Tutorial

Systematic Methodology for Big Data in Industrial Applications

Professor Jay Lee, Director of IMS Center , University of Cincinnati
Dr. David Siegel, Chief Technology Officer at Predictronics
Dr. Brian Weiss, Project Leader, National Institute of Standard and Technology (NIST)
Dr. Hossein Davari, Associate Director of IMS Center, University of Cincinnati

IEEE Big Data Conference
Dec. 11-14, 2017, Boston, MA

CENTER FOR INTELLIGENT MAINTENANCE SYSTEMS

WE KNOW WHAT HAPPENS NEXT

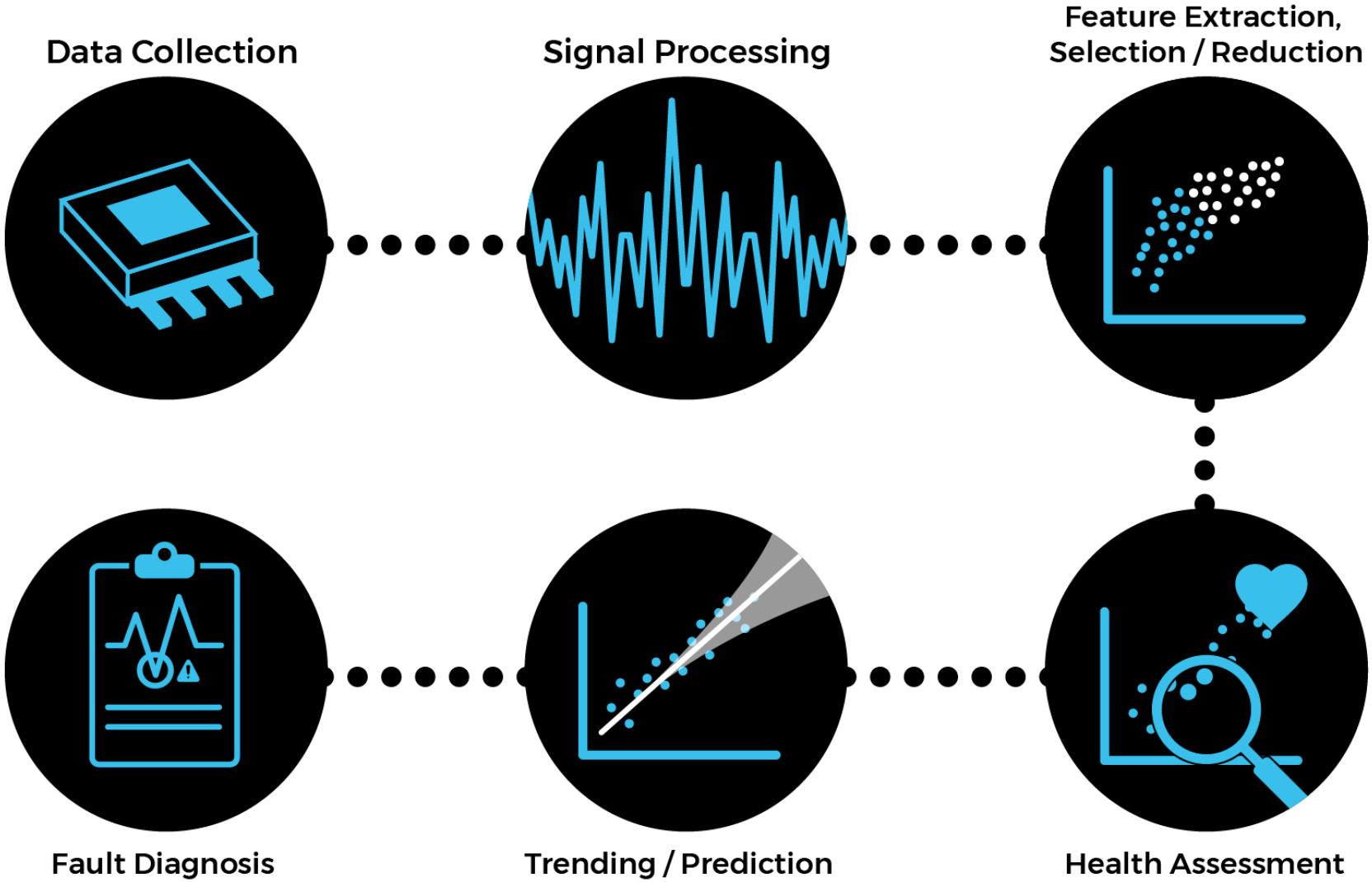
Outline

- » Introduction to Industrial Big Data – Trends and Recent Advances
- » Industrial Big Data – Data Analysis Methodology
- » Test Methods, Reference Data Sets, Verification and Validation
- » Industrial Case Studies
- » Concluding Remarks

Introduction to Industrial Big Data – Trends and Recent Advances

Industrial Big Data – Data Analysis Methodology

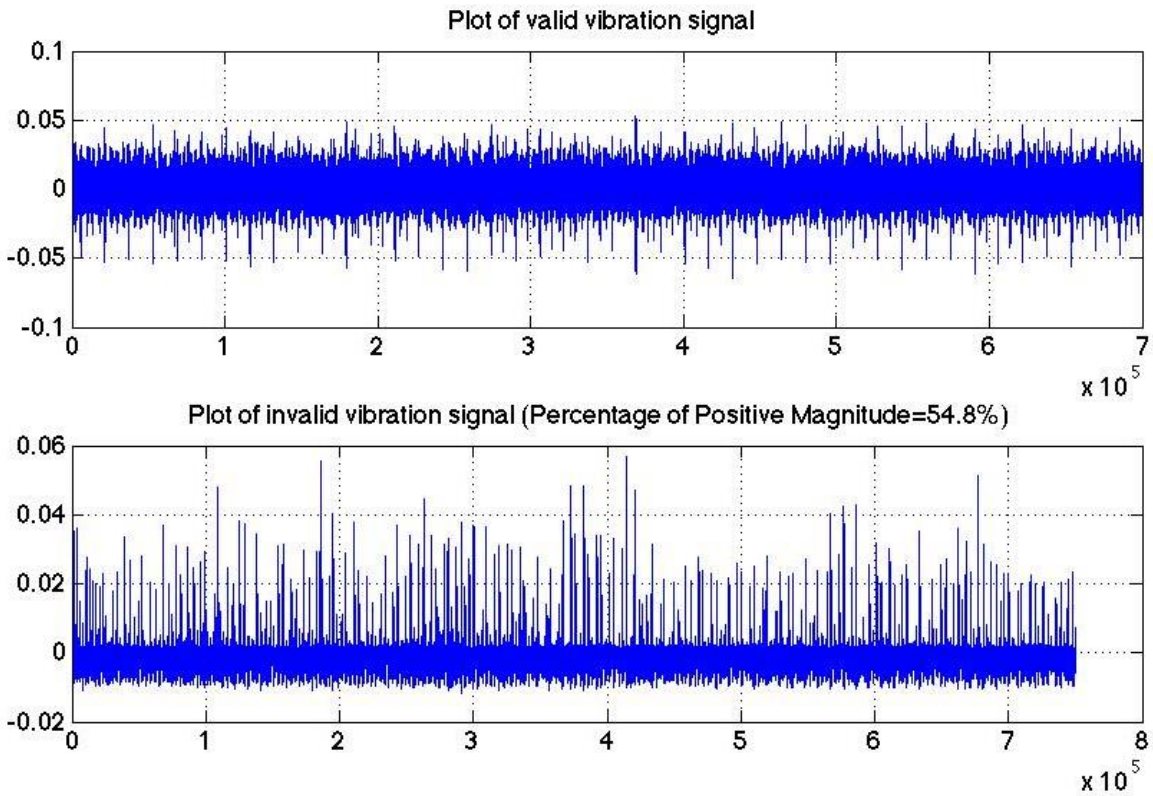
Industrial Big Data – Data Analysis Methodology



Data Pre-Processing Aspects

Real world data could have outliers / incorrect readings	Outlier rules / signal quality checks (domain specific)
To separate the noise from the true signal	Improving the signal to noise ratio can improve model accuracy and reduce false alarms
To combine multiple data sources with the same time stamp (interpolation)	In some applications, various sensors or data sources could be collected at different sampling rates.
Organization of health class / fault labels	It is important for training and validation of the models to have correct labels on the health condition of the asset

Data Quality Check (Example 2 – Vibration Data Quality Check)



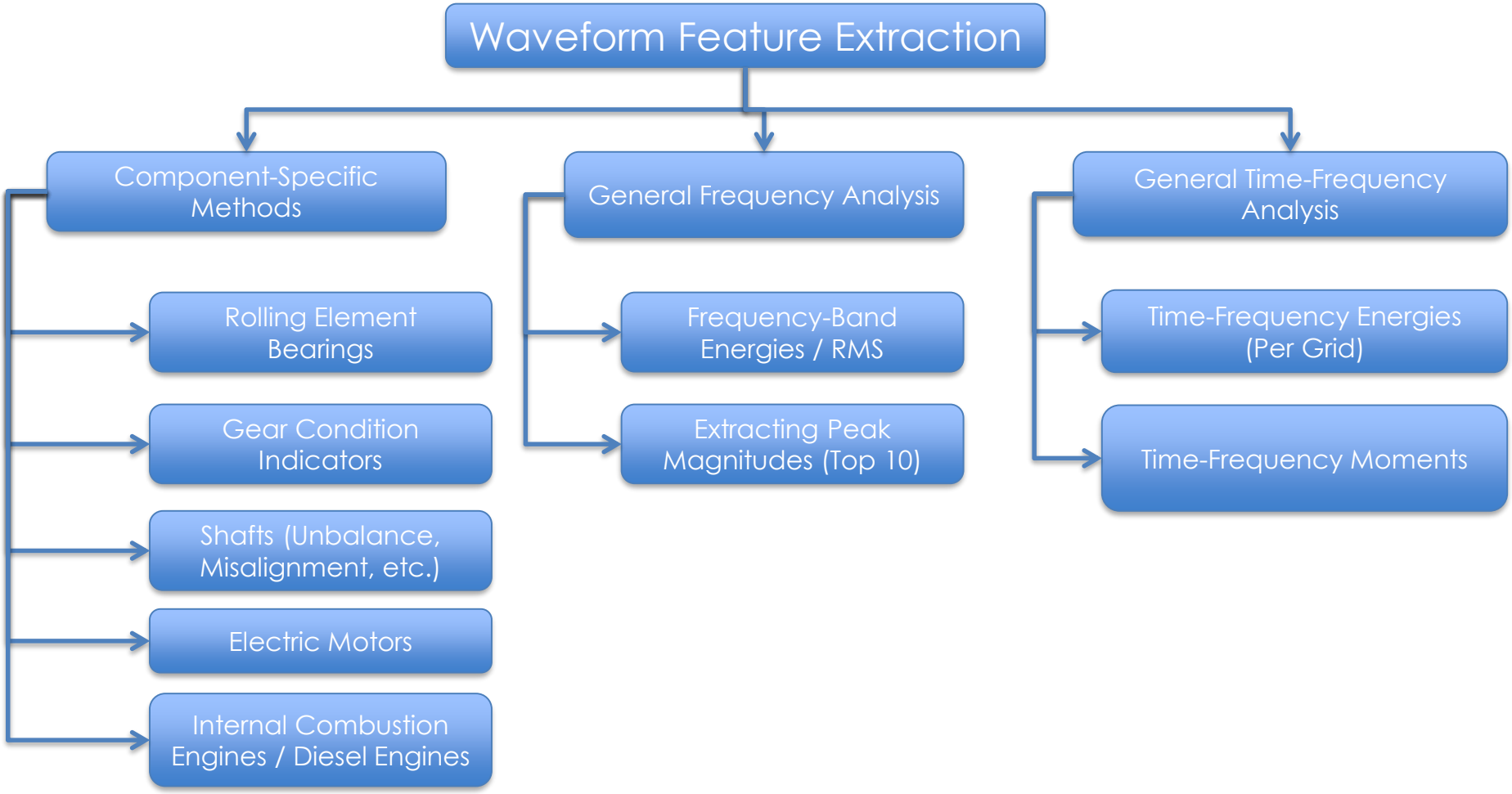
» The positive and negative percentage rule requires that the number of positive magnitude and negative magnitude points should be evenly divided

» The threshold for the larger percentage number (positive or negative) was <52%

A. Jablonski, T. Barszcz, and M. Bielecka, Automatic validation of vibration signals in wind farm distributed monitoring systems, Measurement 44 (2011) 1954-1967'

Feature Engineering

Waveform Feature extraction Overview



» For well studied and monitored components, there is already many established features that work reasonably well.

Bearing envelope analysis (example flow chart)

Vibration Signal

↓

Band pass filter around excited natural frequency

↓

Take Hilbert Transform of band pass signal

↓

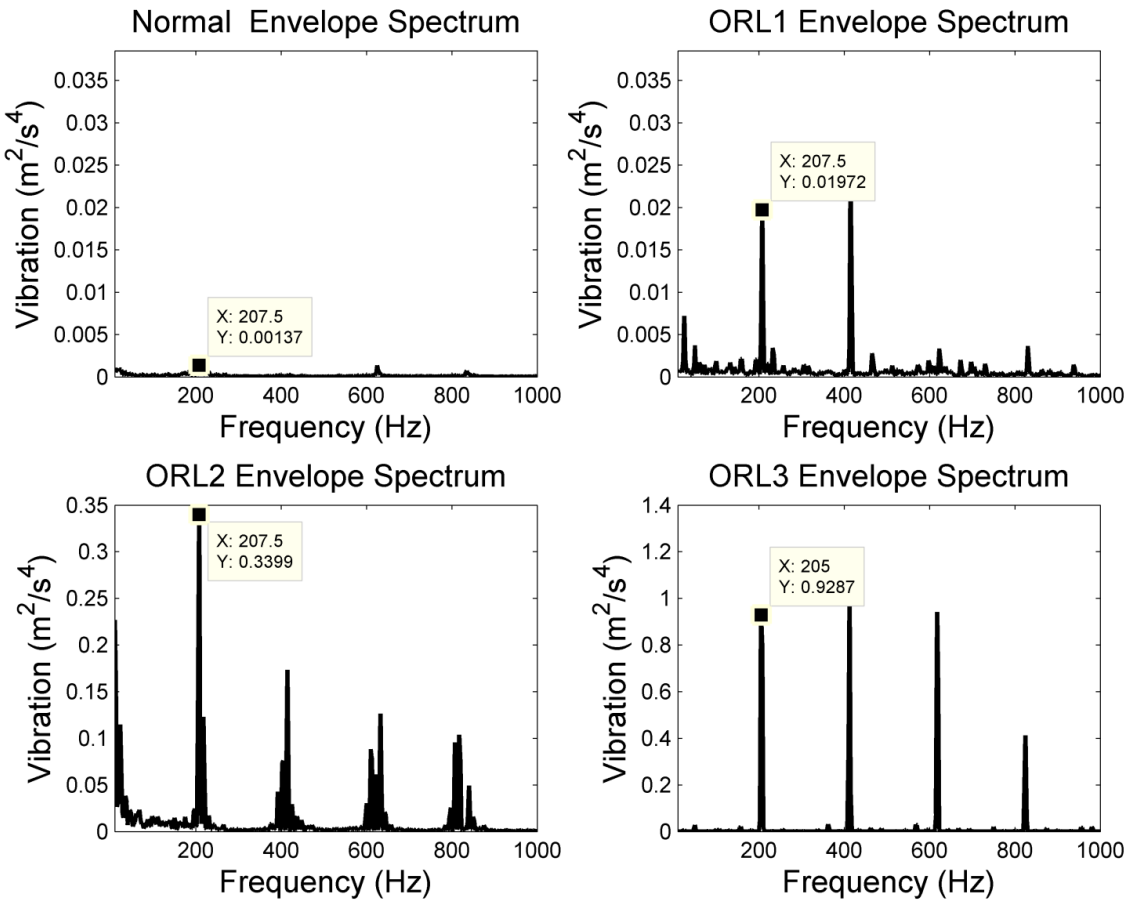
Take magnitude of analytical signal to obtain envelope signal

↓

Take Fast Fourier Transform (FFT) of envelope signal

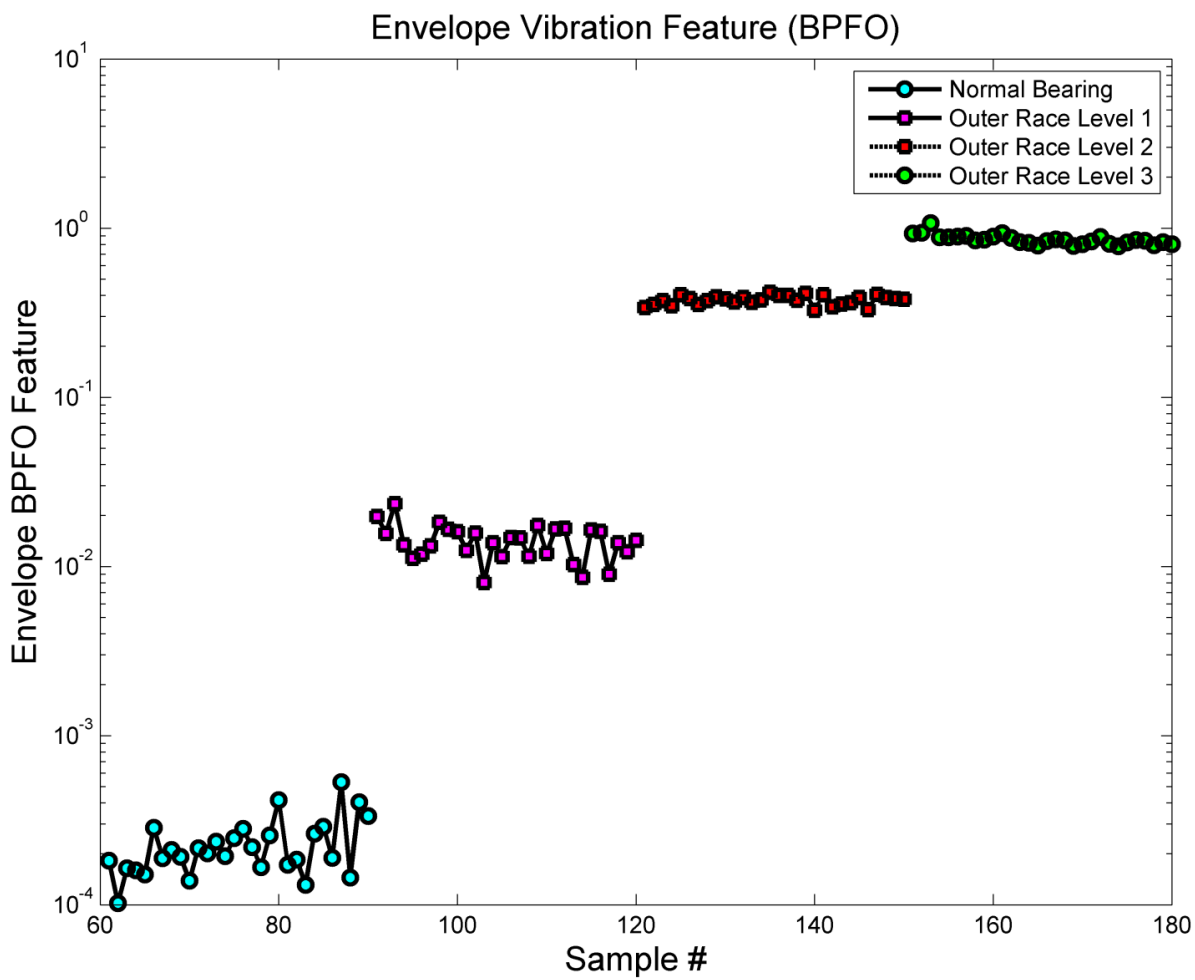
↓

Analyze envelope spectrum at bearing fault frequencies



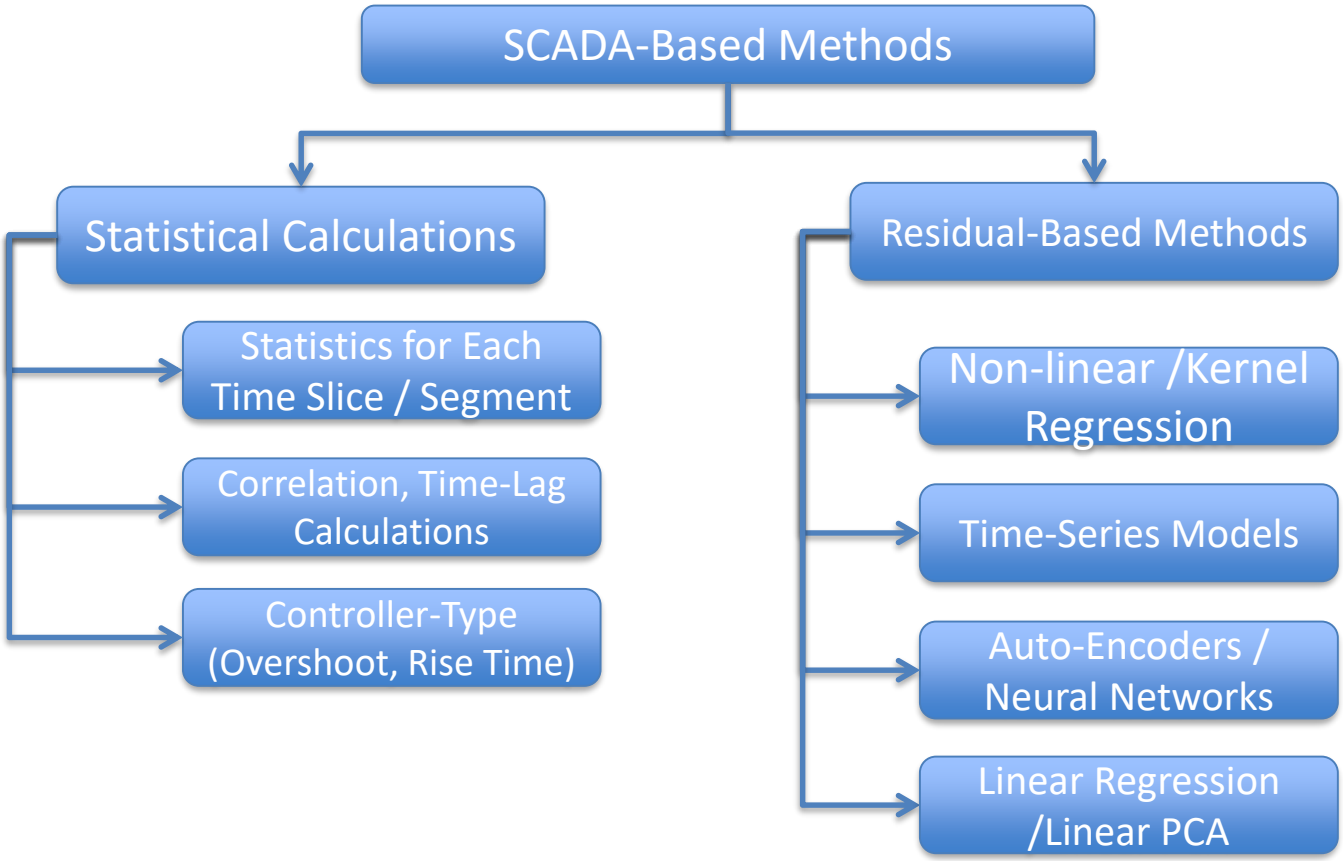
D. Siegel, H. Al-Atat, V. Shauche, L. Liao, J. Snyder, J. Lee, Novel method for rolling element bearing health assessment – A tachometer-less synchronously averaged envelope feature extraction technique, Mechanical Systems and Signal Processing, Volume 29, pp. 362-276, 2012

Example Feature from Envelope Analysis



» The envelope method provide a good feature that can distinguish between a normal bearing and the different levels of outer race degradation.

SCADA Feature extraction Techniques Overview



» The SCADA based feature extraction method includes a variety of residual based analysis methods along with more straightforward statistical based calculations.

Feature Selection and Dimension Reduction Techniques

Dimension Reduction

» Purpose

- Improve algorithm estimation accuracy: "Curse of dimensionality"
- Provide a better understanding of the underlying process
- Reduce the number of sensors installed
- Speed up learning process

Variable selection

Instance selection

Samples	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	. . .	. . .	. . .
1	8546.9	0.2327	15296.9	0.3589	17203.1	0.0998	18453	. . .	. . .	. . .
2	8531.2	0.0786	15203.1	0.2417	17171.9	0.0939	18422	. . .	. . .	. . .
3	8562.5	0.2087	15328.1	0.4016	17218.8	0.0623	18469	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

Category of Dimension Reduction

- Instance selection
  - Clustering/classification
  - Re-sampling
- Variable selection (a.k.a. feature selection, sensor selection)
  - Select a subset of variables
  - “Maximize relevance and minimize redundancy”
- Principal Component Analysis /Independent Component Analysis
  - Transform the original data into low dimensional data (not to reduce the dimensionality of the original data)
  - Remove correlation / redundancy among the variables

Principal Component Analysis

Strength and Weaknesses of PCA

» Strength

– Powerful linear scheme for compressing a set of high dimensional vectors into lower dimensional vectors

» Weaknesses

– PCA fails when the largest variances do not correspond to meaningful axes.

– The method is not useful if the features are highly uncorrelated.

Summary of Steps

Raw Data

	Feature 1	Feature 2	Feature 3	Feature 4	
Sample 1	7	26	6	60	
Sample 2	1	29	15	52	
Sample 3	11	56	8	20	
Sample 4	11	31	8	47	
Sample 5	7	52	6	33	
Sample 6	11	55	9	22	
Sample 7	3	71	17	6	
Sample 8	1	31	22	44	
Sample 9	2	54	18	22	
Sample 10	21	47	4	26	
Sample 11	1	40	23	34	
Sample 12	11	66	9	12	
Sample 13	10	68	8	12	
Mean	7.46	48.15	11.77	30.00	
Variance	34.60	242.14	41.03	280.17	597.94

Standardize the Data

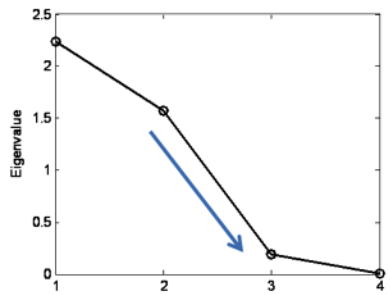
	Feature 1	Feature 2	Feature 3	Feature 4
Sample 1	-0.078	-1.424	-0.901	1.792
Sample 2	-1.098	-1.231	0.504	1.314
Sample 3	0.602	0.504	-0.588	-0.597
Sample 4	0.602	-1.102	-0.588	1.016
Sample 5	-0.078	0.247	-0.901	0.179
Sample 6	0.602	0.440	-0.432	-0.478
Sample 7	-0.758	1.468	0.817	-1.434
Sample 8	-1.098	-1.102	1.597	0.836
Sample 9	-0.928	0.376	0.973	-0.478
Sample 10	2.302	-0.074	-1.213	-0.239
Sample 11	-1.098	-0.524	1.753	0.239
Sample 12	0.602	1.147	-0.432	-1.075
Sample 13	0.432	1.275	-0.588	-1.075
Mean	0.000	0.000	0.000	0.000
Variance	1	1	1	1

Eigenvalues & Eigenvectors

Eigenvalues	2.24	1.58	0.19	0.00
Eigenvectors	-0.476	0.509	0.676	0.241
	-0.564	-0.414	-0.314	0.642
	0.394	-0.605	0.638	0.268
	0.548	0.451	-0.195	0.677

Choose PC to Retain

	PC 1	PC 2	PC 3	PC 4
Sample 1	-36.82	6.87	4.59	0.40
Sample 2	-29.61	-4.61	2.25	-0.40
Sample 3	12.98	4.20	-0.90	-1.13
Sample 4	-23.71	6.63	-1.85	-0.38
Sample 5	0.55	4.46	6.09	0.14
Sample 6	10.81	3.65	-0.91	-0.13
Sample 7	32.59	-8.98	1.61	0.08
Sample 8	-22.61	-10.73	-3.24	0.32
Sample 9	9.26	-8.99	0.02	-0.54
Sample 10	3.28	14.16	-7.05	0.34
Sample 11	-9.22	-12.39	-3.43	0.44
Sample 12	25.58	2.78	0.39	0.45
Sample 13	26.90	2.93	2.45	0.41



Eig 1	2.2357
Eig 2	1.5761
Eig 3	0.1866
Eig 4	0.0016

		Variance per PC	Cumulative Variance
Eig 1	2.2357	55.89%	55.89%
Eig 2	1.5761	39.40%	95.29%
Eig 3	0.1866	4.67%	99.96%
Eig 4	0.0016	0.04%	100.00%

Select Number of PC to Retain

Overview of Machine Learning

Some examples of tasks best solved by learning

» Recognizing patterns:
- Objects in real scenes
- Facial identities or facial expressions
- Spoken words

» Recognizing anomalies:
- Unusual sequences of credit card transactions
- Unusual patterns of sensor readings in a nuclear power plant

» Prediction:
- Future stock prices or currency exchange rates
- Which movies will a person like?
- March Madness bracket?
- Machine failure in the future?

The machine learning framework

$$y = f(\mathbf{x})$$

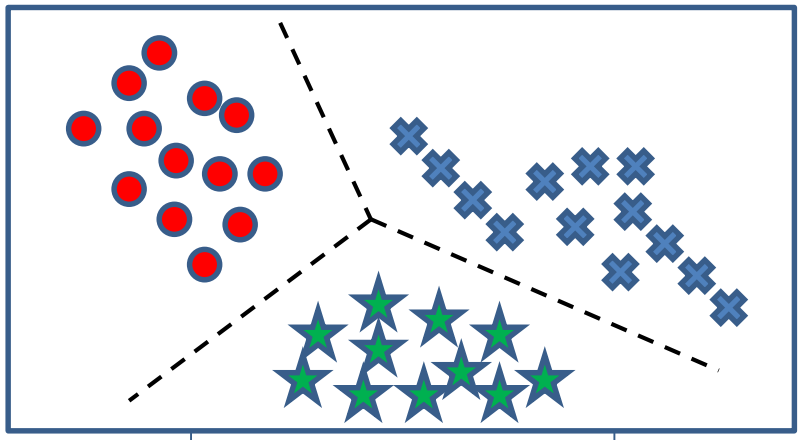
output — prediction function — Image feature

- **Training:** given a *training set* of labeled examples $\{(\mathbf{x}_1,y_1), \ldots, (\mathbf{x}_N,y_N)\}$, estimate the prediction function f by minimizing the prediction error on the training set
- **Testing:** apply f to a never before seen *test example* $\mathbf{x}$ and output the predicted value $y = f(\mathbf{x})$

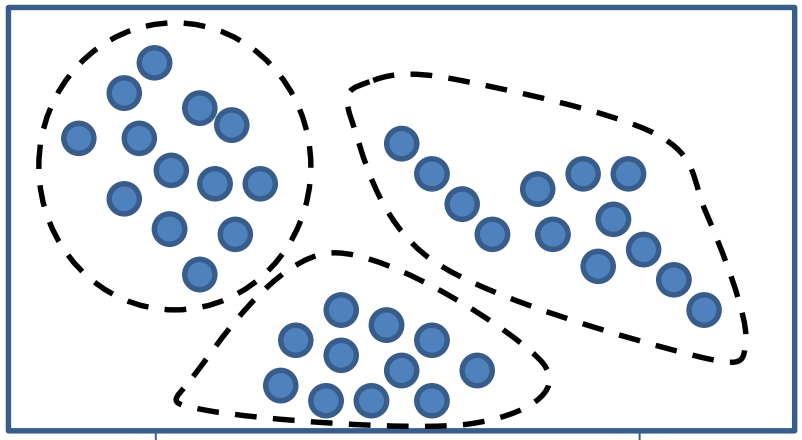
Slide credit: L. Lazebnik

Slide credit: L. Lazebnik

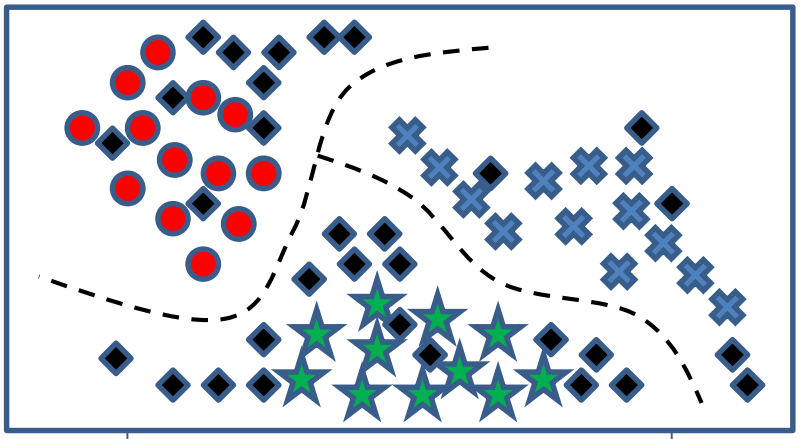
Types of Learning



Supervised learning



Unsupervised learning



Semi-supervised learning

Slide credit: Yi-Fan Chang

Generalization

Training set (labels known)

Test set (labels unknown)

» How well does a learned model generalize from the data it was trained on to a new test set?

