictive Maintenance
- I don’t have big data
- I don’t have any data
- I don’t have a computing cluster
- I need a simpler solution

Workflow

Access and Explore Data

- Files
- Sensors

Preprocess Data

- Visualizing Data
- Data Reduction/ Transformation

Develop Predictive Models

- Machine learning
- Model Validation

Integrate Analytics with Systems

- Embedded Devices/Hardware
- Enterprise Scale Systems

MATLAB Excel .NET C/C++ .exe Java .dll

Workflow

Access and Explore Data → **Preprocess Data** → **Develop Predictive Models** → **Integrate Analytics with Systems**

Files

Sensors

Visualizing Data

Data Reduction/ Transformation

Machine learning

Model Validation

Embedded Devices/Hardware

Enterprise Scale Systems

MATLAB Excel .NET C/C++ .exe Java .dll

Predictive Maintenance of Turbofan Engine

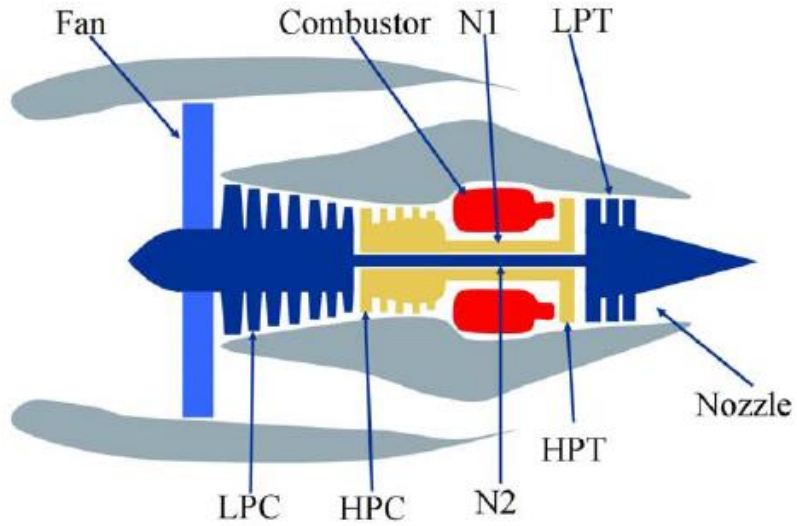
Sensor data from 100 engines of the same model

- Maintenance scheduled every 125 cycles
- Only 4 engines needed maintenance after 1st round

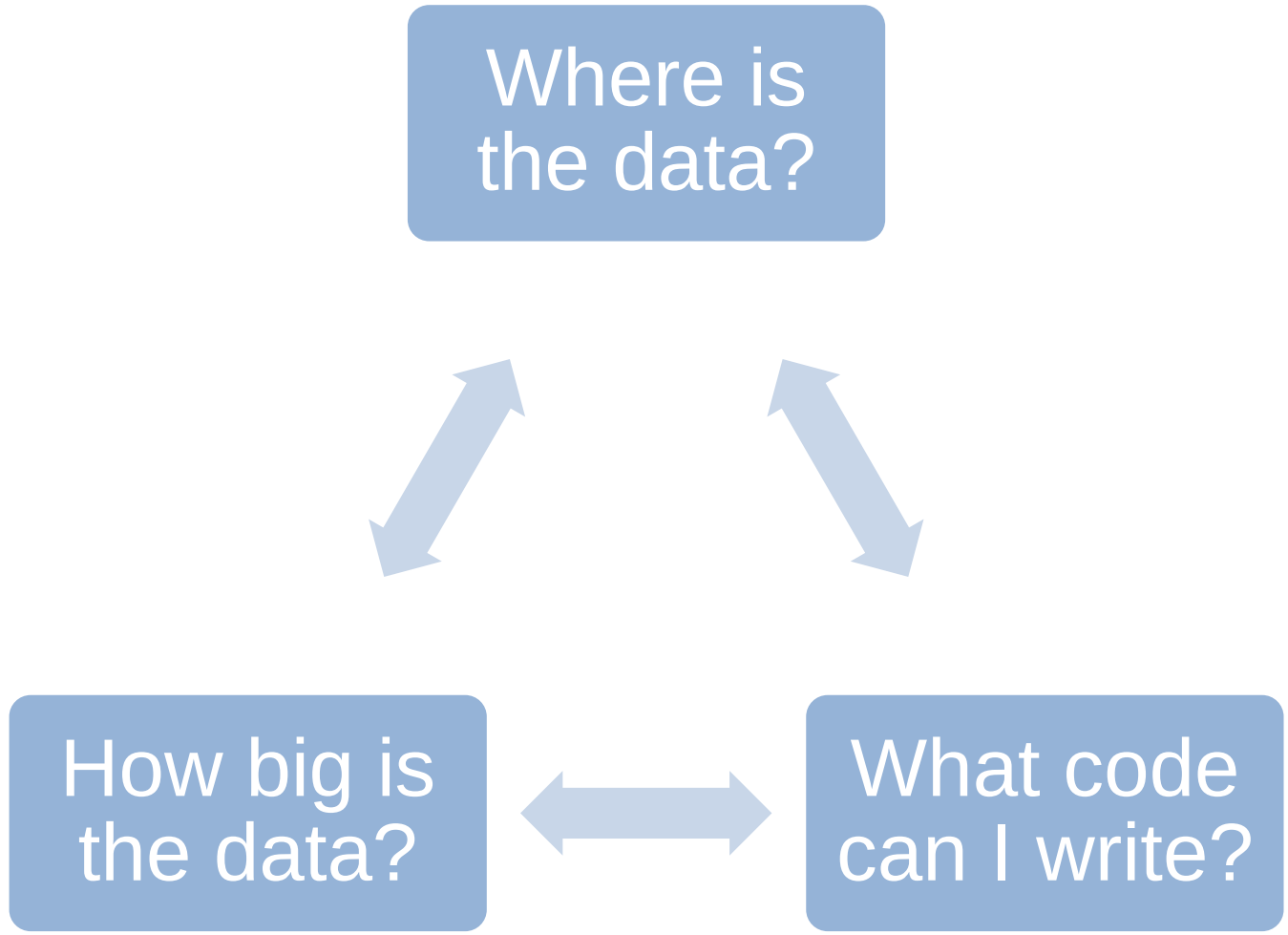
Predict and fix failures before they arise

- Import and analyze historical sensor data
- Train model to predict when failures will occur
- Deploy model to run on live sensor data
- Predict failures in real time

Data provided by NASA PCoE
http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/



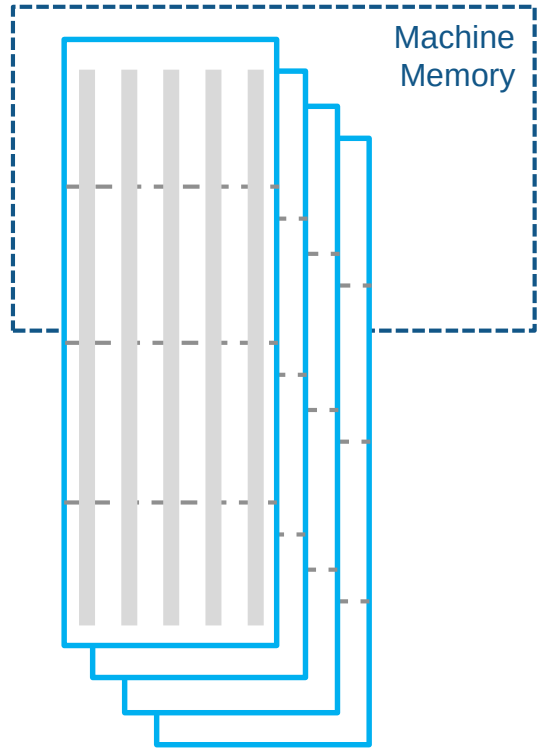
Working with Big Data



Tall Arrays

Scaling your code to big data

- Automatically optimize data access bottlenecks
  - Write code the same way you've always written it
  - MATLAB automatically reorders operations to minimize disk access
- Applicable when:
  - Data is **columnar** – with **many** rows
  - Overall data size is **too big to fit into memory**
  - Operations are mathematical/statistical in nature
- Statistical and machine learning applications
  - Hundreds of functions supported in MATLAB and Statistics and Machine Learning Toolbox



Tall Data

Filtering Data

```
% Point to where the data lives. Could be large text files, large collections
% of small files, or pageable databases.
ds = datastore('.\Data\*.csv');
% Inform MATLAB that we will treat this data as a tall array.
% We could a tall array from a local variable for prototyping.
engineData = tall(ds);
% Assume maintenance is being done regardless of condition after 125 cycles
engineData = engineData(engineData.Time <= 125,:)
```

```
engineData =

  M×16 tall table
```

