Industrial Analytics with IoT- A Game Changing Opportunity

Alan Rezazadeh

https://www.linkedin.com/in/alanrezazadeh/

Samuel Rostam

https://www.linkedin.com/in/samuelrostam/

Acknowledgments

CFREF

(Canada First Research Excellence Fund)

For providing resources to research industrial data analytics.

NYIT

New York Institute Of Technology

For providing access to IEEE Research and Tools

EDP Renewable Energy

Wind Farm, Spain

For making the operational wind turbine data freely available as open-data for research and education.

Arcitura Education

Professional opportunities with global enterprise clients, key industries and promoting vendor neutral, best practices; catalog of Design Patterns

https://www.linkedin.com/in/alanrezazadeh/

Alan.Rezazadeh@sait.ca

https://www.linkedin.com/in/samuelrostam/

srostam@nyit.edu

Agenda

Preamble: Digital Transformation - Industry 4.0

- Analytics Maturity & Frameworks
- Motivation, Bits & Atoms
- An IoT System Framework

Case Study: Industrial Big Data Analytics (Wind Turbine Case Study)

- Wind Turbine Design and Conventional Performance Measures
- Operating Mode Detection - Unsupervised Machine Learning (Clustering)
- Further Optimization with the Operating Modes
- Generator Bearing Health Monitoring using Temperature fluctuations

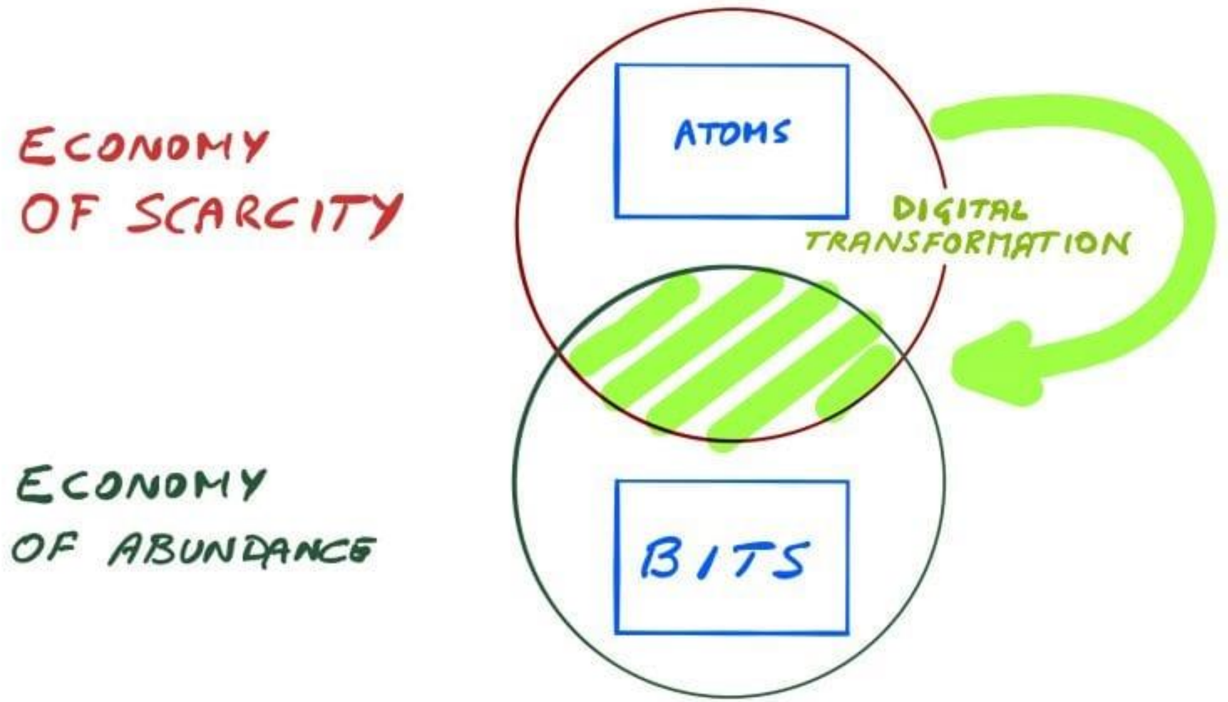
The Methodology:

- Architectural Model with Design Patterns
- Challenges & Opportunities
- Lessons Learned

Digital Transformation – Bits & Atoms

"From Bits to Atoms "
2018 Purdue University Engineering
Distinguished Lecture Series
Prof. Neil Gershfeld, Director at MIT

The real goal is not $$;
it is benefit to the society".



"Customers become Producers"

https://cmte.ieee.org/futuredirections/2019/01/26/digital-transformation-flanking-bits-to-atoms

Industry 4.0

Industrial Revolutions:

- 1st: 1760s
    - Textile manufacturing / Iron industry
- 2nd: 1871
    - Railroad and telegraph networks
- 3rd: 1940s
    - Computers / Microprocessors
- 4th: Now
    - Interconnectedness / IoT / Clouds

"The economy of bits is much more recent, although immaterial economy has roots that go far back in the past, like the economy of knowledge that was hyped in the last decade".

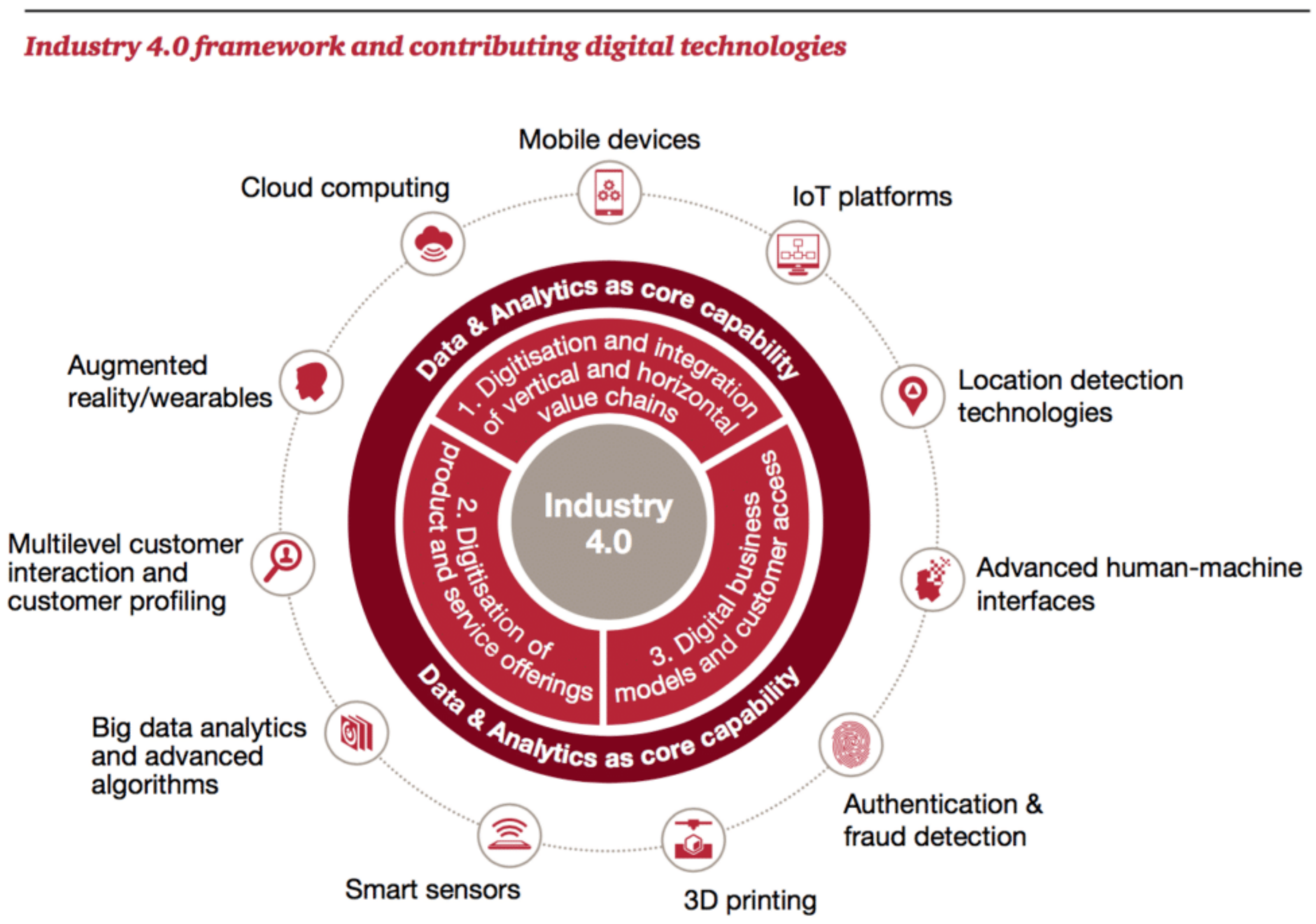
Industry 4.0 Framework - *Defined!*

Potential Markets

- Horizontal Markets:
- Vertical Markets :