Slide credit: L. Lazebnik

Generalization

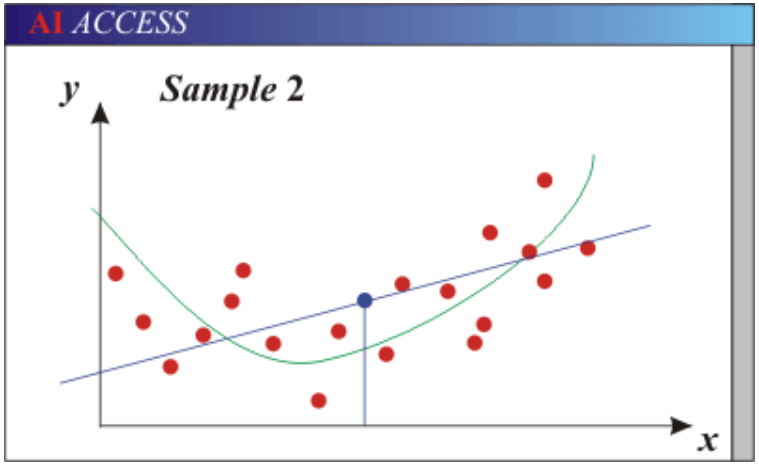
- Components of generalization error
  - Bias: how much the average model over all training sets differ from the true model?
    - Error due to inaccurate assumptions/simplifications made by the model
  - Variance: how much models estimated from different training sets differ from each other

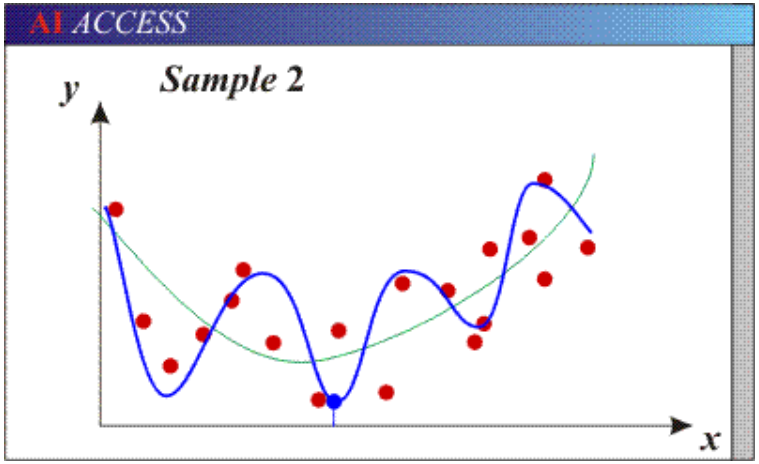
- Underfitting: model is too "simple" to represent all the relevant class characteristics
  - High bias and low variance
  - High training error and high test error

- Overfitting: model is too "complex" and fits irrelevant characteristics (noise) in the data
  - Low bias and high variance
  - Low training error and high test error

Slide credit: L. Lazebnik

Bias-Variance Trade-off





» Models with too few parameters are inaccurate because of a large bias (not enough flexibility).

» Models with too many parameters are inaccurate because of a large variance (too much sensitivity to the sample).

Slide credit: D. Hoiem

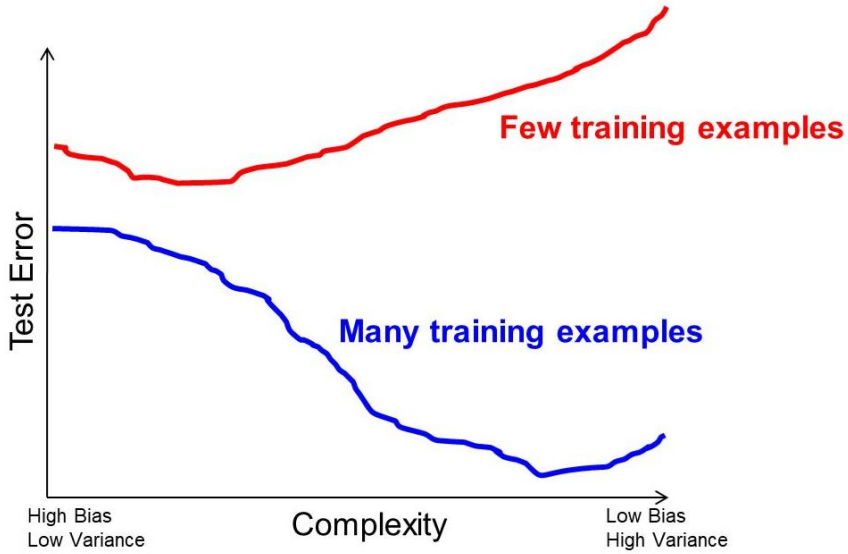
Slide credit: D. Hoiem

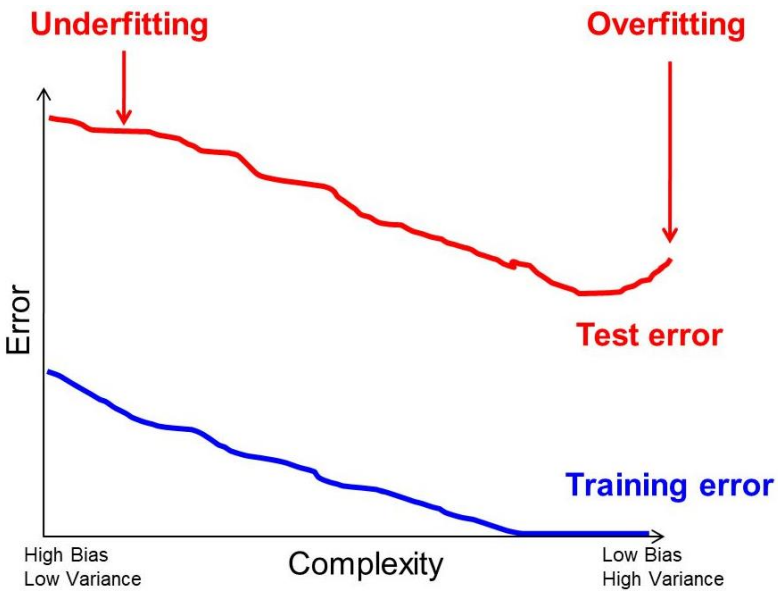
How to reduce variance?

» Choose a simpler classifier

» Regularize the parameters

» Get more training data

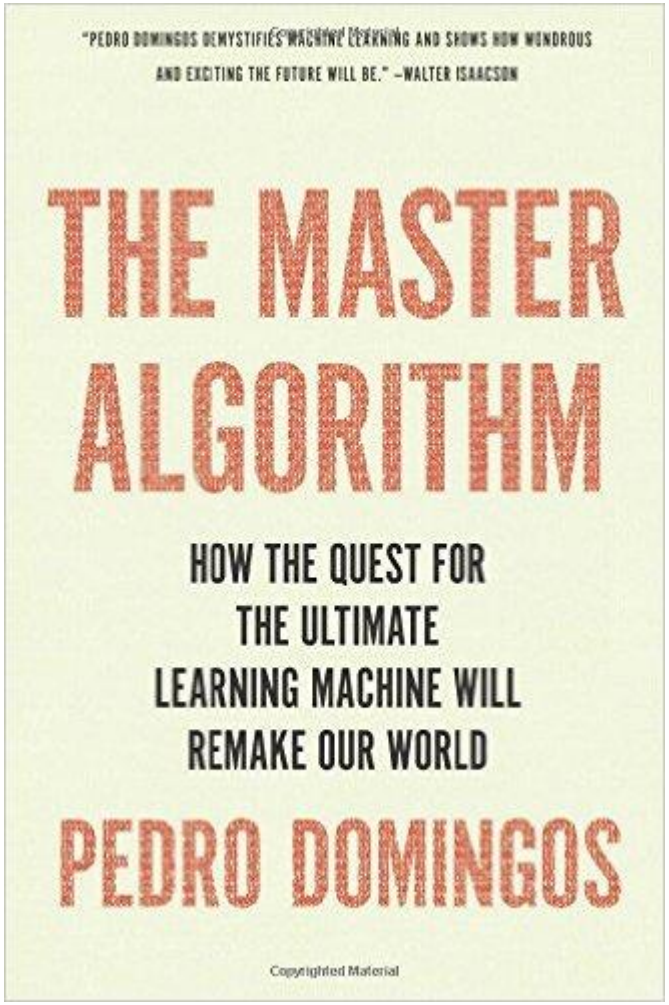




Book Suggestion: The Master Algorithm

» Hundreds of new algorithms are invented every year, but they are all based on the same few basic ideas.

» Rival Schools of thought in machine learning:
- Symbolists: take ideas from philosophy, psychology and logic.
- Connectionists: they reverse engineer the brain and are inspired by neuroscience and physics.
- Evolutionaries: simulate evolution on the computer and draw on genetics and evolutionary biology.
- Bayesians: believe learning is a form of probabilistic inference and have their roots in statistics.
- Analogizers: learn by extrapolating from similarity judgments and are influenced by psychology and mathematical optimization.



Anomaly Detection / Health Index Algorithms

Anomaly Index / Health Index Algorithm Comments

Consider your Objective	Early Detection, Diagnosis, Failure Prediction
Algorithm Selection	Select the algorithm that offers the best performance (if several are similar, choose the most simple one)
Understand the Performance Metrics and Evaluation Criteria	Real world problems always have some tradeoff between sensitivity and false alarm rate
Start with Baseline/Anomaly Detection Methods	Failure events are rare and its unlikely to have enough of them to train any machine learning algorithm

Algorithms for Anomaly Detection – Summary /Comments

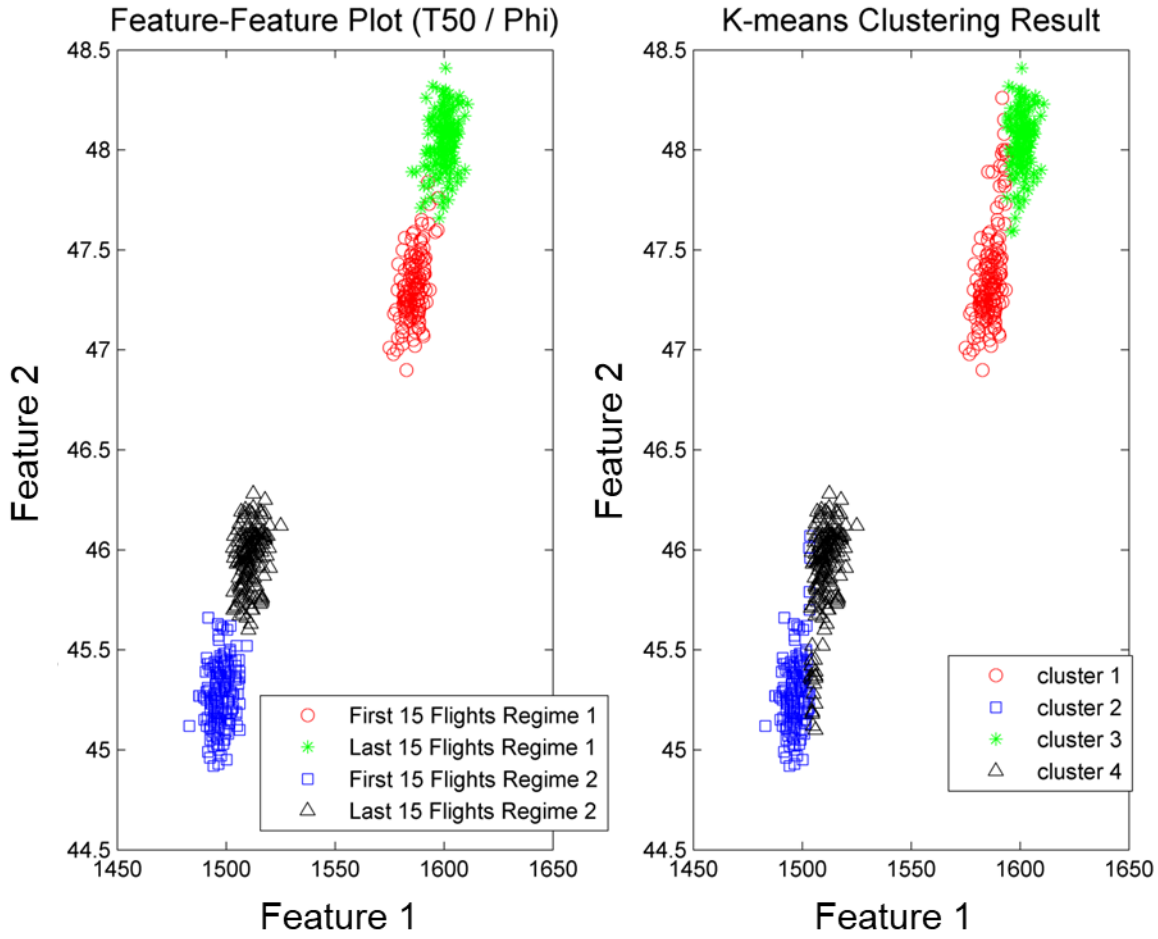
Algorithm	Advantages	Disadvantages
Auto-Associative Kernel Regression	Works very well with correlated signals, training is much easier then auto-encoder.	Does not work well for high-dimensions, sensitive to bandwidth setting.
Vector Auto Regression	Very effective for time-series anomalies, matrix formulation.	Difficult to use for large dimension data set, sensitive to selection of lag setting.
Principal Component Analysis (T^2 and SPE)	Matrix formulation, sensitive to both correlation and magnitude changes.	Does not model time-series or non-linear patterns well.
Self-Organizing Map (MQE/Distance Metric)	Built in dimension reduction/clustering.	Not well suited for time-series anomaly detection – training is more computational intensive.
Single Spectrum Analysis	Matrix formulation, good for time-series anomaly detection.	Does not model non-linear time series patterns.
Distribution (KL Divergence Metric)	Well suited for batch/distribution type data.	Sensitive to moving window size.
K-Nearest Neighbor Distance	Simple approach to anomaly detection.	Distance metrics do not work well with high-dimensions.
Auto-Encoder (Reconstruction Error)	Non-linear model, can work with large dimension data sets.	Training is more complicated, more configurable settings.

Diagnostics Approaches

Health Diagnosis Algorithm Tabular Comparison

Algorithm	Pros	Cons
Self Organizing Maps	Good for visualization and could be use for multiple classes (failure modes or severity levels).	Does not incorporate expert knowledge and not suited for discrete variables.
Support Vector Machine	Very good non-linear classifier, ideal for two class classification problems.	Requires multiple SVM's to be trained for multi-class classification, choice of kernel function is not straightforward.
Bayesian Belief Network	Can incorporate prior experience for designing BBN network, and provides a probability of a particular failure mode occurring given the current condition.	Designing a practical BBN requires experience, also requires one to make the continuous variables discrete.
Neural Network	Can be used for multi-class problems, also non-linear classifier.	Many parameters to configure, training the algorithm is more difficult, also can result in over fitting.

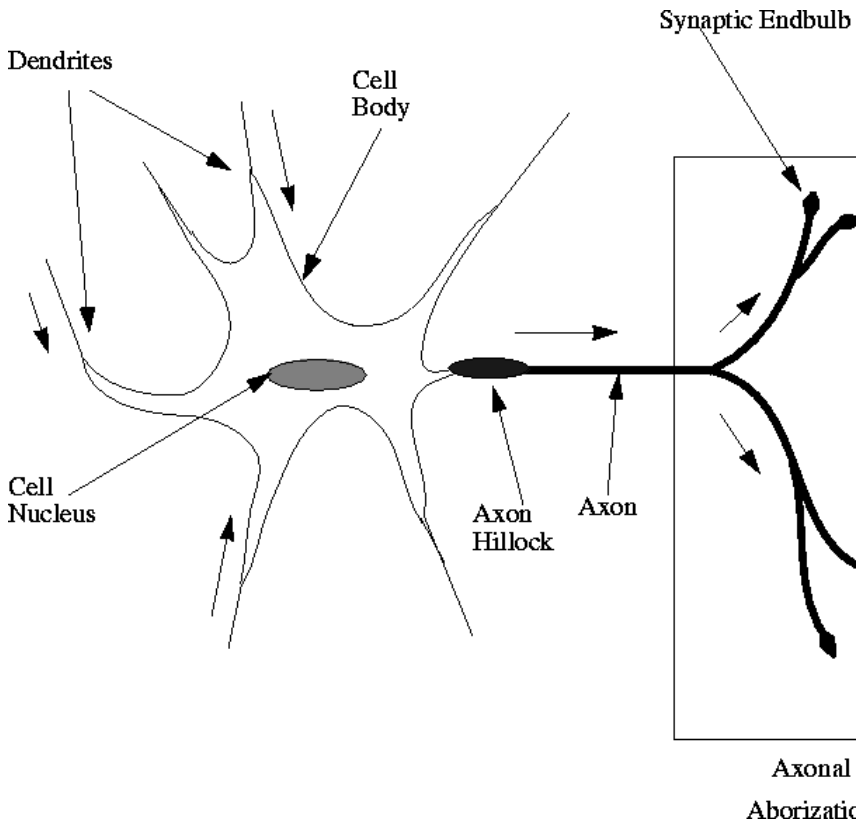
Clustering Example - K-Means

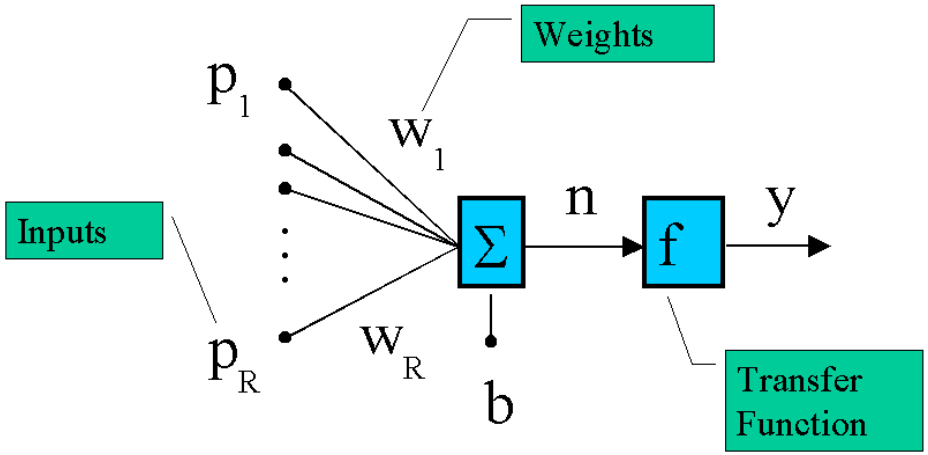


» This is an example data set in which four groups of data are provided (Two regimes and a new engine vs. an engine right before it is retired).

» K-means can cluster the data relatively well into the four groups.

Single Neuron

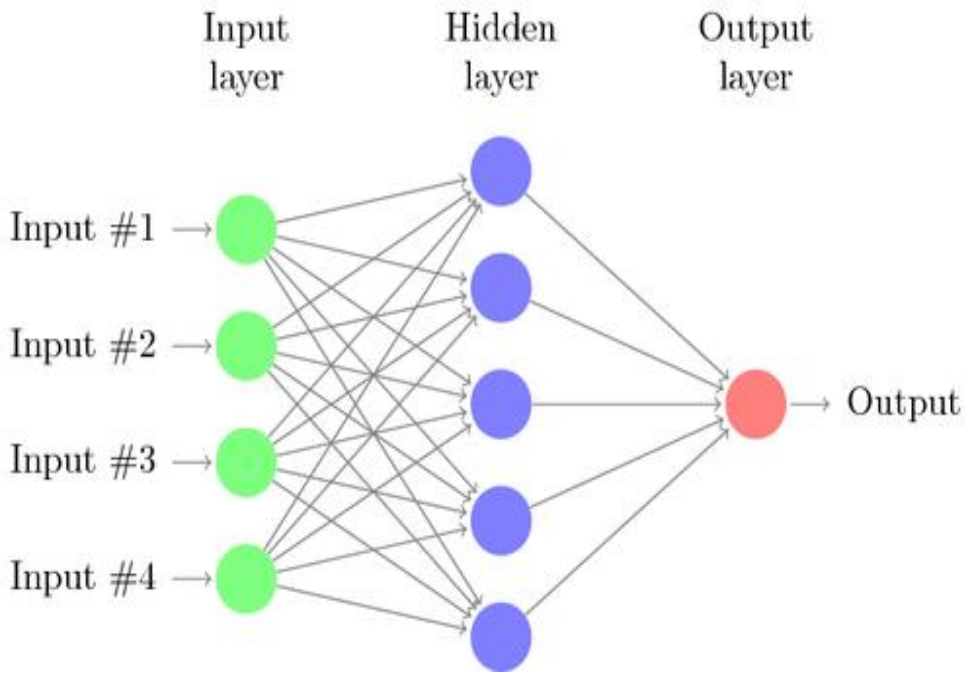




$$y = f(w'p + b)$$

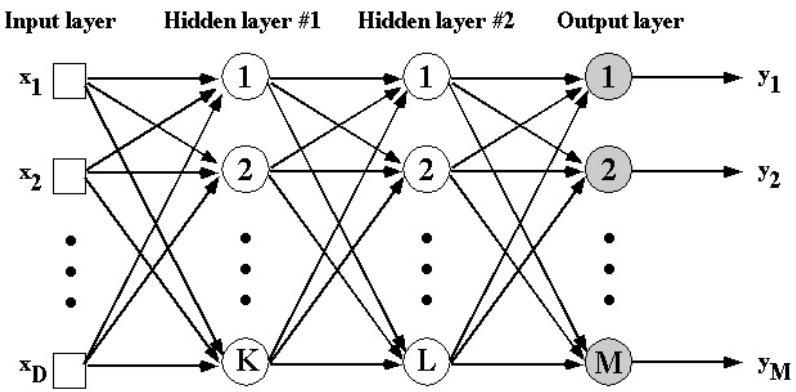
- » $p_1,\ldots,p_R$ – inputs, y – output,
- » $w_1,\ldots,w_R$ – weights, b – threshold/offset
- » f – activation function (transfer function)

Multi-layer Feed Forward



Single hidden layer NN

Multiple hidden layers NN

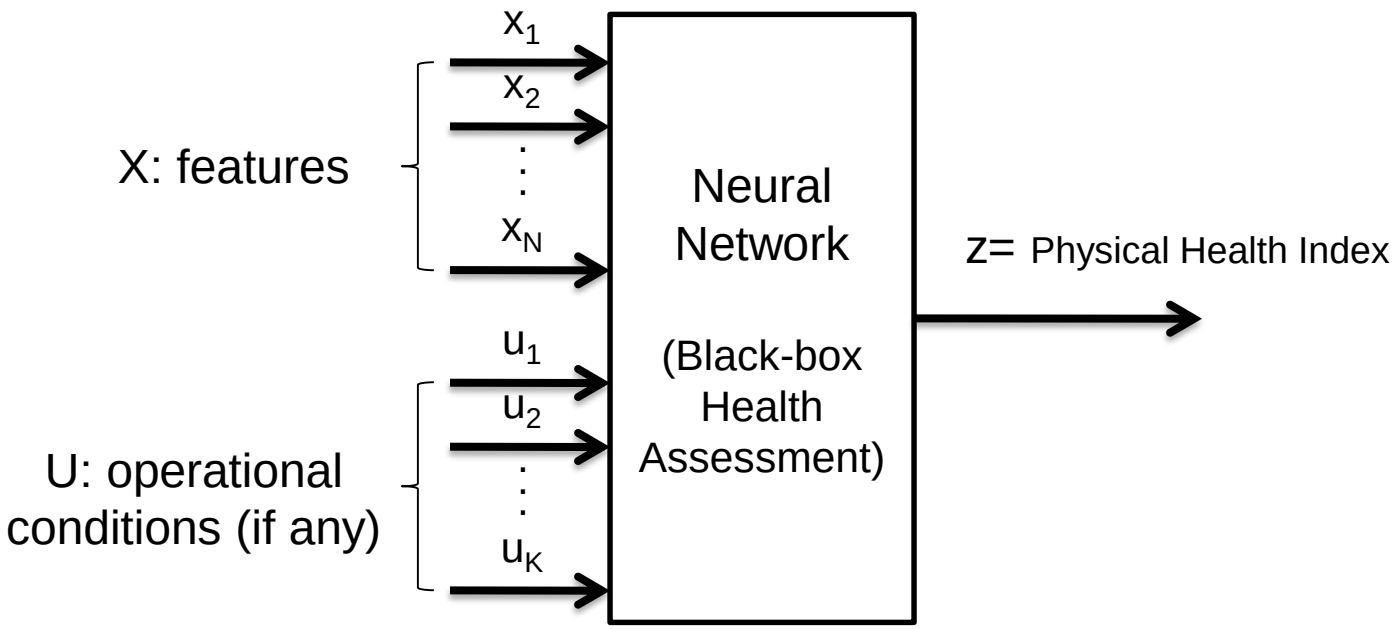


$$a^2 = f^2(W^2 f^1(W^1 p + b^1) + b^2))$$

$$a^3 = f^3(W^3 f^2(W^2 f^1(W^1 p + b^1) + b^2) + b^3)$$

Neural Network for Health Assessment

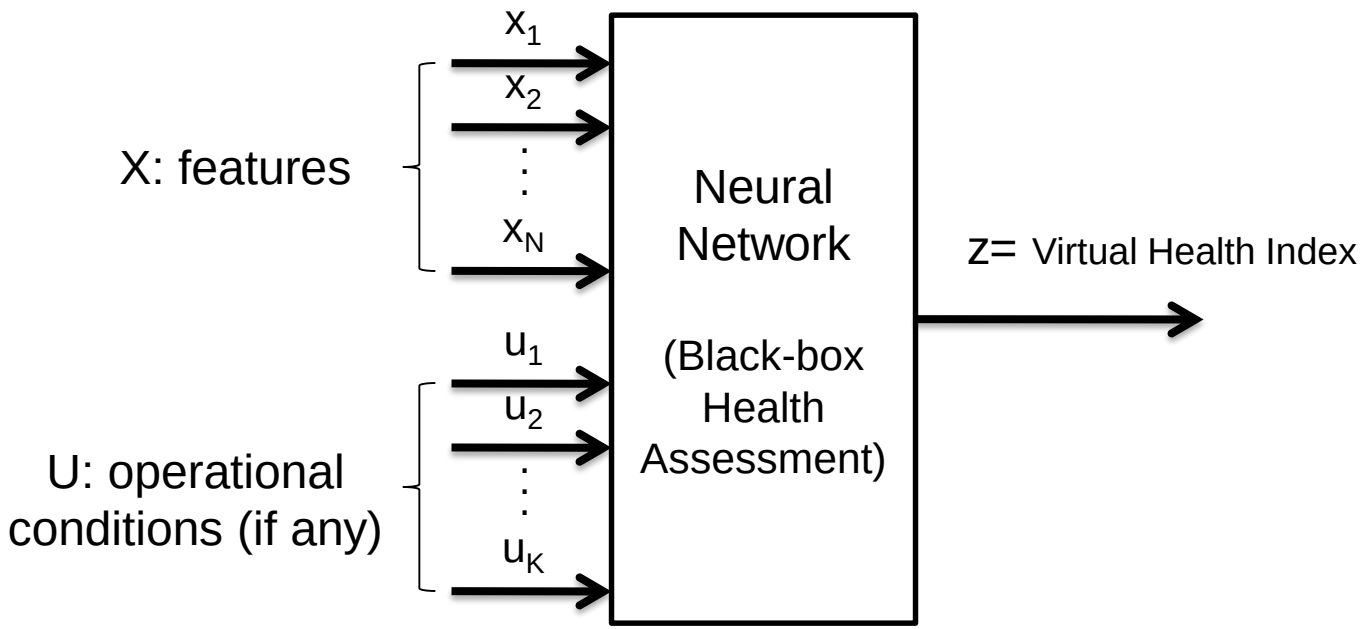
» Scenario 1: Evaluate a physical health index
- A physical property of fault / degradation is measurable in offline experiments
- Online condition data are used to estimate the physical property
- E.g. tool wear, which is quantifiable by the cutter geometry
- E.g. crack length on a turbine blade
- Training: use all the measurements {(X, U, z)}