Unit	Time	LPCOutletTemp	HPCOutletTemp	LPTOutletTemp	TotalHPCOutletPres
1.00	5.00	642.21	1587.03	1403.21	554.16
1.00	6.00	642.26	1585.98	1402.76	554.23
1.00	7.00	642.33	1586.08	1401.69	554.34
1.00	8.00	642.37	1585.08	1401.04	554.26
1.00	9.00	642.33	1586.72	1399.63	554.11
1.00	10.00	642.19	1588.39	1398.47	554.03
1.00	11.00	642.23	1587.85	1398.93	554.00
1.00	12.00	642.15	1586.07	1399.40	554.04
:	:	:	:	:	:
:	:	:	:	:	:

Workflow

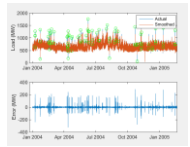
Access and Explore Data

Files

Sensors

Preprocess Data

Visualizing Data



Data Reduction/ Transformation

Develop Predictive Models

Machine learning

Model Validation

Integrate Analytics with Systems

Embedded Devices/Hardware

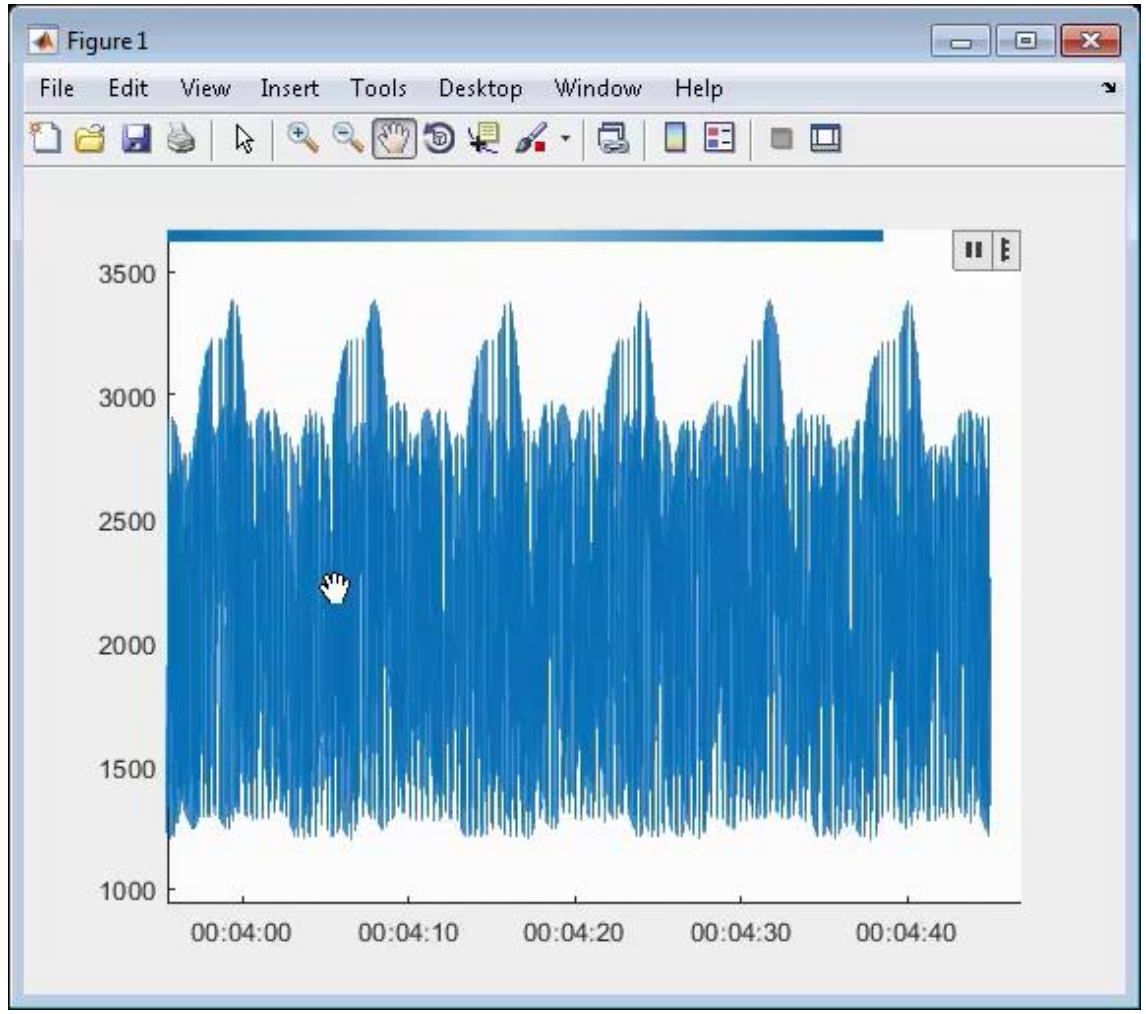
Enterprise Scale Systems

MATLAB Excel .NET C/C++ .exe Java .dll

Visualizing Big Data Using `tall`

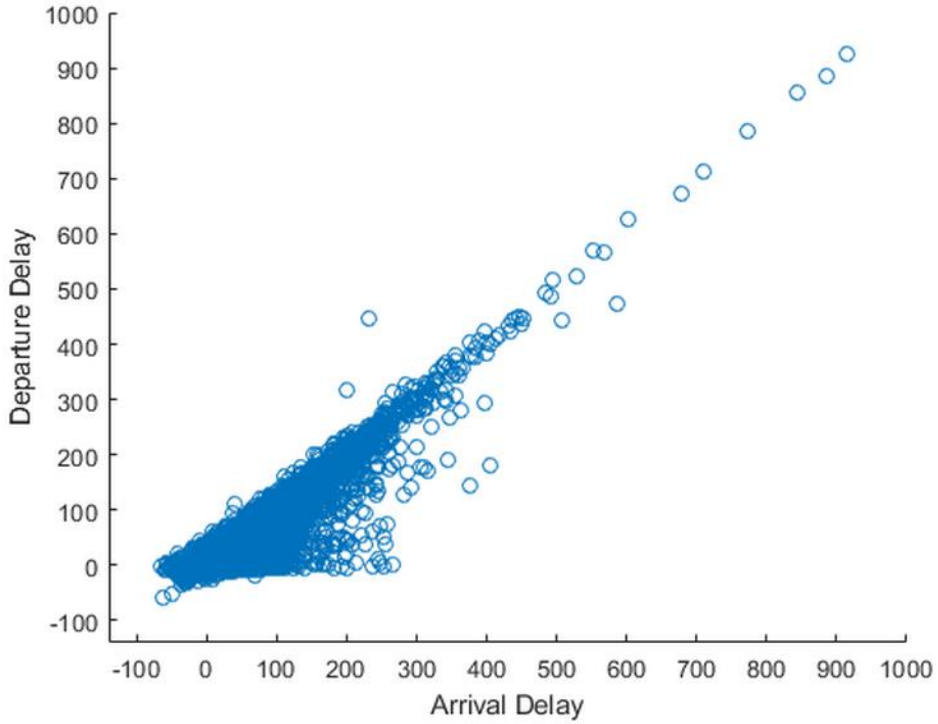
- Support for:
  - `histogram`
  - `histogram2`
  - `ksdensity`
  - `plot`
  - `scatter`
  - `binscatter`

R2017b

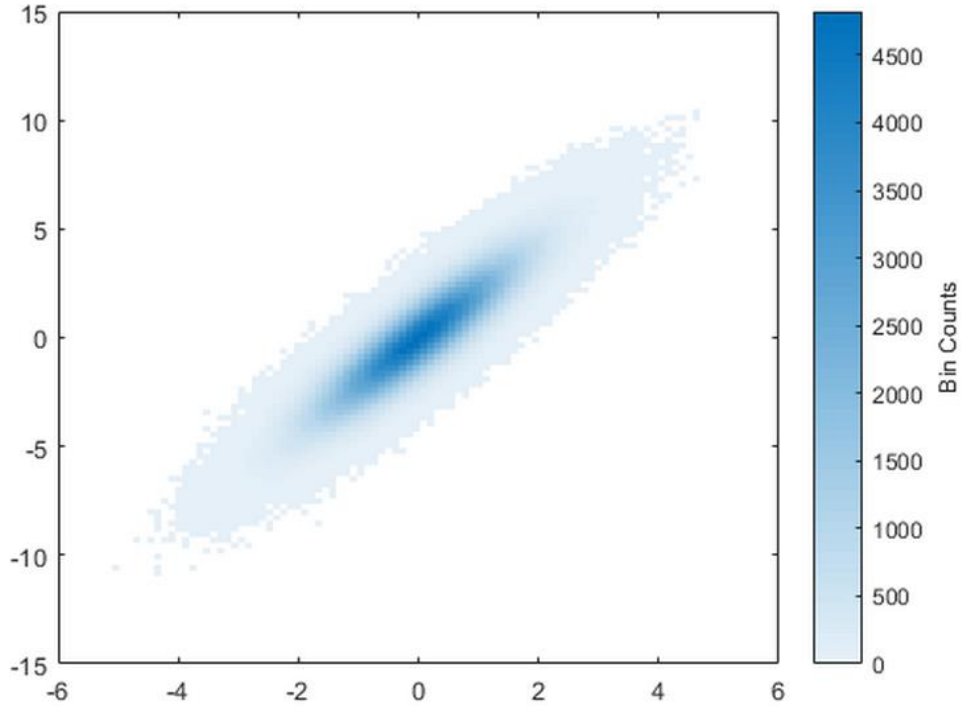


Visualizing Big Data Using `tall`

`scatter`



`binscatter`



Standardizing Data

```
% Pull out the sensor data, ignoring the unit and timestamp, and
% format it as an array
Xtrain = table2array(engineData(:,3:end));
% Give all sensors mean of zero and standard deviation of one
XtrainMeanTall = mean(Xtrain); % mean of each signal
XtrainStdTall = std(Xtrain); % standard deviation of each signal
% Uses implicit expansion
XtrainStandard = (Xtrain - XtrainMeanTall)./XtrainStdTall
```

```
XtrainStandard =

  M×14 tall array

    ?    ?    ?   ...
    ?    ?    ?   ...
    ?    ?    ?   ...
    :    :    :
    :    :    :

Preview deferred. Learn more.
```