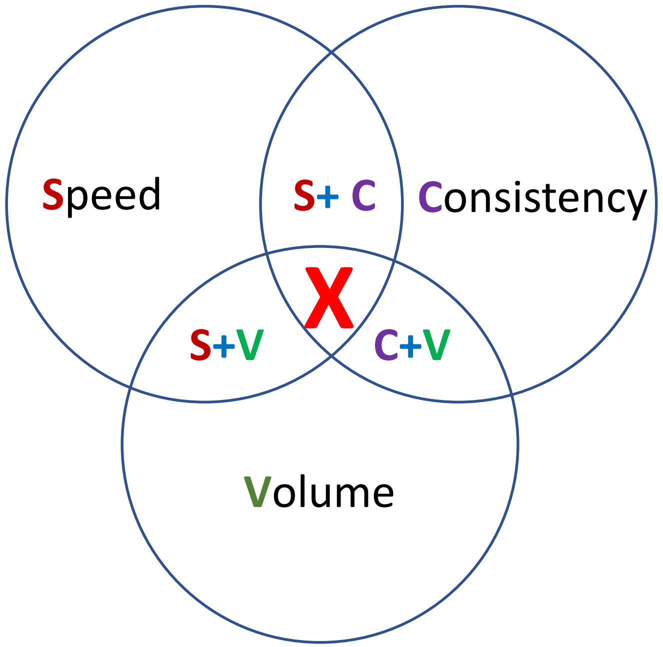
Services specific to trade & professions, telecom, healthcare, manufacturing, banking, ..

Complexity is magnified!

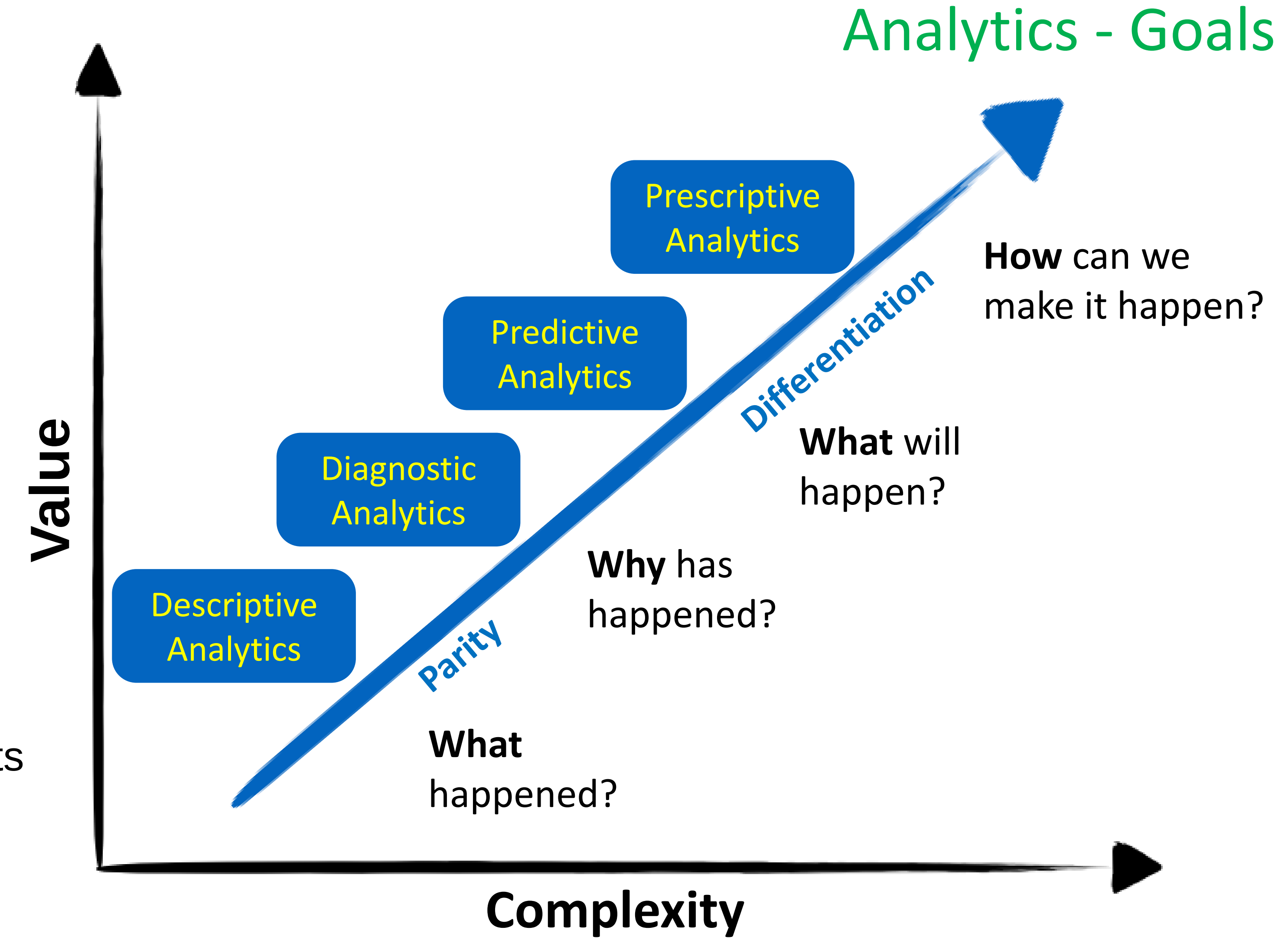


BCNET CONNECT

Big data Analytics – *Business Value & Maturity*



A Principle (related to CAP Theorem)
If speed (S) and consistency (C) are required, it is not possible to process high volumes of data (V) because large amounts of data slow down data processing.