Neural Network for Health Assessment

» Scenario 2: Evaluate a virtual health index

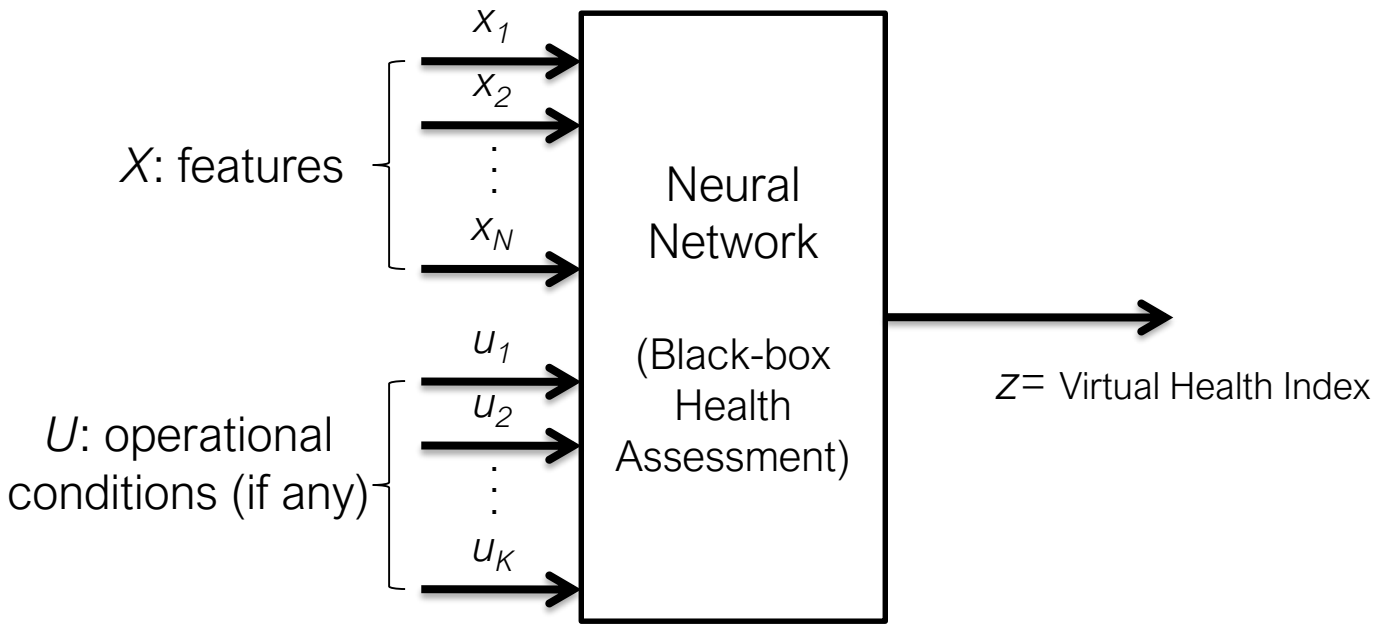
- No physical properties of fault / degradation are measured
- Condition data from normal and faulty condition are available
- Need to quantify the current health condition through a certain (virtual) health index
- Training: normal data (early life data) $\{(X, U, z)_{normal}\}$, with z=1
  faulty data (end life data) $\{(X, U, z)_{faulty}\}$, with z=0



Neural Network for Health Diagnosis

» Scenario 3: Evaluate a class label (classification)

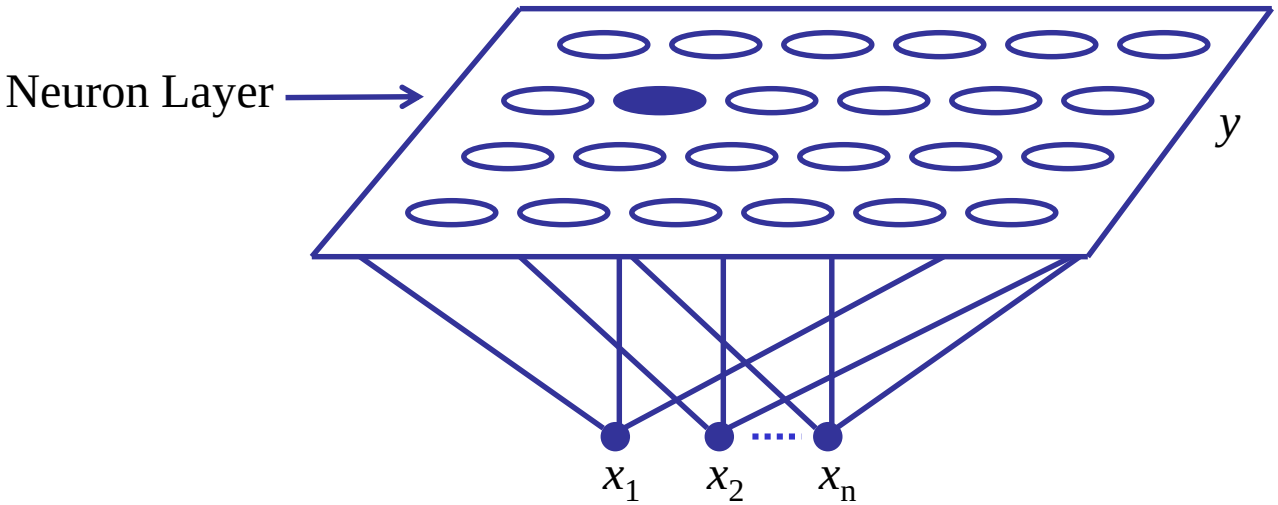
- Data from system in different conditions (e.g. healthy, fault 1, fault 2 etc.) is available.
- There is a need to identify the fault type or the faulty component within a system
- Training: portion of data from each class $\{(X, U, z)_{normal}\}$, with z=1(or [1,0,0])
  fault 1 data $\{(X, U, z)_{fault\ 1}\}$, with z=2(or [0,1,0])
  fault 2 data $\{(X, U, z)_{fault\ 2}\}$, with z=3(or [0,0,1])



Self-Organized Maps (SOM)

Kohonen, 1982

Self-organized feature maps use competitive learning in a network whose neurons are arranged in a 2-dimensional grid.



$r_i =$ position of unit i on the grid.

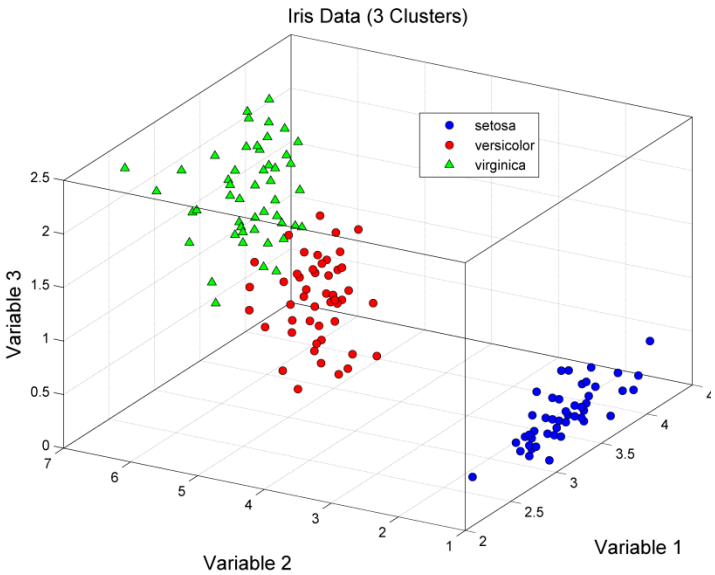
$w_i = [w_{i1} \quad w_{i2} \quad . \; . \; . \quad w_{in}]^T$

SOM Map

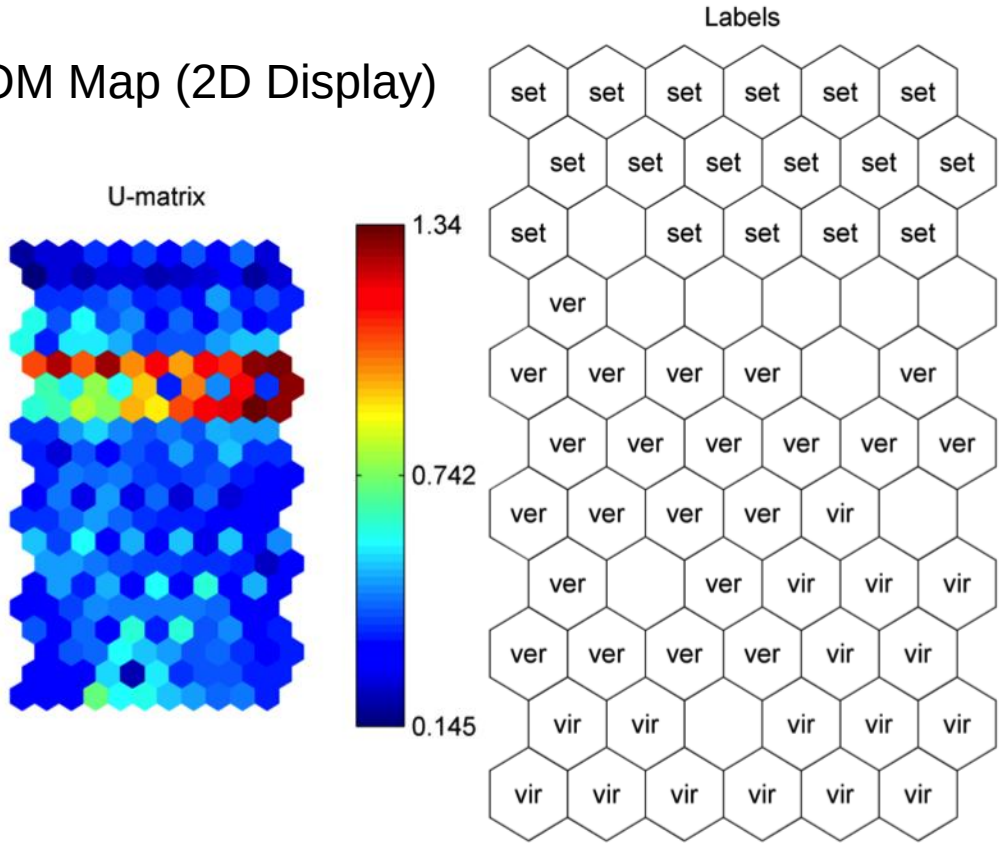
» Visualization of multidimensional data in a 2-D display

» Unsupervised and supervised Clustering

Original Data

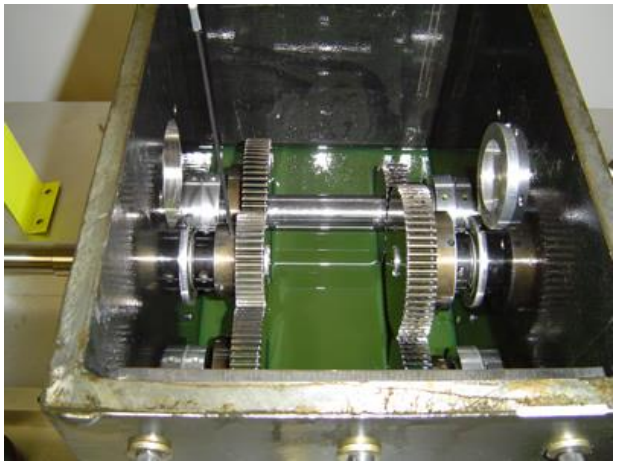


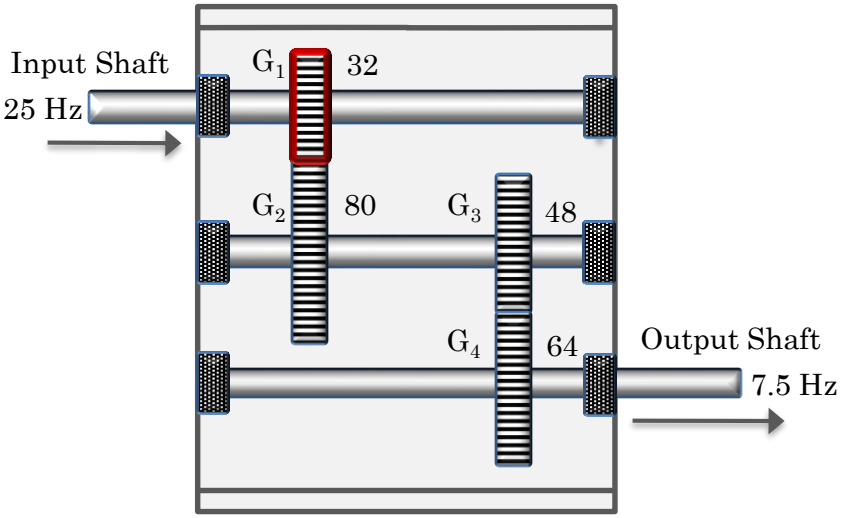
SOM Map (2D Display)



Example: Gearbox Fault Diagnosis

Tekniker
Generación y aplicación de la tecnología y el conocimiento
ik4 research alliance



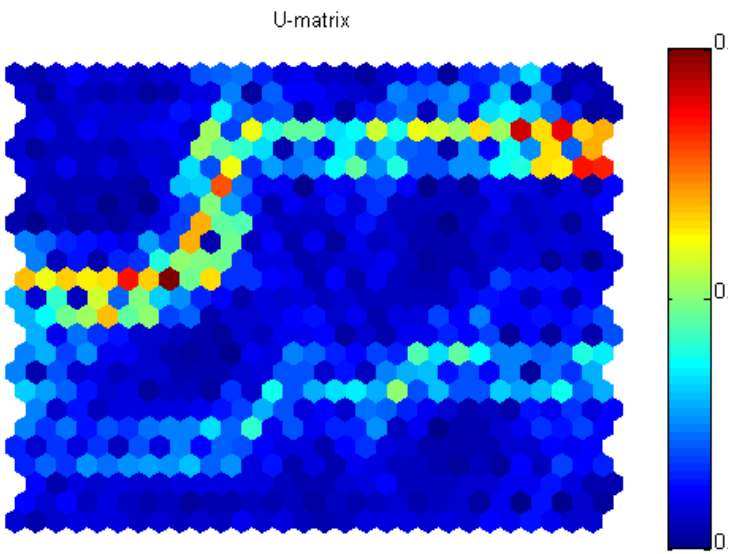


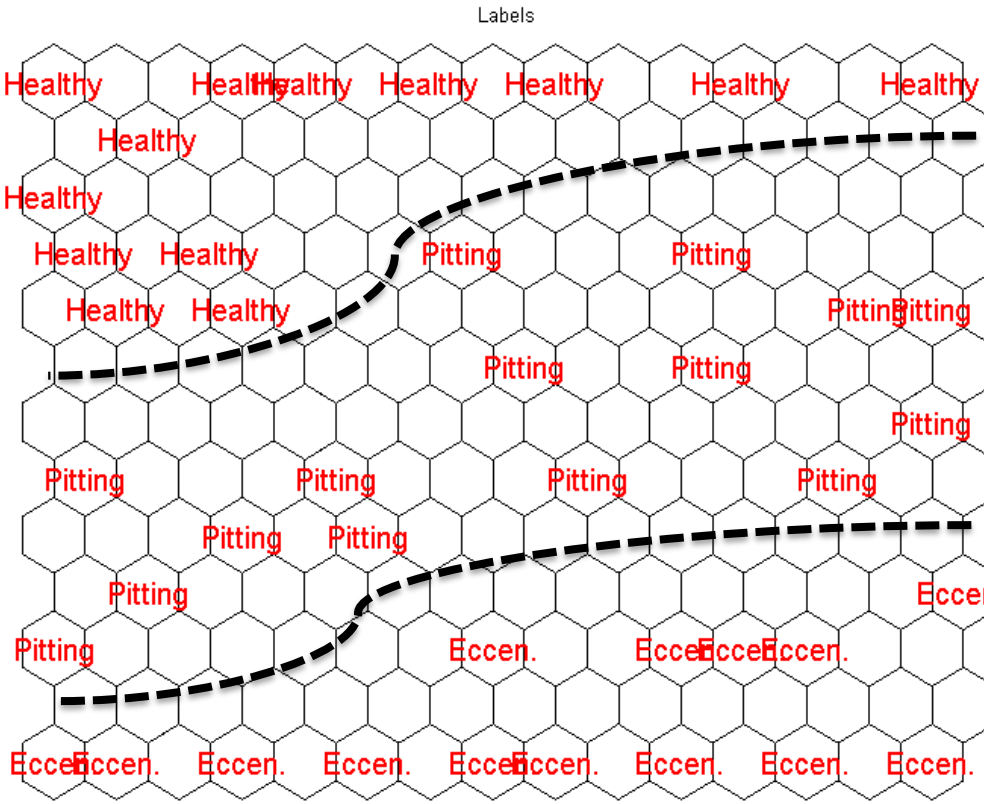
» Faults induced in Gear G1: Eccentricity and Pitting

Fault Classification Using Self-Organizing Map (SOM)

» Fisher criteria was used to select the features that could best discriminate between the three classes of healthy, pitting defect and eccentricity defect.

» SOM could successfully diagnose the faults with 100% accuracy.

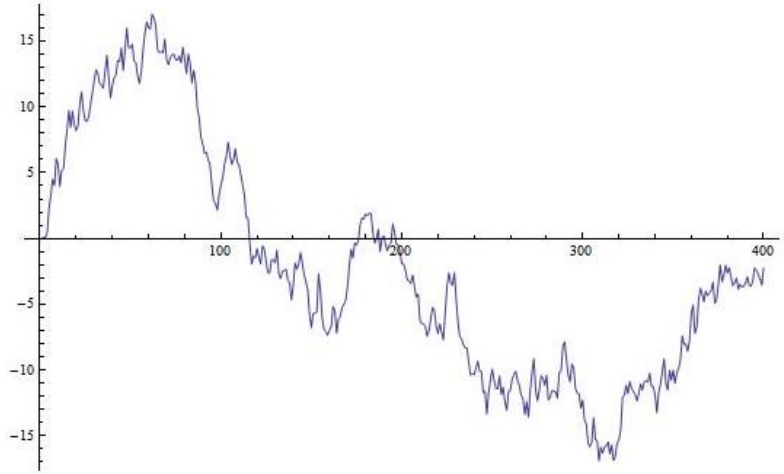




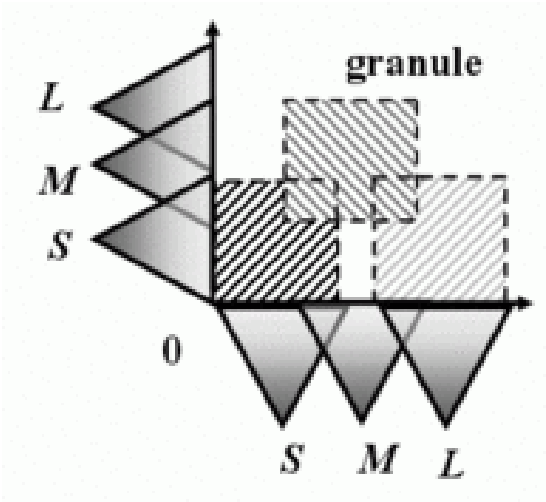
Prognostics Approaches

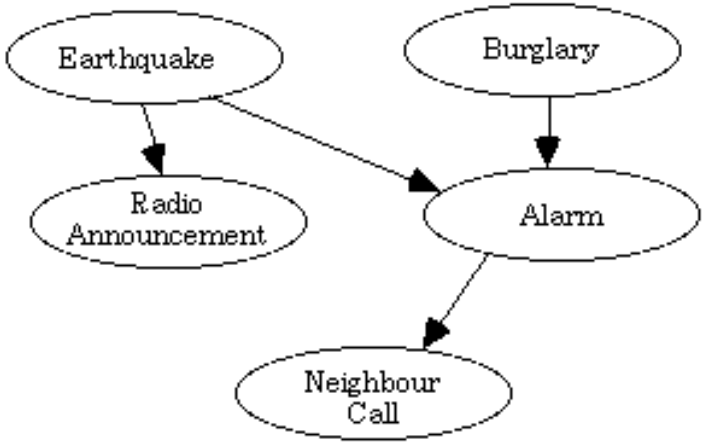
Performance Diagnosis and Prediction

Autoregressive Moving Average (ARMA)

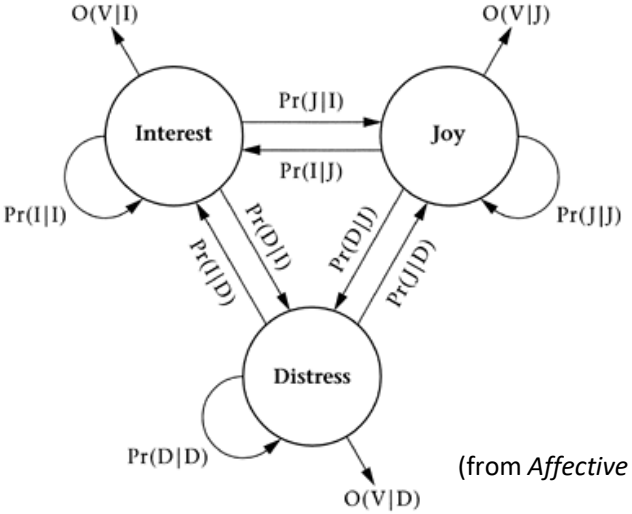


Fuzzy Logic





Bayesian Belief Networks

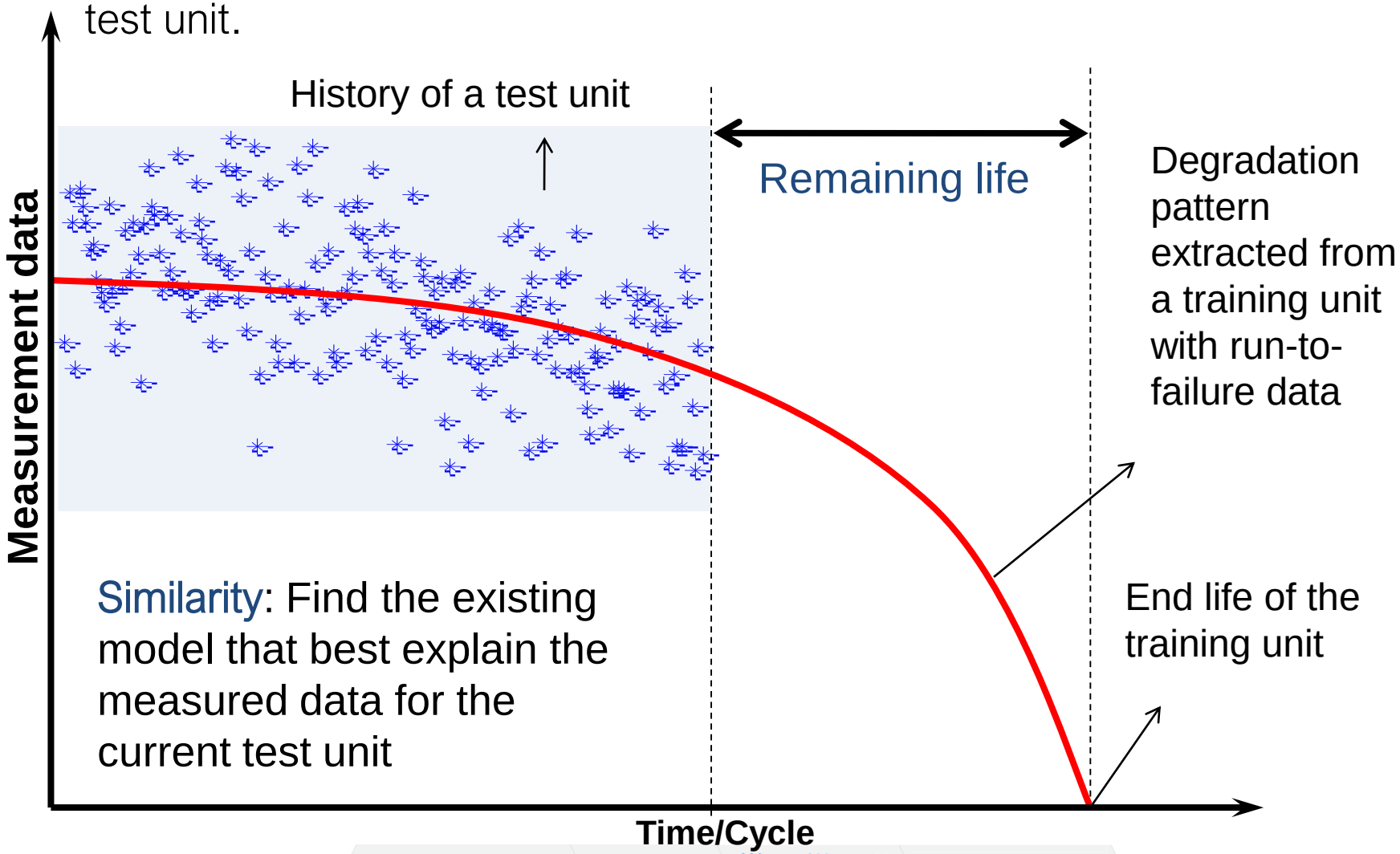


(from *Affective Computing*)

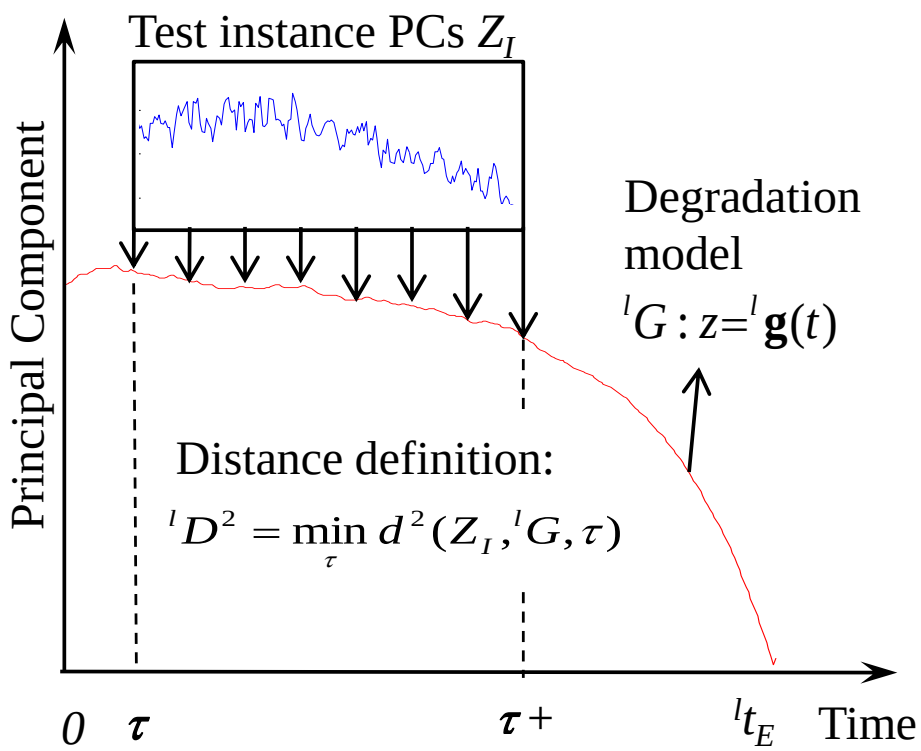
Hidden Markov Model

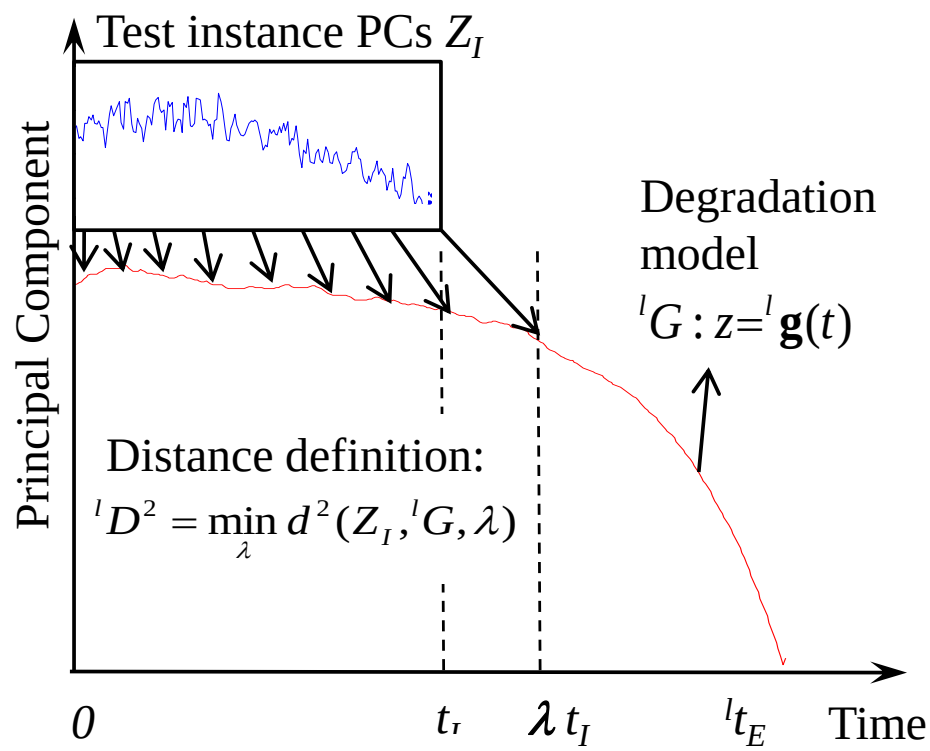
Trajectory Similarity-based Prediction (TSBP)

» Philosophy of Instance-base approaches: the RUL of a test unit can be estimated by the actual life of a training unit which behaves similar to the test unit.



