# A Tutorial on Feature Extraction Methods

Tianyi Wang
GE Global Research

Subrat Nanda
GE Power & Water

September 24, 2012

GE imagination at work

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

# Where Feature Extraction fits in a PHM System

Advisory Generation (AG)

↑

Prognostics Assessment (PA)

↑

Health Assessment (HA)

↑

State Detection (SD) — a.k.a. ***Feature Extraction*** in data-driven PHM solutions

↑

Data Manipulation (DM) — such as normalization, smoothing, outlier removal, missing data imputation, ...

↑

Data Acquisition (DA)

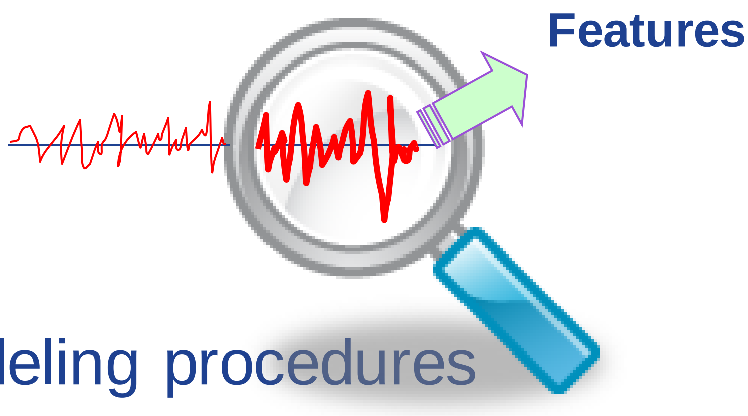
GE imagination at work

source: MIMOSA OSA CBM architecture

# Feature extraction: what and why

## What:

Feature extraction transforms raw signals into more informative signatures or fingerprints of a system




## Why:

- Extract information from data
- Serve the need of follow-up modeling procedures
- Achieve intended objectives

GE imagination at work

# Example of feature extraction

**Problem**: bearing health assessment

**Data**: vibration (from accelerometers)

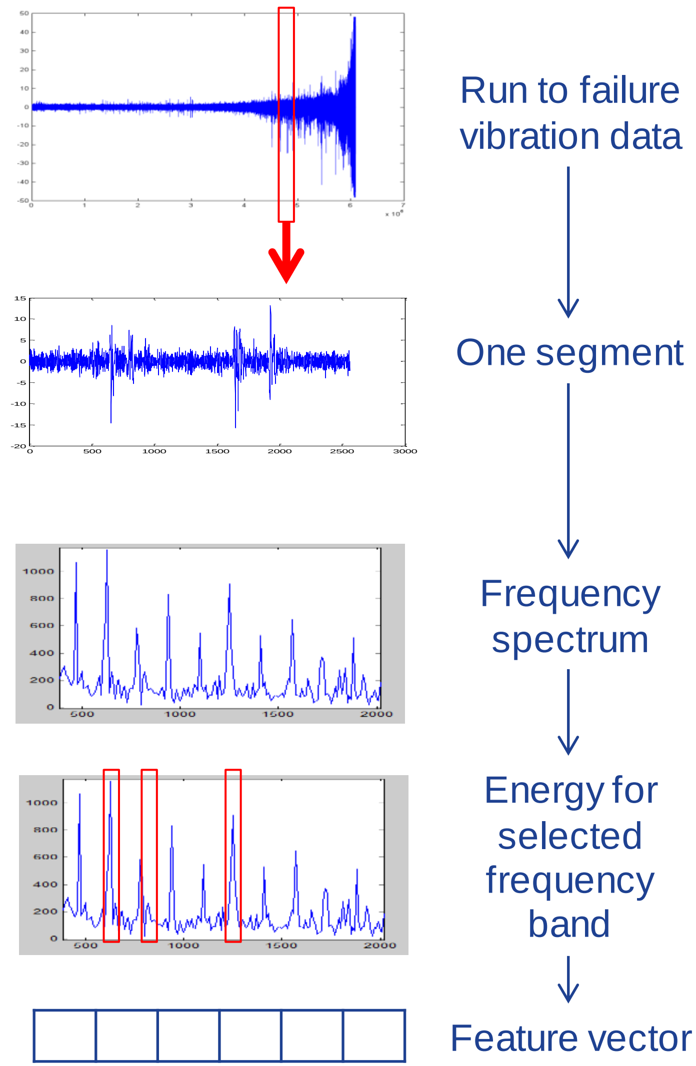
**Extract frequency domain features:**

- Segment the data with a certain time window
- Transform each segment into frequency spectrum with FFT
- Calculate energy for each frequency band around interested frequency F

$$E_F = \sum_{|f-F|<\Delta} A_f^2$$

where $A_f$ is the amplitude of frequency f

- Obtain feature vector [$E_{F1}$, $E_{F2}$, …]




GE imagination at work

# Feature extraction process

Raw data/ preprocessed data

Data representation:
- Varying format: 1-D, 2-D time series, events, ...
- Potential heterogeneous

**Feature generation**

- With exhaustive or ad hoc approach
- Dimensionality may reduce or increase
- Incorporate domain knowledge
- Underlying physical phenomenon

Feature set

Feature representation:
- A scalar or vector per feature
- One vector concatenating all features
- One matrix holding all samples of features