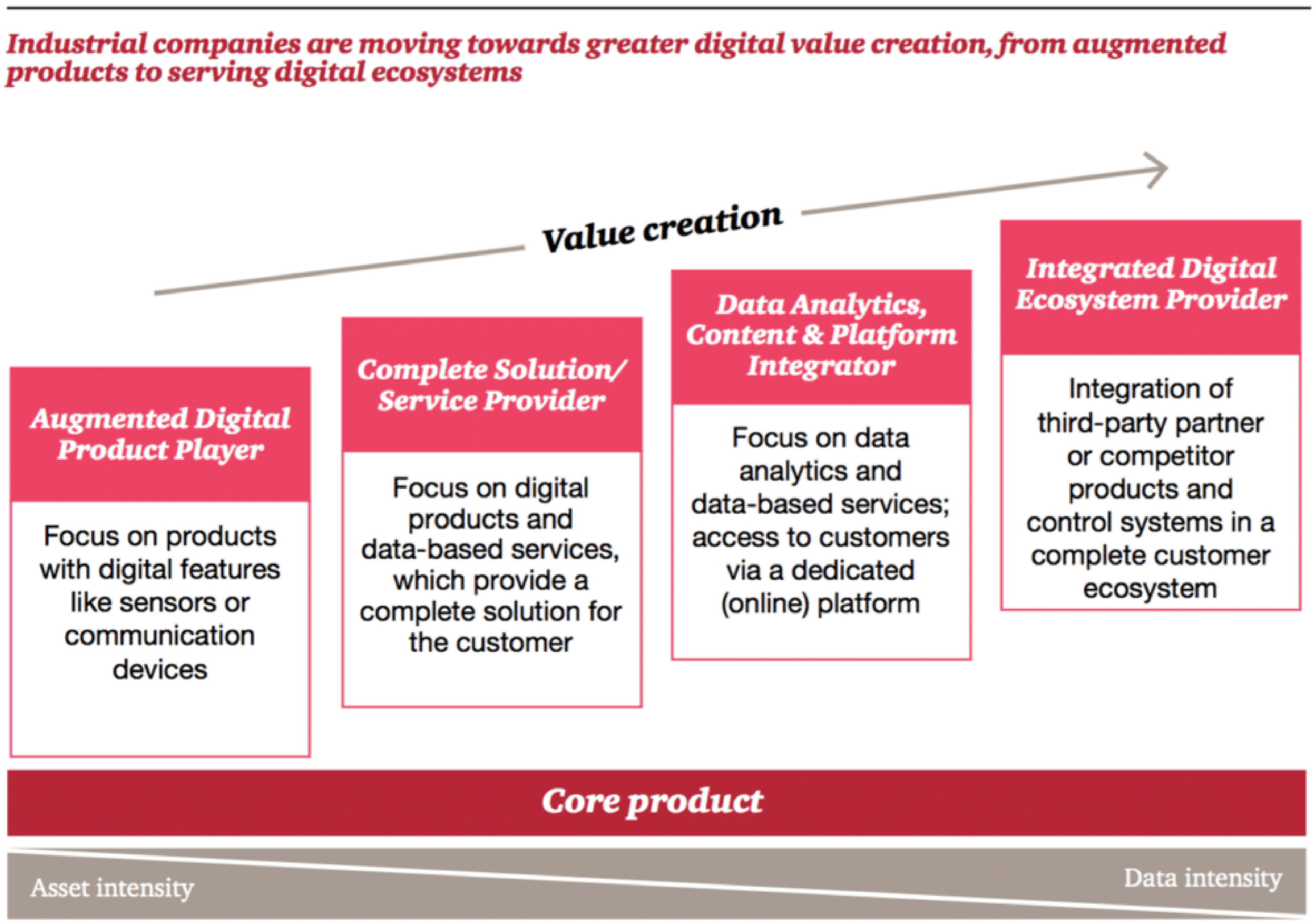
BCNET CONNECT

Generally, a real-time data analysis system will either be S+V or C+ V

Industry 4.0 Framework - *Goals!*

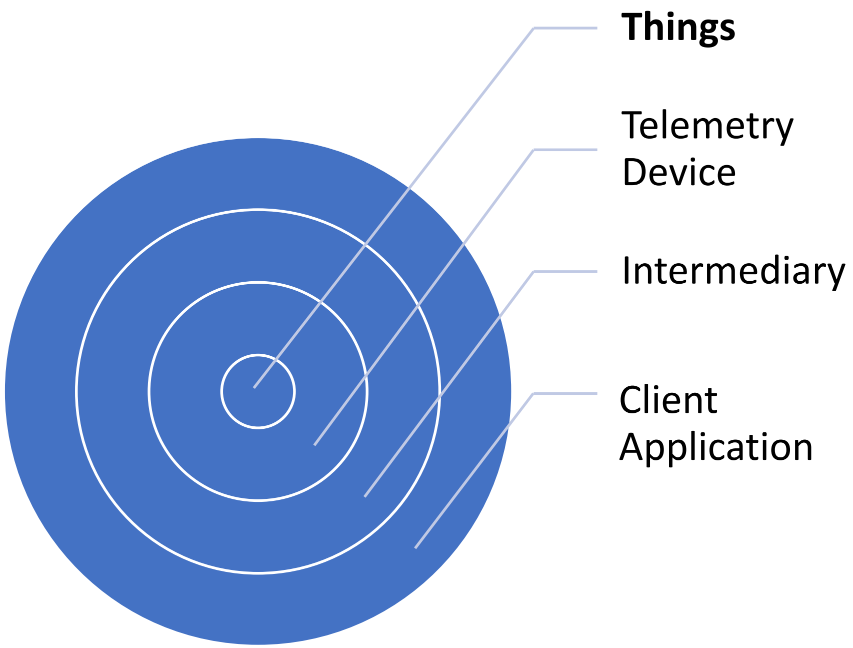
Is all for the sake of the Digital World?

Perhaps serving **analog ecosystems** as well!



IoT – *Components*

IoT evolved from machine-to-machine (M2M) Communication.

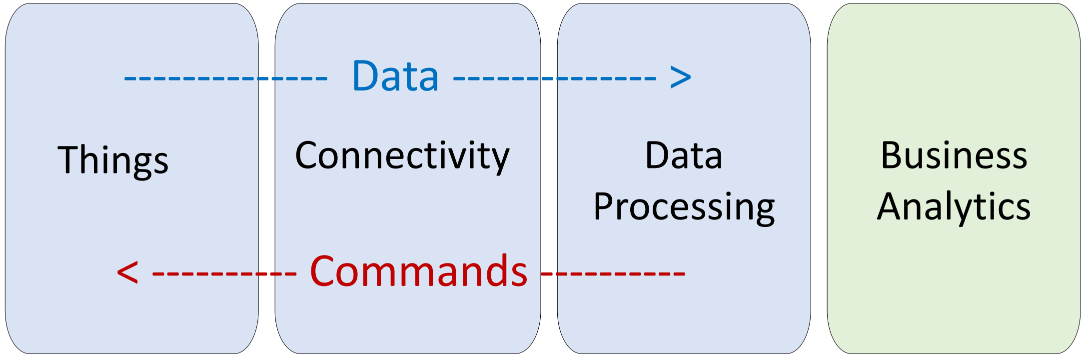


IoT System:

- An IoT system consists of a collection of devices connected by a network that may rely on gateways, telecommunications towers and satellites to cover a geographic range.

IoT Device:

- A typical IoT device is a physical hardware component with a power source and a built-in modem.
- Depending on its purpose, the device can contain one or more sensors and/or actuators.
- Some devices contain control logic that allows them to act autonomously, others are designed with an API that enables them to receive and act upon commands sent from an external source.



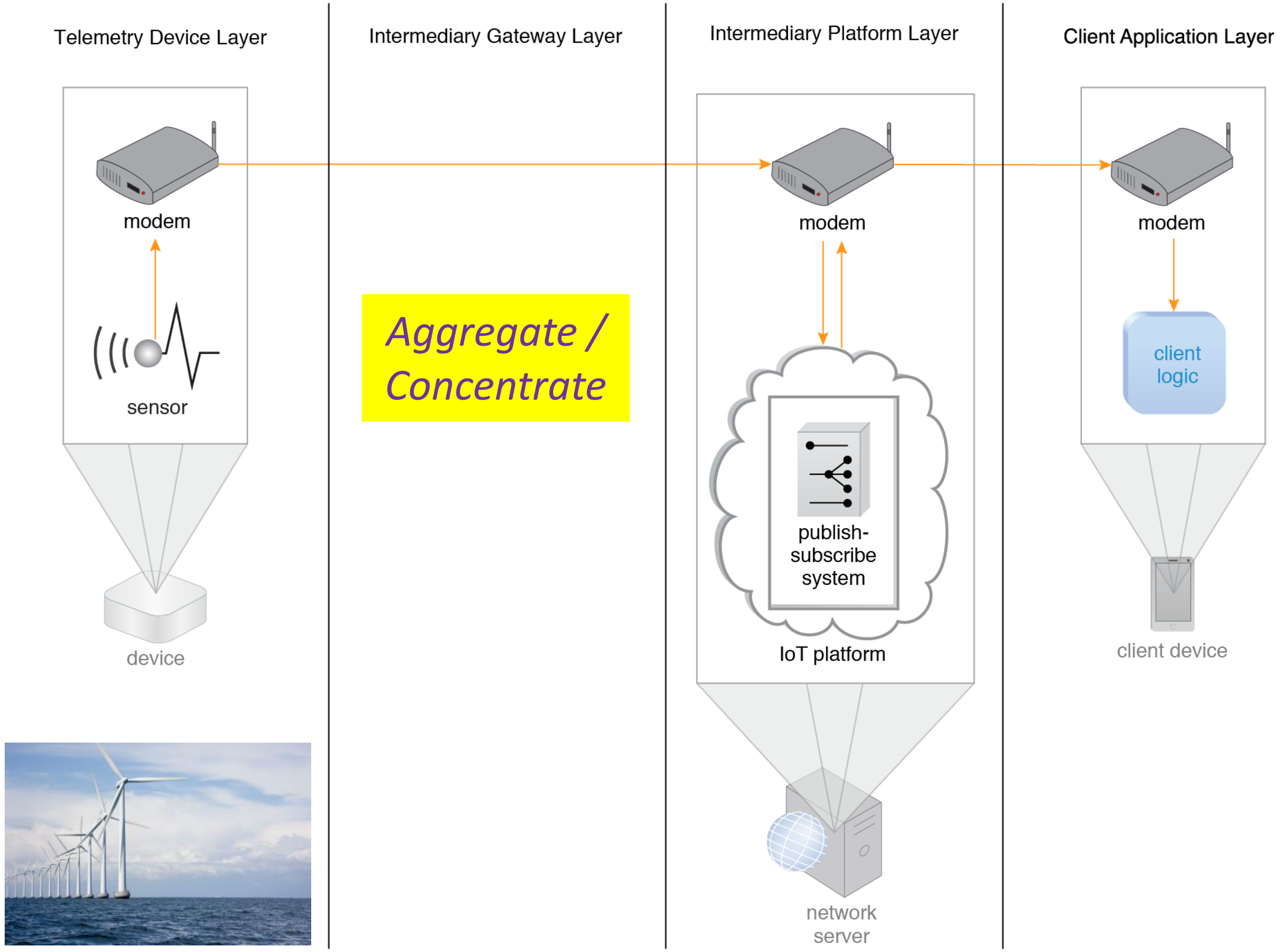
System Components

IoT – *A framework* – *Wind Turbines*

A diagram depict 4 basic layers with common parts of an IoT Solution.