Similarity Evaluation



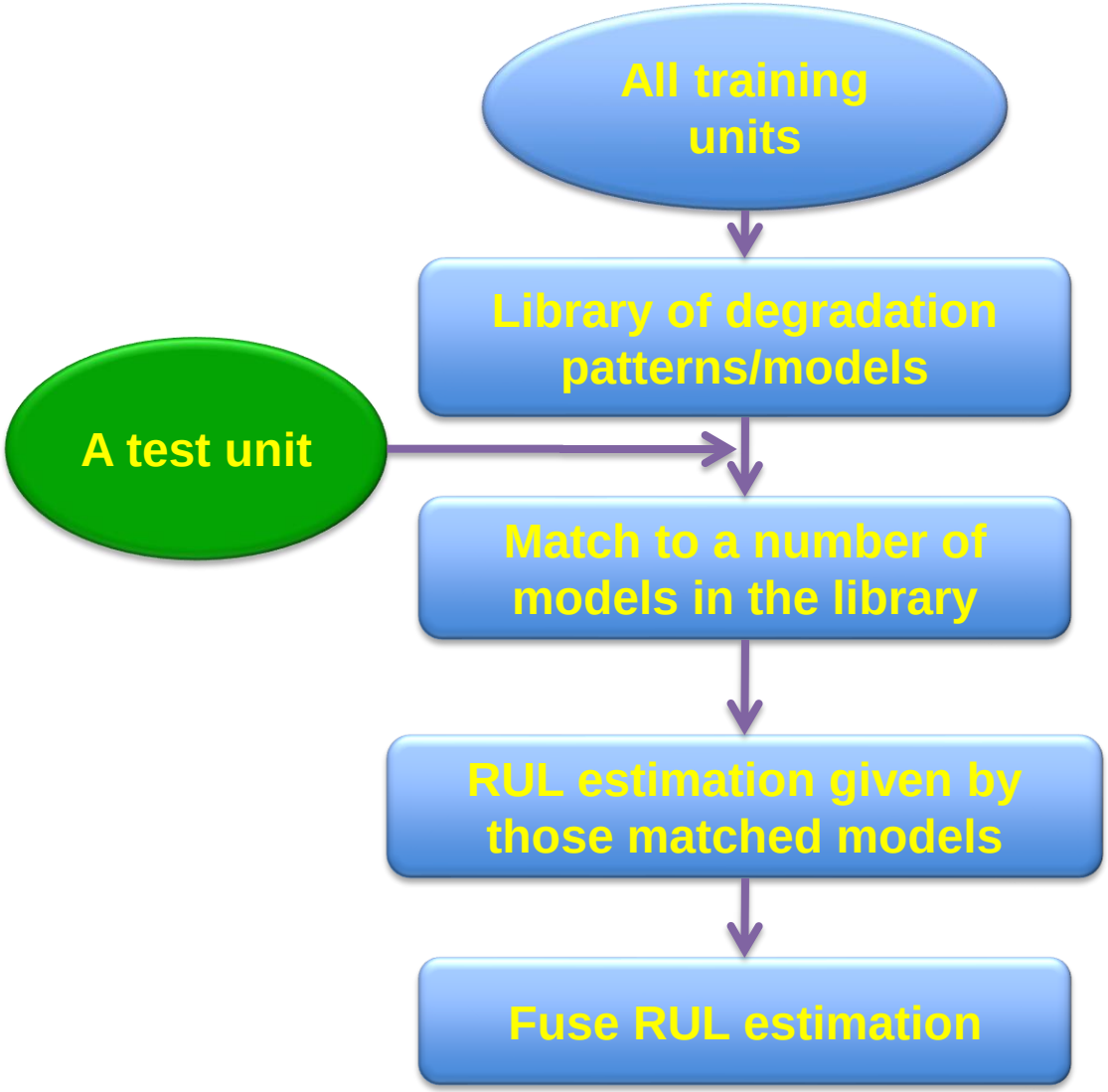


Scenario 1: Minimal Euclidean Distance with Time Lag

Scenario 2: Minimal Euclidean Distance with Degradation Acceleration

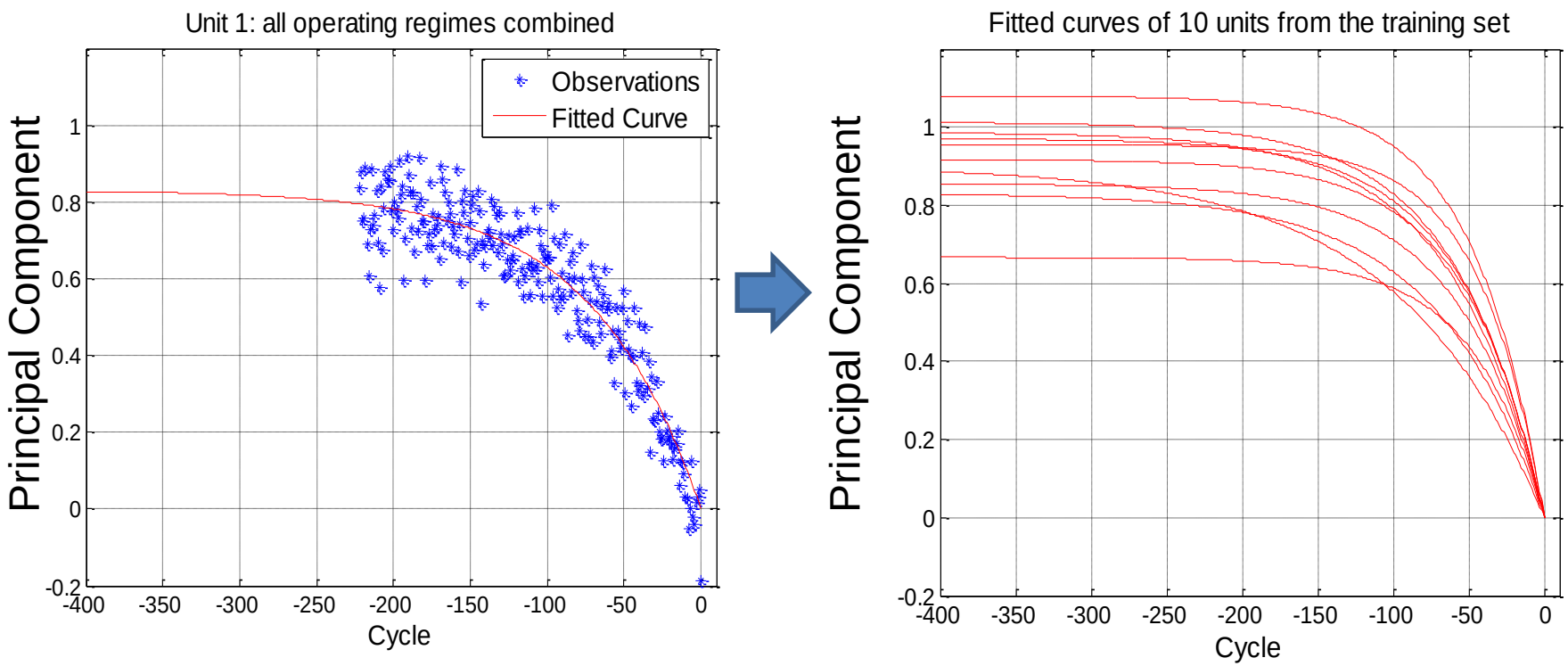
Scenario 3: Minimal Euclidean Distance with Time and Degradation Acceleration

Overview of TSBP



Assumption:
The training data covers a representative set of units of the system/ component

Degradation Trajectory Abstraction

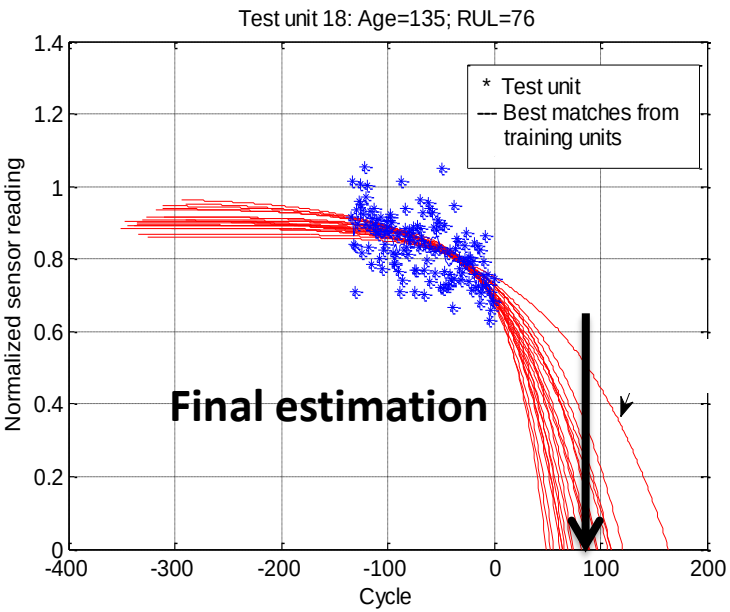


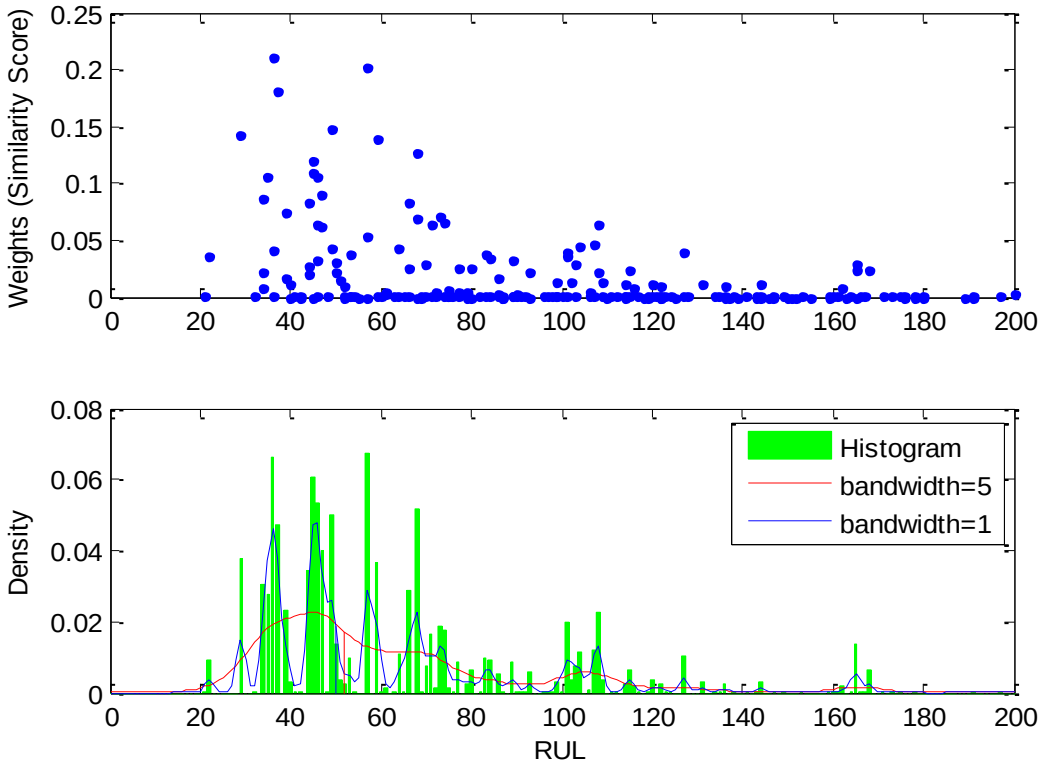
» Extract degradation trajectory by smoothing the PC data

» Create a library of degradation patterns/models $\{^{l}G\}$

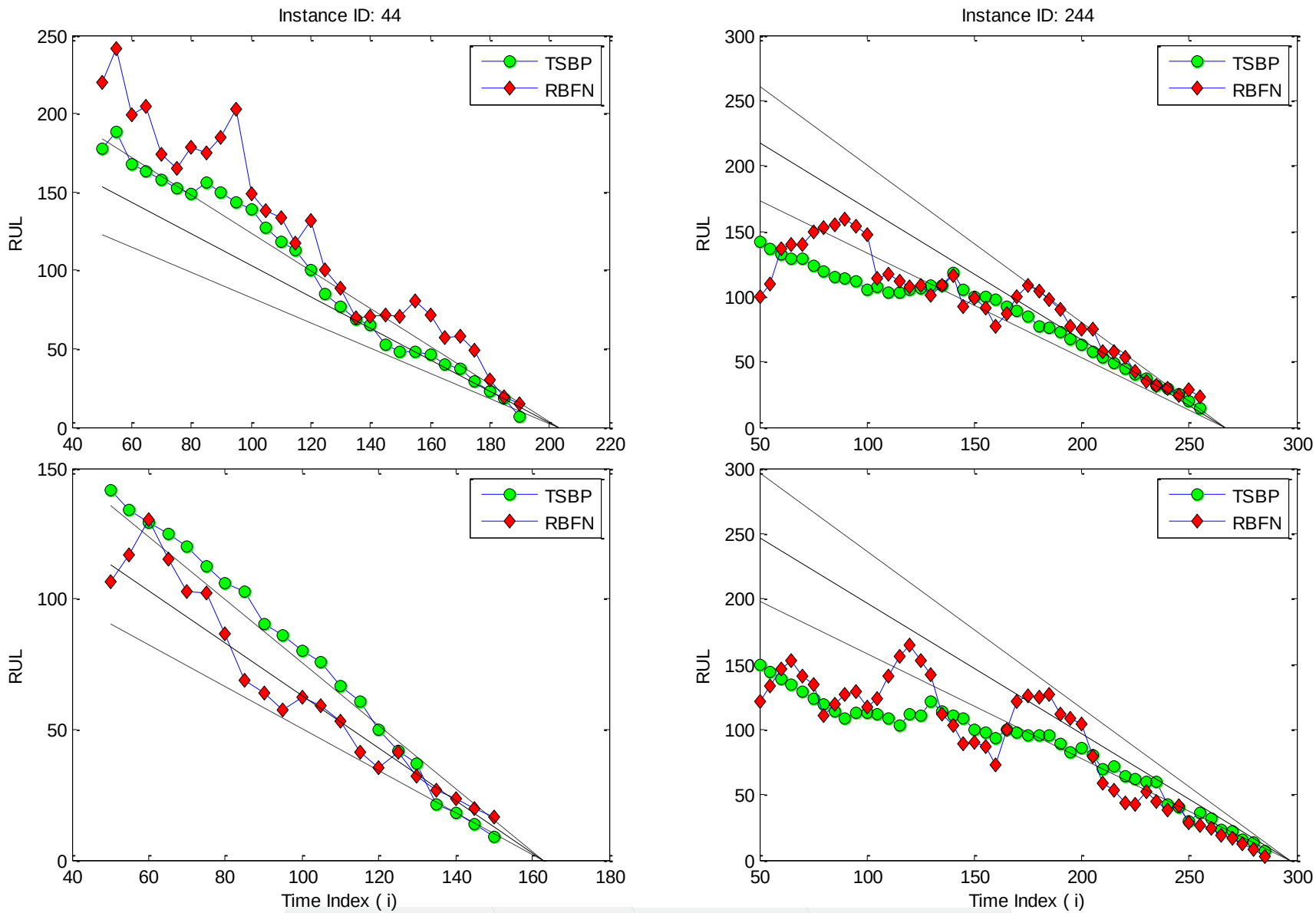
Model Aggregation

» An RUL ${}^{l}r_{I}$ with similarity score is given by each of the l models,





Comparison with Neural Network based Prediction



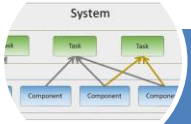
Test Methods, Verification and Validation, and Example Case Studies

DEVELOPING MEASUREMENT SCIENCE…

Research efforts are aimed at developing *Measurement Science Products* for robust sensing, diagnostics, prognostics, and control that enable manufacturers to respond to planned and un-planned performance changes. These products are aimed to be device-agnostic and broadly applicable within the manufacturing community

Standards and Guidelines

Reference Datasets and Software Tools

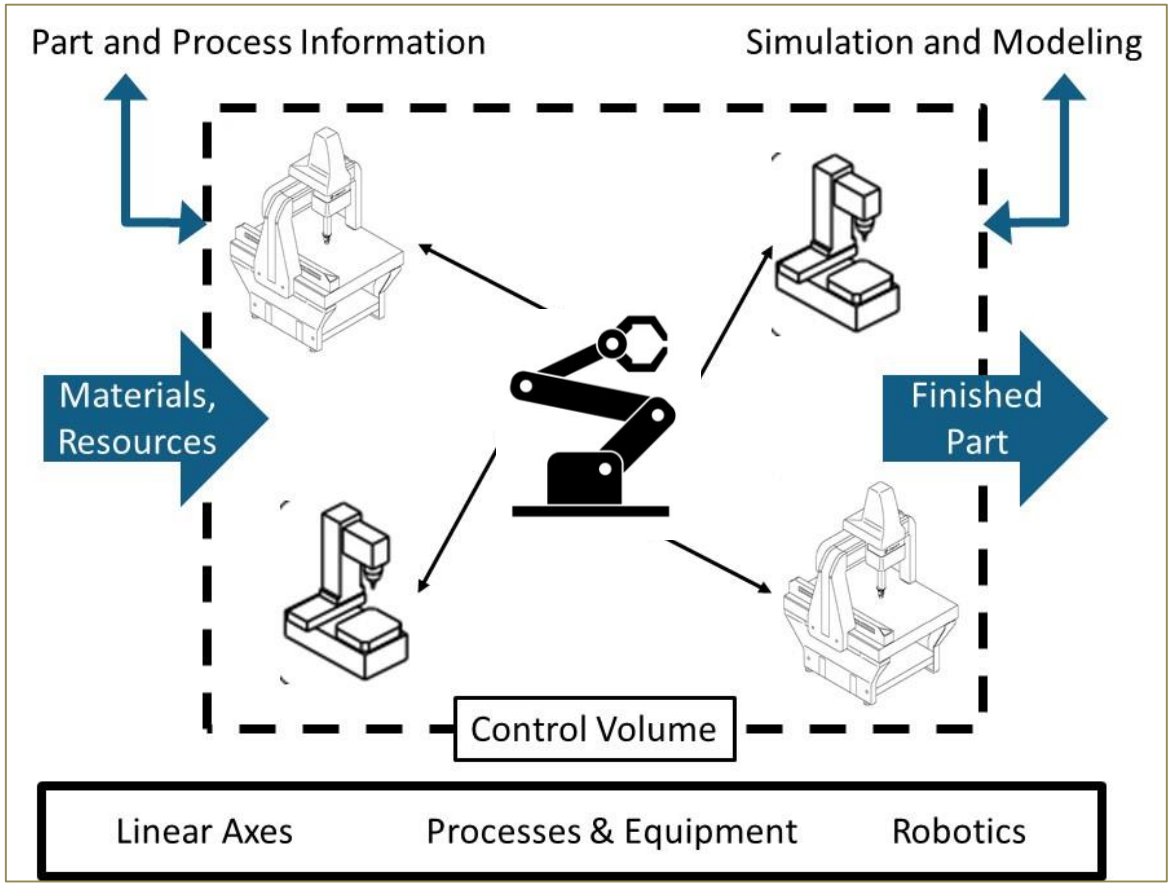


Test Methods & Performance Metrics

Use Cases and Test Scenarios

Roadmaps & Case Studies

USE CASE – DYNAMIC CONTROL of a PRODUCTION CELL



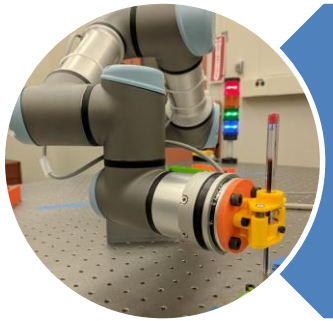
AREAS of IMPACT

- Planning and scheduling support
- Maintenance planning & spare part provisions
- Request for proposals
- Resource budgeting
- Workforce augmentation
- Automation

RESEARCH LEVELS

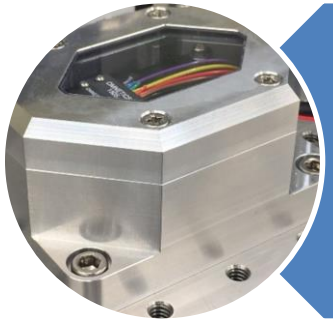
Manufacturing Process and Equipment Monitoring

- System-Level Research
- Smart Manufacturing Systems Test bed



Health and Control Management for Robot Work Cells

- Work Cell-Level Research
- PHM for Robot Systems Lab/Test bed



Machine Tool Linear Axes Diagnostics and Prognostics

- Component-Level Research
- Linear Axis Test bed & ‘Shops’ Machine Tools

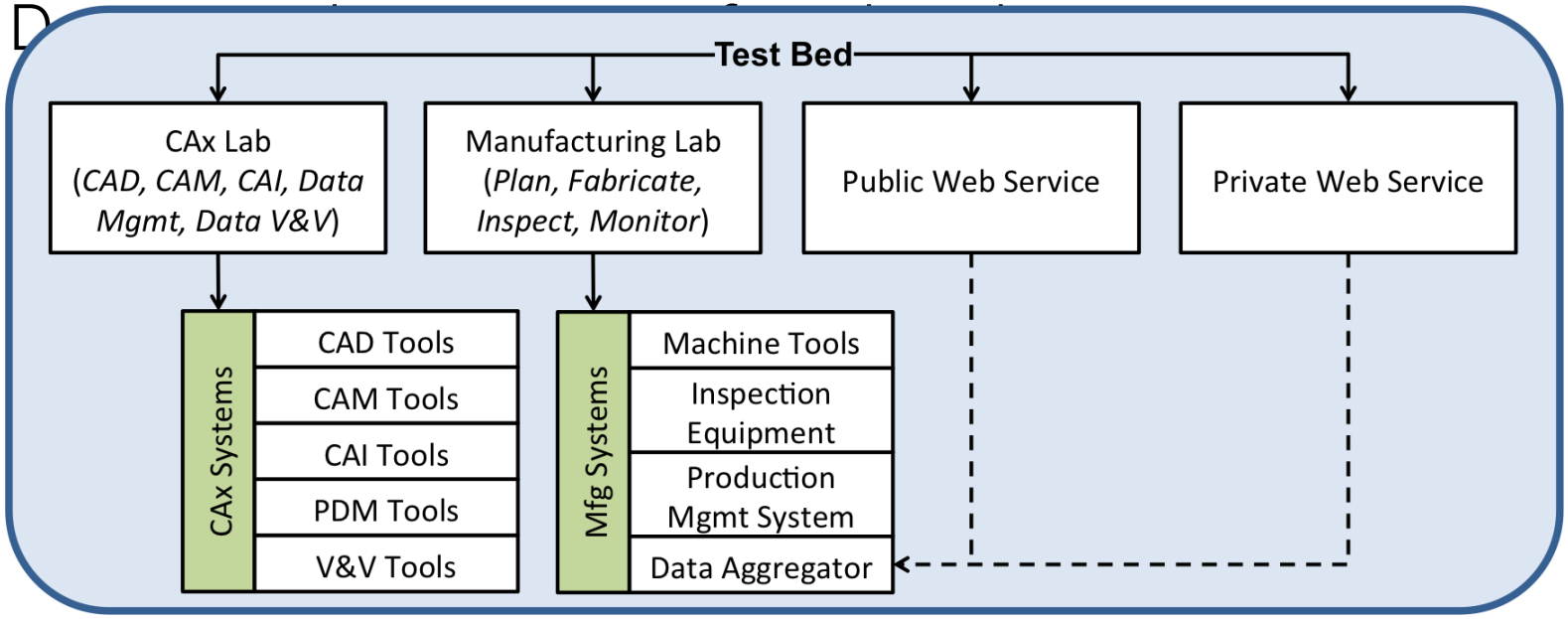
MANUFACTURING PROCESS and EQUIPMENT MONITORING

- Research Objective - Coordinate and control activities of a manufacturing system based on the measured and predicted capability of the system and its components and additional part and process information
- Key Output to Date
  - Designed and implemented architecture to collect data from shop-floor equipment and systems
  - Developed rules engine to identify key events in manufacturing system based on device-level data that informs system control (e.g., planning, scheduling, routing)
  - Proposed enhancements to MTConnect standard to support part and process modeling for supplying device-level data to MES/ERP, enabling traceability, and enabling production control and engineering analysis
- Impact
  - Documented architecture and SMS Test Bed implementation guides manufacturing community to deploy smart manufacturing effectively and cheaply
  - Standardized interfaces ease and democratize systems integration
  - Maintenance and PHM may be performed in the context of production scheduling to minimize any planned or unplanned disruptions

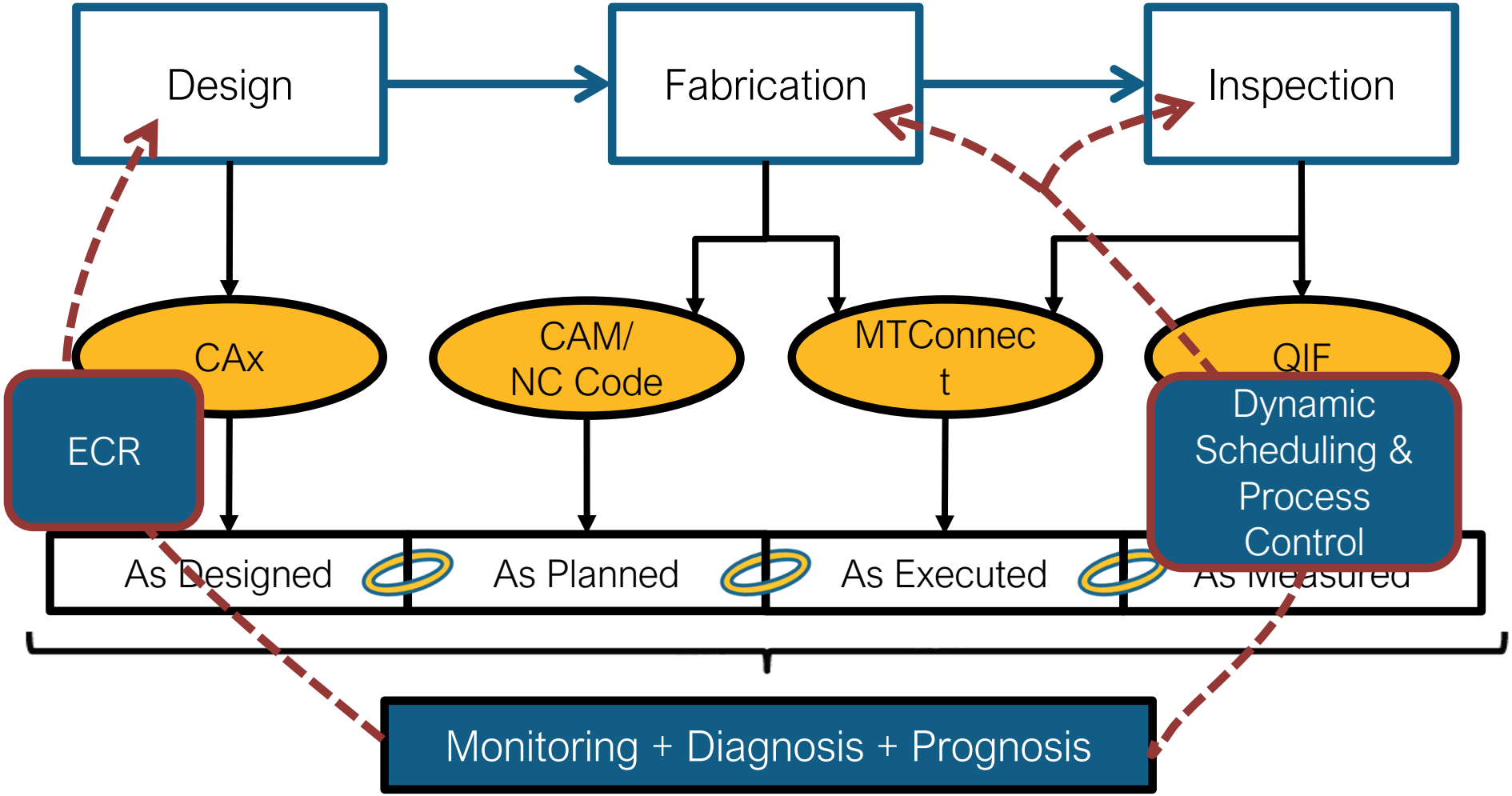
NIST SMART MFG. SYSTEMS TEST BED

Goals:

- Reference architecture and implementation
- Rich source of data for fundamental research
- Physical infrastructure for standards & technology development
- D



DATA COLLECTION and AGGREGATION

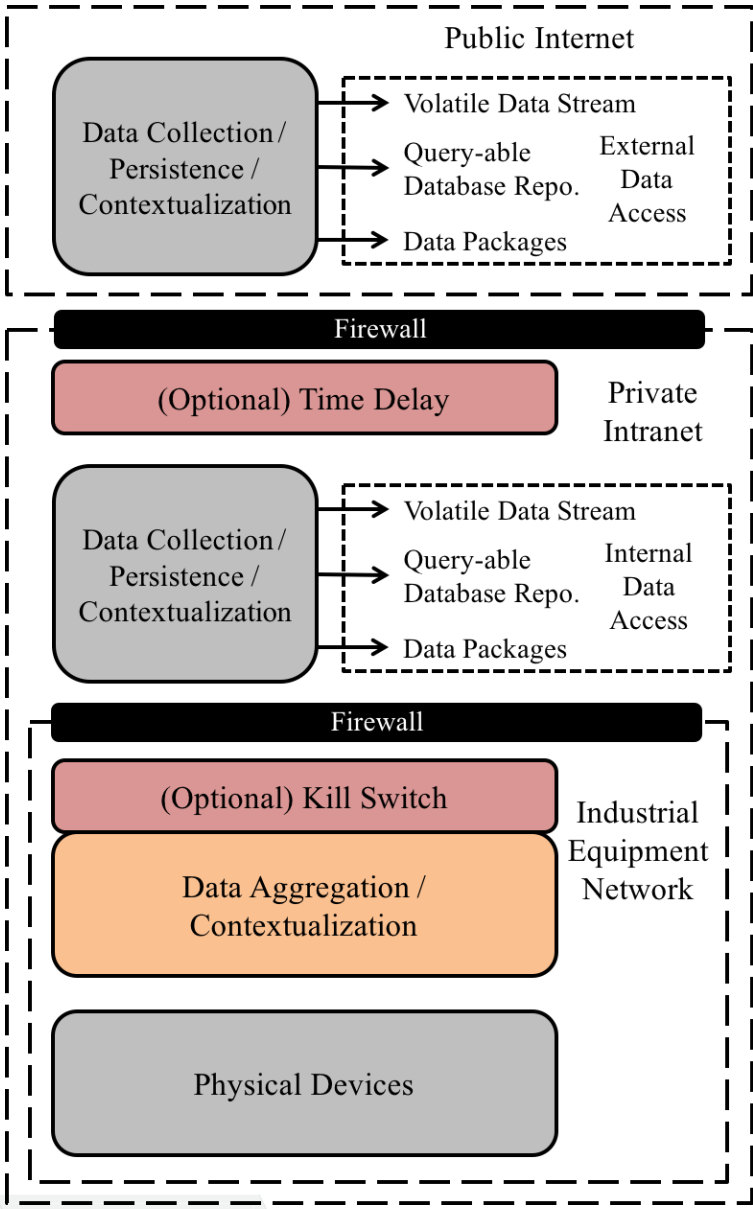


MANUFACTURING RESOURCES

- Equipment:
  - 6x Three-Axis Mill
  - 1x Five-Axis Mill
  - 2x Turn
  - 2x Mill-Turn
  - 2x EDM
  - 1x CMM
  - 1x Surface Profilometer
  - Digital Micrometers
- Production Mgmt:
  - ShopTech E2
- Additional Sensors:
  - Accelerometers
  - Imaging
  - Powermeters
  - Thermocouples

MANUFACTURING DATA ARCHITECTURE

- Designed as a four-tier architecture
- Implemented across three networks
- Provides segregated access to internal and external clients

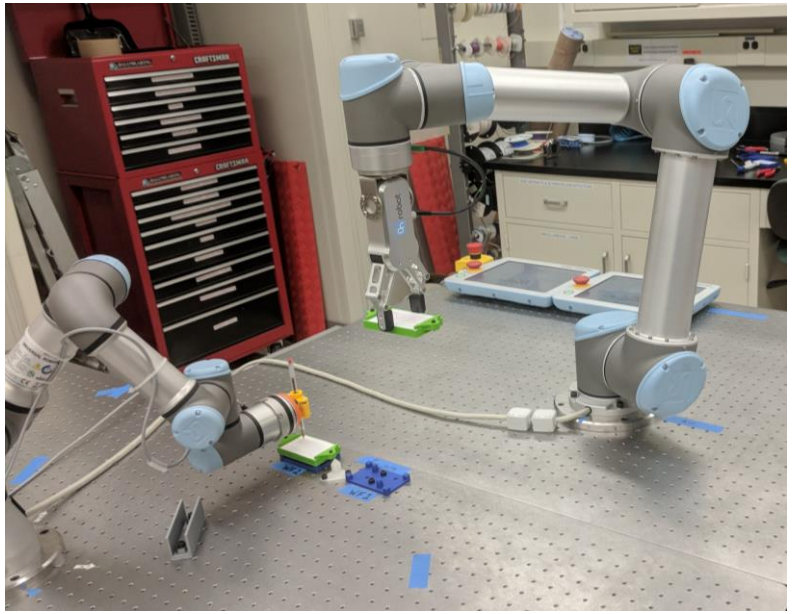


DEMONSTRATIONS

- » Virtual Data Stream (VDS)
  - Current
  - Sample
  - Probe
- » Query-able Data Repository (QDR)
- » Technical Data Packages (TDP)
- » VIMANA ENRICH