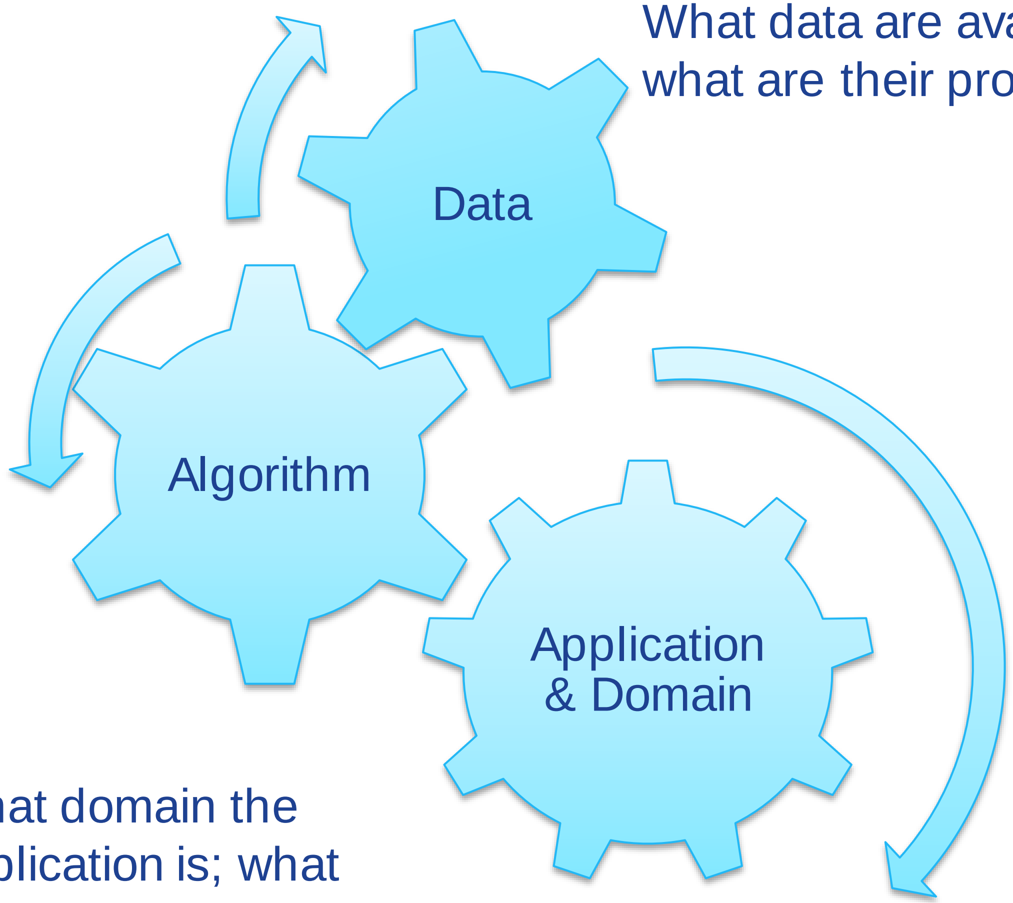
**Feature dimensionality reduction**

Two approaches
- Select a subset of generated features
- Transform the features to another space with lower dimensions

# What features to extract? Factors to consider...

What data are available and what are their properties

What feature extraction algorithms are available and applicable



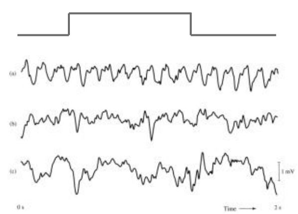

What domain the application is; what knowledge and requirements are present

GE imagination at work

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

# Data (signal) properties



**Time dependency**
- Time variant
- Time invariant (meta data such as asset ID)

**Sampling (time discretization)**
- Transaction/event (push)
- Sensor (pull)
  - Evenly sampled
  - Unevenly sampled

**Physics nature**
- Temperature
- Pressure
- Current
- Voltage
- Speed
- Acceleration
- ...

**Sample dimension** (not counting time)
- Scalar
- Vector
- Matrix
- ...

**Dynamics (relative to sampling)**
- Stationary
- Cyclic (non periodic)
- Waveform (periodic)
- Stochastic (non cyclic)

**Sample value discretization**
- Binary
- Discrete nominal (categorical)
- Discrete ordinal (integer)
- Continuous (real number)

GE imagination at work

# Data sampling (time discretization)

## Transaction/event (data are "pushed" by data originator)

- Data records occur only at the specified time stamp.
- Data between the time stamps (interpolation) are undefined.

## Sensor (data are "pulled" from data originator)

- Data samples are acquired only at the specified time stamp
- Data between the time stamps are just not observed.
- Sampling rate
  - Evenly sampled – controlled (e.g. 100 Hz)
  - Unevenly sampled - triggered

# Sample value discretization

## Binary

- Events status, on/off sensor

## Discrete nominal (categorical)

- Event code, operating mode, asset ID

## Discrete ordinal (integer)

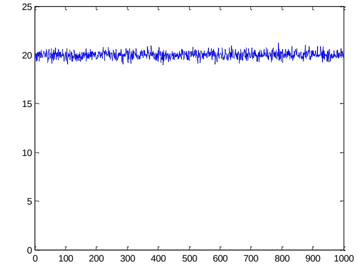
- If interpolation is meaningful, treat as continuous; otherwise, treat as discrete nominal

## Continuous (real number)

- Most sensors

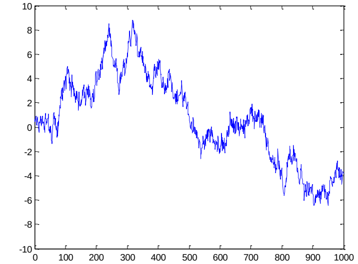
# Signal dynamics (relative to sampling)

## Stationary (constant + white noise)



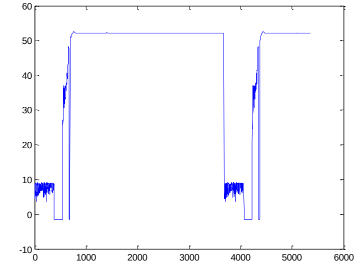
- Power, speed, temperature in steady state of motors, gas turbines, etc.