- The Intermediary Gateway Layer may include additional components not limited to modems, control logics and network gateways.
- The Intermediary Platform Layer may include various Data Management, Integration enablement, and Analytics capabilities.
- The Client Application Layer may include a range of business automation solution or be a form of decision enabler leveraging AI or machine learning.

Separation of Concerns Principles



Edge Or Fog Computing

Cloud Computing

https://patterns.arcitura.com/internet-of-things-iot-patterns/non-controlling-sensor-device

BCNET CONNECT

Industrial Analytics – Typical Scope and Definition

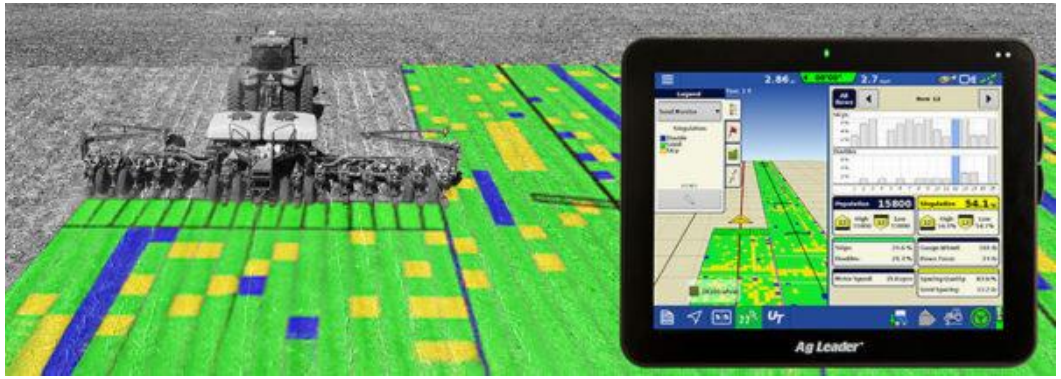
Large time-series data produced by sensors (IoT) used for Process Optimization, Knowledge Discovery and Decision Making.

Scope:

- Interconnected sensors, and actuators (control systems)
- Connected with computers' industrial applications
- Industries:
  - Agriculture
  - Manufacturing
  - Energy

Objectives:

- Optimizing Operations
  - Visibility into KPIs
  - Increasing productivity
  - Decreasing non-productive time
- Condition Monitoring
  - Monitoring degradation
  - Earliest "Initial Point of Detection"

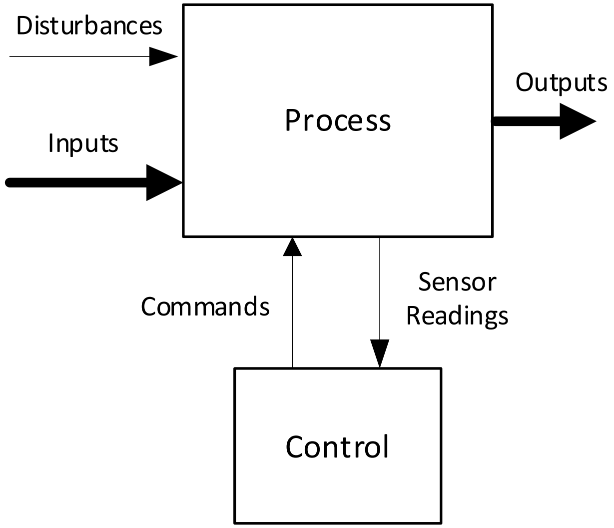


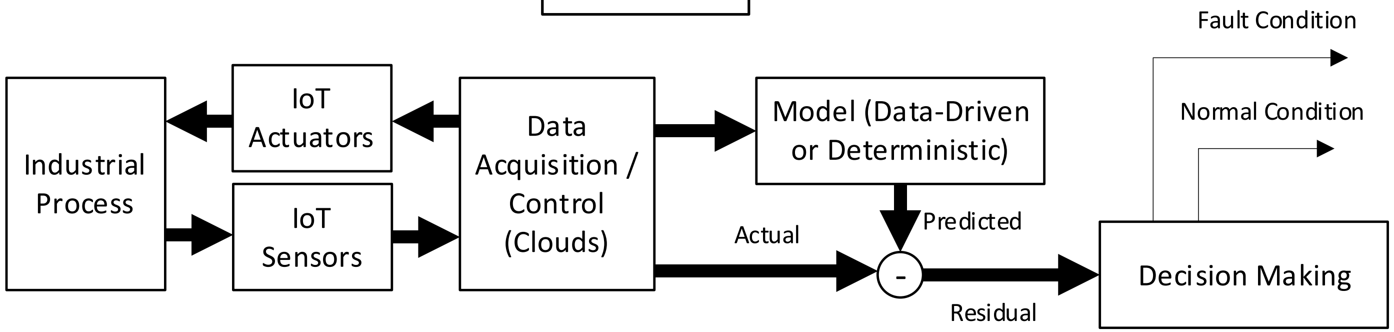
Industrial Analytics – Process Perspective

An industrial equipment is a process with a number of inputs and outputs

Definitions:

- Condition Monitoring:
  - Maintenance approach
  - Predicting machine health
  - Using sensor data and analytics
- Two main approaches:
  - Data-Driven (Stochastic)
    - Governing Laws are complex or **unknown**
    - E.g., industrial equipment, fluid dynamics
    - Larger and more complex systems
  - Model-Based (Deterministic)
    - Governing laws are **known**





Data Science in Heavy Industry and the Internet of Things. (2020)
Harvard Data Science Review https://doi.org/10.1162/99608f92.834c6595

Test Case: Wind Turbine Data Set

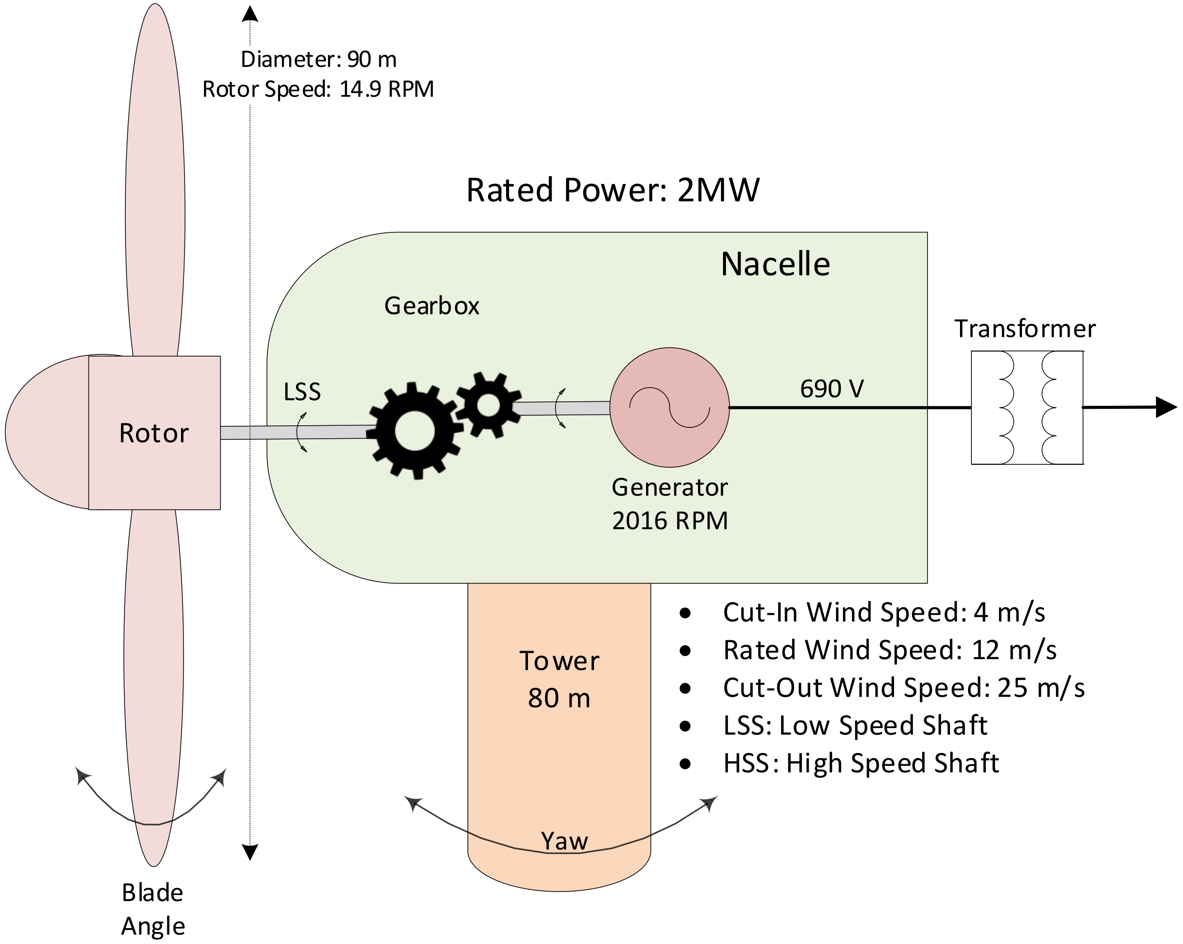
Wind Farm Data Set (EDP Energy):