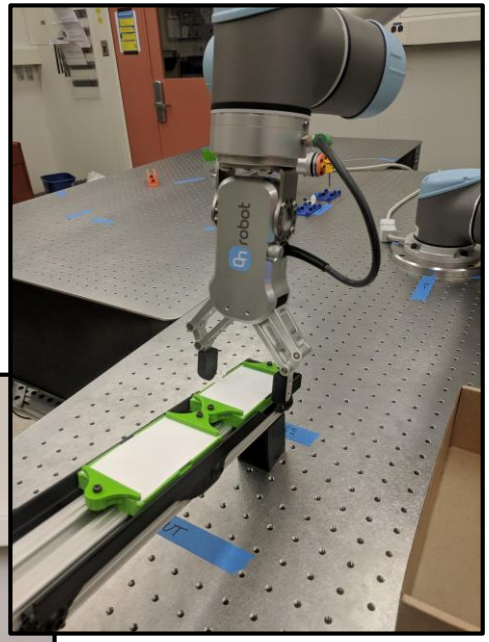
HEALTH and CONTROL MANAGEMENT of ROBOT SYSTEMS

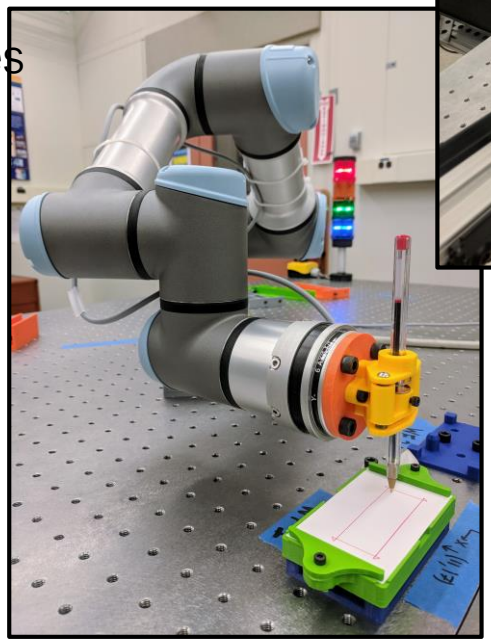
- Research Objectives
  - Develop a quick health assessment methodology to quickly assess a robot's health degradation, including the tool center position accuracy degradation.
  - Verify and validate monitoring, diagnostic, and prognostic technologies implemented in a robot work cell



ROBOTIC WORK CELL – TEST BED

- Problem:
  - Robot work cells are made up of complex combinations of automation hardware and software to perform a specific operation.
  - No existing method to verify and validate PHM on robot work cells
- Goal:
  - Develop the necessary measurement science to enable the V&V of monitoring, diagnostic, and prognostic technologies within a manufacturing robot work cell

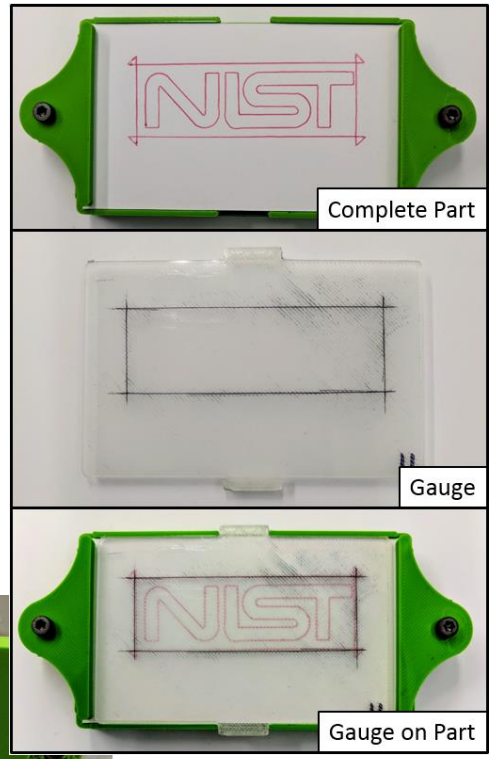


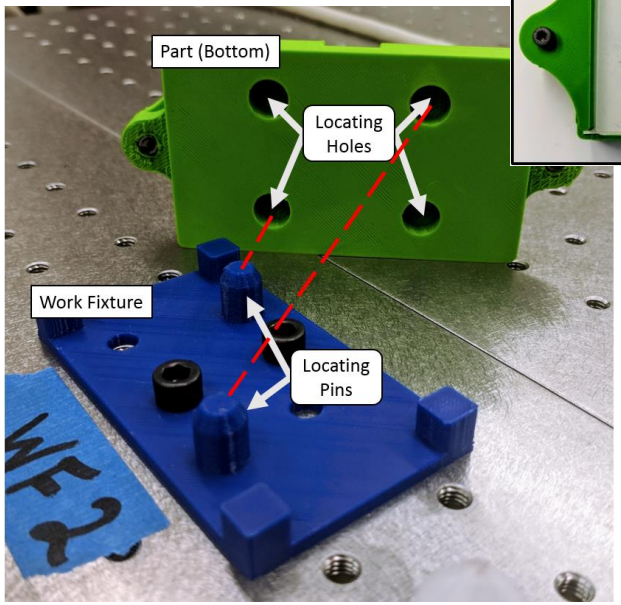


ROBOTIC WORK CELL - USE CASE

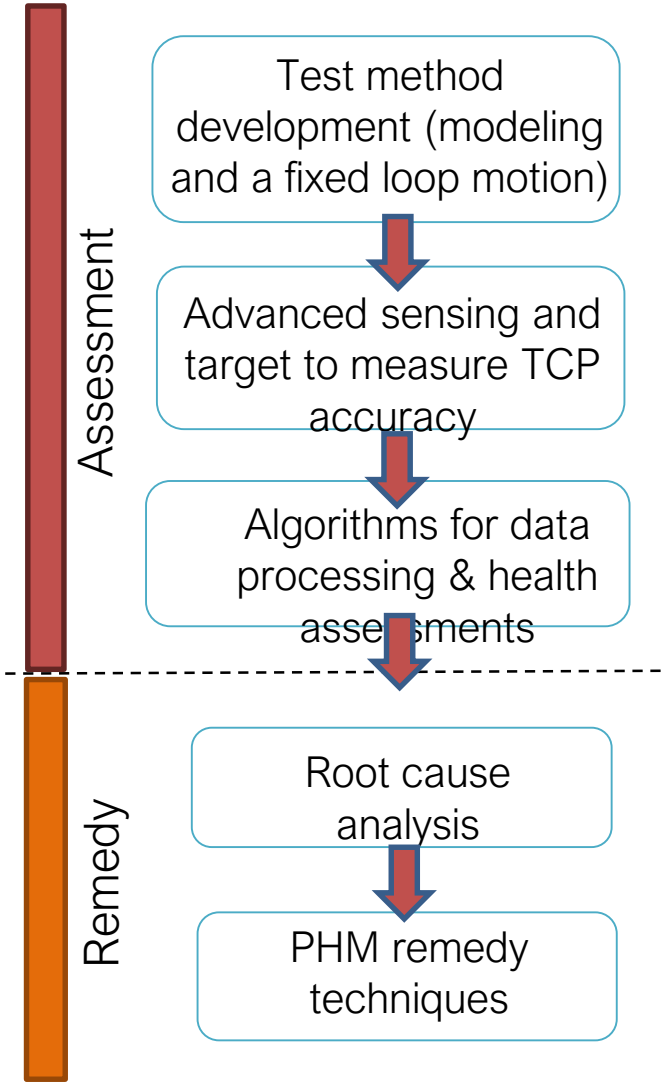
ROBOTIC WORK CELL – PHM METHOD V&V

- Test methods under development to inject degradations into the system
  - Input part quality issues
  - Fixture wear
  - Robot joint deviations from nominal (artificial)
- Reference PHM methods under development
  - Verification of positioning (degradation) of components within work cell





ROBOT – WORKFLOW



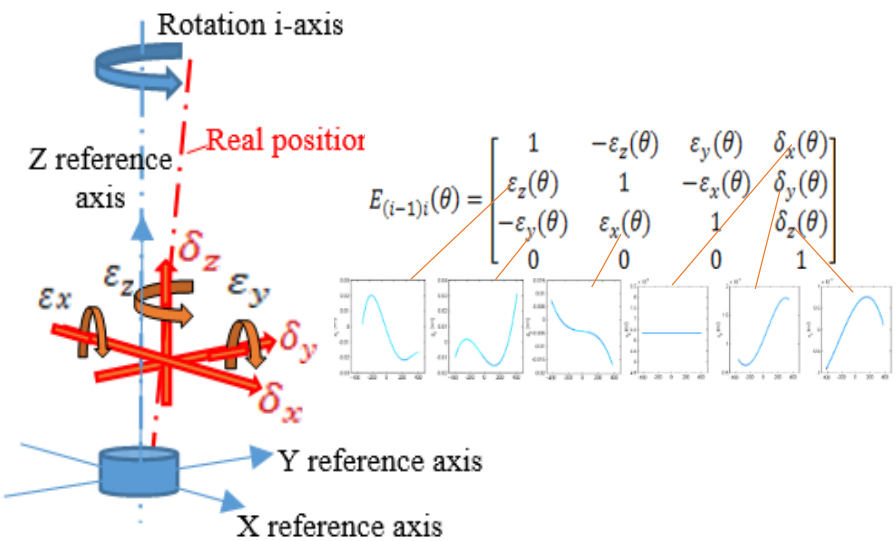
How to measure – design fixed loop motion, that will be performed routinely

7-D measurement system and the special target

Volumetric error assessment in robot workspace

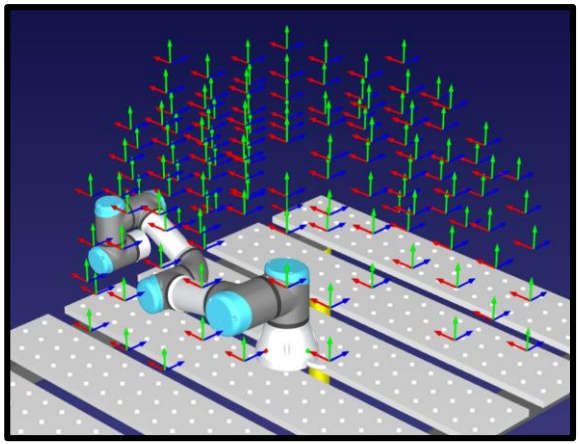
ROBOT – TEST METHOD MODELING

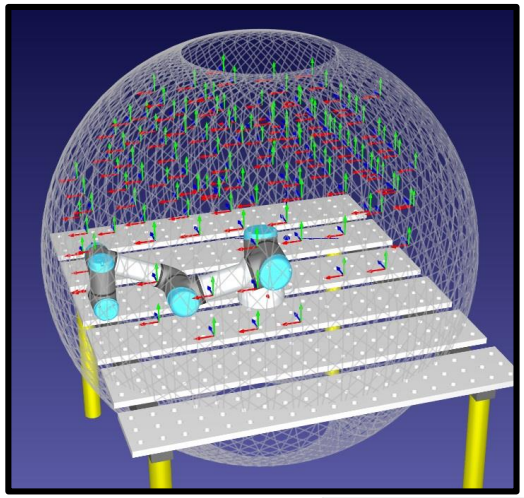
- Traditional modeling methods assume that joint motion is ideal, and the geometric relationships between the joints are constant. Non-geometric errors are present, such as the non-ideal motion of joints, and deflections of the structure and joints due to external loading or gravity, backlash, etc.
- Each tool center position in the Cartesian space could have multiple inverse kinematic solutions; errors change by choosing different solutions.

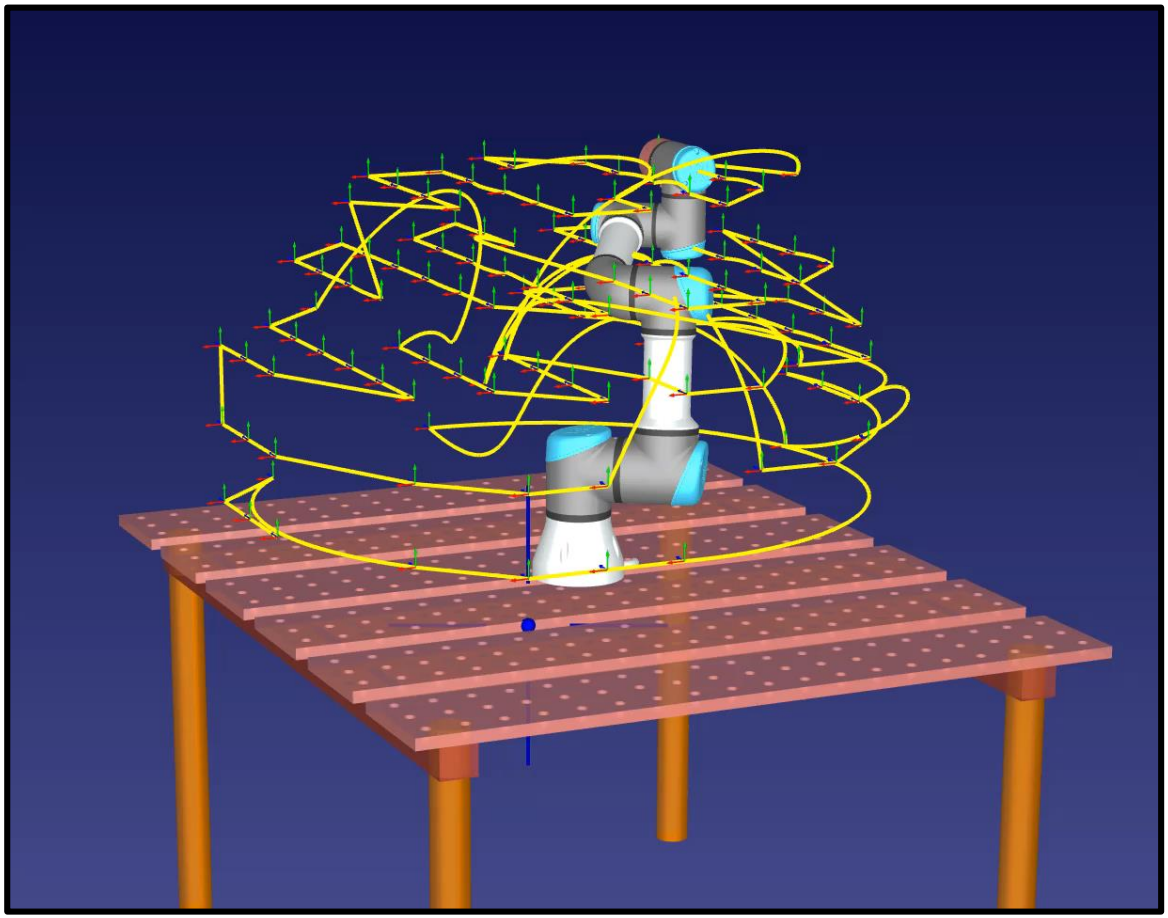


Modeling for the test method

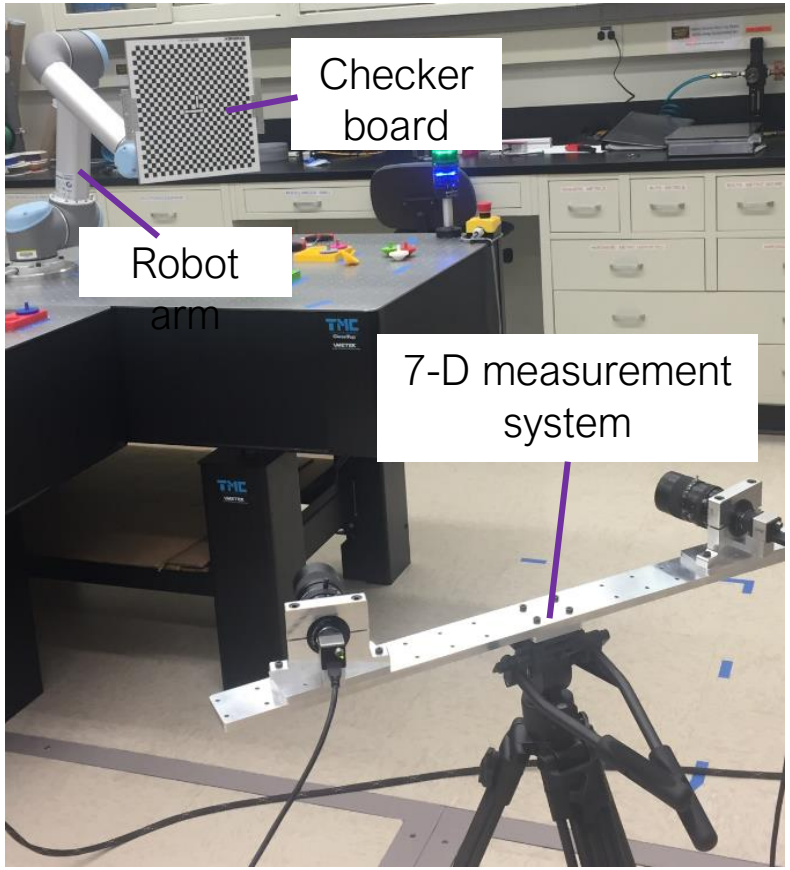
ROBOT – FIXED LOOP MOTION DESIGN

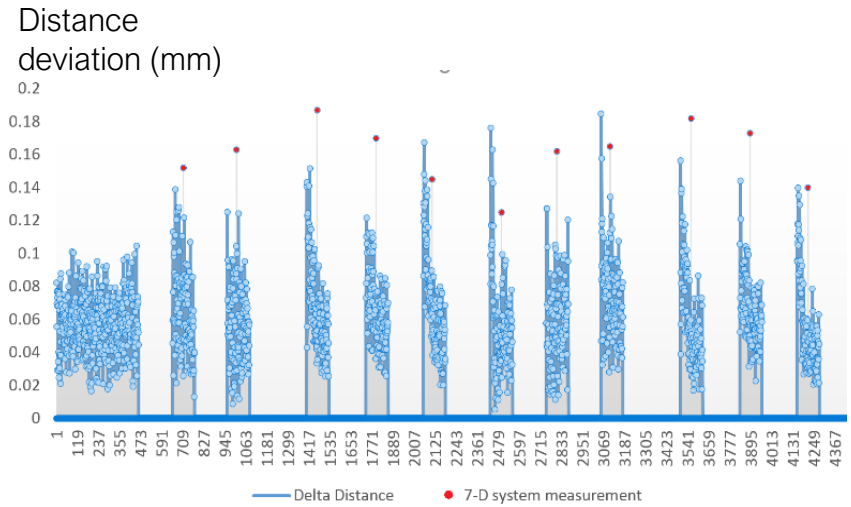






ROBOT – REFERENCE DATA SET COLLECTION and ANALYSIS





TCP deviations: 7-D system measured vs. calculated deviations from controller actual joint positions minus target joint positions

Reference data set URL:
https://www.nist.gov/el/intelligent-systems-division-73500/cognition-and-collaboration-systems/degradation-measurement

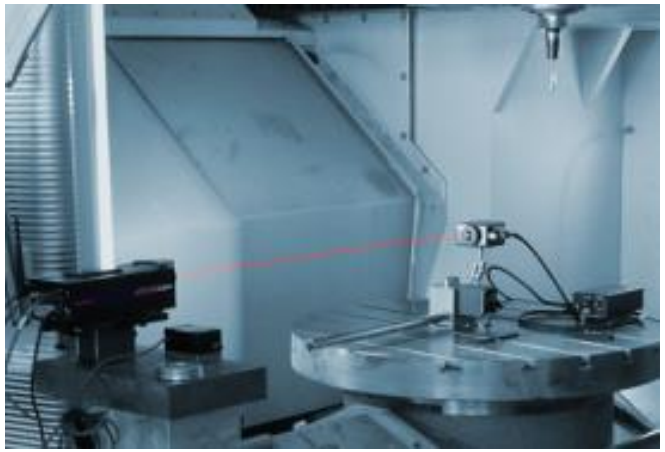
ROBOT – VERIFICATION and VALIDATION EFFORTS

- Verification
  - Using high accuracy checkerboard artifacts for measurement instrument accuracy verification
  - The checkerboard has +/- 0.025mm dimensional accuracy
- Validation
  - Actively in discussions to obtain feedback in a real industrial environment where robots are being used layout composite material to construct 3D parts. NIST's quick health assessment methodology will be tested in the robot work cell with the critical requirements on monitoring robot system performance degradation

LINEAR AXES DIAGNOSTICS & PROGNOSTICS - INDUSTRY CHALLENGE

- "Machine tool health in 5 min?"
  - Faults/failures → 10s of $Billions (> new machines!)
  - Routine tracking of performance can be expensive in time & equipment

Laser → 1-2 days



http://www.apisensor.com/user-story-carter-ats/

Ballbar → 30-60 min

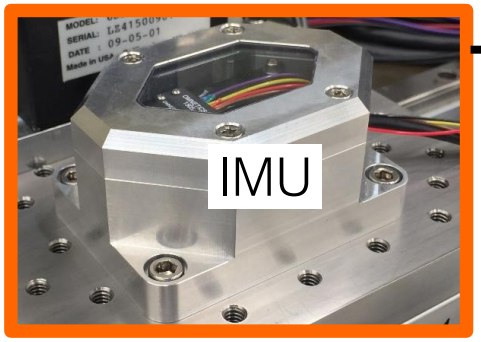
http://www.renishaw.com/en/qc20-w-ballbar-system--11075

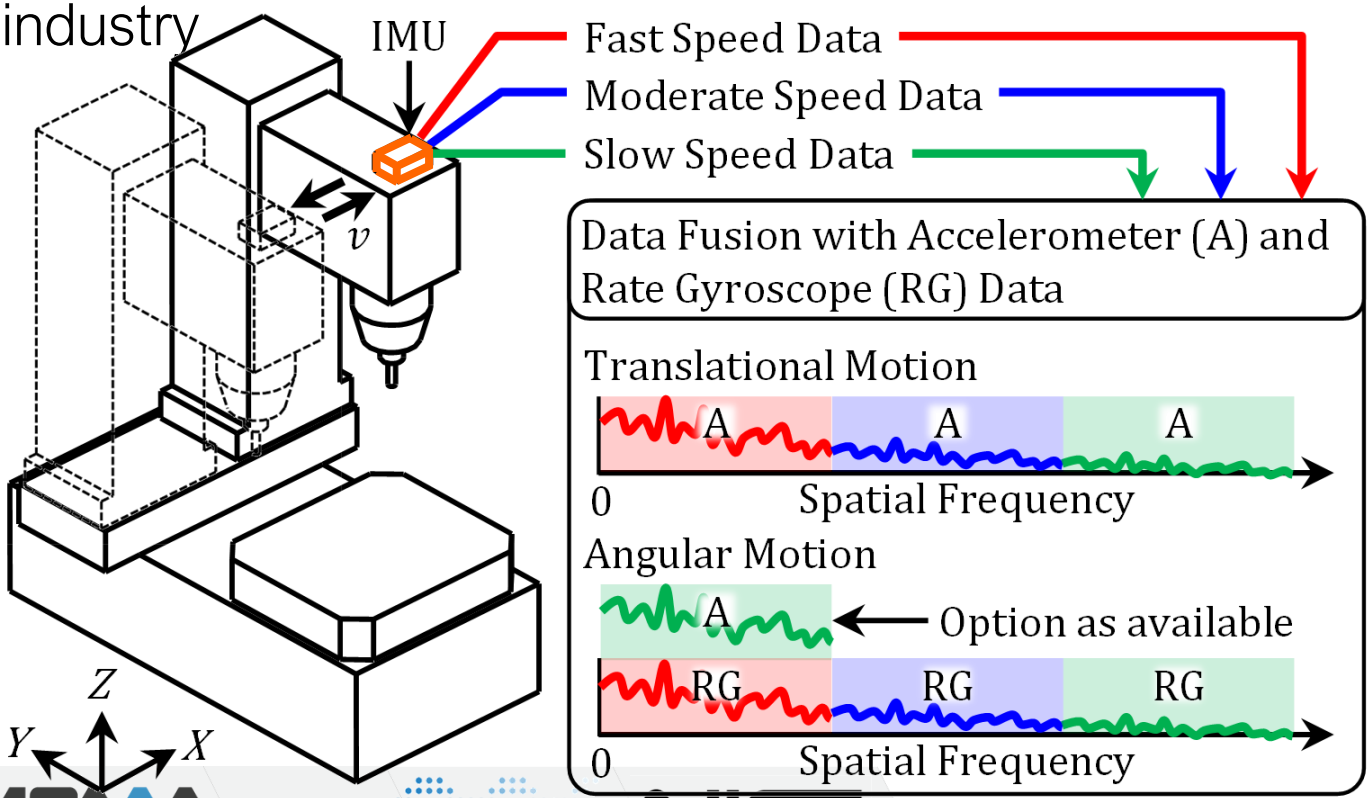
New method? → 5 min

?

TECHNICAL SOLUTION

- Inertial measurement unit (IMU) measures changes in error motions
- 6-DOF
- IMU could answer industry challenge
  - Non-intrusive
  - Quick

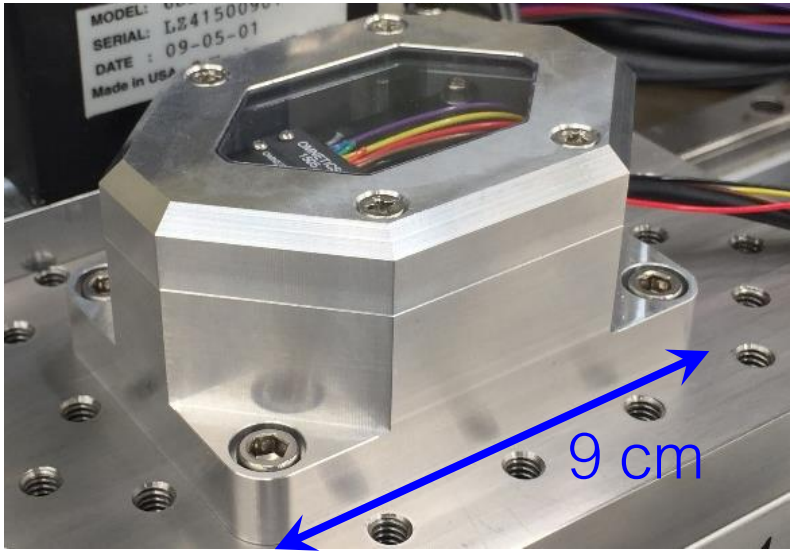


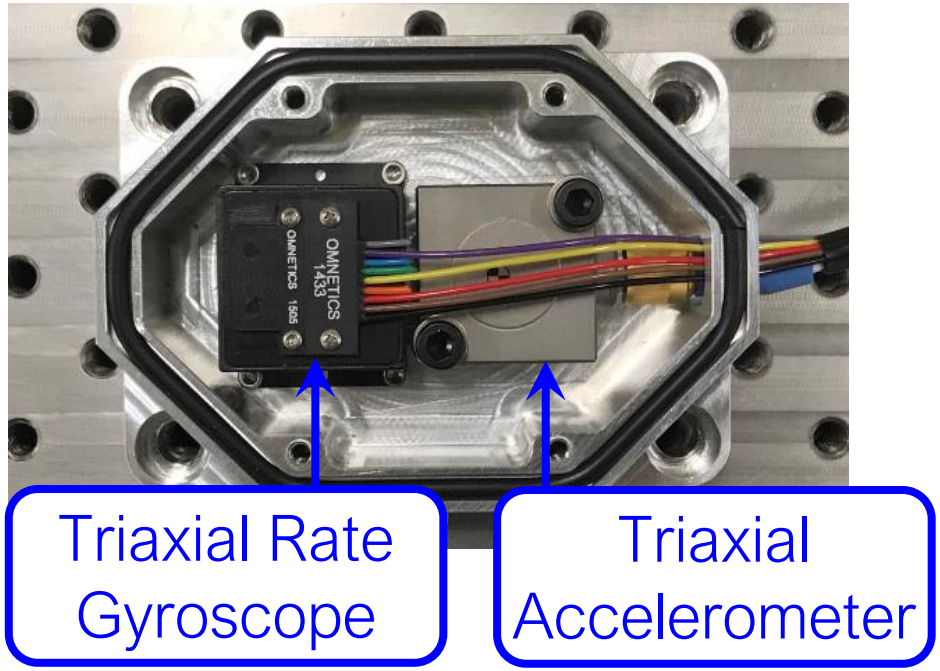


Ref: Vogl G.W., Donmez M.A., and Archenti A. (2016) Diagnostics for geometric performance of machine tool linear axes. *Annals of the CIRP* 65(1):

IMU SENSORS

IMU uses precision MEMS inertial sensors



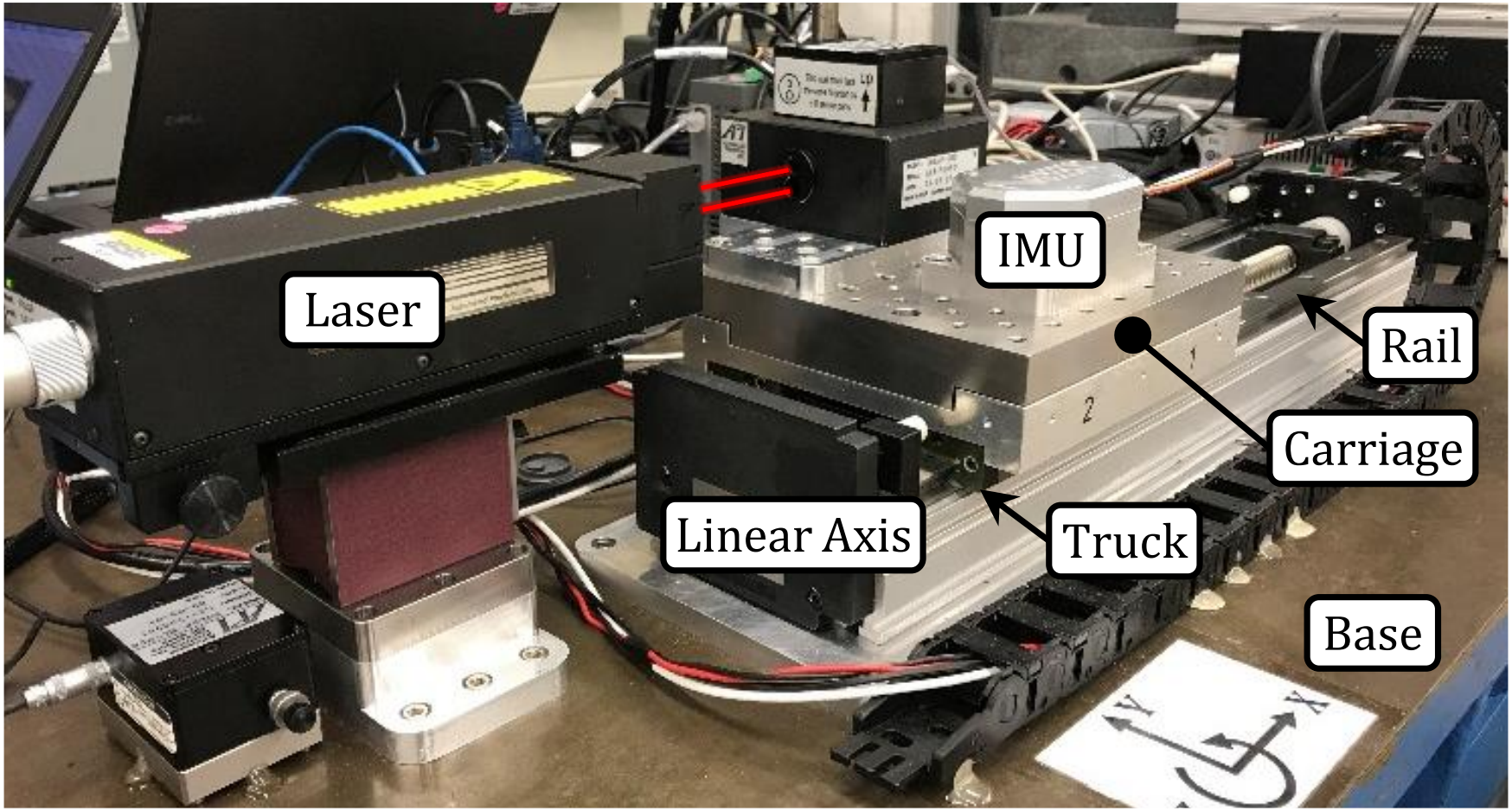


Sensor	Bandwidth[a]	Noise
Accelerometer	0 Hz to 400 Hz	69 (μm/s^2)/√Hz
Rate Gyroscope	0 Hz to 200 Hz	35 (μrad/s)/√Hz