Deferred evaluation and gathering

```
% read in data, assuming here that one data file can fit into memory
sensorData = gather( engineData( engineData.Unit == 1, : ) )
```

```
Evaluating tall expression using the Local MATLAB Session:
- Pass 1 of 1: Completed in 7 sec
Evaluation completed in 7 sec
  sensorData =  121×16 table
```

	Unit	Time	LPCOutletTemp	HPCOutletTemp	LPTOutletTemp	TotalHPCOutle...
1	1.00	5.00	642.21	1587.03	1403.21	554.16
2	1.00	6.00	642.26	1585.98	1402.76	554.23
3	1.00	7.00	642.33	1586.08	1401.69	554.34
4	1.00	8.00	642.37	1585.08	1401.04	554.26
5	1.00	9.00	642.33	1586.72	1399.63	554.11
6	1.00	10.00	642.19	1588.39	1398.47	554.03
7	1.00	11.00	642.23	1587.85	1398.93	554.00
8	1.00	12.00	642.15	1586.07	1399.40	554.04
9	1.00	13.00	642.25	1585.91	1399.38	553.96

What does “gather” do?

1. **Evaluate** any pending operations
2. **Collect** the partitioned data into MATLAB main memory
3. **Unwrap** the data into an array or table

Workflow

Access and Explore Data

Files

Sensors

Preprocess Data

Visualizing Data

Data Reduction/ Transformation

Develop Predictive Models

Machine learning

Model Validation

Integrate Analytics with Systems

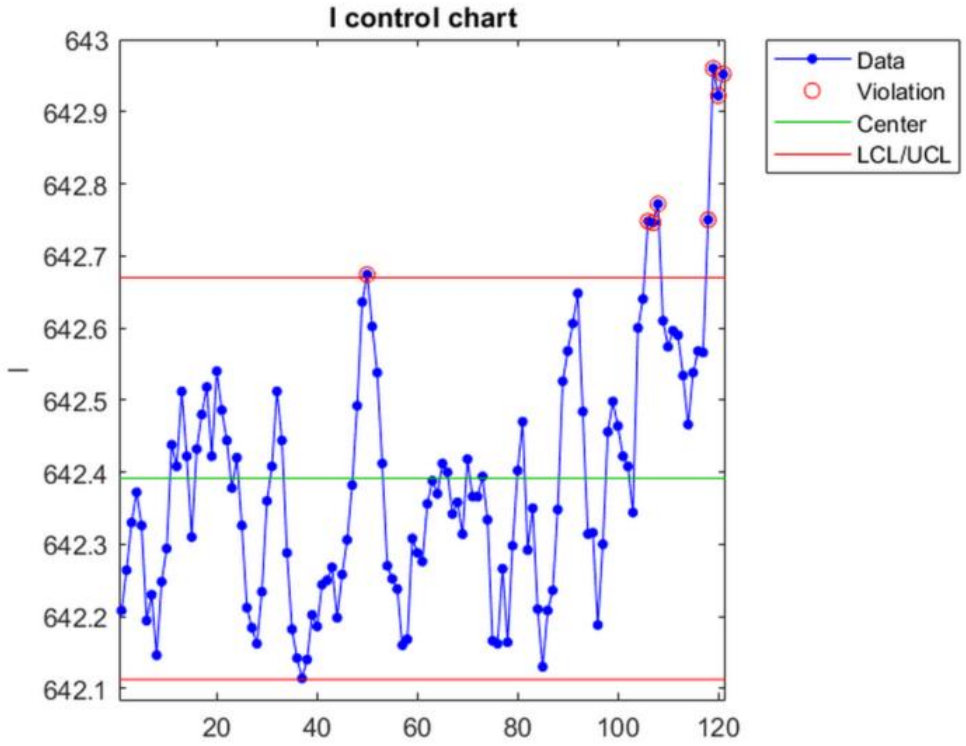
Embedded Devices/Hardware

Enterprise Scale Systems

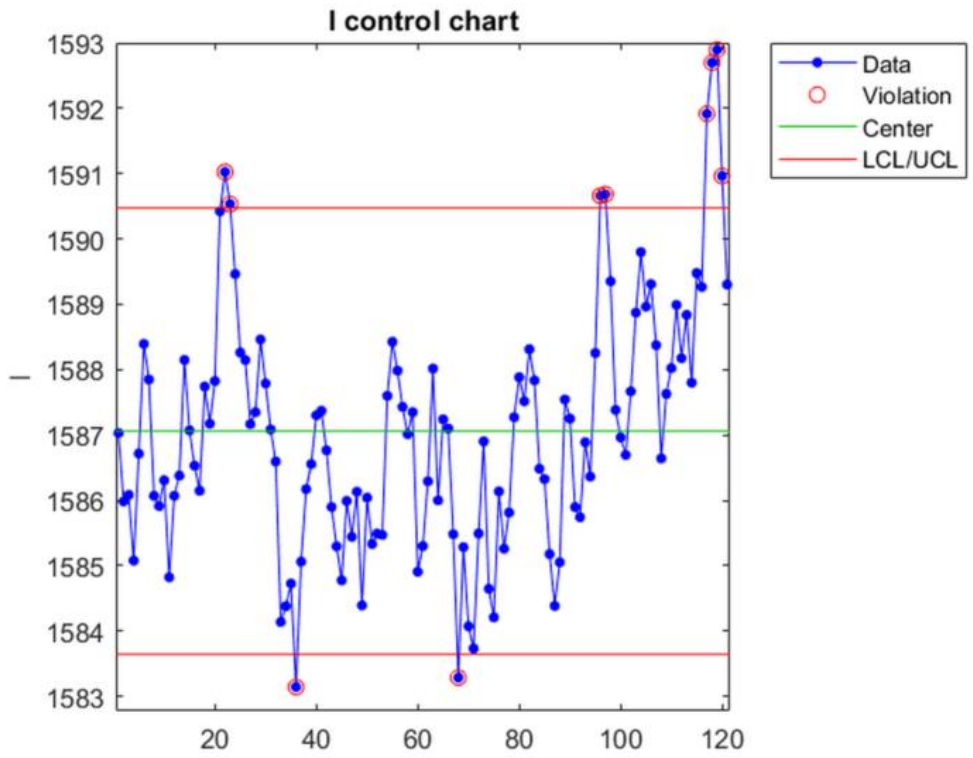
MATLAB EXPO 2017

Traditional Approaches

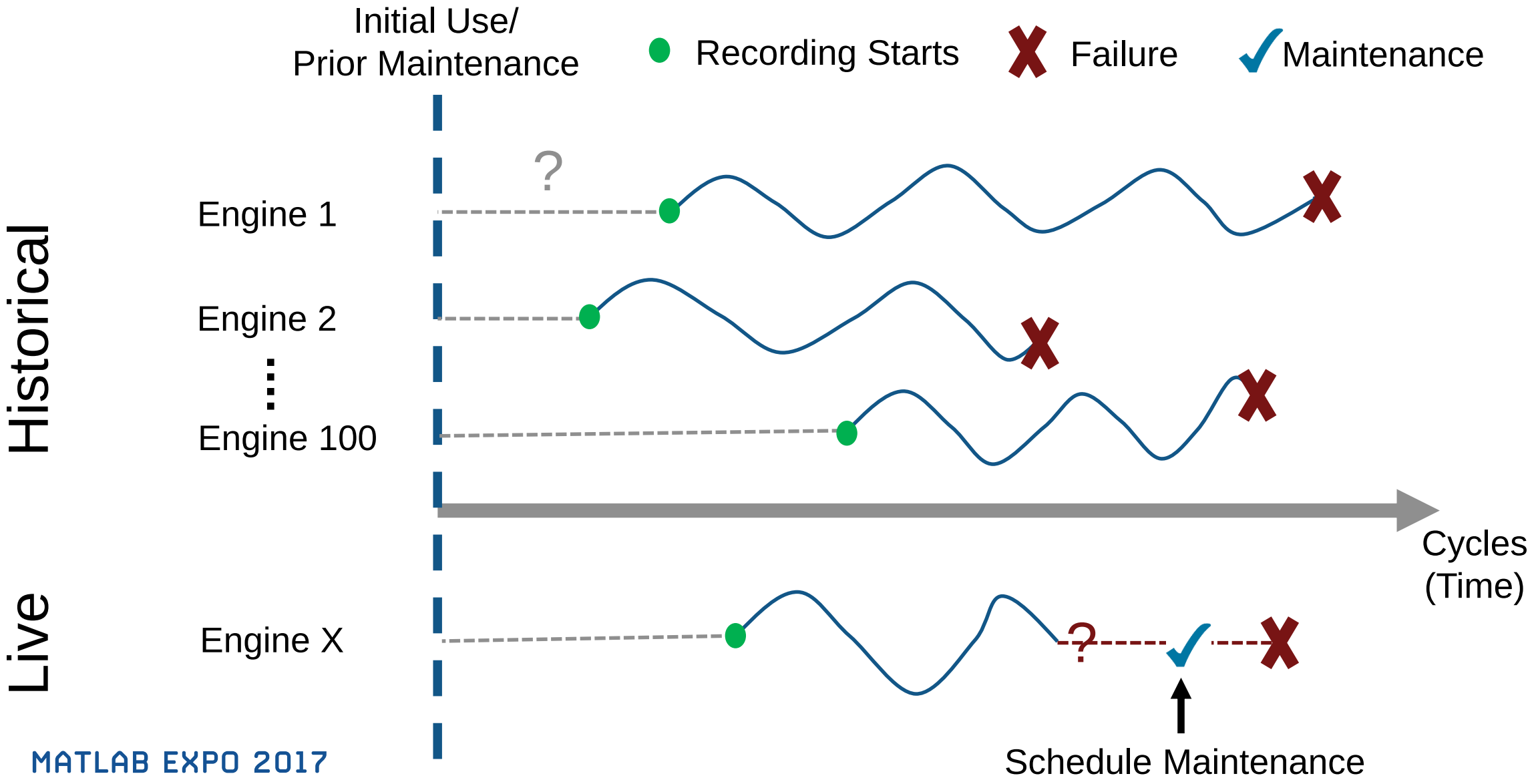
```
controlchart(sensorData.LPCOutletTemp(sensorData.Time<=125),'chart','i')
```



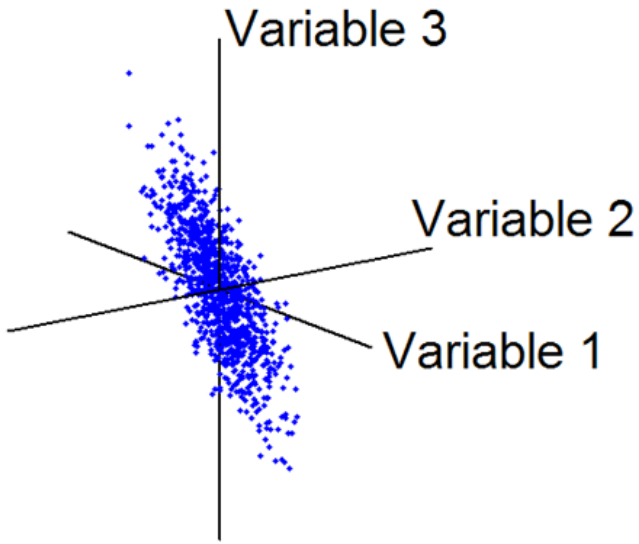

```
controlchart(sensorData.HPCOutletTemp(sensorData.Time<=125),'chart','i')
```

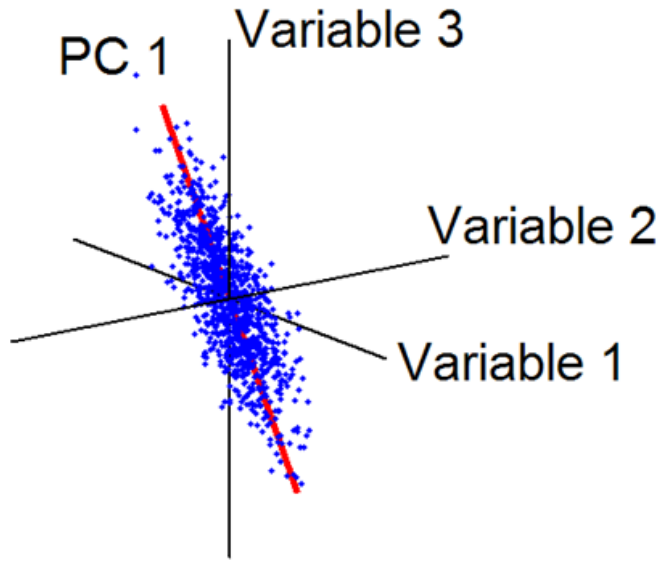


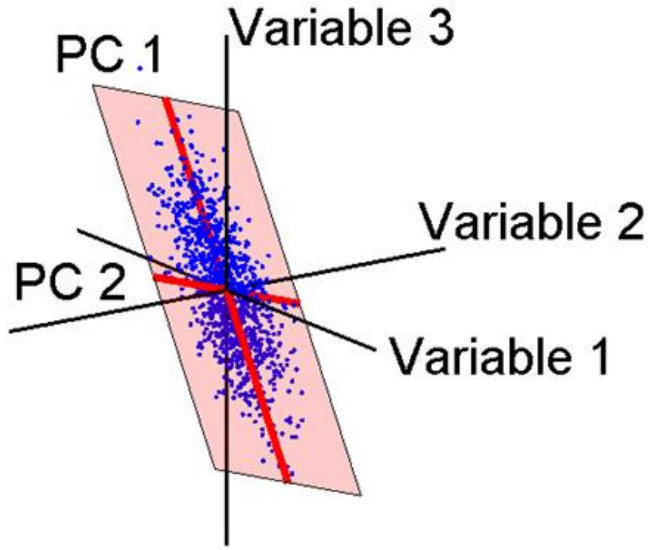
Use historical data to predict when failures will occur