- 5 turbines
- 2 years of monitoring
- 10 minutes time interval
- 104 K observations per turbine (total 520,000)
- 157 data elements (factors)

Actual work was performed with a partner on:

- 5 wind turbine data set
- 8 years of data
- 10 minutes time intervals
- 450 K observations per turbine (total 2.1 million)
- 450 data elements (factors)

https://opendata.edp.com/pages/homepage/



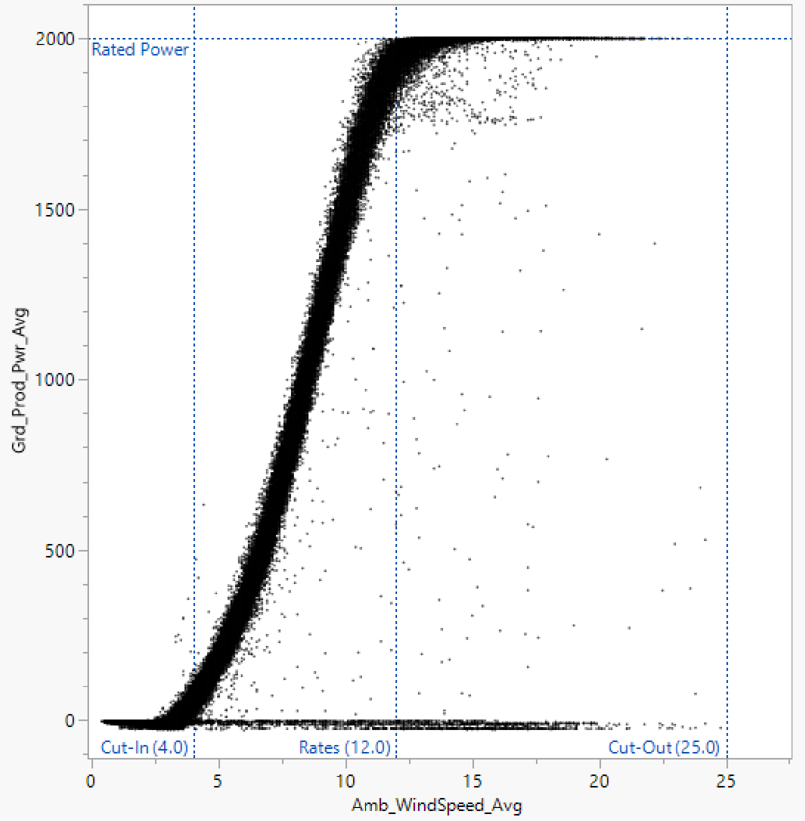
- Cut-In Wind Speed: 4 m/s
- Rated Wind Speed: 12 m/s
- Cut-Out Wind Speed: 25 m/s
- LSS: Low Speed Shaft
- HSS: High Speed Shaft

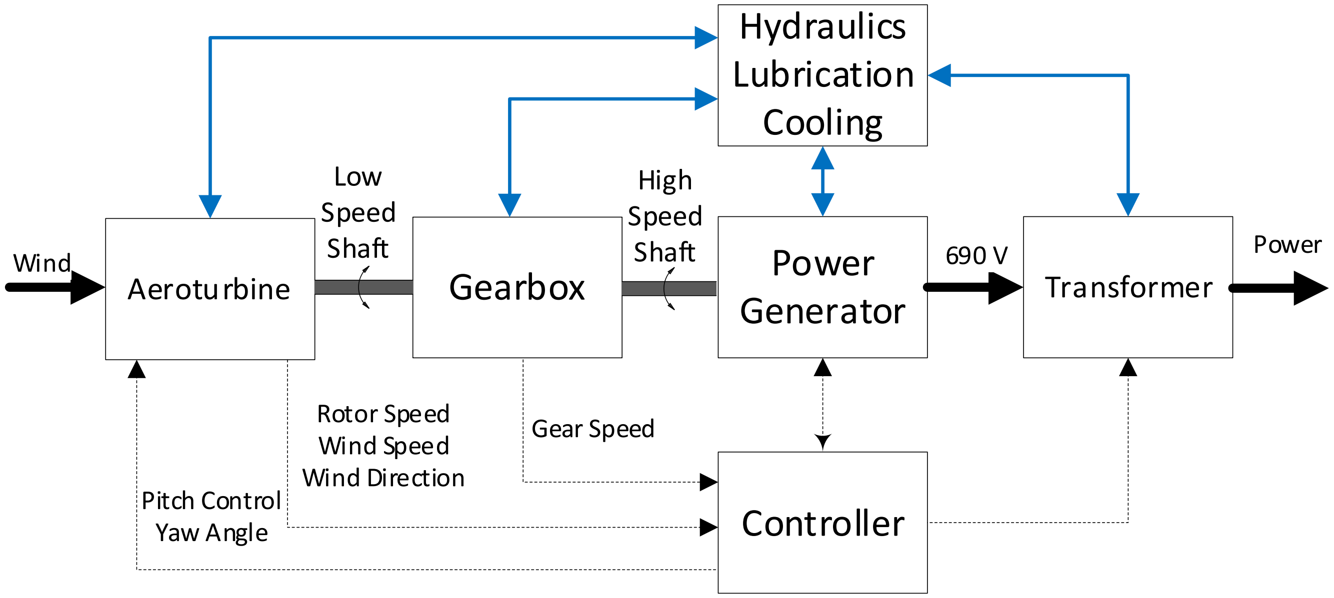
An open-data set is used due to confidentiality and legal reasons.
Open-data was used throughout the research as a comparison benchmark.

Understanding Complexity - (Wind Turbine)

Power Curve Displays (Model-Based):

- Produced power (output)
- Wind speed (input)





Notes:

- A complex industrial system is built of multiple components
- Each component has interaction with others
- Condition monitoring of a specific component is more successful than the whole system

Cloud-Integrated Cyber-Physical Systems for Complex Industrial Applications. Mobile Networks and Applications [*https://doi.org/10.1007/s11036-015-0664-6*](https://doi.org/10.1007/s11036-015-0664-6)

BCNET CONNECT

Data Preprocessing Steps

Missing Values

Are zeroes actual 0 or missing value?

Analyze patterns of missing values

Outliers

Remove any value physically not possible

Avoid changing values to “what they should be”

Date/Time Conversions

Finding time unit that makes sense (day/week/month)

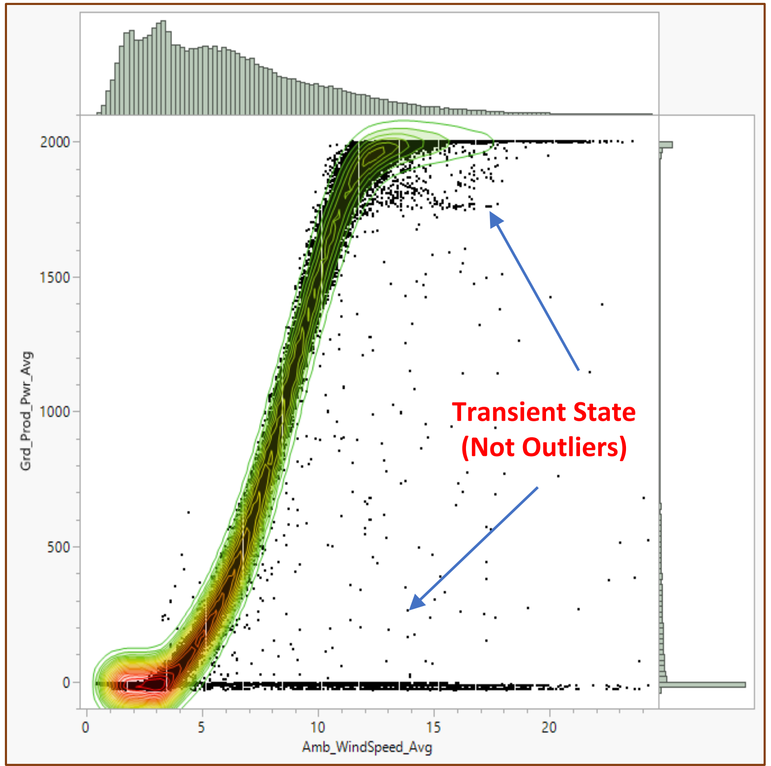
Time unit affects standard deviation calculations

Feature imputation:

Unnecessary imputations usually reduce model accuracy.