[a] frequencies correspond to half-power points, also known as 3 dB points

LINEAR AXIS TEST BED @ NIST

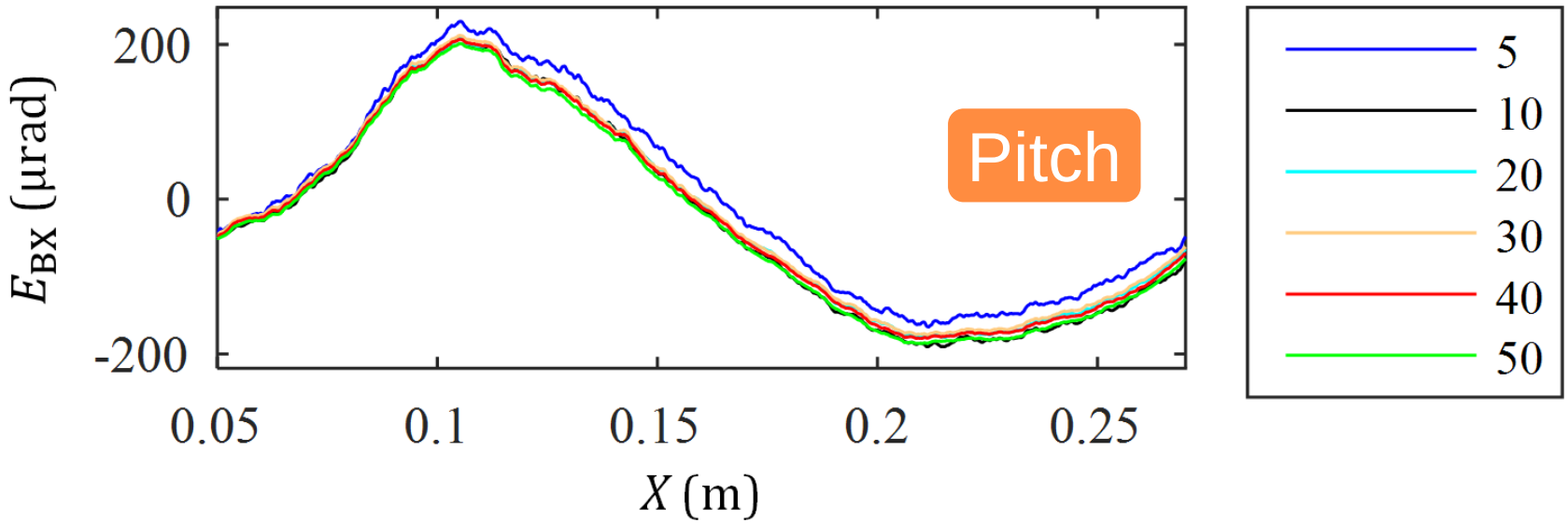
Compare IMU-based results to laser-based results



CONVERGENCE via AVERAGING

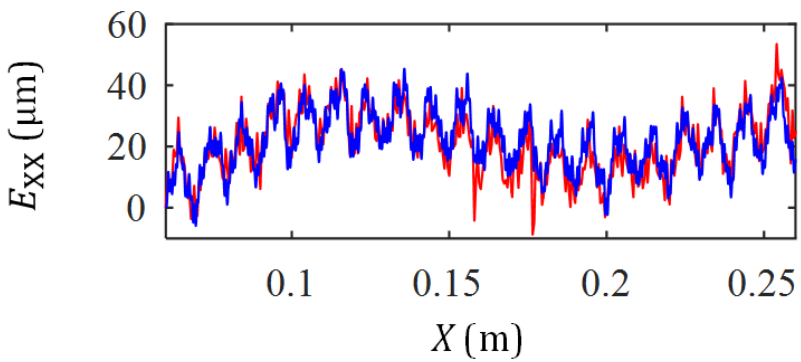
- Results are averaged across runs (5, 10, ..., or 50)
- Convergence within 5 μm or 15 μrad for < 10 runs

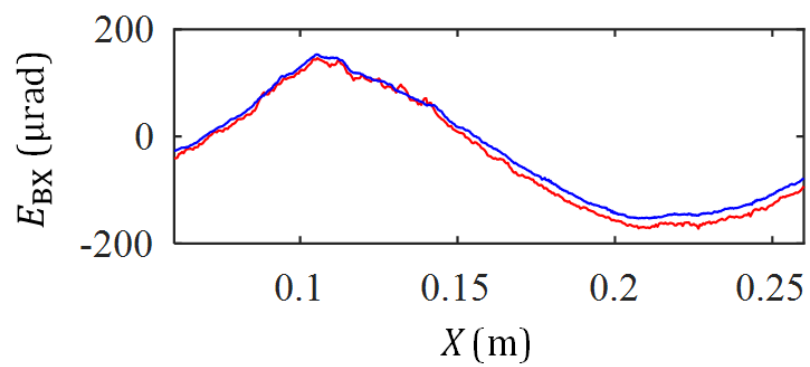
Data is collected over again for same path to achieve convergence

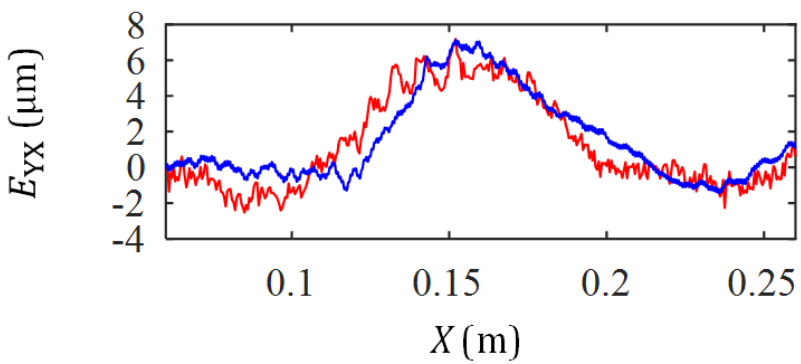


GENERAL RESULTS

- Example: Converged error motions from IMU match those of laser-based system



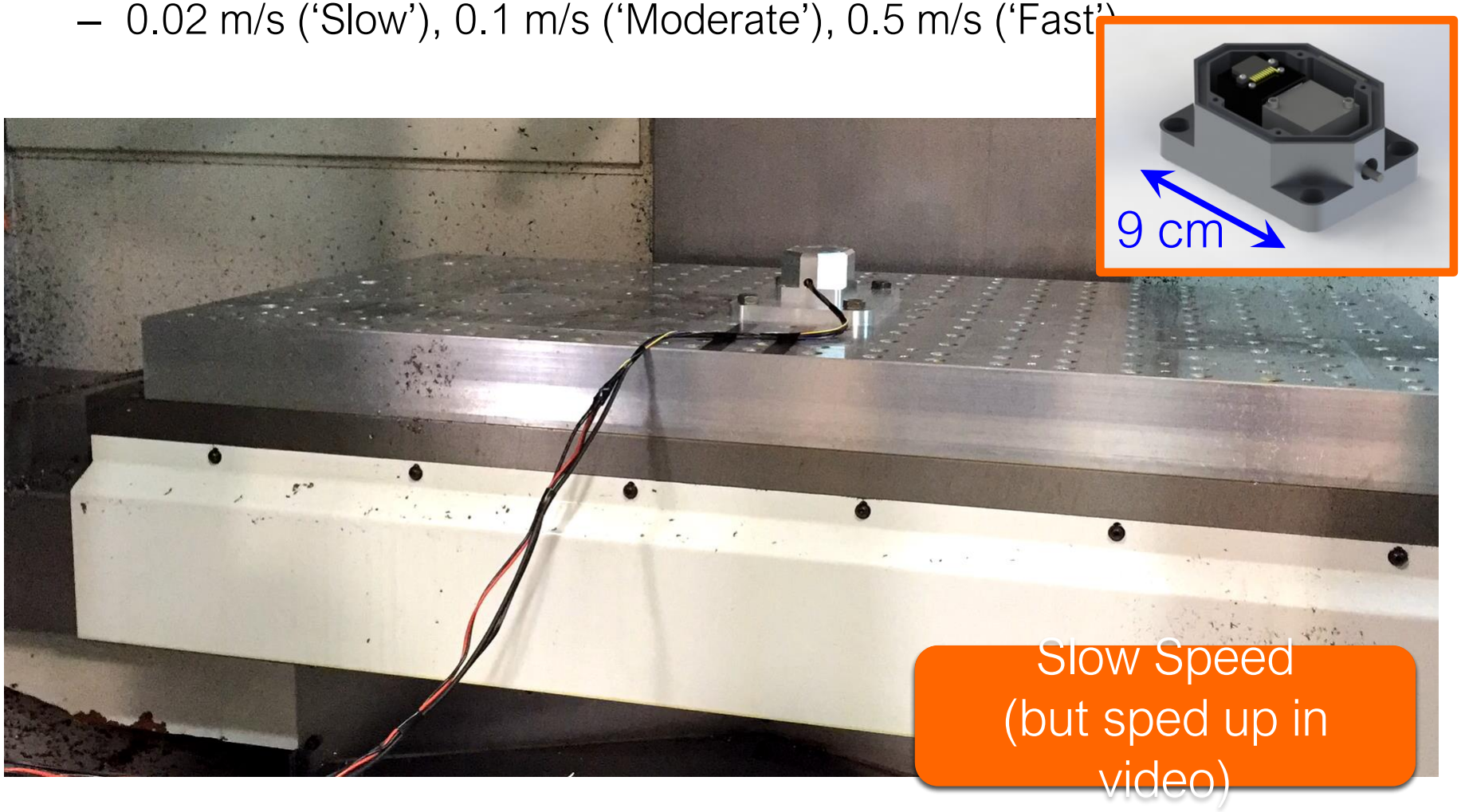




Errors within ±11 μm, ±2.3 μm, and ±13 μrad (k = 1) for positioning, straightness, and angular error motions

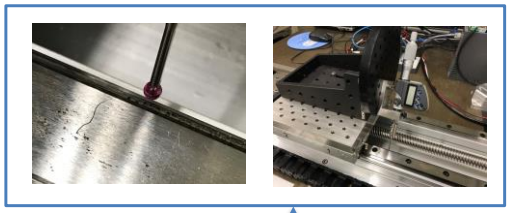
IMU on MACHINE TOOL

- Data collected for 3 different axis speeds
  - 0.02 m/s ('Slow'), 0.1 m/s ('Moderate'), 0.5 m/s ('Fast')



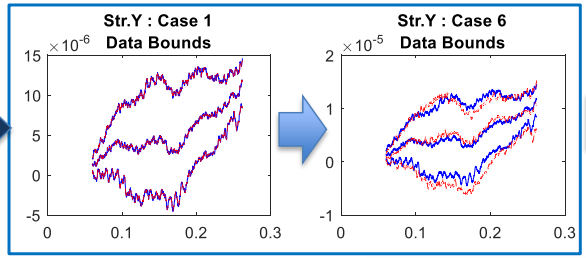
PHM DATA MODELING

- Health and Control Management of Robot Systems
  - Design of joint motor anomaly detection and health quantification algorithms
  - Exploration into general data requirements
- Linear Axes Diagnostics and Prognostics
  - Progressive manually articulated degradation data set analyzed
  - Testing of isolation of bearing degradation from rail faults



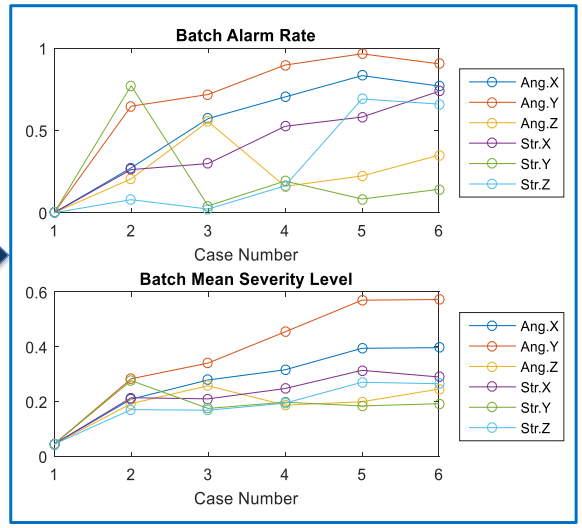
Experimental Setup

Data Collection



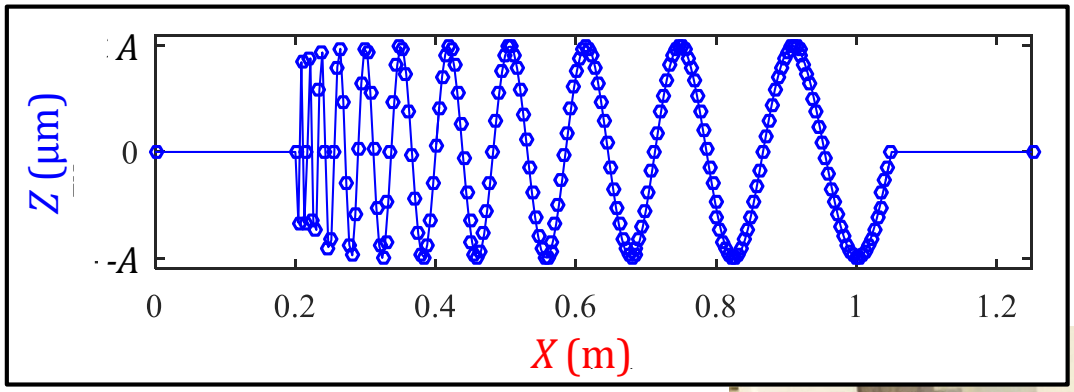
Data Processing

Anomaly Detection and Quantification

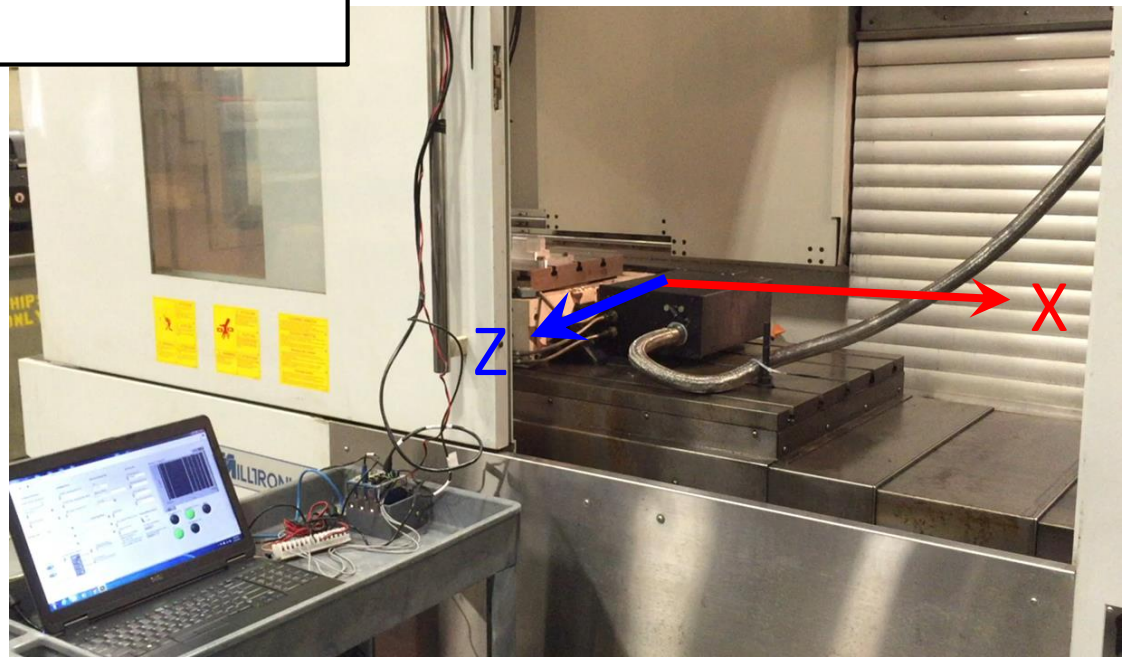


CASE STUDY 1: SIMULATED DEGRADATION

- Mechanically simulate degradation via 2-axis motion

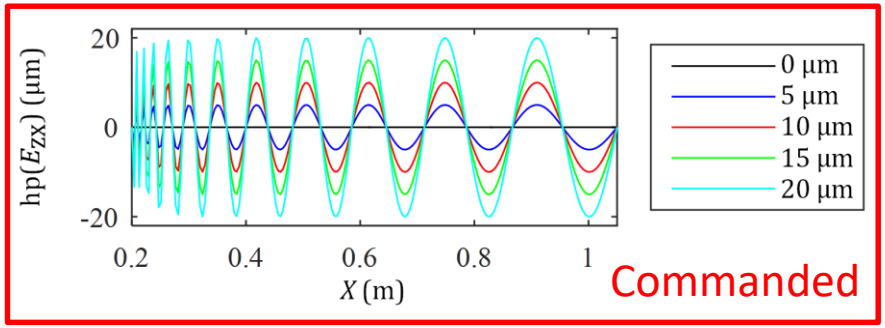


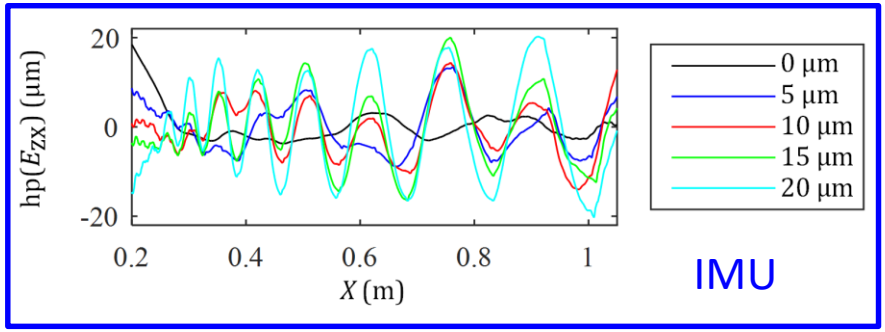
» A varies from 0 μm to 20 μm



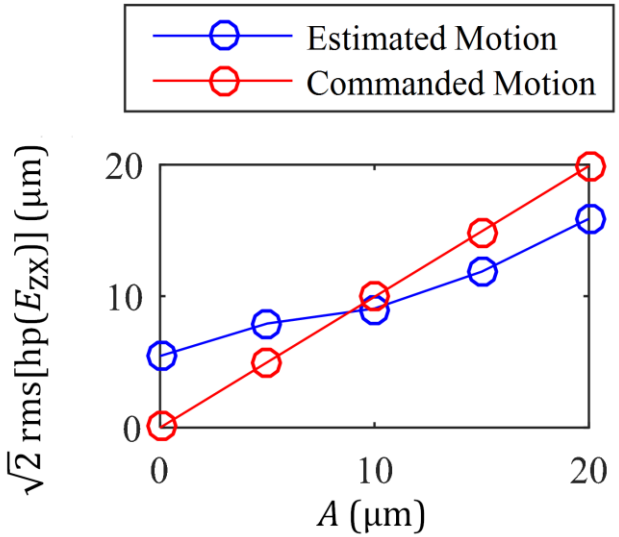
CASE STUDY 1: SIMULATED DEGRADATION

- Simple metric used to track linear axis degradation at μm levels



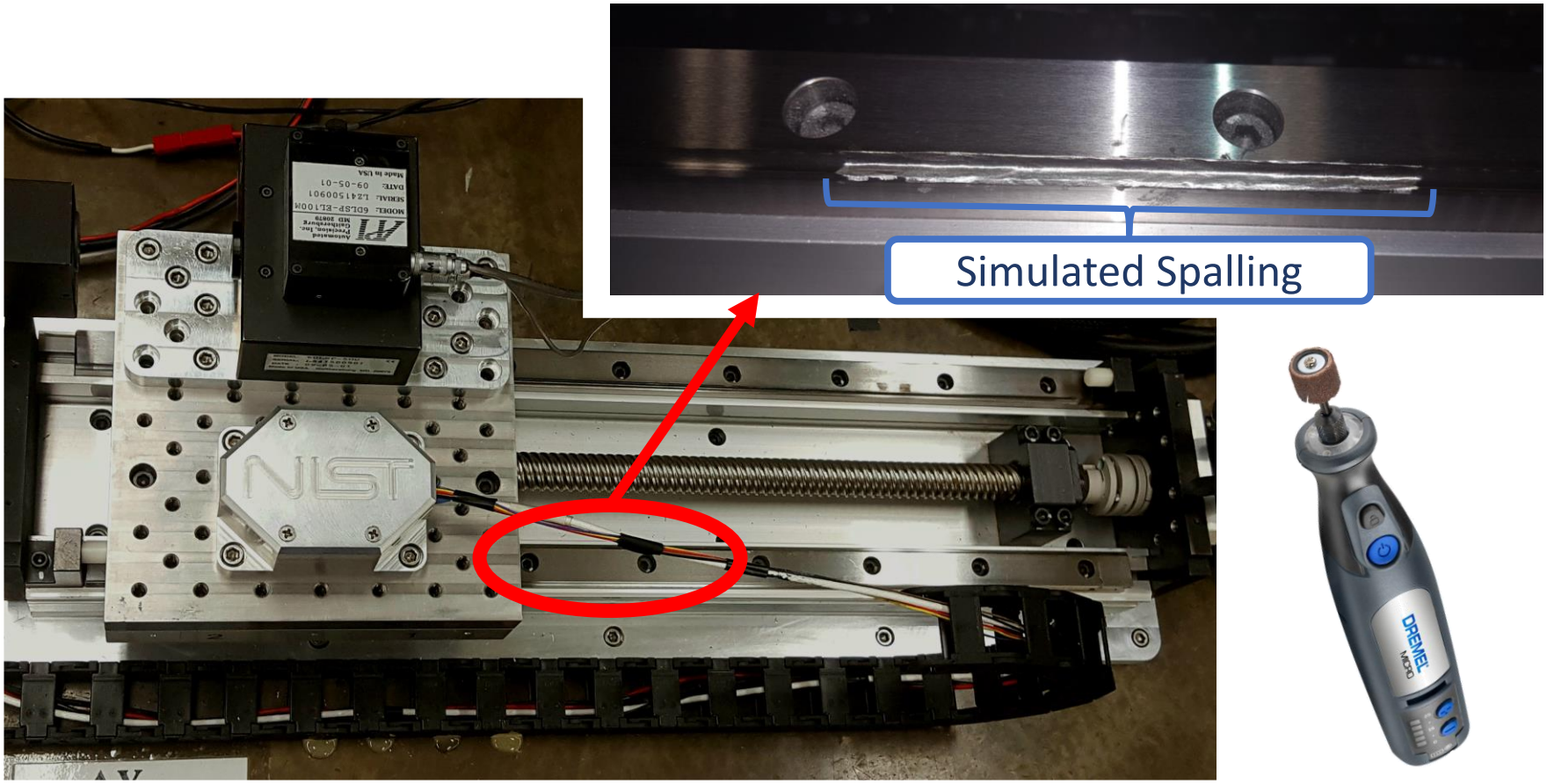


$\sqrt{2} * \text{rms}$



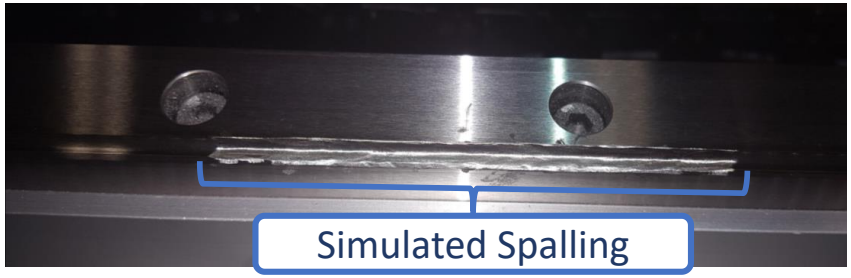
CASE STUDY 2: RAIL DEGRADATION

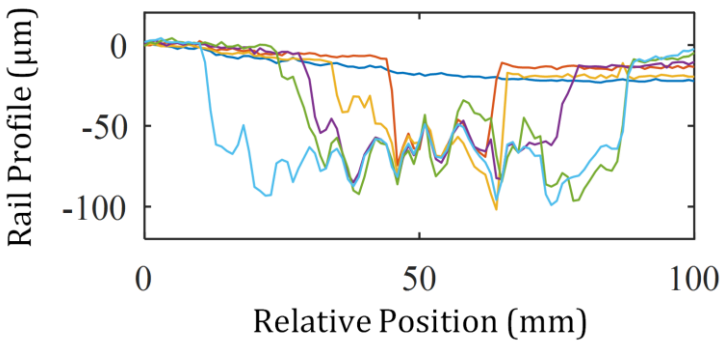
Rail was degraded to represent typical spalling



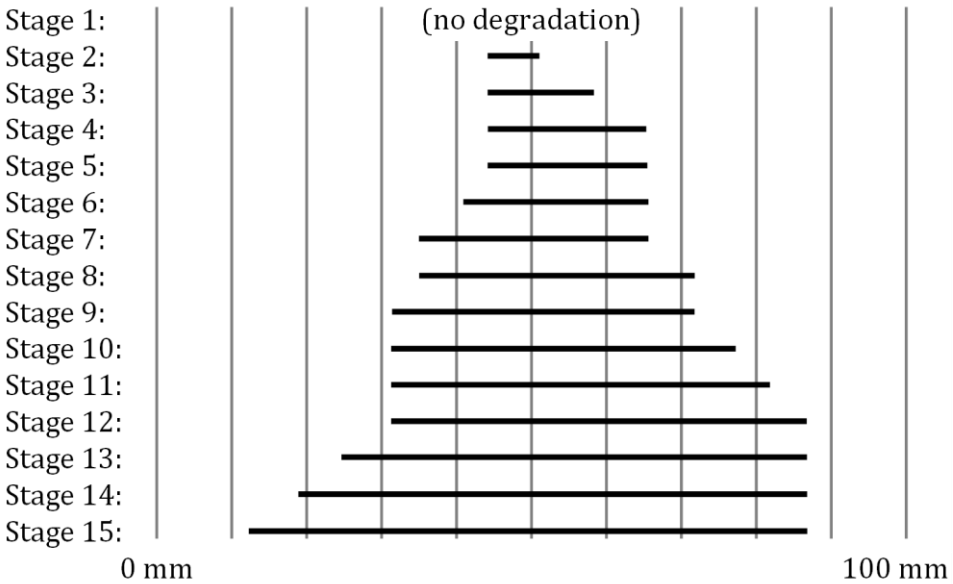
CASE STUDY 2: RAIL DEGRADATION

- Rail was degraded to represent typical spalling





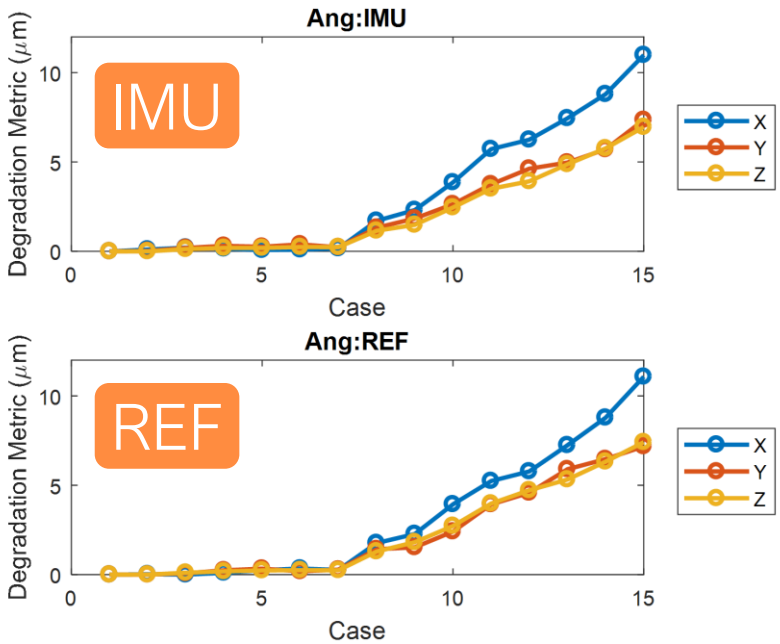
Depth

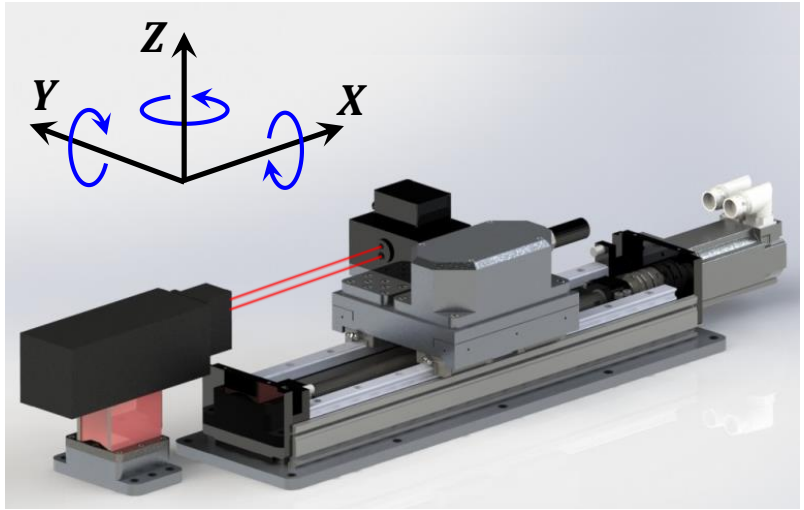


Length

CASE STUDY 2: RAIL DEGRADATION

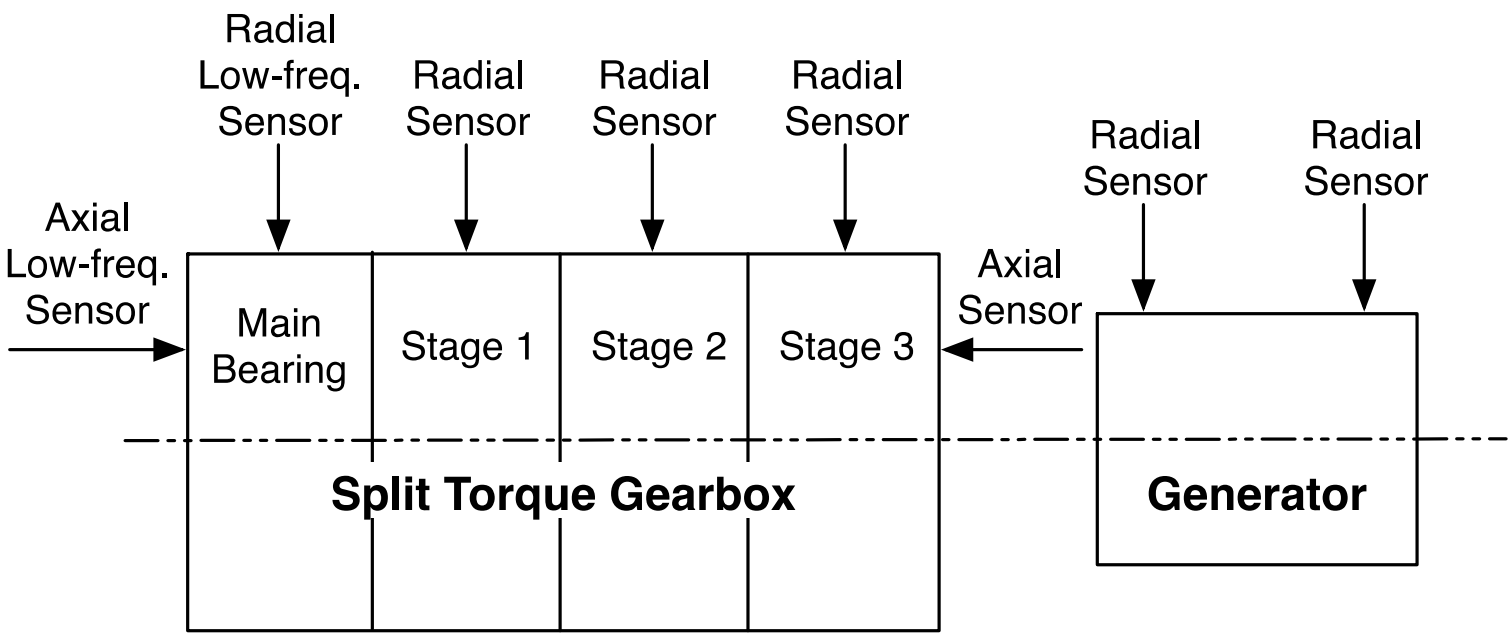
Metric of angular errors shows increasing “spalling”





Case Study:
Wind Turbine Predictive Monitoring Example

OFFSHORE WIND TURBINE CONDITION MONITORING CASE STUDY



» This case study is for an offshore wind turbine, in which vibration and SCADA data is provided for a 15 month time period.