## Stochastic (non-cyclic)



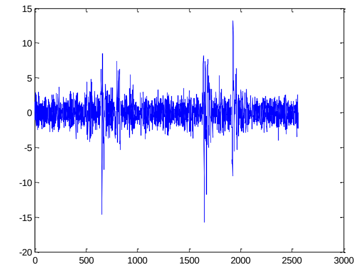
- Power, speed in wind turbine operation

## Cyclic (consider each period individually)



- Power, speed, pressure in manufacturing process, gas turbine startup, etc.

## Waveform (consider multiple period together)



- Vibration sensors, acoustic sensors

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

# Properties of extracted features

Properties

Presence

**Remarks**

Ease of Explainability

Handle Missing Data

Ease of Maintenance

Uncertainty Handling

Subjective Opinions

Data Property

Domain Understanding

Sample rate

Yes

Rule Out Black Box Methods
Spatial Space Features
Statistical Moments
Bayesian Models /Rules

No

Feature Space Transforms
Pattern analysis methods

# Application domain

Category

- Mechanical, structural, thermal, electrical, chemical, ...

Systems

- Machine tool, vehicle, aircraft, locomotive, wind turbine, construction machinery, ...

Common components

- Bearing, gearbox, motor, pump, engine, gas turbine, battery, ...

Many features extraction methods and data processing procedures come from domain know-how

GE imagination at work

# Domain specific feature extraction

**Failure Mode:** depending upon the failure type, certain rations, differences, DFEs, etc. are extracted for tracking over time

**Operating Mode:** specific sensors can be more/less critical in different operating conditions of machines…

- raw sensors to be used for feature extraction…
- variances under different conditions itself can form basis for further feature extraction

**Component Function:** Features extracted on basis of knowledge about specific components for which PHM desired…

**Known Relations:** Certain relation types can be assumed between variables of interest…this can affect features calculated for those relations

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

# Feature extraction method overview

- Data descriptive statistics
  - For sensors: RMS, variance, kurtosis, crest factor, correlation coefficient, ...
  - For events: count, occurrence rate, duration, time delays, ...
- Data descriptive models
  - Distribution models: Parametric distributions, histogram, ...
  - Information-based models: mutual information, minimal description length, ...
  - Regression models (use model parameters or modeling errors): curve fitting, AR models, ...
  - Classification/clustering models (use class label as feature), sequence matching likelihood
- Time-independent transforms
  - Explicit mathematical operations: difference, summation, ratio, logarithm, power n, ...
  - Principal component analysis, Independent component analysis, etc.
- Time series transforms (mainly for waveform signal)
  - Frequency domain, time-frequency domain, wavelet domain, EMD
- Domain dependent feature extraction
  - Physics based features: expected input-output or output-output relations, derived hidden states, etc.
  - Special procedures for data processing: operational regime segmentations, envelop analysis, etc.

# Data descriptive statistics

For sensors:

- One variable: RMS, mean, variance, kurtosis, crest factor, peak2peak, auto correlation...

$$\text{crest factor} = \frac{0.5(x_{\max} - x_{\min})}{\text{RMS}}$$

- Two variables: cross correlation

For events:

- Count, occurrence rate, duration, time delays, ...

# Data descriptive models

Distribution models:

- Parametric distributions, histogram, ...

Information-based models:

- mutual information, minimal description length, ...

Regression models (use model parameters or modeling errors):

- Curve fitting (linear, exponential, etc.), AR models, ...

Classification/clustering models (use class label as feature):

- Any pattern classifiers (Fisher discriminant, Bayes, etc.)
- Sequence matching likelihood

# Time-independent transforms

Explicit mathematical operations:

- Difference, summation, ratio, logarithm, power n, ...

Data dimension reduction transforms:

- Principal component analysis, Independent component analysis, etc.

Notes: These transforms

- Do not alter the number of samples
- Are usually used to produce feature from features

GE imagination at work

# Time series transforms

## Methods mainly for vibration analysis/waveform data

### Stationary signals

#### Frequency domain

- Spectral analysis
- Envelope analysis
- Cepstrum analysis
- Higher order spectrum