	Col 1	Col 2	Col 3
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Missing Value



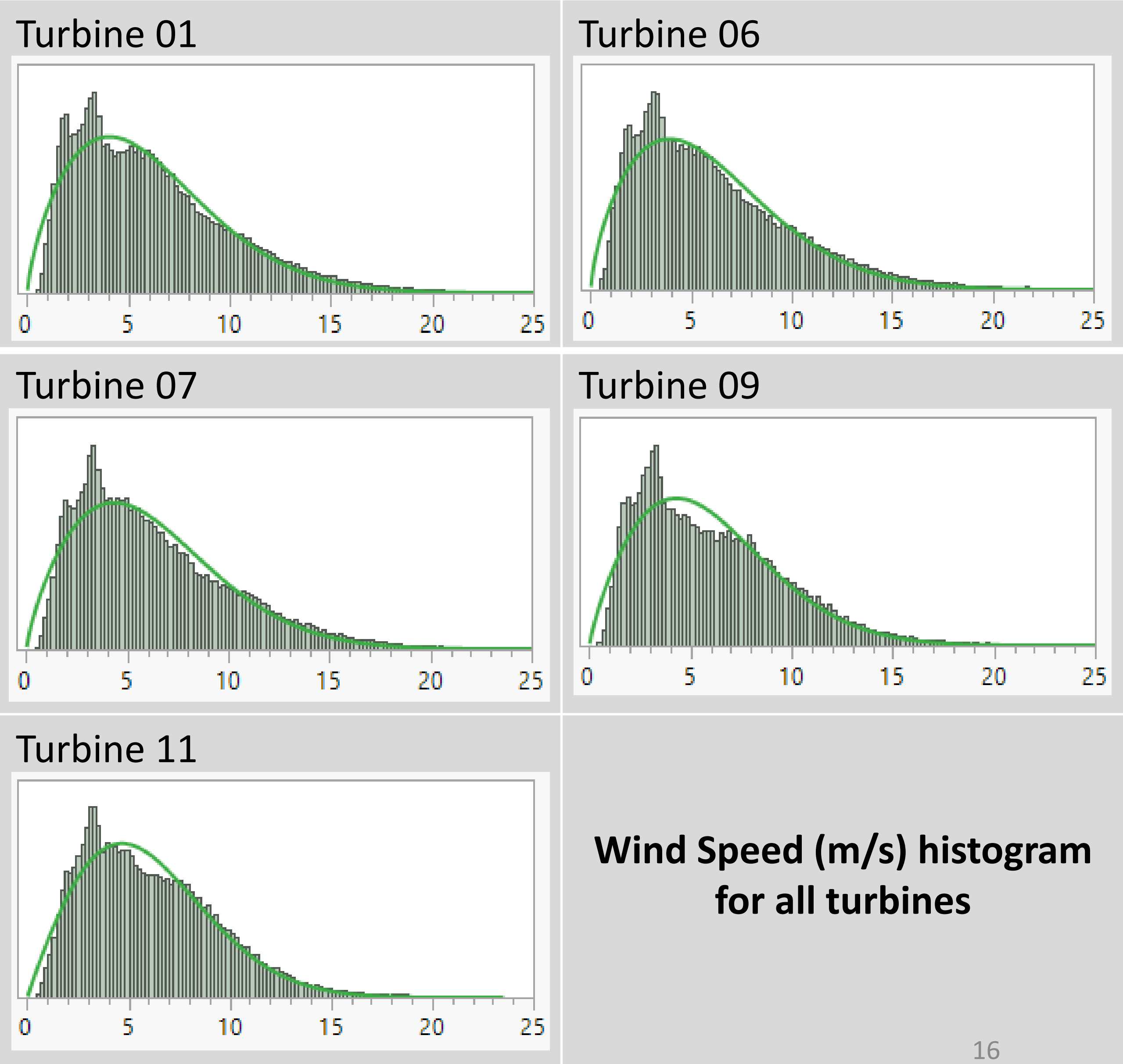
- Time period must cover a full cycle (24 hour)
- Duration of a full cycle must allow comparison of statistical measures (std dev, mean, …)
- For instance, a 24-hour duration includes both day and night temperature fluctuations.

Data preprocessing and intelligent data analysis. Intelligent Data Analysis, https://doi.org/10.1016/s1088-467x(98)00007-9

Univariate (One Variable) Analysis – Understanding Data

- Analyzing each variable, separately, for distributions, variance and missing values
- Validating data quality by comparing against governing laws (if exist)
- Wind speed usually conforms to Weibull distribution (model driven validation)
- The plots against Weibull indicate distortions exist, which could be due to interaction (wake) effect of turbines

Each deviation from physical laws, may offer an improvement opportunity



Wind Speed (m/s) histogram for all turbines

A Survey. IEEE Internet of Things Journal, 7(7).
https://doi.org/10.1109/jiot.2019.2958185

BCNET CONNECT

Operational Modes

Operational modes (steady-states) are:

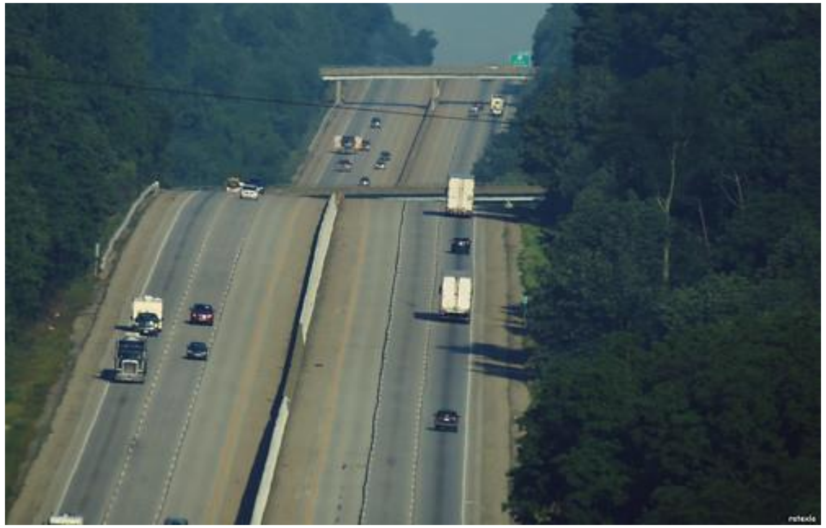
- Usually known by operators of process and equipment
- Can be discovered using clustering methods

Operational modes are dynamic:

- Due to change in the process inputs
- E.g., Ambient (weather) conditions, power demand,…

Optimization Opportunities:

- Analysis of transient states
- Smaller variance for steady states
- Detection of potentially smaller that are not expected



Prediction of Performance Deterioration:

- Performance to be monitored with the similar modes
- Examples of modes (steady-state operating modes)
  - Start up / Shutdown
  - Full/Partial Loads

Bagherzade Ghazvini, M., Sànchez-Marrè, M., Bahilo, E., & Angulo, C. (2021). Operational Modes Detection in Industrial Gas Turbines Using an Ensemble of Clustering Methods. Sensors, 21(23), 8047. https://doi.org/10.3390/s21238047

Identification of Operational Modes (Clustering)

Used Normal Mixture clustering:

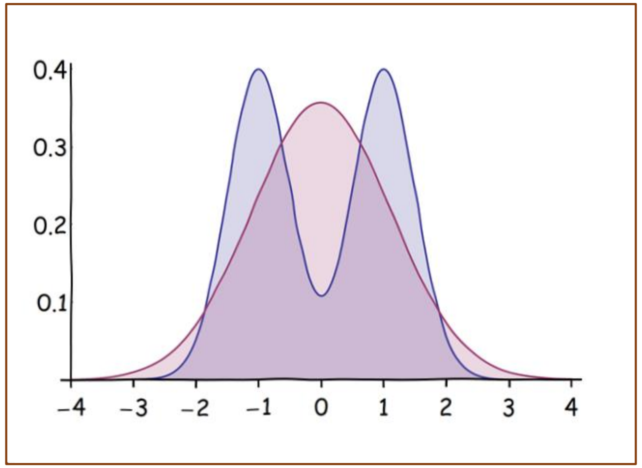
- Clusters can overlap
- Clusters are convex
- There is one high peak
- Probabilistic assignment of data points to clusters

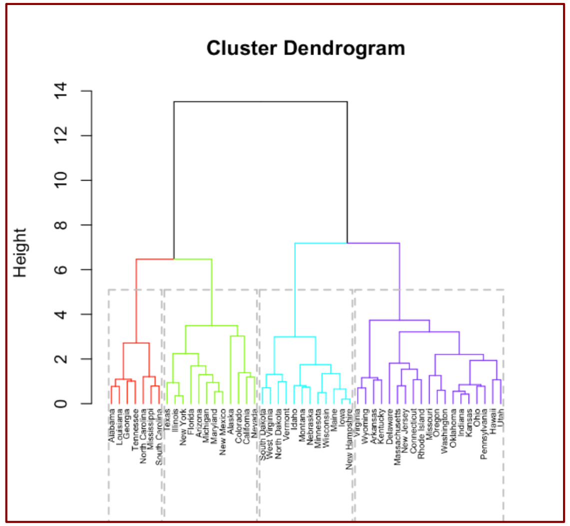
C. Biernacki, G. Celeux and G. Govaert, "Assessing a mixture model for clustering with the integrated completed likelihood," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 7, pp. 719-725, July 2000, doi: 10.1109/34.865189.