» The wind turbine develops a problem in the rotor shaft, in which downtime occurs and the turbine is not operational for a period of time.

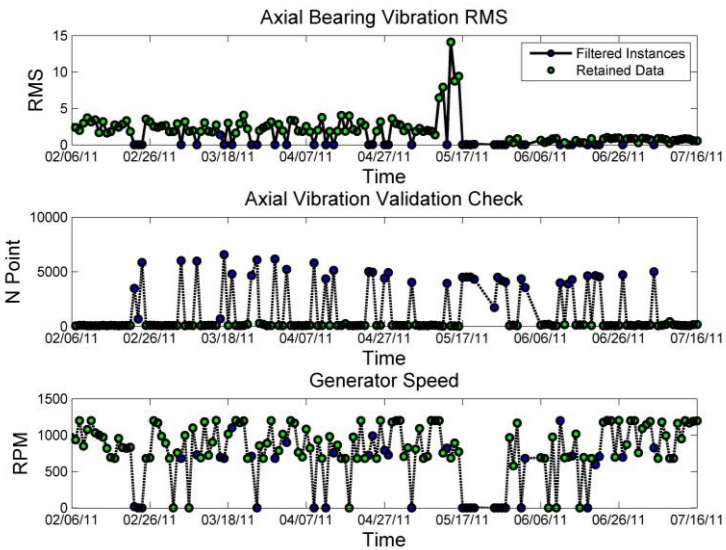
» The schematic of the drivetrain and the 8 accelerometer locations are shown in the diagram above.

Siegel, D. (2013). Prognostics and Health Assessment of a Multi-Regime System using a Residual Clustering Health Monitoring Approach. Doctoral Dissertation. University of Cincinnati.

DATA PRE-PROCESSING / DATA CLEANING

» Does not typically require advanced algorithms but a necessary step:

- Simple equations or rules to make sure all the sensor readings are in a normal range.
- Removing instances when the machine is not running.
- Confirming that the known past failure information is accurate (important for validation).
- The example on the right is from the wind turbine case study that will be presented later; a vibration data quality check was used to filter out readings when the turbine was idle.



Raw and Filtered Out Samples

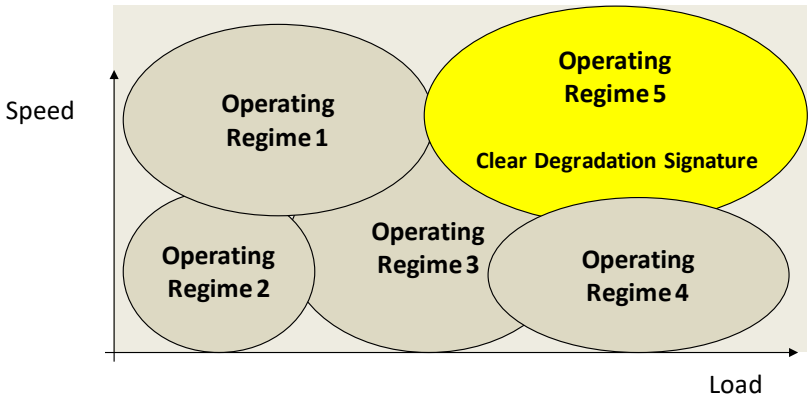
KEY SUBSYSTEM VARIABLES FOR ROTOR SHAFT

#	Accelerometer	Signal Processing	Statistic
1	Axial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	Peak to Peak
2	Axial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	RMS
3	Axial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	Kurtosis
4	Radial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	Peak to Peak
5	Radial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	RMS
6	Radial Accelerometer Near Main Bearing	Kurtosis-Based Filtering	Kurtosis

» A kurtosis-based filtering method and time statistics for accelerometer 1-2 (closest to rotor shaft) were used as the inputs into the subsystem health model.

REGIME SEGMENTATION

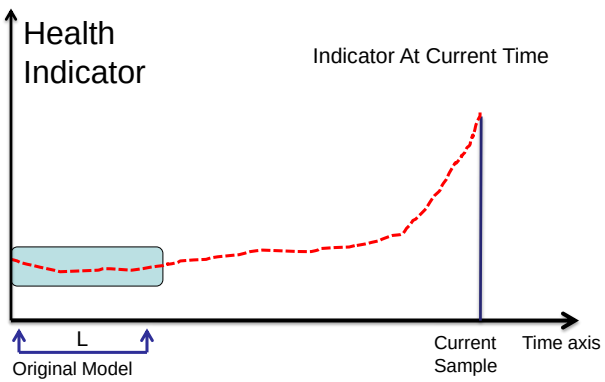
» Regime Consideration:
- The Number of Regimes and which regime is best for condition monitoring.
- Regime Variables
  - Output Power
  - Rotational Speed
  - Wind Speed
  - Ambient Temperature



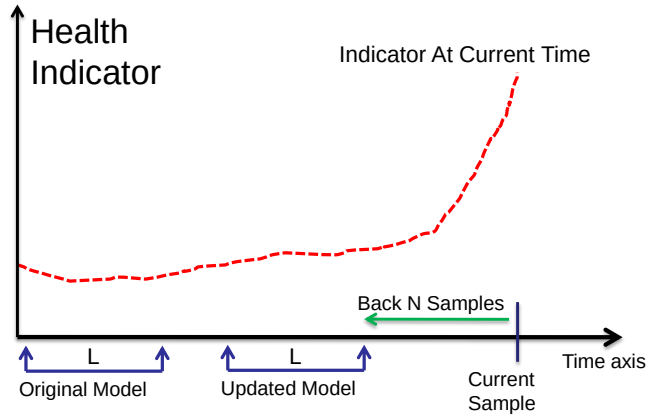
- For the wind turbine case study, we just considered the output power and the rotational speed as the two variables related to the operating regime.

BASELINE CONSIDERATIONS

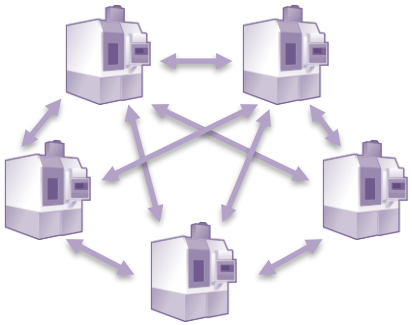
Static and Individual Baseline



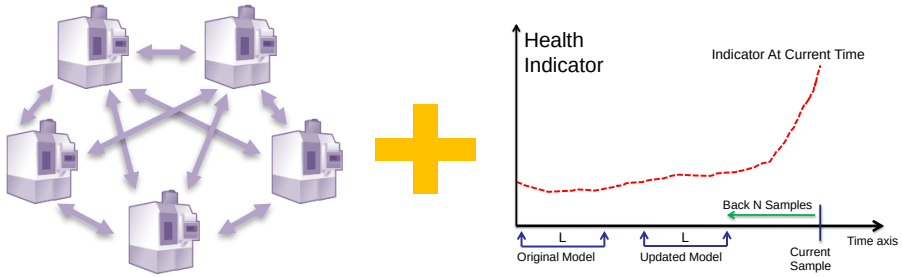
Individual and Updating Baseline



Global Baseline from Peer Units



Global Baseline and Updating



» The current approach is a baseline that is based on individual unit baselines and a baseline that is updated after major maintenance occurs.

SUBSYSTEM MODELS VS. OVERALL SYSTEM HEALTH MODELS

Approach 1

Overall System Health Model

Use large variable set for anomaly detection and only 1 health index

Advantages:
Simplifies monitoring system and number of outputs to consider; contribution plots can help determine what subsystem is the cause

Disadvantages:
In many cases, not all the variables are well correlated and overall system model might lack sensitivity to correlation changes.

Approach 2

Subsystem Health Models

Have 3-6 subsystem health models that are based on subsystem variable sets – results in 3-6 health indices to monitor.

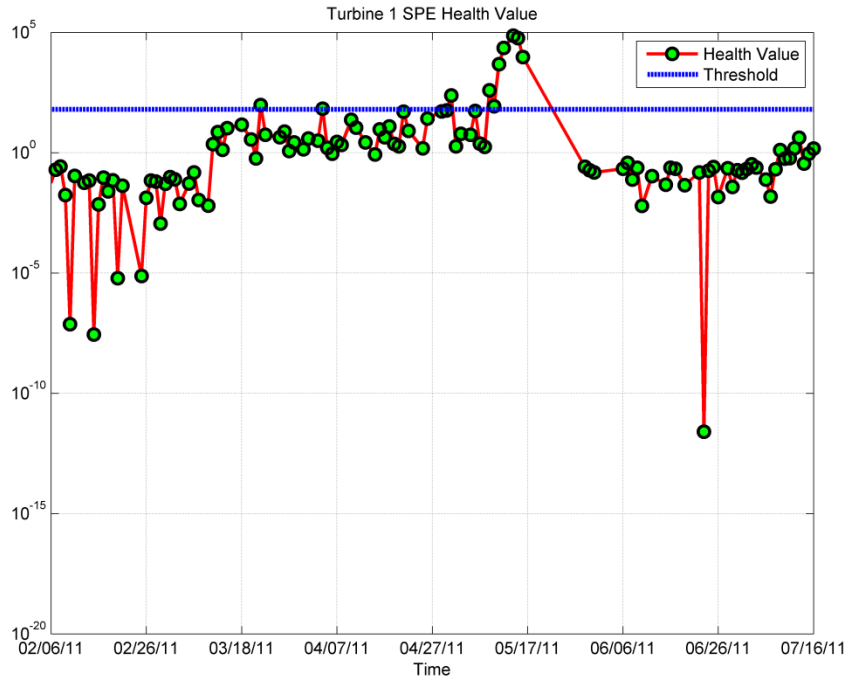
Advantages:
Subsystem variables should be more correlated with each other and health models should have more detailed accuracy.

Disadvantages:
There are more health values to monitor and set thresholds for; also more effort is needed in maintaining and updating health models.

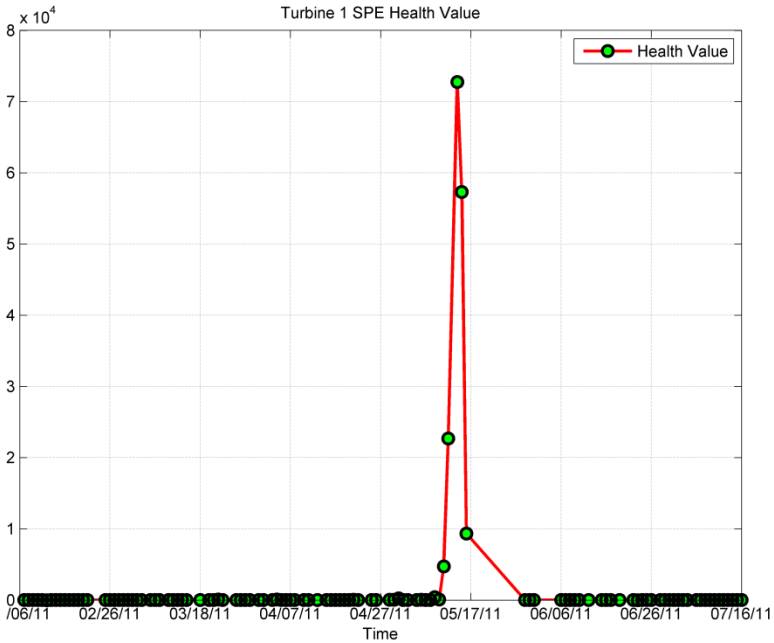
» From our experience, we found overall health models to be very difficult with a large signal set; subsystem health models with better correlated variables has shown better results.

HEALTH VALUE RESULTS / AUTOMATED DETECTION

Log-Scale View of Result



Health Value Result – Normal Scale



» The initial detection of the problem is several days ahead of time but is quite intermittent; however the health values 5-7 days ahead of the event are well beyond the threshold.

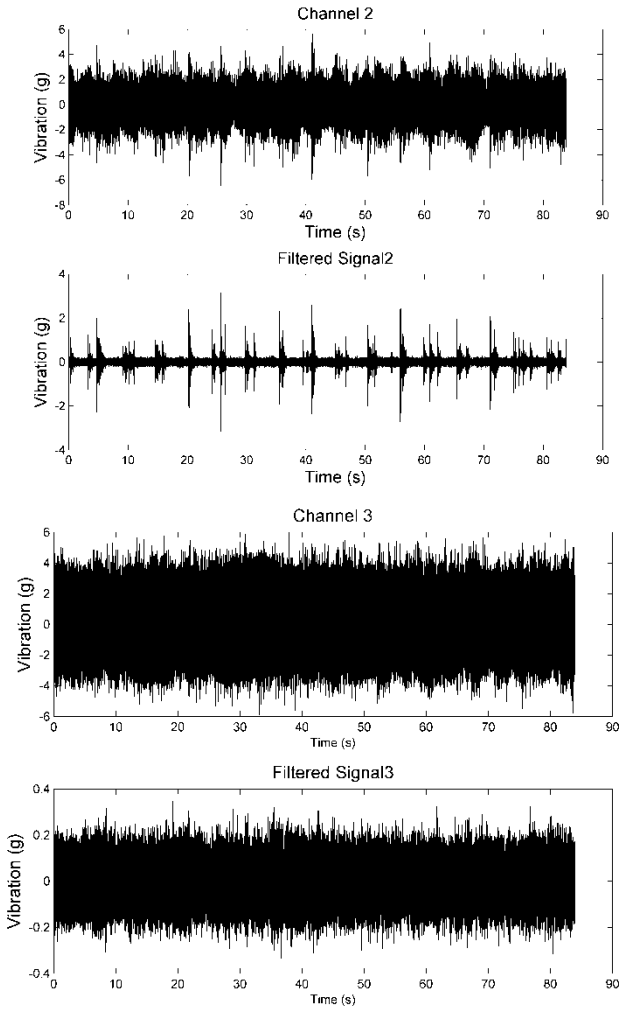
DETECTION SUMMARY

Method	Time When Health Value First Exceeds Threshold	Advanced Warning (Days)	Time When Health Value Never Drops Below Threshold	Advanced Warning (Days)
Health Value	3/22/2011	55 (8 Weeks)	5/9/2011	7 Days (1-Week)

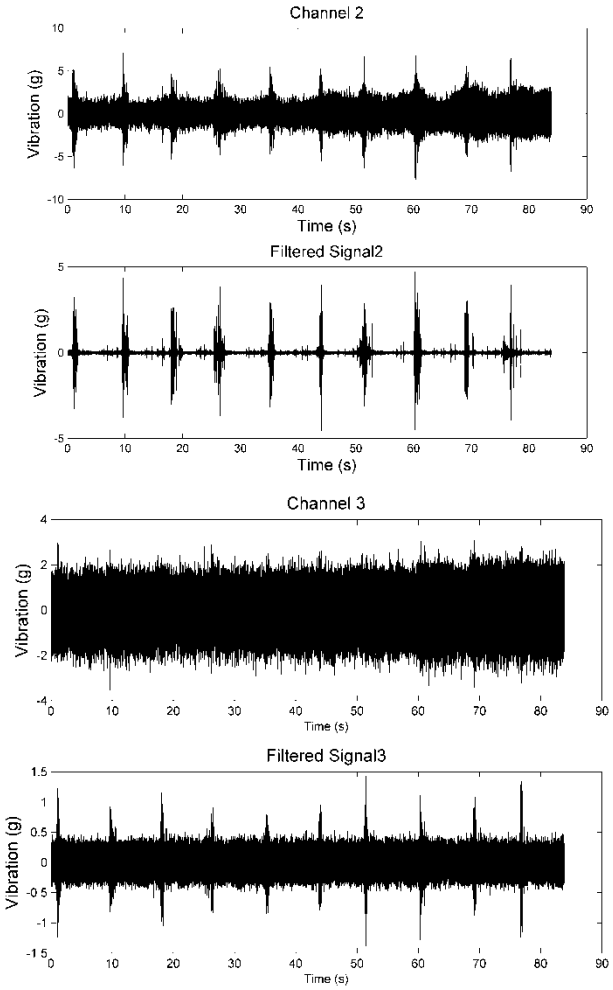
- The instance that the health value first passes the threshold is 55 days ahead of the downtime.
- However, a consistent health value that is always above the threshold is only obtained 7 days ahead of time.
- This indicates that the method can provide advanced warning, but having a health trend with a more consistent trend could be an area for improvement.

FURTHER INSPECTION OF THE VIBRATION SIGNATURE – FILTERED SIGNAL

63 Days Before the Maintenance Event



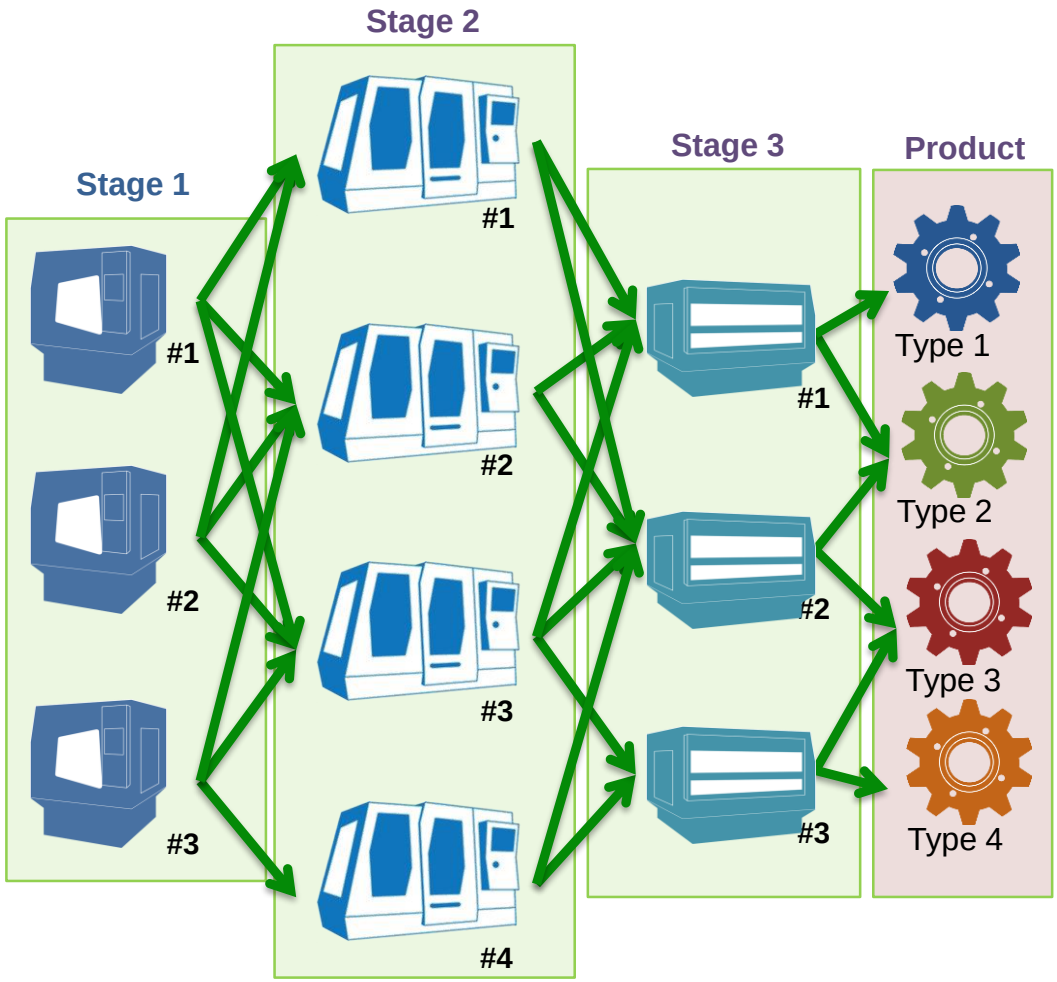
2 Days Before the Maintenance Event



Case Study:

A Minimal-sensing Technique for Monitoring Multi-stage Manufacturing Processes

The Scope of the Work



- An example of a three-stage manufacturing process
- The available data consists of the route of each product along with multiple quality parameters measured from each product.
- Multiple models of a product with different specs are being manufactured.
- The objective is to determine the source of variations in the quality of the products.

General Concept

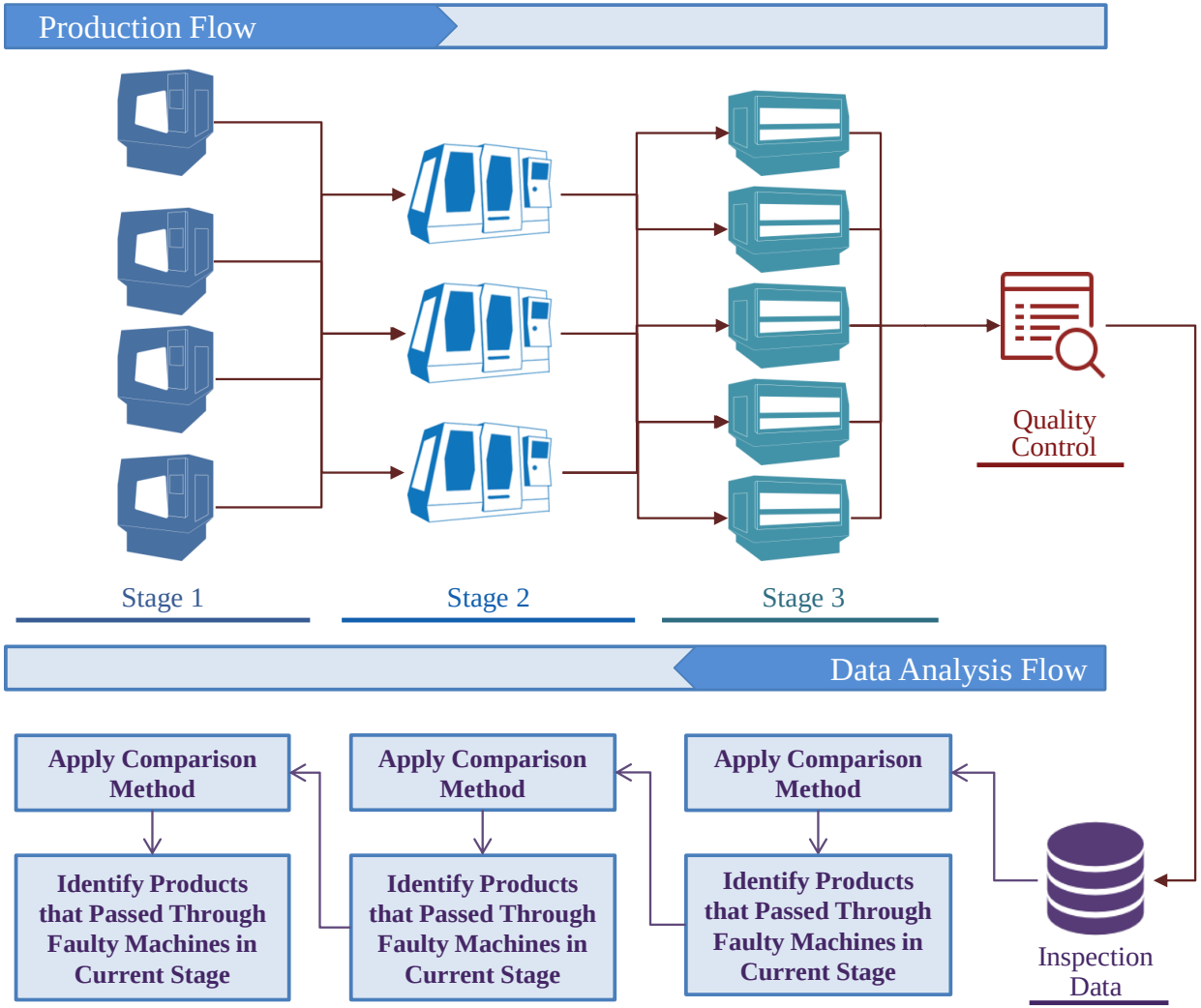
» Assumptions:

–The distribution of products from one stage to another is random. This enables us to implement similarity-based techniques.

–Majority of machines function properly. That ensures the establishment of a reliable baseline.

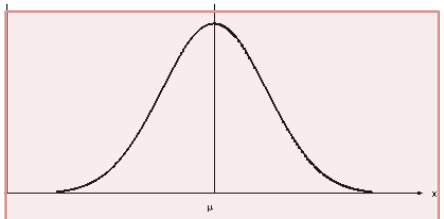
–Data includes the following:

- Quality measurements for each or a sample of products
- Manufacturing route for each product

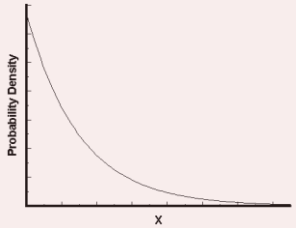


Machine Performance Metric Development

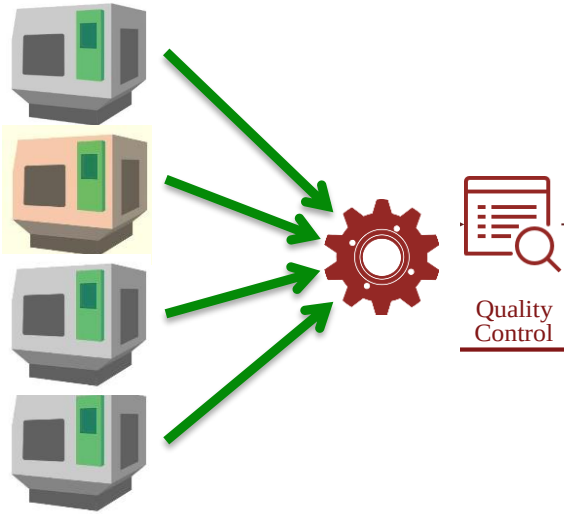
- A standard health value, ranging from 0 to 1, is designed as a metric for machine performance measurement.
- A moving window is applied to the data and the distribution of quality parameters over each window was used as the basis of the metric development:
  - Distribution of parameter "A" for one machine is compared with the distribution of the same parameter from all the other machines
- In this example, Gamma distribution is the best option based on the shape of the parameter distributions.
- The PDF curves is then used for metric development.