### Non-stationary signals

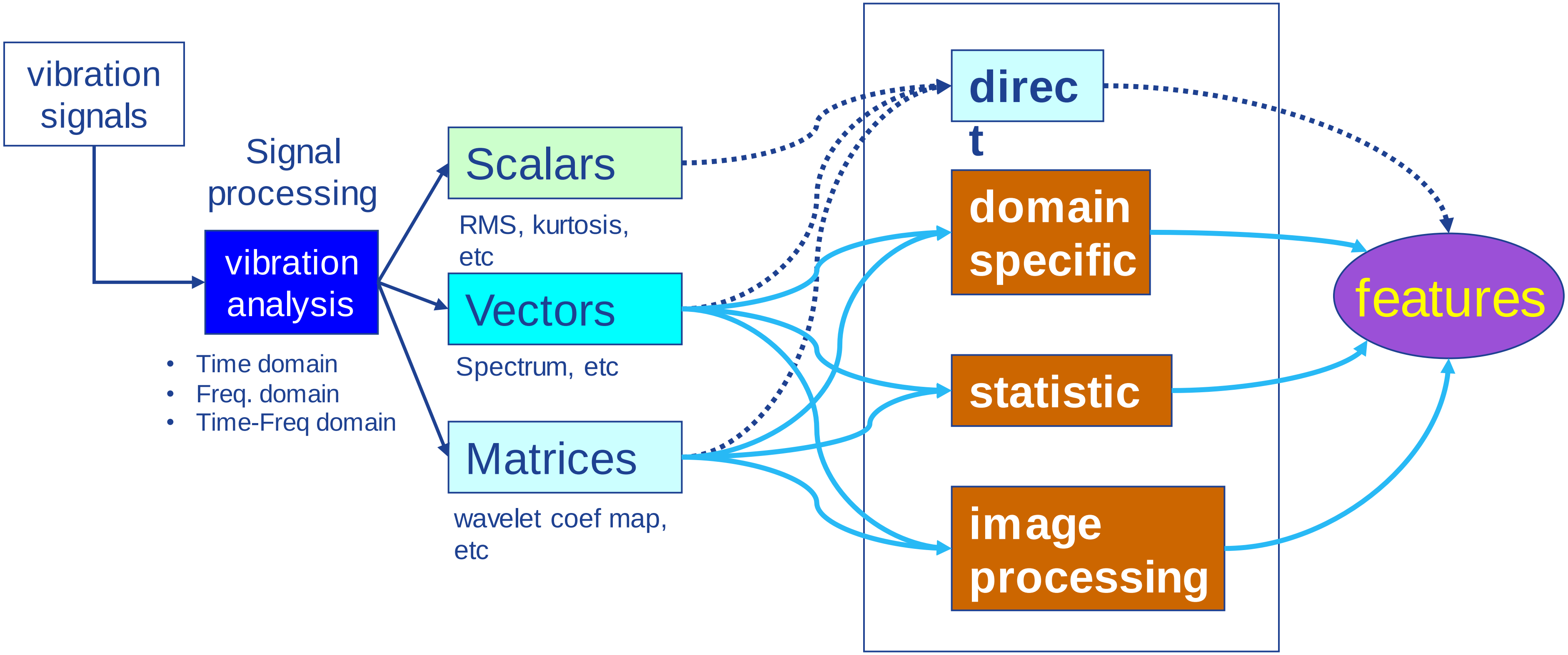
#### Time-frequency

- Short-time Fourier Transform (STFT)
- Wigner-Ville distribution (WVD)
- Empirical mode decomposition (EMD)
- Spectral kurtosis
- Cyclostationary analysis

#### Wavelets

- Continuous wavelet transform (CWT)
- Discrete wavelet transform (DWT)
- Wavelet packet transform
- Morlet wavelet
- Hilbert-Huang transform

# Feature extraction ≠ vibration analysis




GE imagination at work

# Domain dependent feature extraction

## Physics based features

- Simple input-output or output-output relations
- Errors between model output and observations
- Estimated unobservable states
- System identification parameters

Model based FDI approaches

## Special procedures for data preprocessing

- Time synchronous averaging
- Enveloping/demodulation
- Operational regime segmentation
- ...

GE imagination at work

# Domain dependent feature extraction: an example for bearing

## Bearing characteristic frequencies

$$Outer\ Race\ (BPFO) = \frac{N}{2}\left(1 - \frac{D_b}{D_p}\cos\theta\right) \times f_{sh}$$

$$Inner\ Race\ (BPFI) = \frac{N}{2}\left(1 + \frac{D_b}{D_p}\cos\theta\right) \times f_{sh}$$

$$Ball\ /\ Roller\ (BSF) = \frac{D_p}{2D_b}\left(1 - \left(\frac{D_b}{D_p}\cos\theta\right)^2\right) \times f_{sh}$$

$$Cage\ (FTF) = \frac{1}{2}\left(1 - \frac{D_b}{D_p}\cos\theta\right) \times f_{sh}$$