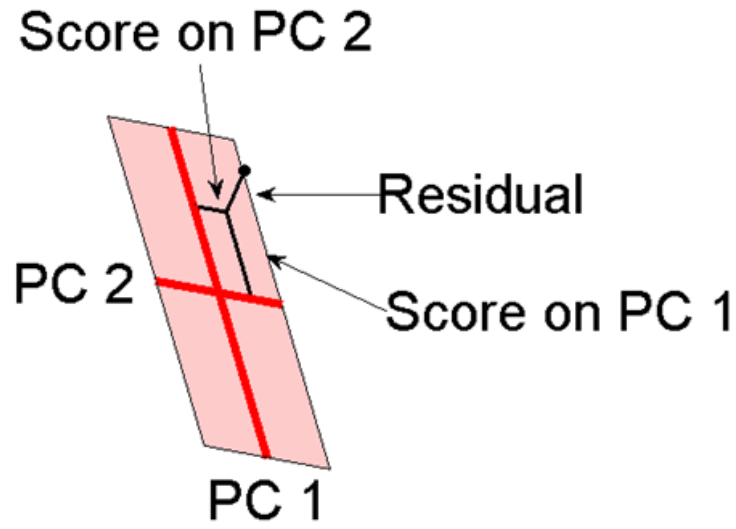


Principal Components Analysis



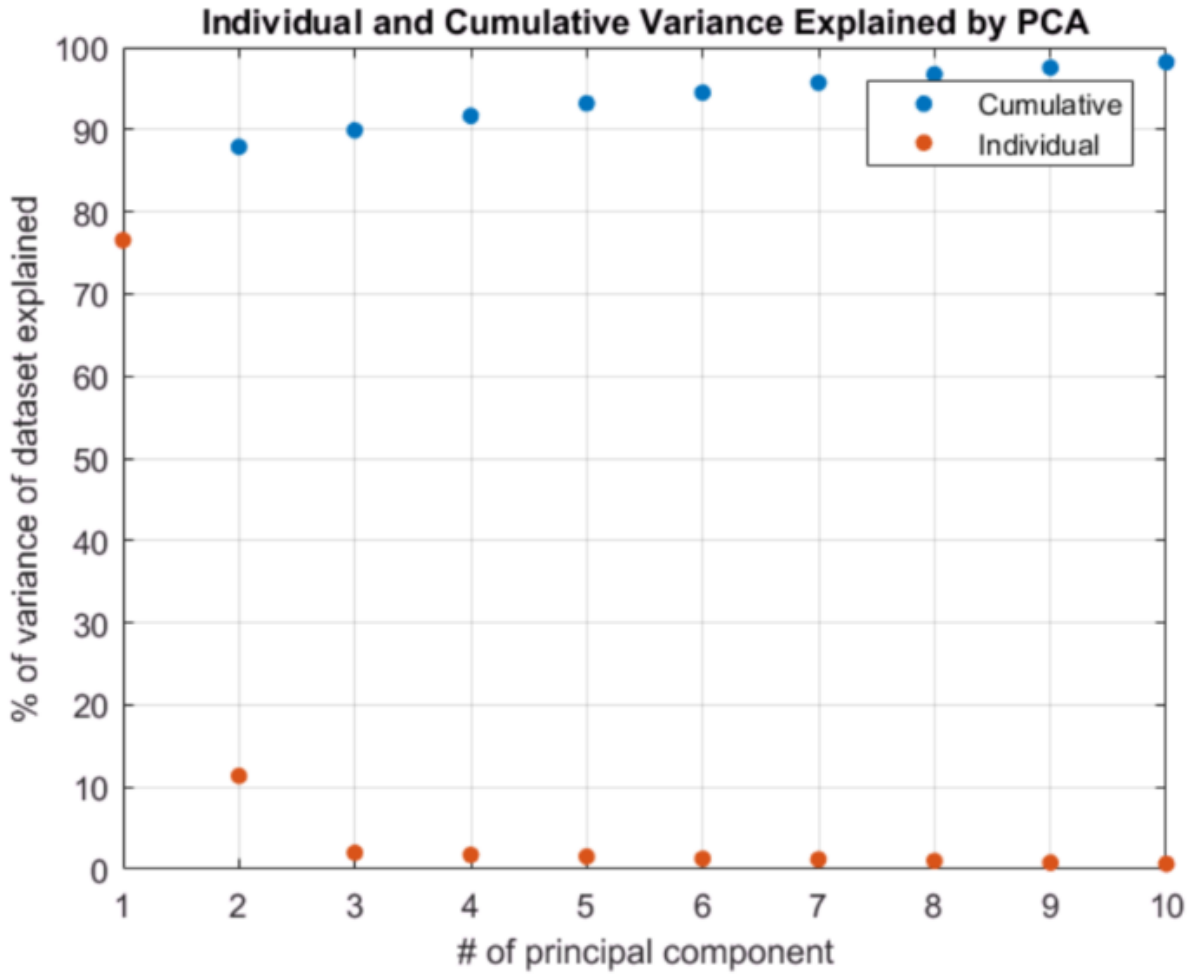






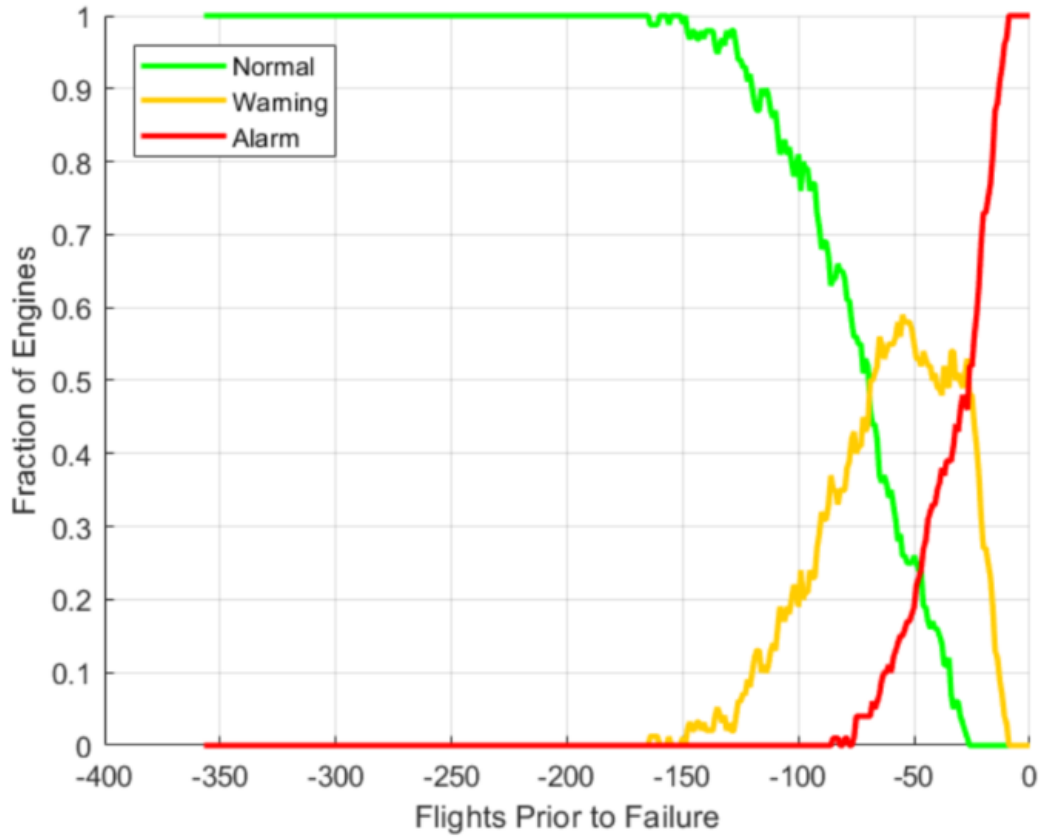
Dimensionality Reduction with PCA

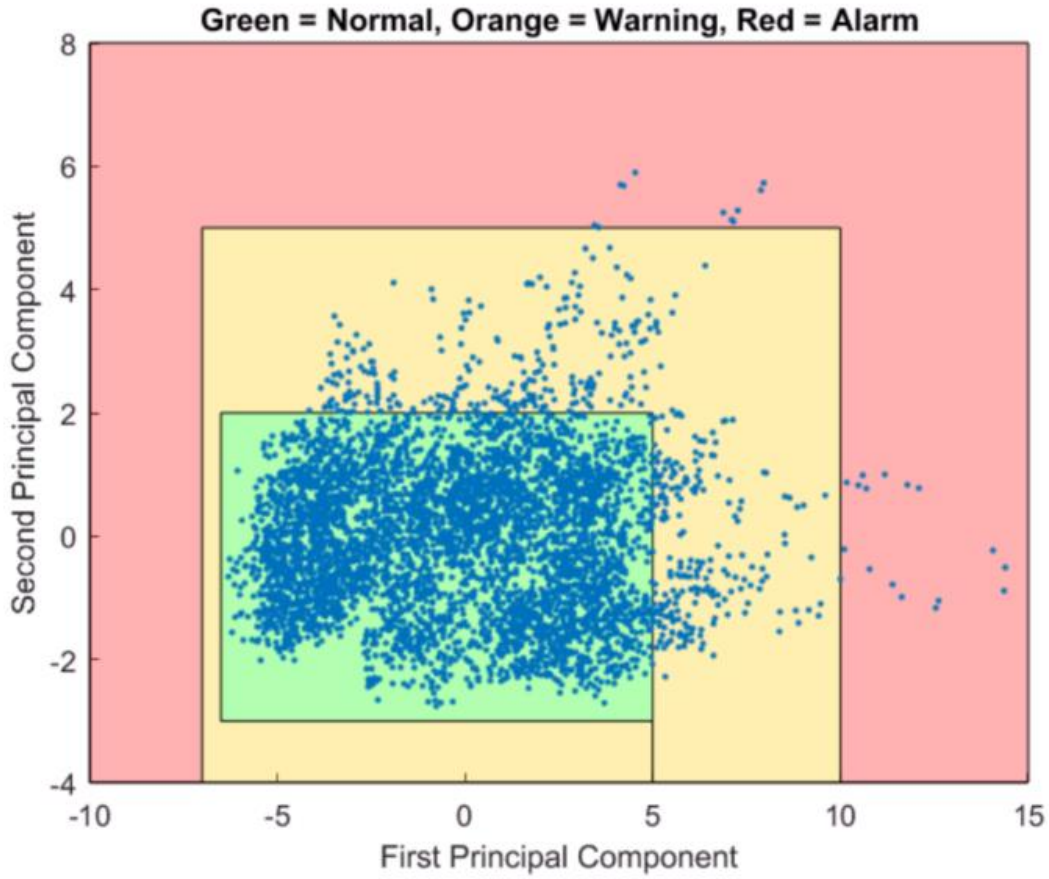
```
[coeff,score,latent] = pca(XtrainStandard);
```