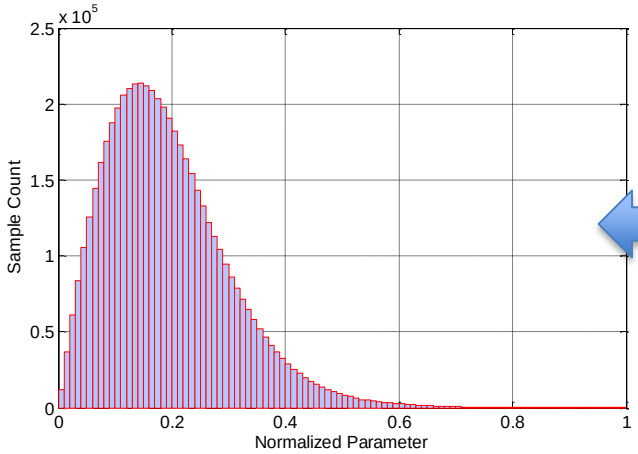


Gaussian Distribution

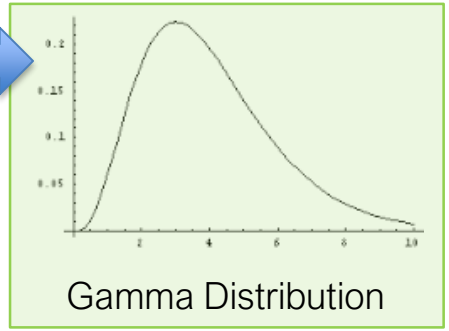


Exponential Distribution



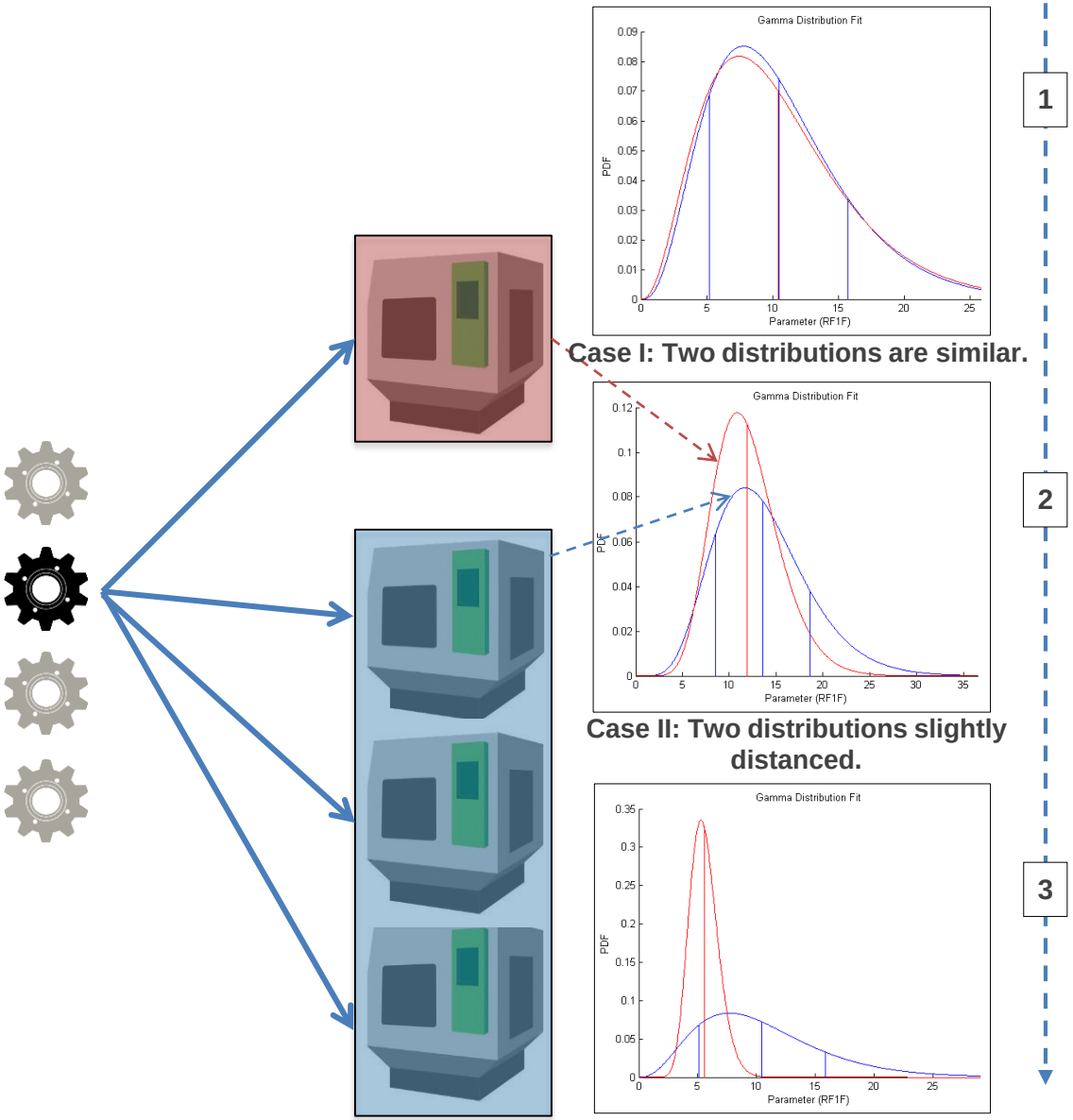


The histogram of a product quality parameter

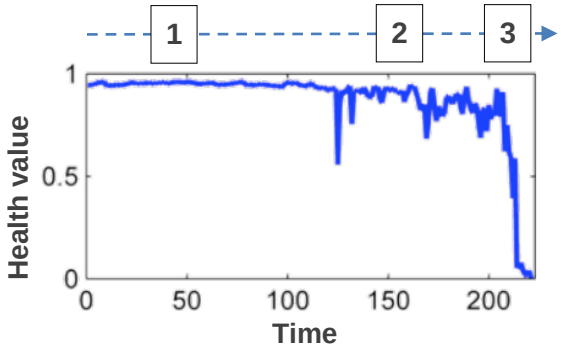


Gamma Distribution

Similarity-based Metric Development

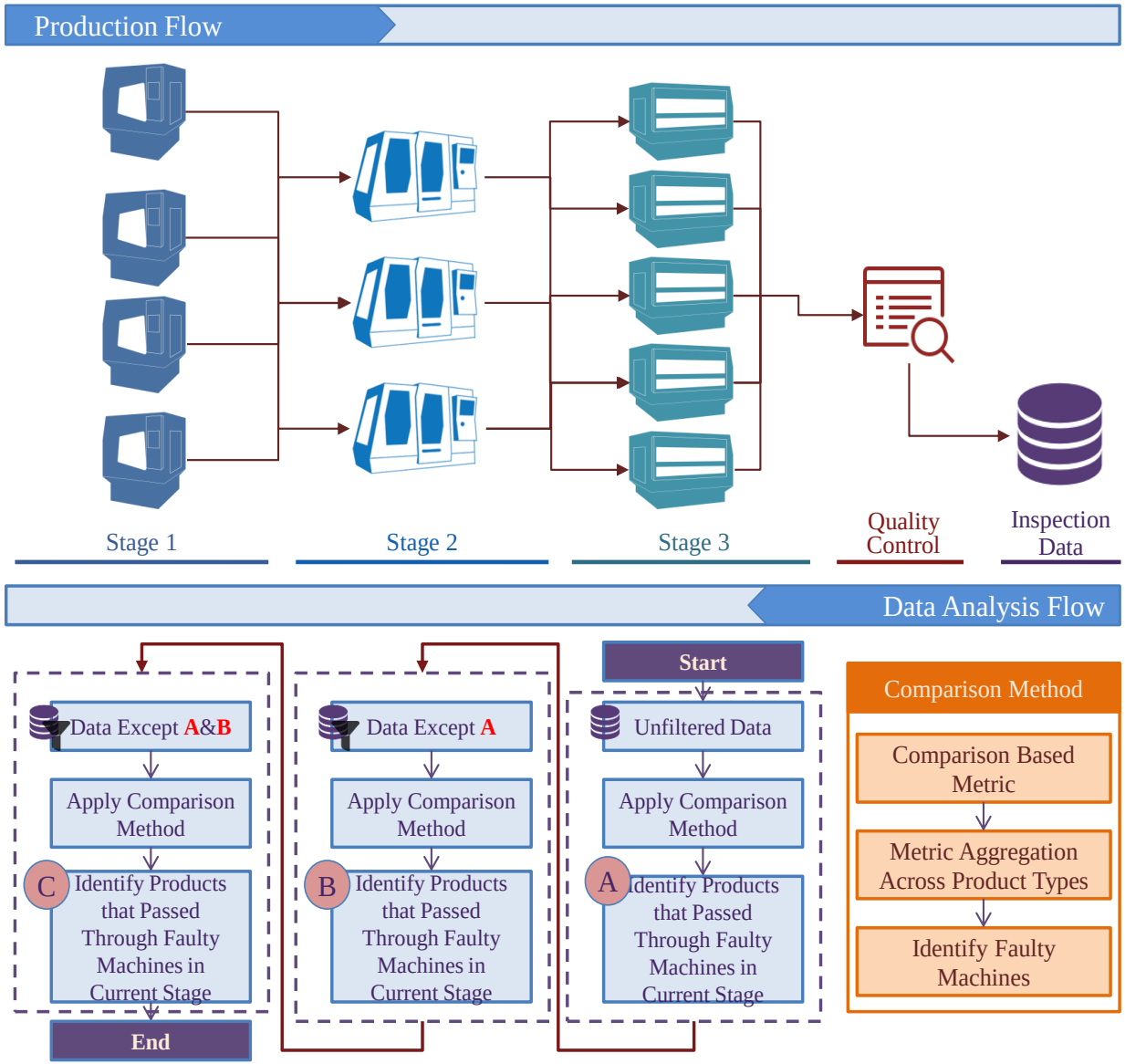


- Within each time window, the distribution of the parameters for one machine versus all other machines are measured.
- The method introduced for metric development was used to calculate the health value for the machine under study.



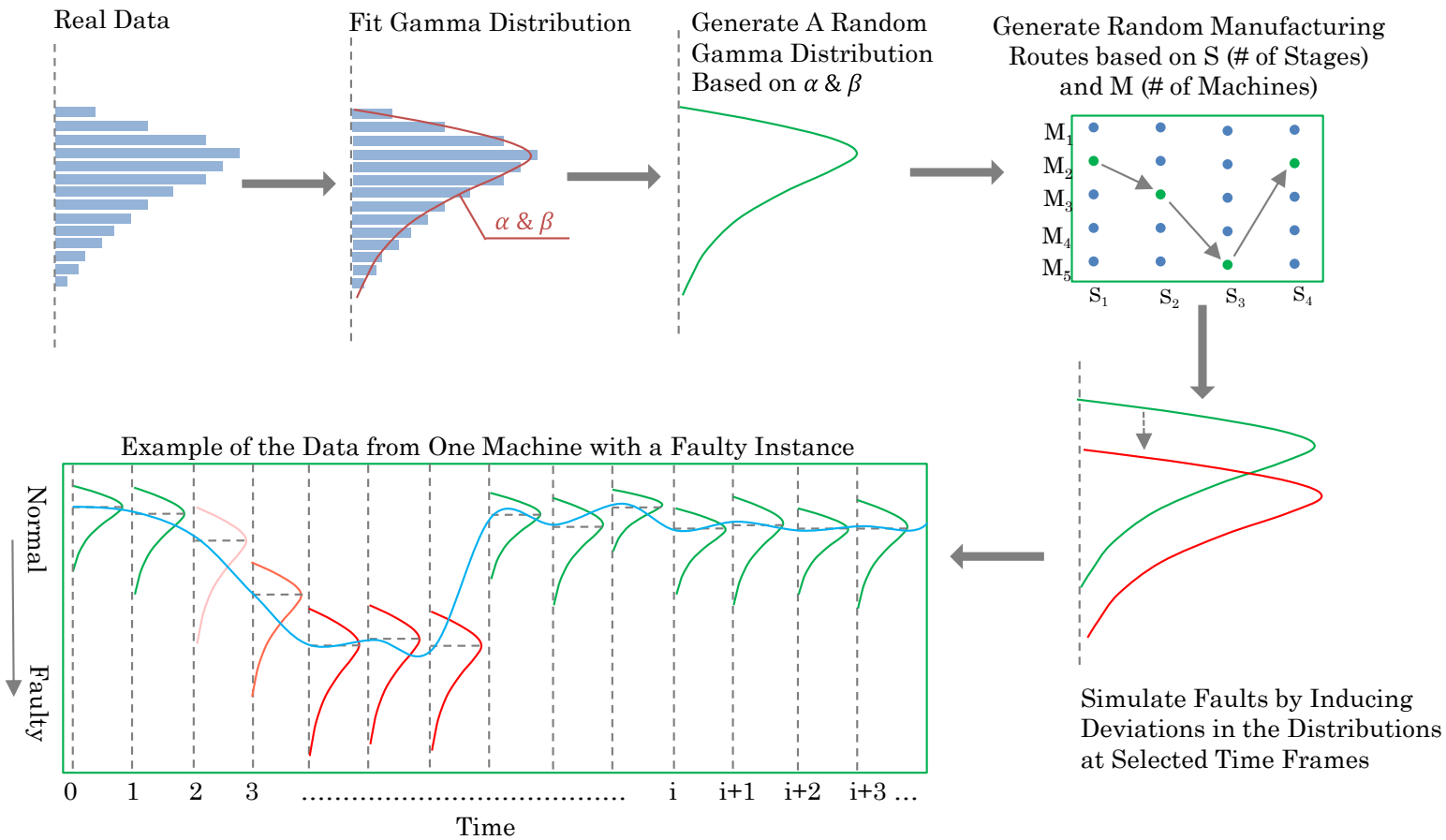
Expanding the Method to Other Stages

- It is assumed that products are being randomly distributed from each stage to the next.
- After identifying the bad machines, these samples are filtered out from the data and the remainder of the data is used to assess the machines performance in the previous stage.
- The threshold is set based on k-means clustering technique.

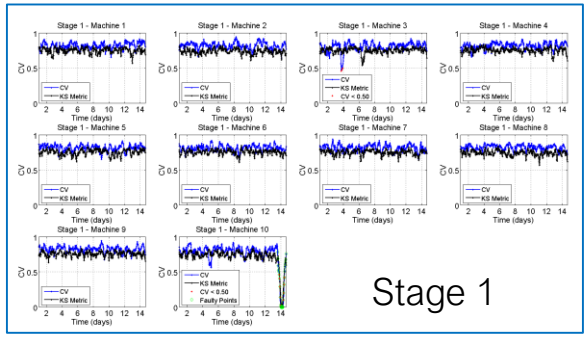


Monte Carlo Simulation

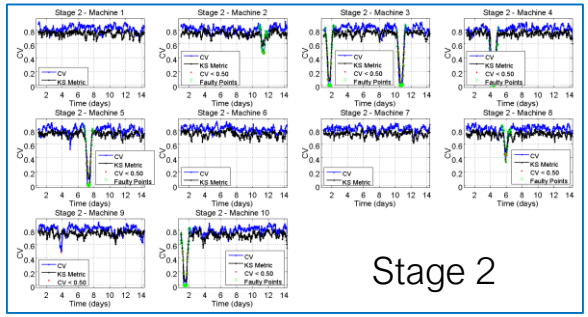
» Monte Carlo technique: a powerful tool for drawing a large number of samples within a distribution defined based on probability parameters



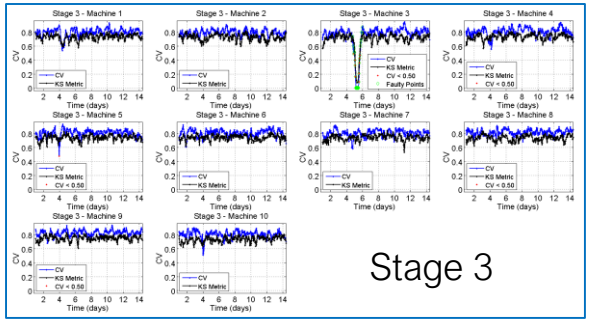
Results for a Process with Five Stages and 10 Machines per Stage



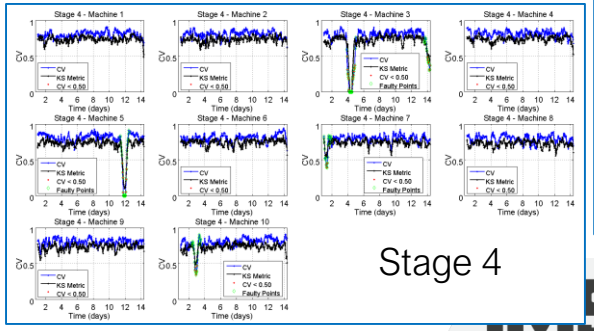
Stage 1



Stage 2

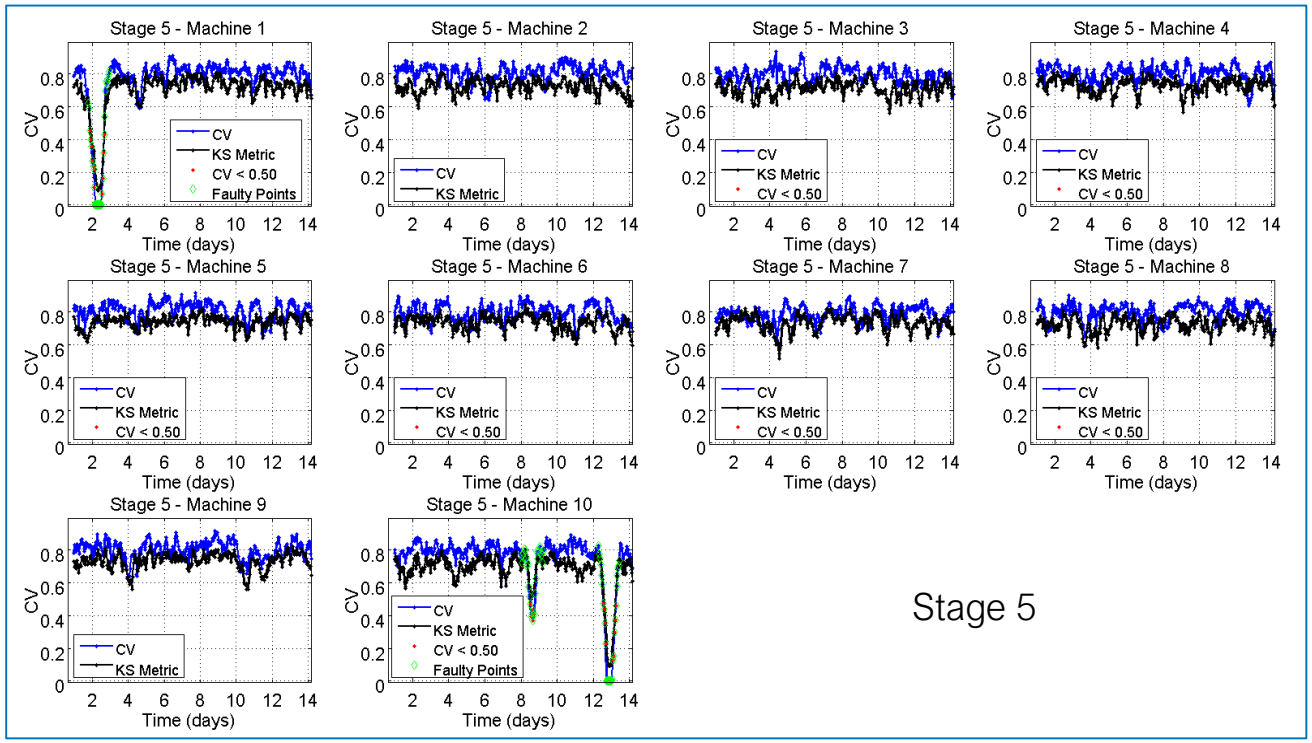


Stage 3



Stage 4

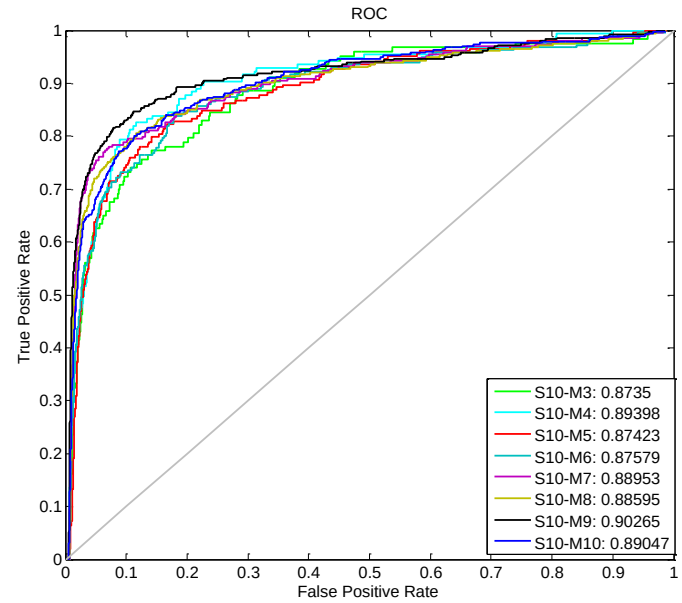
- An example of the proposed metric and KS metric for a simulated data set is provided.
- The method performed consistently well for all the machines and stages.

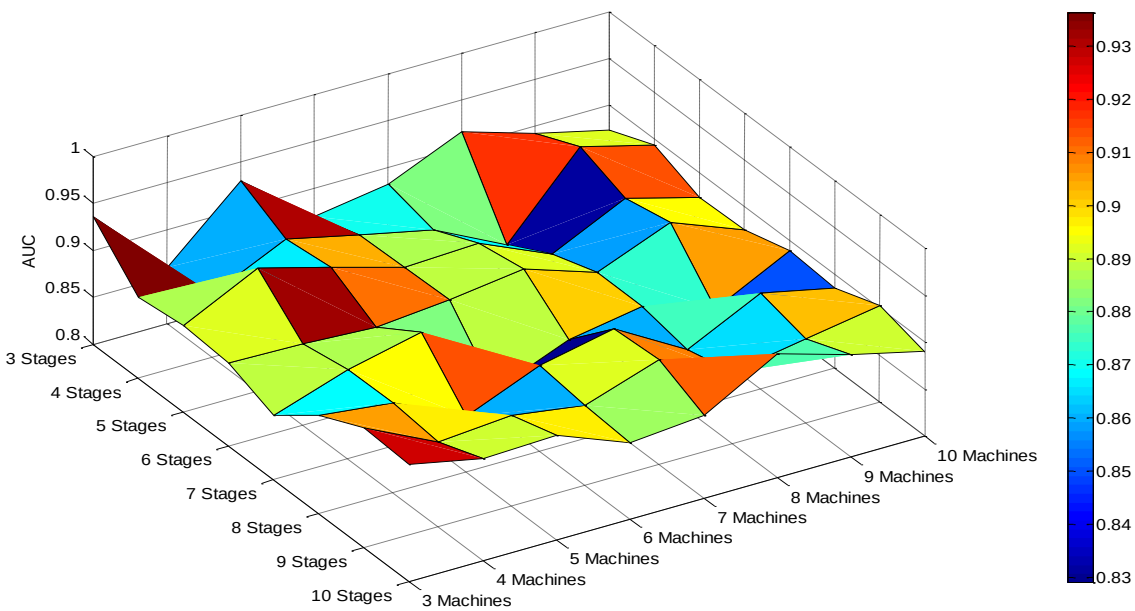


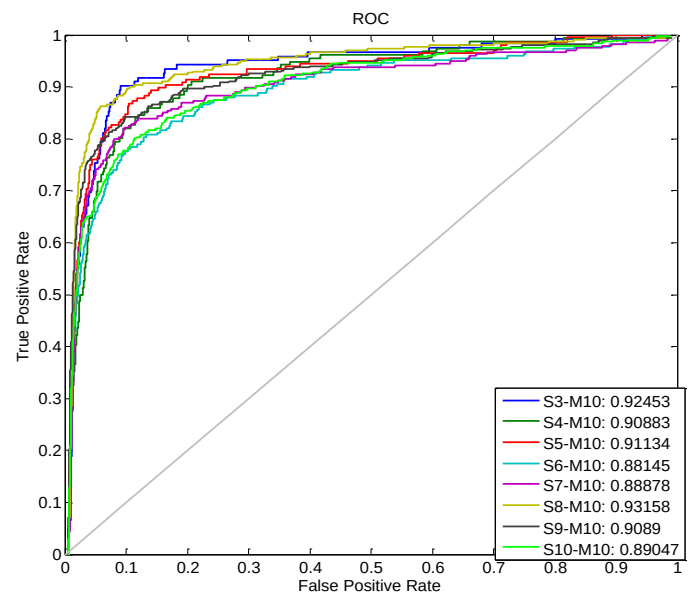
Stage 5

Impact of the Process Scale on Fault Detection Accuracy

- » Results were calculated for processes with number of stages ranging from 3 to 10, and number of machines per stage ranging from 3 to 10.
- » No significant difference was observed in the accuracy of the method as the process scale changed.

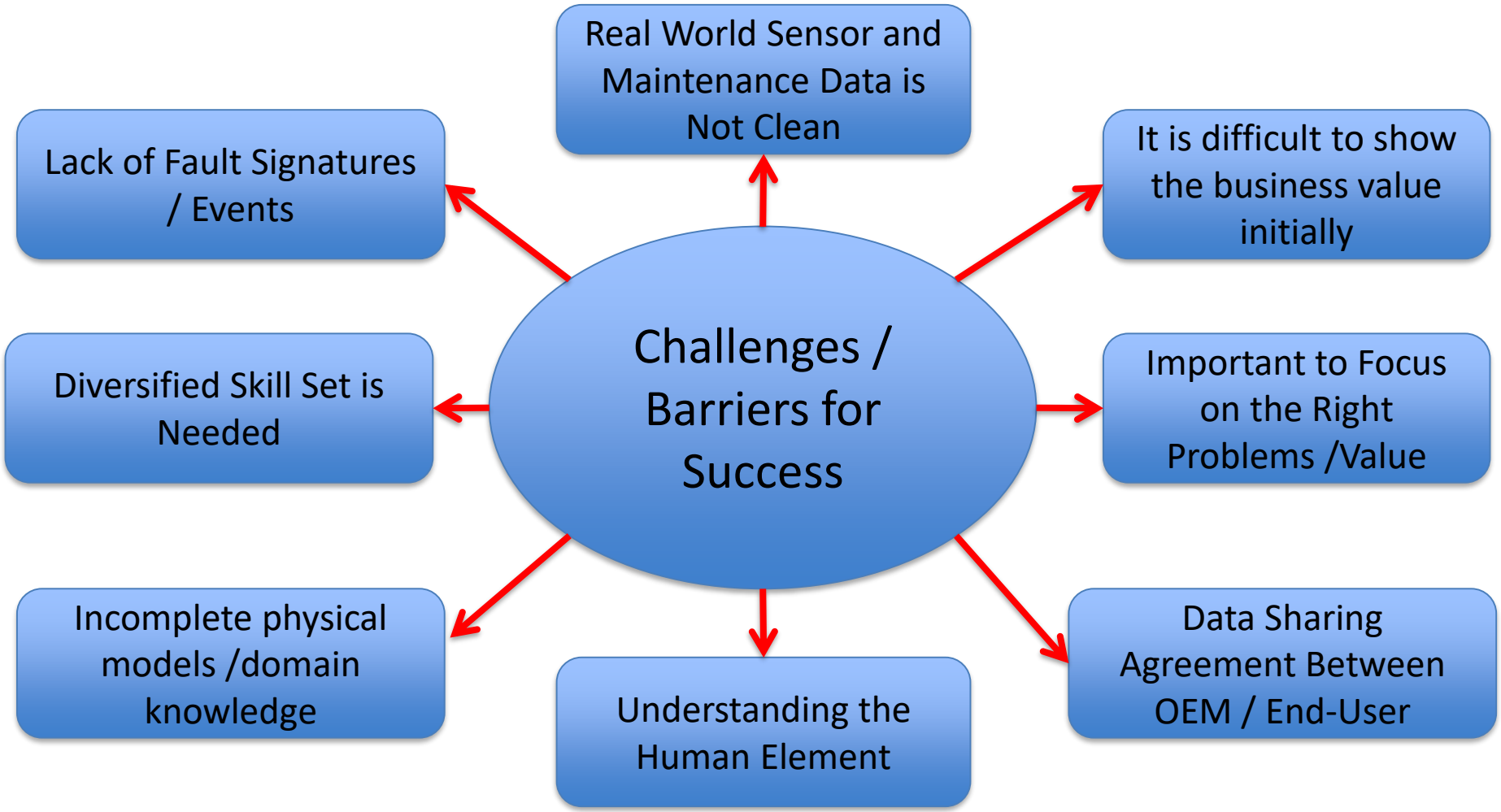




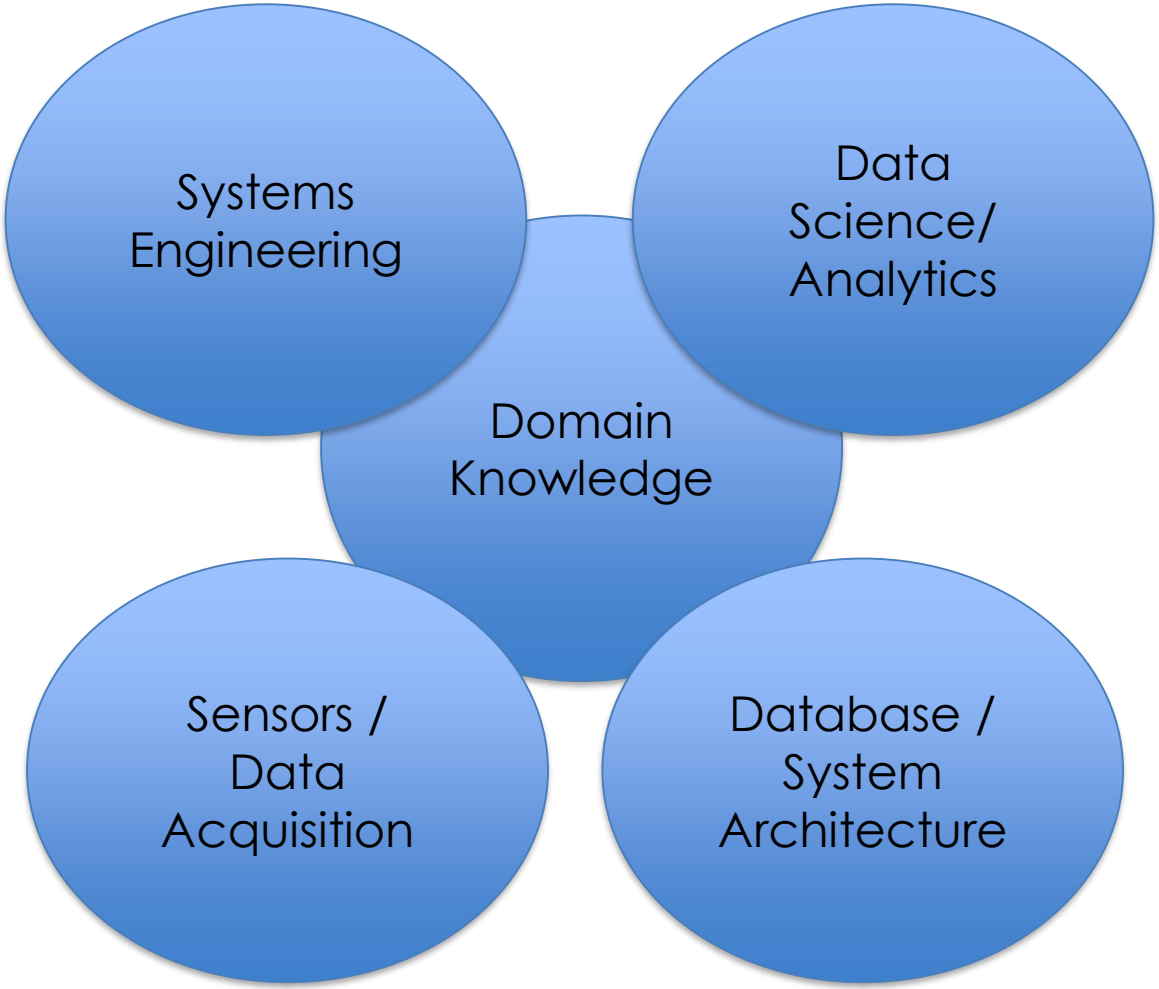


Concluding Remarks

Industrial Big Data Challenges



Skill Sets Needed for Industrial Big Data Solutions



- Very diverse skills are needed, and normally a single person would only have a combination of one or two of these skills.
- Normally one would like someone good at 1 or 2 of these skills, but at least be able to communicate or work with other team-members on the other topics.

Lessons Learned

1. Start small
   i. A proof of value and a subset of the machines is a good first step
   ii. Don't wait too long, the longer you wait, the more likely your competition has already started.

2. Machine Learning and Pattern Recognition is Great but it is not Magic
   i. You still need good engineering experience for problem selection, data collection, and evaluation criteria.

3. Select a high value problem but not the most difficult /challenging application
   i. It's a new technology, you need an early win to build momentum.
   ii. However, if you select something very easy that could be done without this technology, you might also get some resistance.

4. Work with the data you have
   i. In certain cases, using controller data or existing data sources could be enough to get started.
   ii. This represents a lower entry cost to get started with this technology.

Concluding Remarks

» Finding the right problem is half the challenge – one that has business value and that is technically feasible.

» You need the right data also, but it is good to consider using the data you already have.

» A more direct sensor measurement can simplify the analysis algorithm used; there is likely a tradeoff regarding the more precise/direct measurement and implementation cost.

» Start small, find an early win, and build from there.

» There is no magic algorithm that can solve these diverse problems – the right method for the right problem is needed.