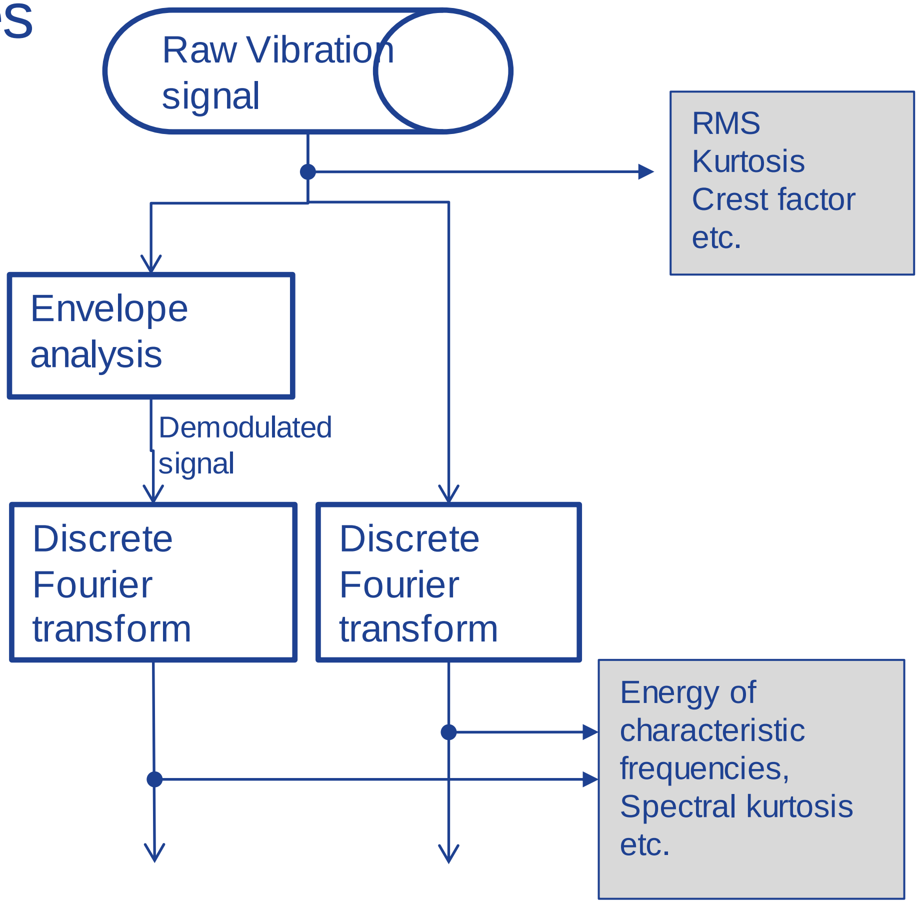
$N$ – *number of rotating elements*

$D_b$ – *rolling element diameter*

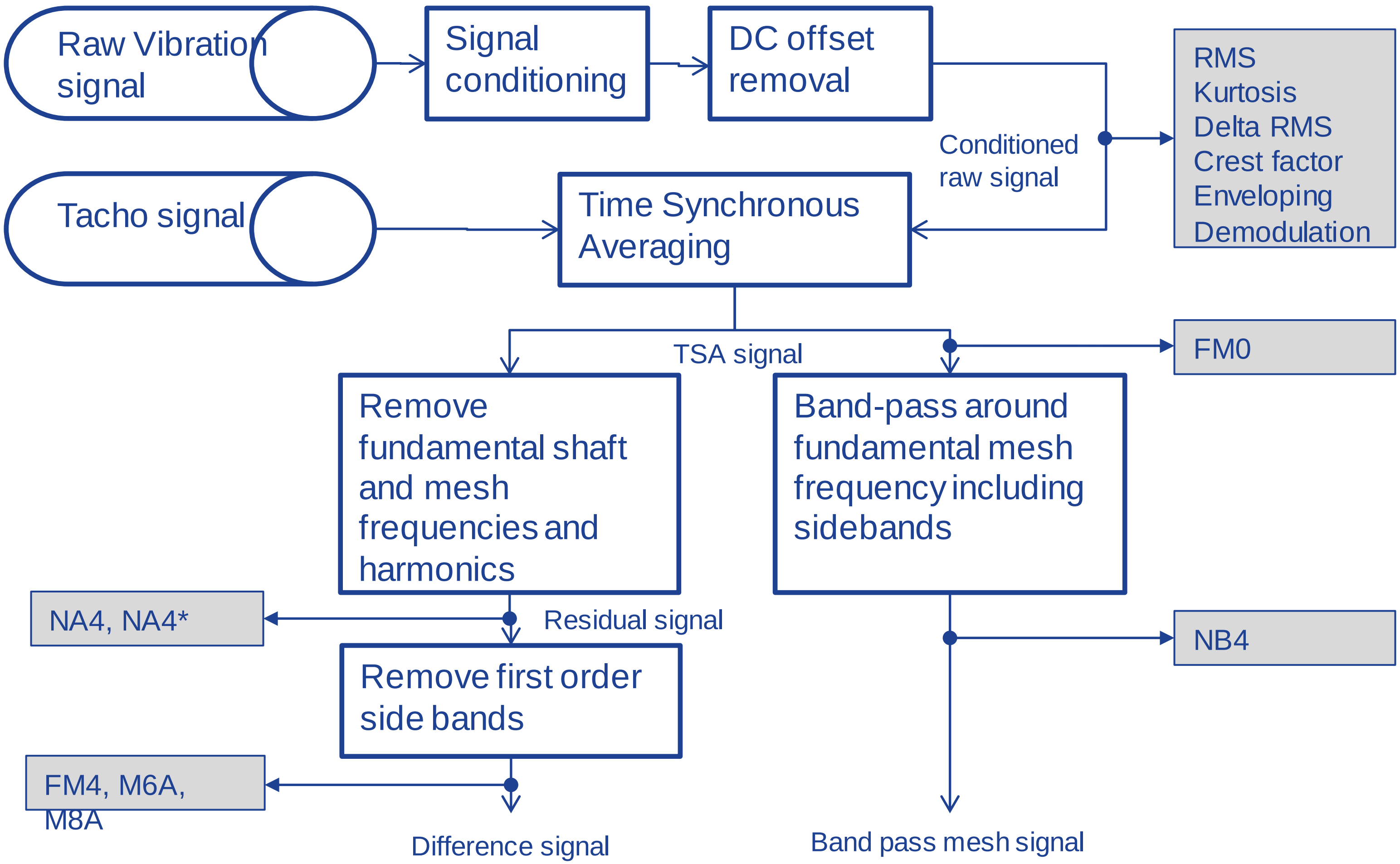
$D_p$ – *pitch diameter of rolling elements*