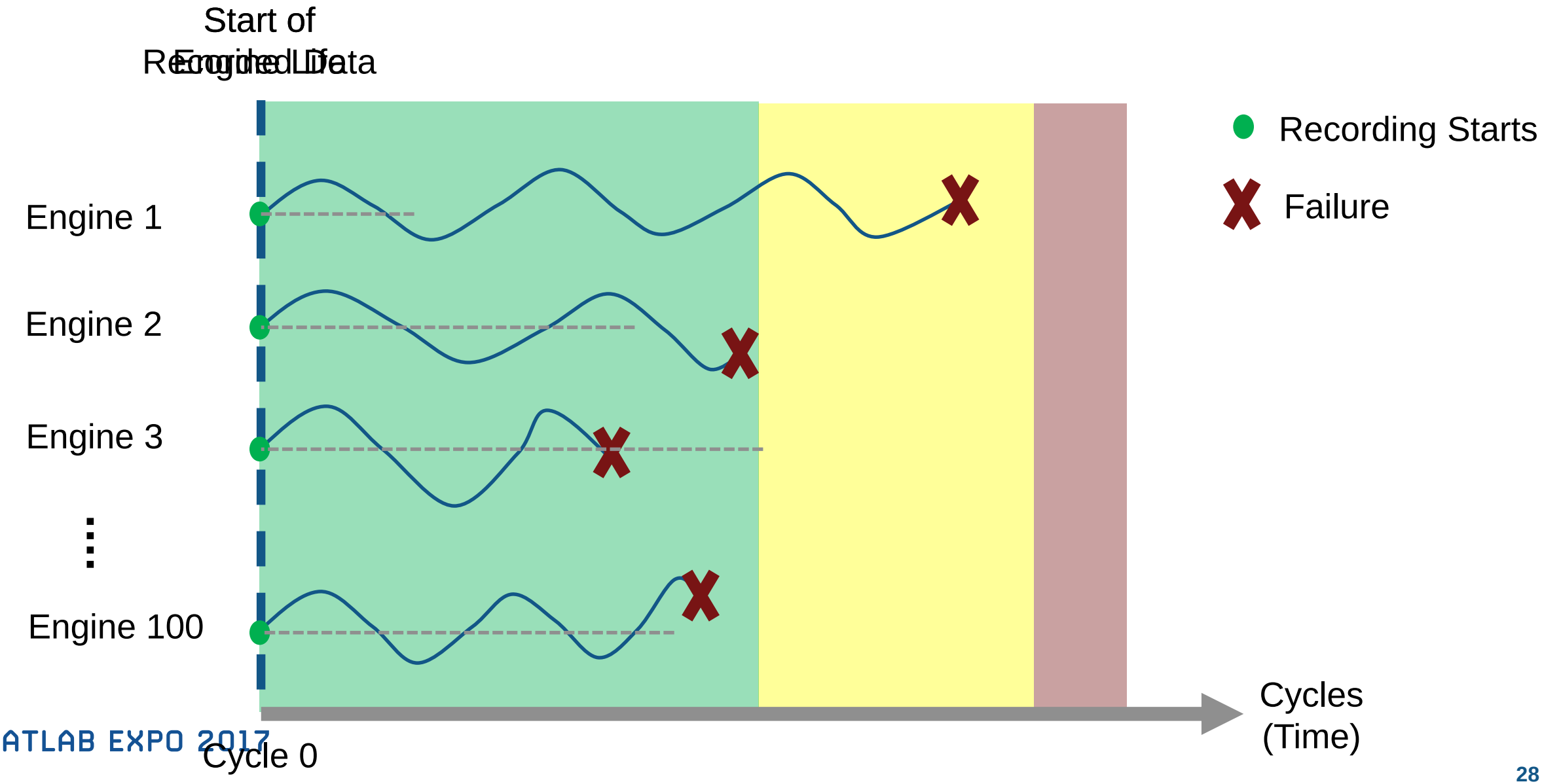
SignalContributions	
LPCOutletTemp	0.27153
HPCOutletTemp	0.25817
LPTOutletTemp	0.28662
TotalHPCOutletPres	-0.2875
PhysFanSpeed	0.29237
PhysCoreSpeed	-0.13822
StaticHPCOutletPres	0.29246
FuelFlowRatio	-0.29173
CorrFanSpeed	0.29192
CorrCoreSpeed	-0.18721
BypassRatio	0.27843
BleedEnthalpy	0.2631
HPTCoolantBleed	-0.27535
LPTCoolantBleed	-0.27706

Early Warning System

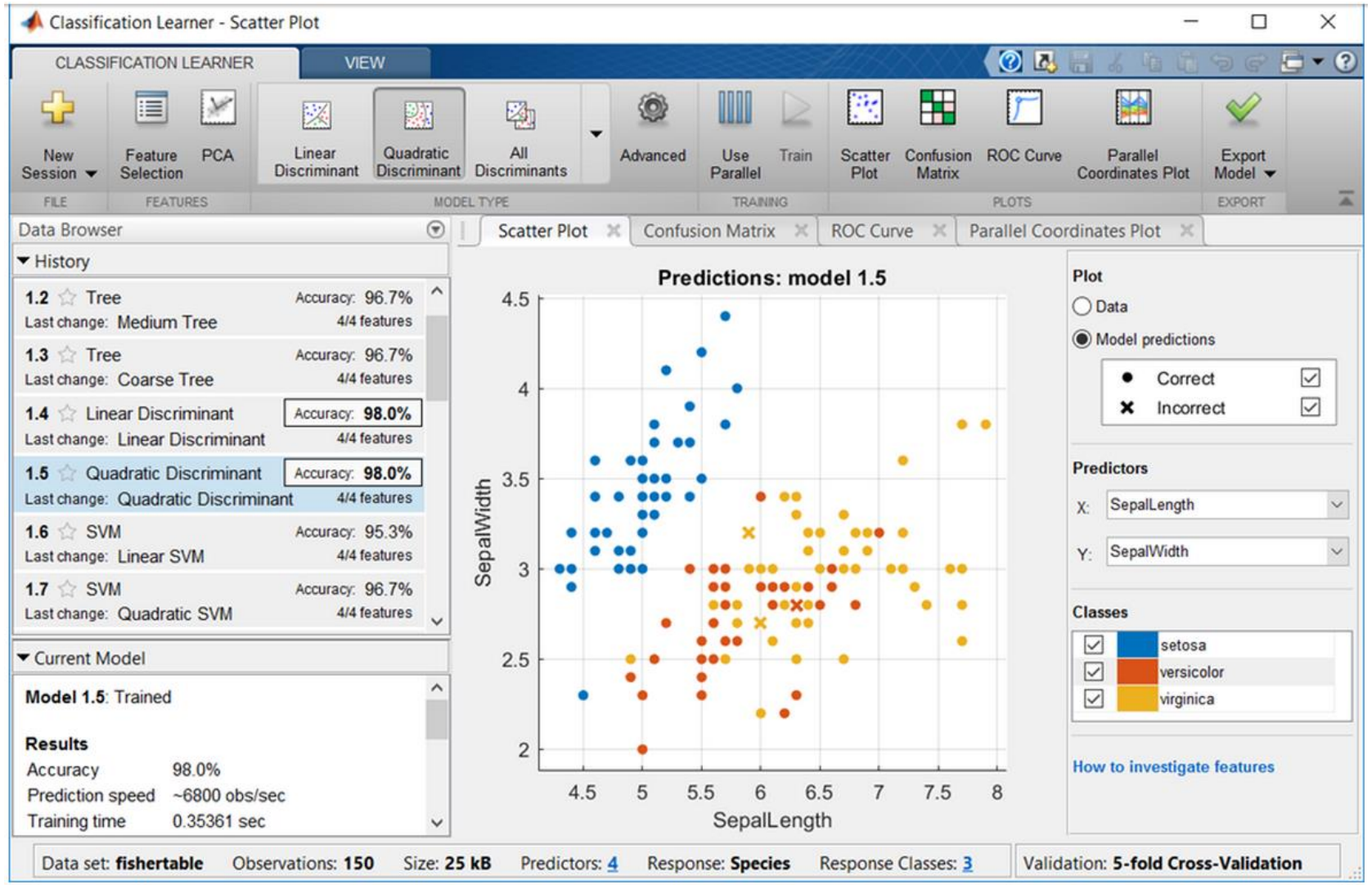




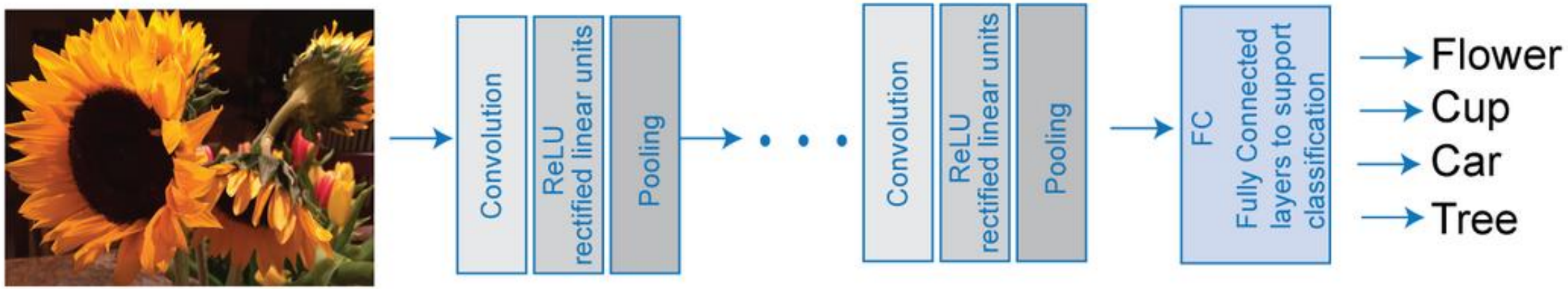
Preprocessing and Classifying our Input Data