Hierarchical
- Used for larger datasets (business cases)
- Gives exact clustering logic

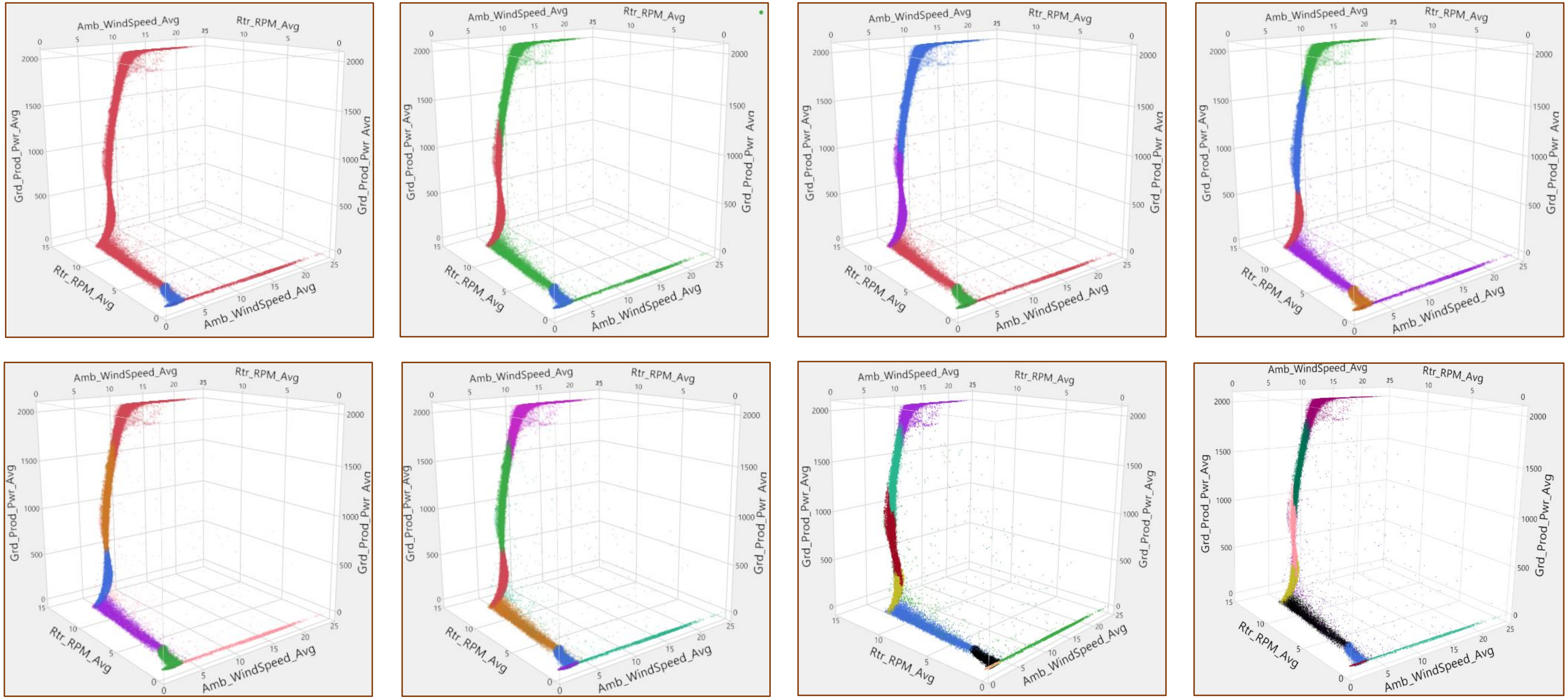
K-Mean
- Computationally intensive
- Not practical for large datasets

Normal Mixture
- Relatively fast
- Used for overlapping clusters





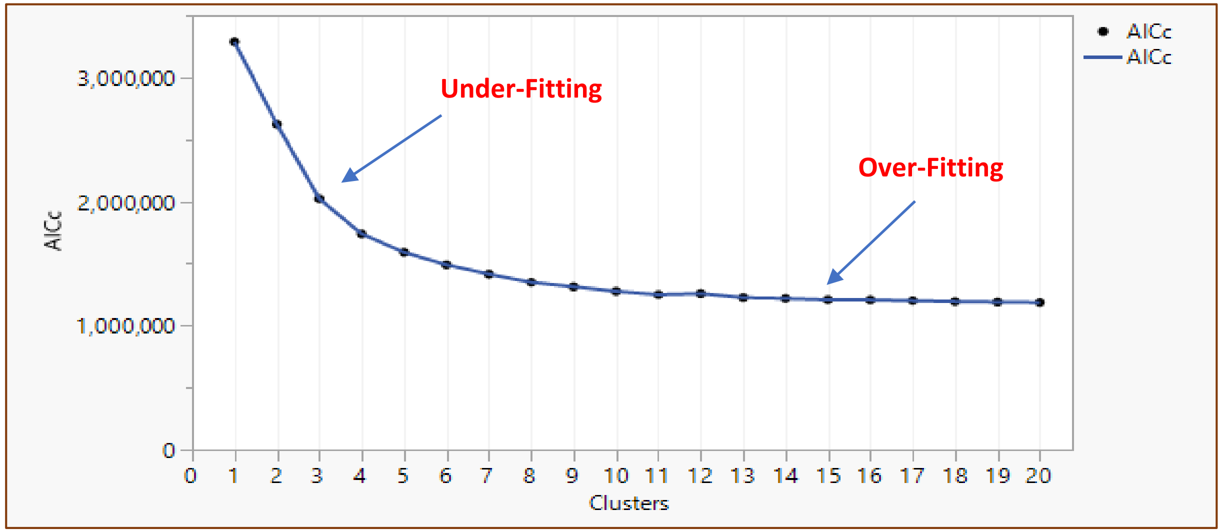
Identification of Operational Modes (Clustering)

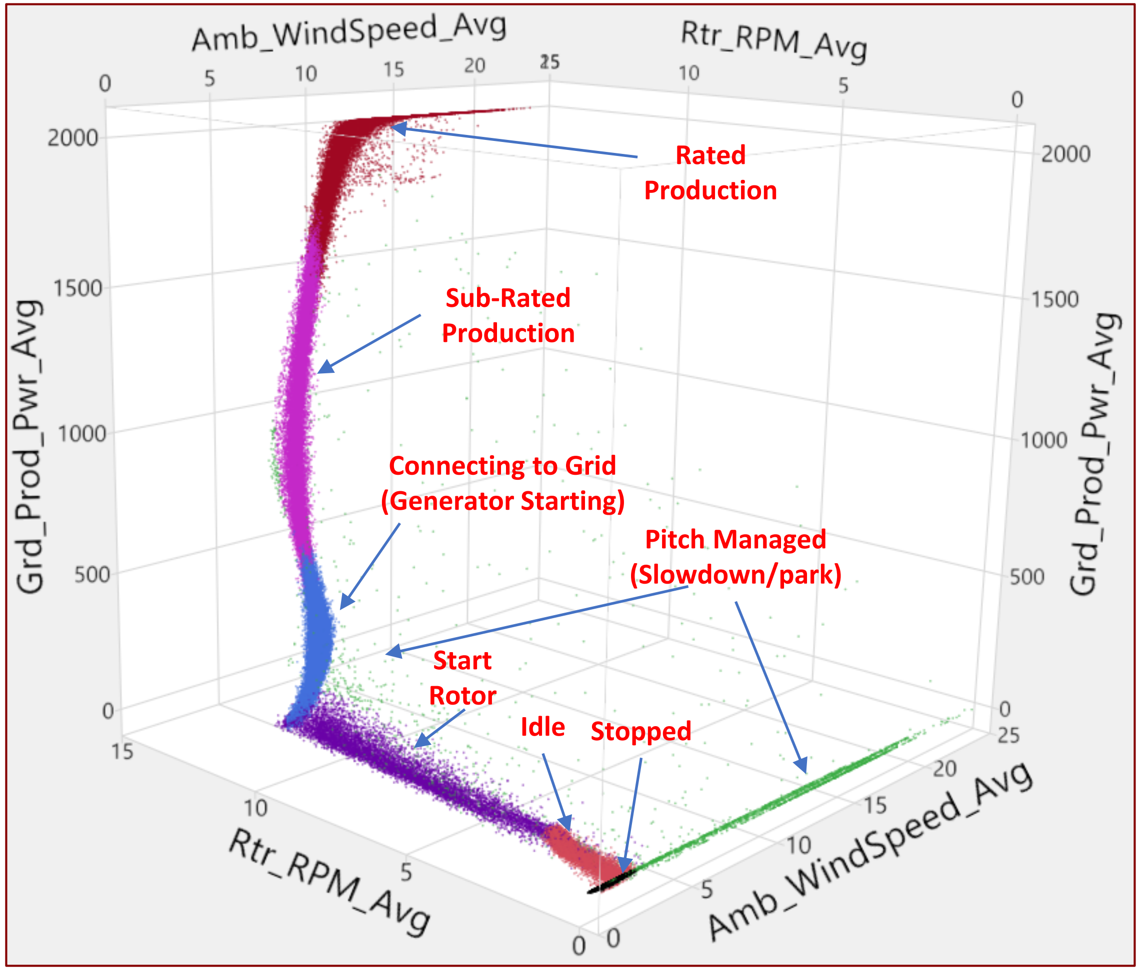


Identification of Operational Modes (Clustering)