$\theta$ – *contact angle*

$f_{sh}$ – *shaft speed (Hz)*



# Domain dependent feature extraction: an example for gearbox




GE imagination at work

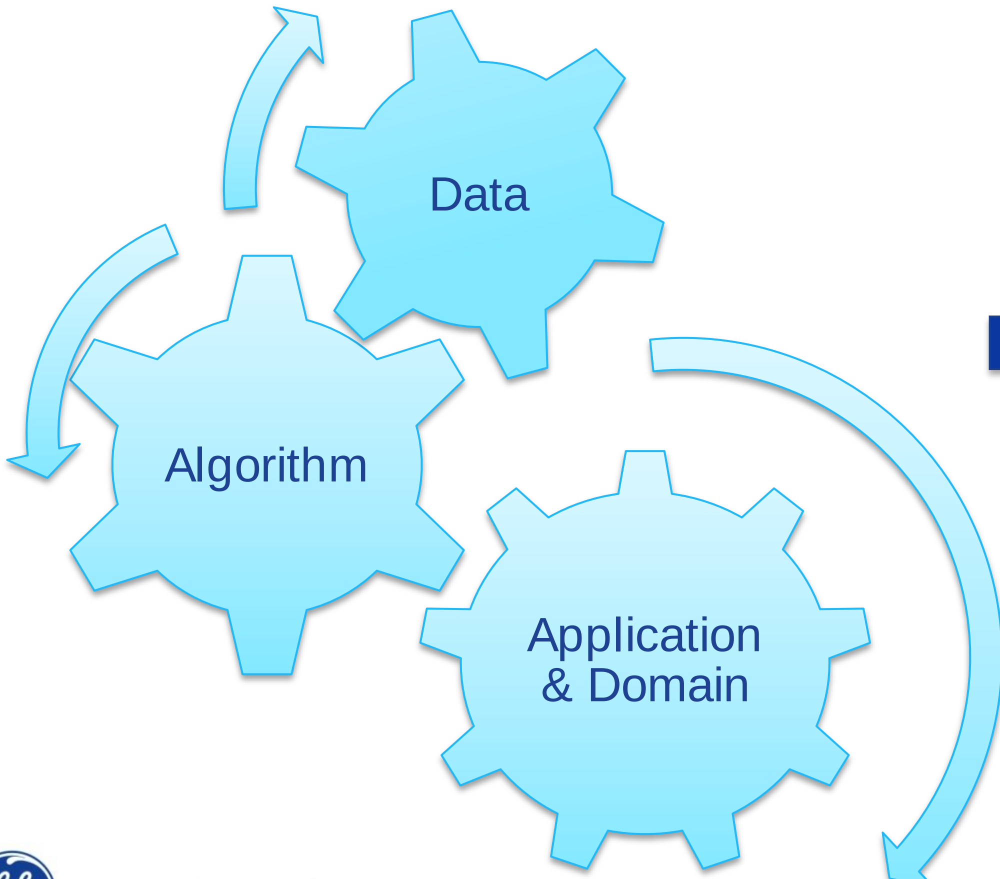

Lebold, M.; McClintic, K.; Campbell, R.; Byington, C.; Maynard, K., Review of Vibration Analysis Methods for Gearbox Diagnostics and Prognostics, Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, May 1-4, 2000, p. 623-634.

# Requirements/limitations of algorithms

Examples of what a feature extraction algorithm may care

- Continuous value?
- Evenly sampled data?
- Missing data handled first?
- Waveform? e.g. frequency domain analysis applicable?
- Presence of special signals? e.g. to apply Time Synchronous Averaging (TSA), Tacho & Vibration signals are required
- One, or two, or more sensors together? e.g. to apply correlation, PCA
- Similar measurements? e.g. to apply mathematical difference

# Exhaustive feature generation




Reasoning/prediction engine

↓

Features to extract

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

GE imagination at work

# Feature extraction process

Raw data/ preprocessed data

Data representation:

- Varying format: 1-D, 2-D time series, events, ...
- With exhaustive or ad hoc approach
- Dimensionality may reduce or increase
- Incorporate domain knowledge
- Underlying physical phenomenon

Feature generation

Feature set

Feature representation:

- A scalar or vector per feature
- One vector concatenating all features
- One matrix holding all samples of features

Feature dimensionality reduction

Two approaches

- Select a subset of generated features
- Transform the features to another space with lower dimensions

# Feature selection: what are good features

Desired characteristics of features

- High relevance to the objective, e.g., anomaly detection, diagnosis, degradation, PoD/FDR, etc.
- Low redundancy (linearly independent) among the features

Additional characteristic that are frequently overlooked

- Low relevance to non-objective factors, e.g. across assets, environment, usage pattern/ operating conditions, etc.