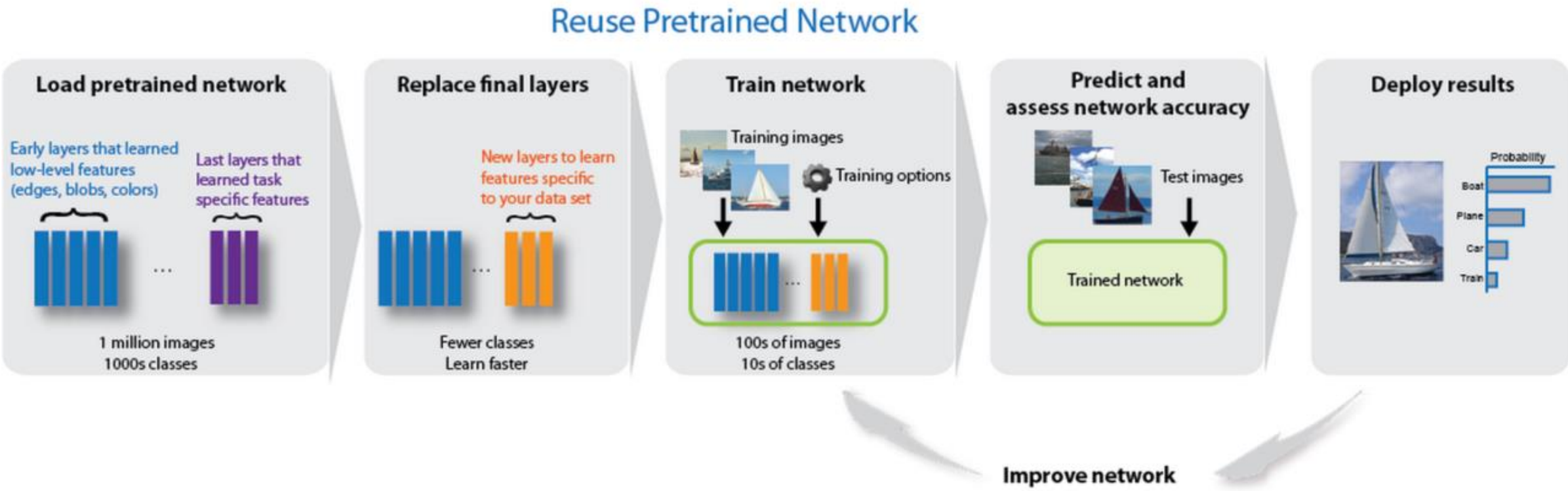
Classification Learner App



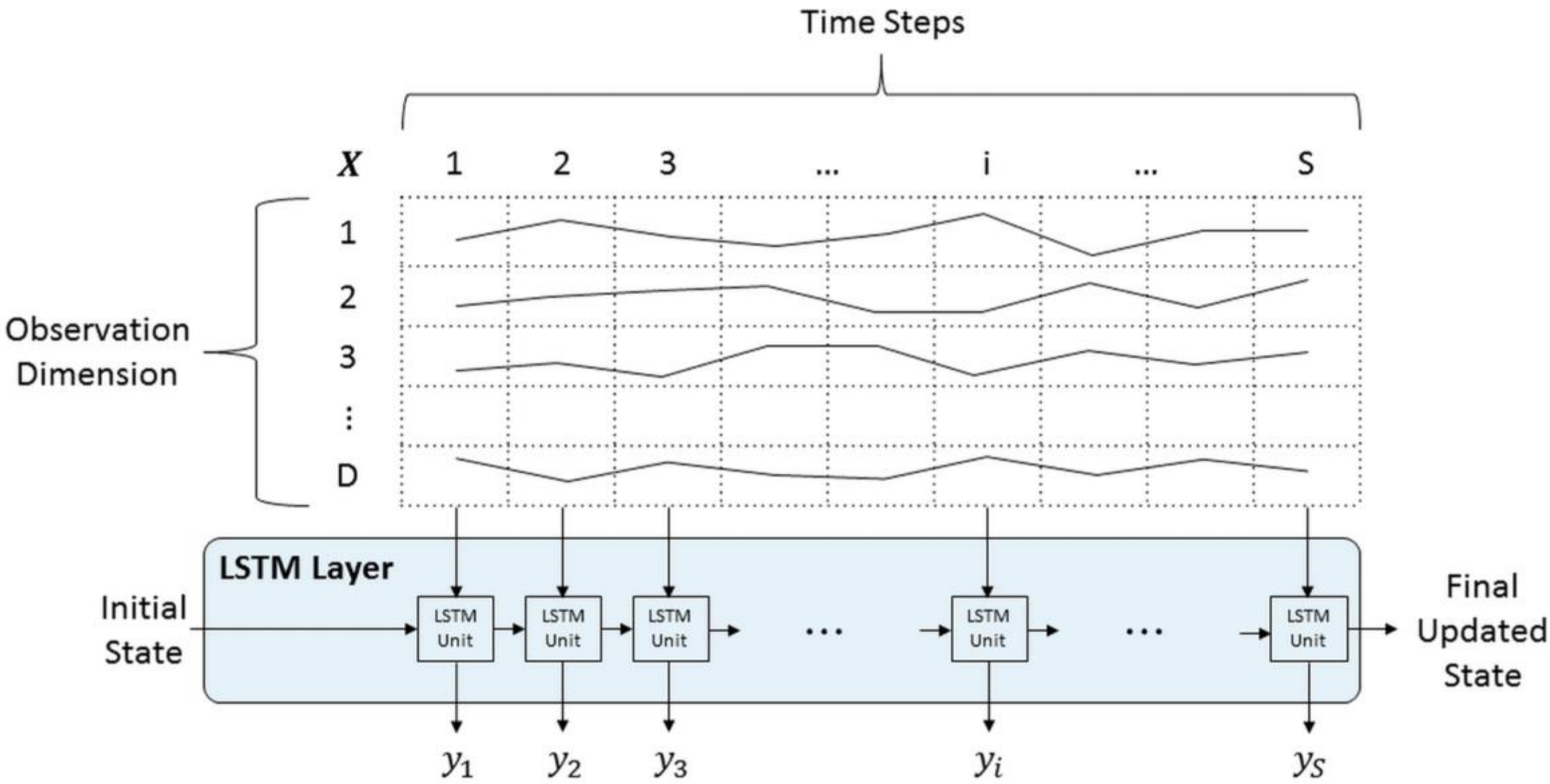
Convolutional Neural Network



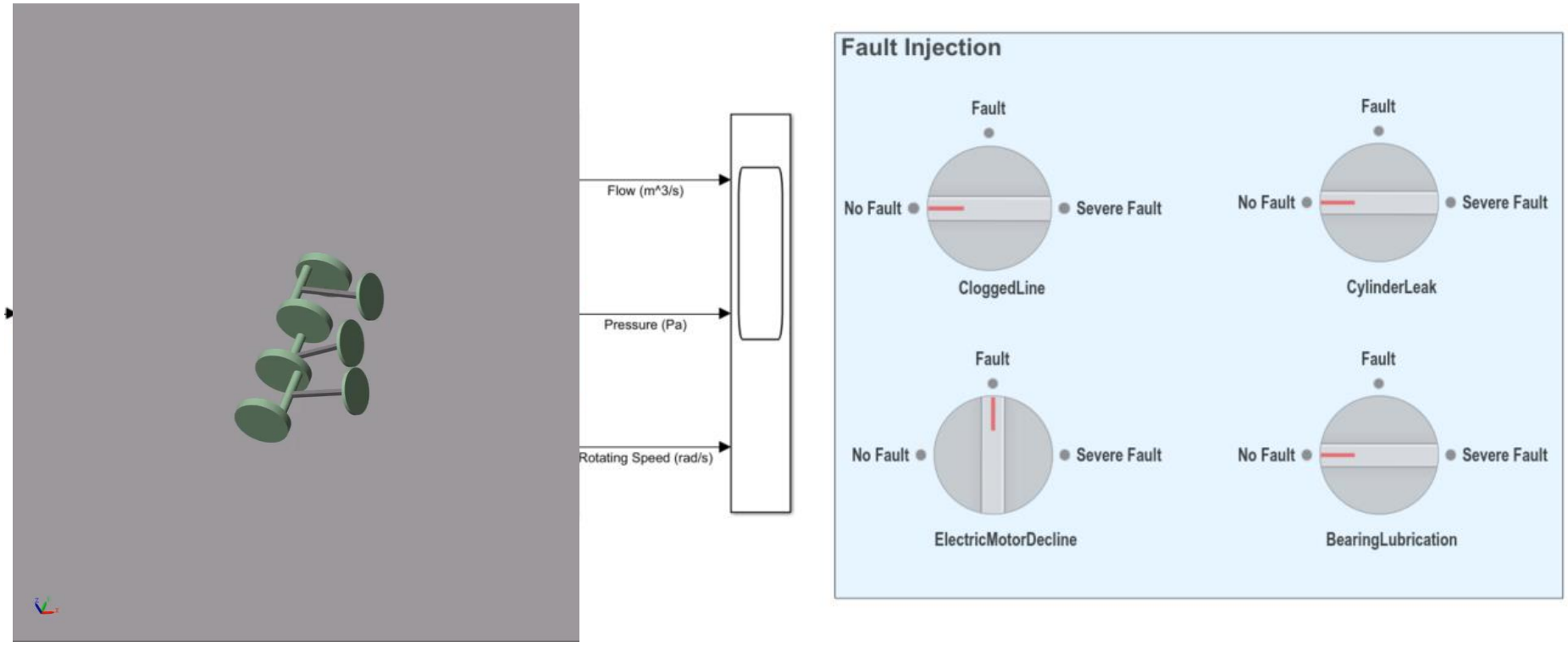
Pretrained Networks



LSTM Network



Useful Life Estimation Simulink Model



Workflow

Access and Explore Data

Files

Sensors

Preprocess Data

Visualizing Data

Data Reduction/ Transformation

Develop Predictive Models

Machine learning

Model Validation

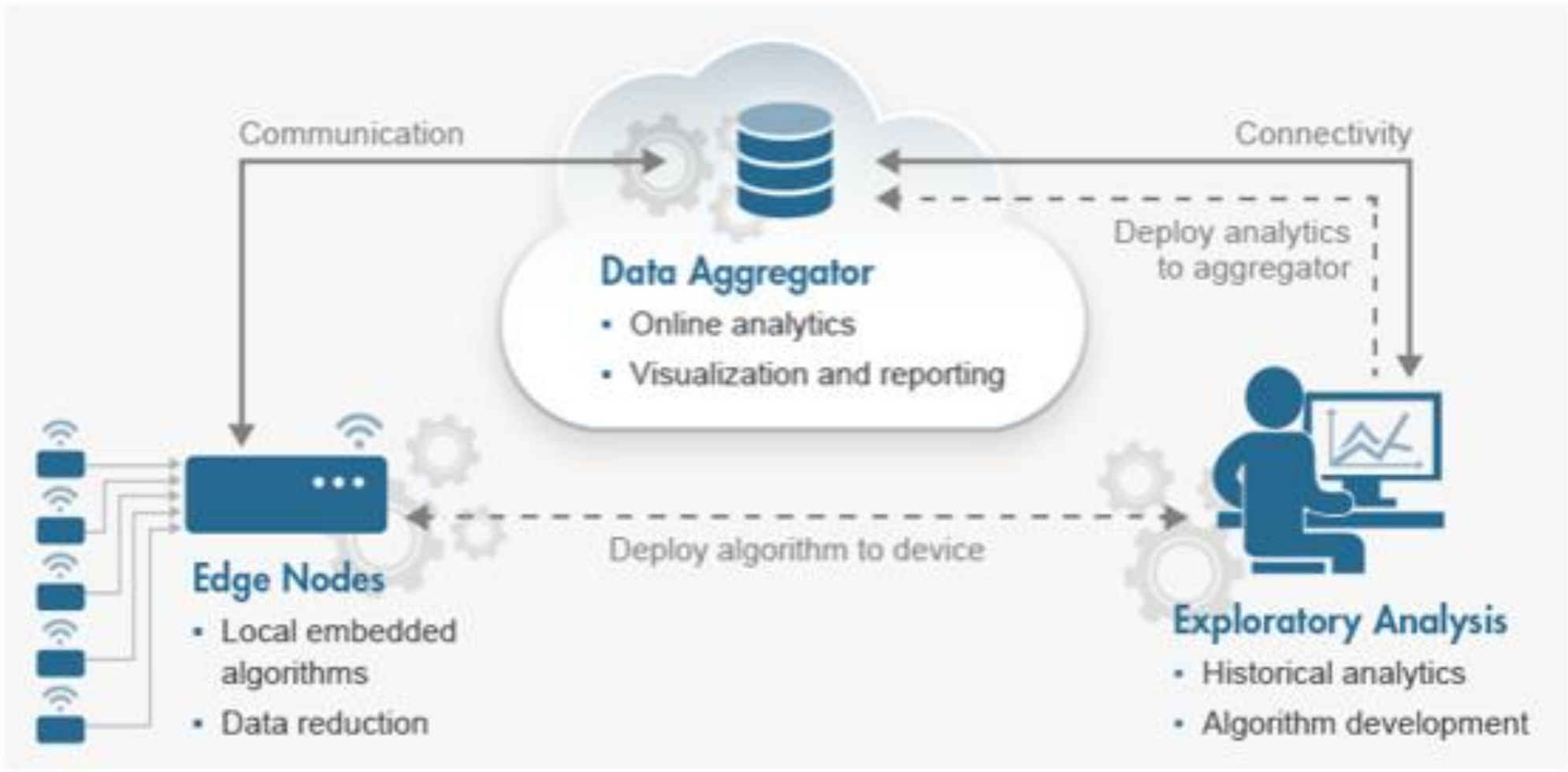
Integrate Analytics with Systems

Embedded Devices/Hardware

Enterprise Scale Systems

MATLAB Excel .NET C/C++ .exe Java .dll

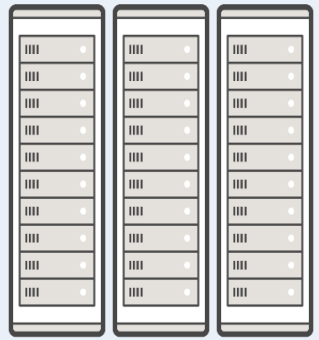
Internet of Things



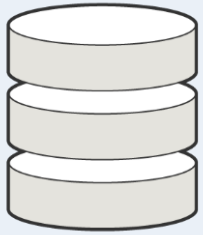
Using Tall Arrays

- **Run in parallel on compute clusters**
  MATLAB Distributed Computing Server

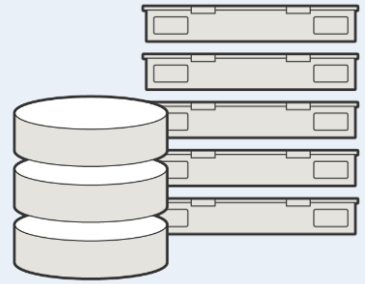
Compute Clusters



Local disk
Shared folders
Databases



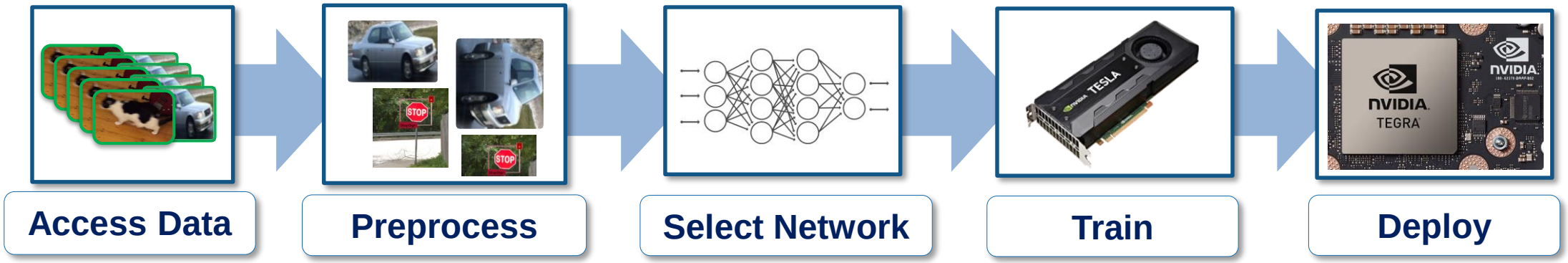
Spark + Hadoop



- **Tall arrays**
  MATLAB
- **100's of functions supported**
  MATLAB
  Statistics and Machine Learning Toolbox
- **Run in parallel**
  Parallel Computing Toolbox

- **Run in parallel on Spark clusters**
  MATLAB Distributed Computing Server
- **Deploy MATLAB applications as standalone applications on Spark clusters**
  MATLAB Compiler

Working with GPU Coder: Deep Learning Workflow



Machine Learning on MATLAB Production Server

Shell analyses big data sets to detect events and abnormalities at downstream chemical plants using predictive analytics with MATLAB®. Multivariate statistical models running on MATLAB Production Server™ are used to do real-time batch and process monitoring, enabling real-time interventions when abnormalities are detected.

Feedback

Big Data and Predictive Analytics at Shell

Amjad Chaudry, Shell

Where Next?

Talks

- MatConvNet: Deep Learning Research in MATLAB
- Introduction to Machine & Deep Learning
- Scaling MATLAB for your Organisation and Beyond

Demo Stations

- Big Data with MATLAB
- Deep Learning with MATLAB
- Predictive Maintenance with MATLAB and Simulink
- Deploying Video Processing Algorithms to Hardware
- Using MATLAB and ThingSpeak to Explore the Internet of Things

Thank You!