# Feature selection strategies

## Filter approach

- Metrics defined using local criteria different from the target models
- Search for ‘Good’ representation of raw data/features
- Computationally less-expensive

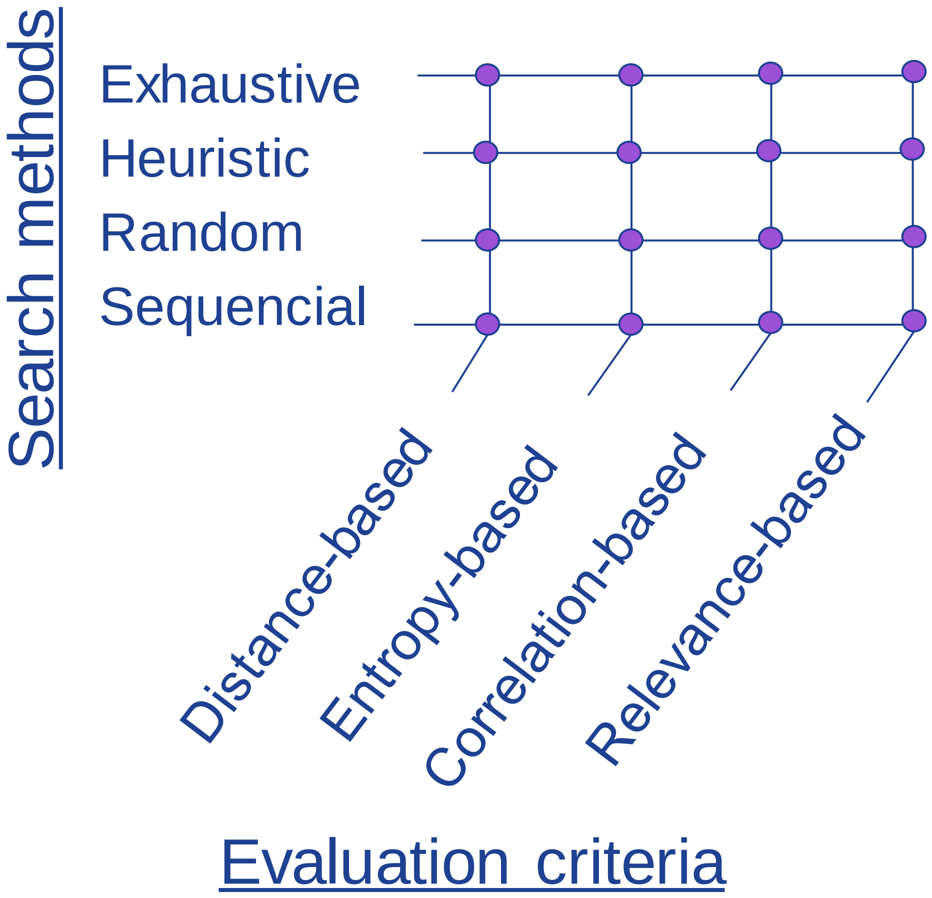
## Wrapper

- Metrics defined by the performance (accuracy) of the target models
- ‘Application’ specific
- Computationally expensive

## Embedded approach

- Feature selection built into the target model
- Regression: sparse regression, LASSO, etc.
- Classification: decision tree, regularized random forest

# Filter approaches




Examples

- mRMR (Minimum-redundancy-maximum-relevance )
- Fisher score
- Gini score
- Kruskal Wallis statistics

GE imagination at work

# Feature transformation

| Linear | Non-linear |
|---|---|
| • PCA (Principal Component Analysis)<br>• ICA (Independent component analysis)<br>• LDA (*Latent Dirichlet Allocation* )<br>• Latent semantic indexing<br>• Genetic Programming | • NPCA or KPCA<br>• NLDA or KLDA<br>• MDS (Multidimensional scaling)<br>• Principal curves<br>• Neural networks<br>• Genetic Programming |

# Outline

- Introduction
- Data characteristics
- Application & domain
- Feature extraction methods
- Feature dimensionality reduction
- Issues in real applications
- Summary

# Issues in real applications

## Issues:

- Features have high inconsistent (seemingly noisy) due to
  - Varying operating conditions
  - Asset-to-asset variations
- Features have low sensitivity to faults or degradation

## Handling methods

- Normalization / Standardization
- Feature of features (find generalizable features)
- Operating condition clustering & time series segmentation
- Use of local models for post-feature-extraction processing

# Example: aircraft engine

Ref: 2008 PHM data challenge

Domain: Aircraft engine

Signals:

- Operational variables: altitude, speed, thrust, ambient temperature
- Measurements: pressure, temperature at multiple location inside the engine

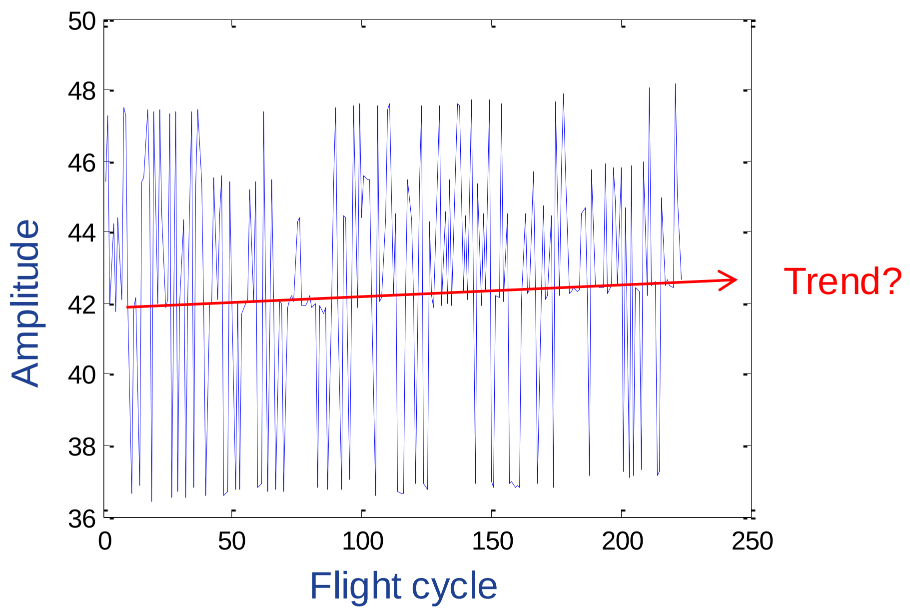
Feature extraction:

- Average of each signal during flight cruise (steady state).
- One feature vector per flight; one scalar per signal channel

# Example: aircraft engine (2)

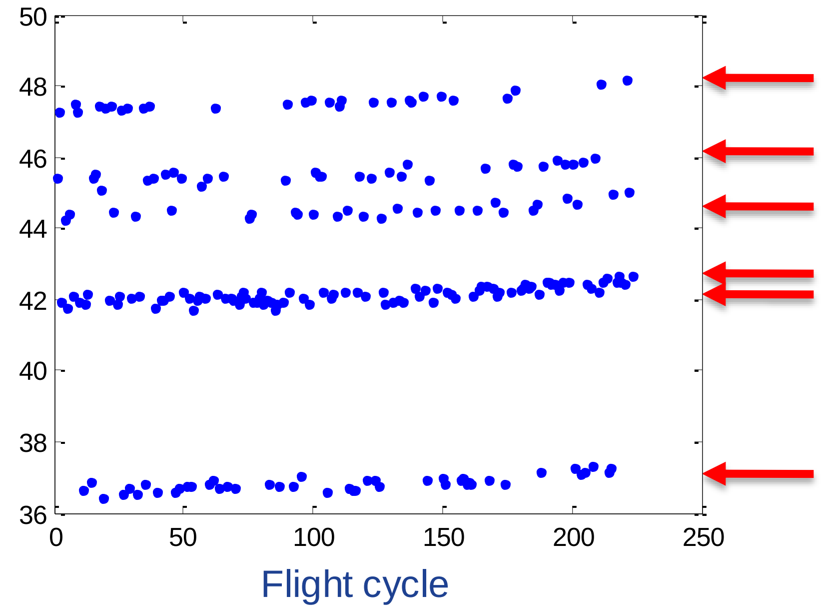
Ref: 2008 PHM data challenge

Run-to-failure time series of one feature: line plot




Seemingly random noise when considering the features time series as a whole

Run-to-failure time series of the same feature : dot plot




Trend more clear under each operating condition

GE imagination at work

# Example: aircraft engine (3)

Ref: 2008 PHM data challenge

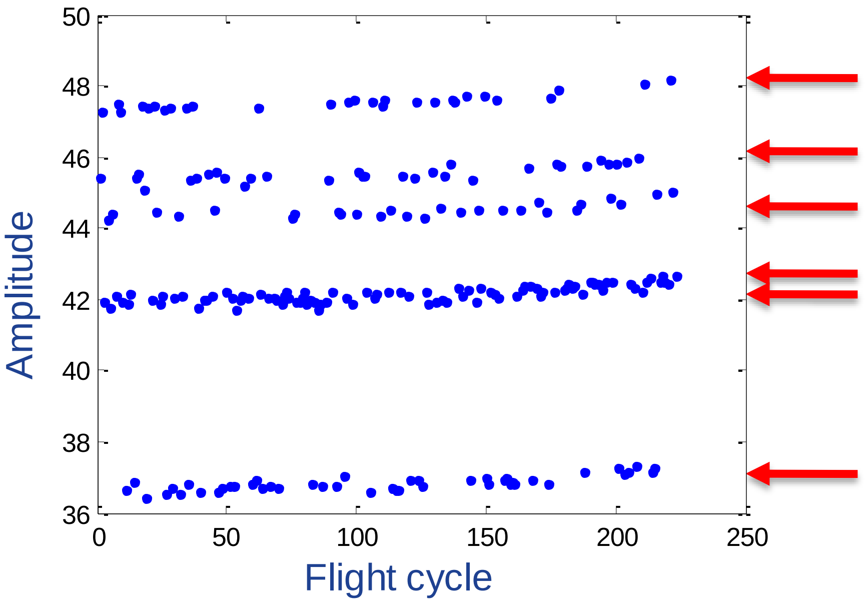
## Handling methods:

- Feature normalization
    - with physics model
    - with data-driven model

- Use of local models /multiple models for follow-up procedures

- Generate feature of features that is invariant to operating conditions

Run-to-failure time series of the same feature : dot plot




Trend more clear under each operating condition

# Key takeaways

- Procedure: feature extraction + dimension reduction
- What to extract: data property vs. application domain vs. algorithm requirements
- Feature extraction vs. signal processing
- Feature goodness: relevance and redundancy
- Feature selection: wrapper approach vs. filter approach
- Feature consistency and sensitivity issues

phmsociety

# A Tutorial on
# Feature Extraction Methods

Tianyi Wang
GE Global Research

Subrat Nanda
GE Power & Water

September 24, 2012

GE imagination at work