Selected Seven Operating Modes (Clusters)

- Selecting minimum number of clusters where the reduction in model error is slowing
- Akaike Information Criterion (AIC) estimator of prediction error





Christensen, W., Model Selection Using Information Criteria (Made Easy in SAS®), University of California, Los Angeles, Paper 2587-2018

"Everything should be made as simple as possible, but not simpler! "
Albert Einstein

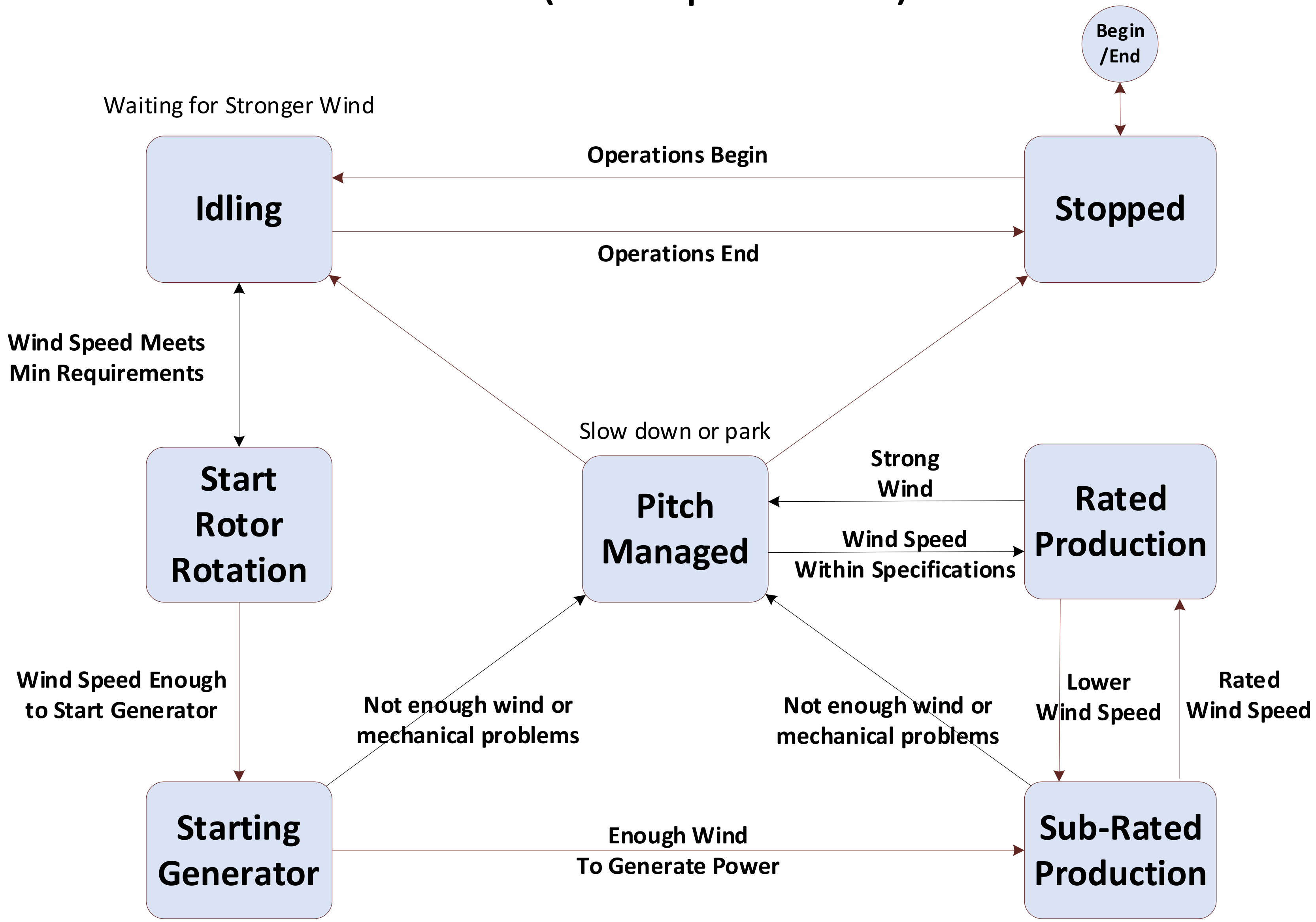
BCNET CONNECT

Data-Driven Operational Modes (Simplified)

Finding minimum parameters describing the process and operating modes:

- Wind Speed
- Power Generated
- Rotor Speed
- Blade Pitch Angle

Zaman, I., Pazouki, K., Norman, R., Younessi, S., & Coleman, S. (2021). DEVELOPMENT OF AUTOMATIC MODE DETECTION SYSTEM BY IMPLEMENTING THE STATISTICAL ANALYSIS OF SHIP DATA TO MONITOR THE PERFORMANCE. International Journal of Maritime Engineering, 159(A3). https://doi.org/10.5750/ijme.v159ia3.1026



Operation Optimization

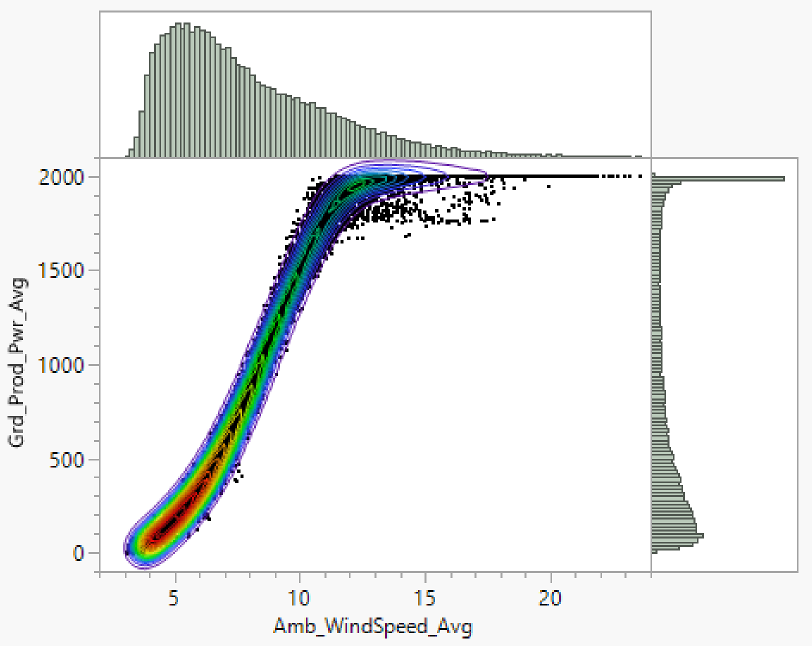
Optimization Objectives are:

- Reducing Rated-Production standard deviation
- Reducing size of Idle/Stopped steady states
- Maximizing Sub-Rated Production mean

		Wind Speed		Rotor RPM		Blades Pitch Angle		Grid Prod Power	
OperatingMode	N	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean
Grid Connecting	29624	0.8	5.2	0.5	11.5	0.7	-1.1	128.3	224.7
Idling	13861	0.4	2.7	0.7	1.6	0.0	24.0	3.6	-6.7
Pitch Managed (break)	3097	5.4	8.6	4.3	2.6	31.3	65.5	268.4	86.2
Rated Prod	13837	2.1	12.7	0.1	14.8	4.6	4.2	140.3	1872.9
Start	8133	0.4	3.4	2.6	7.3	6.8	10.7	24.7	12.3
Stopped	13109	0.4	1.6	0.0	0.0	0.0	24.0	2.0	-4.9
Sub-Rated Prod	23022	1.0	8.2	0.7	14.0	0.3	-1.9	324.7	928.1

Red indicates excluded data modes

Only using productive operational modes for analysis

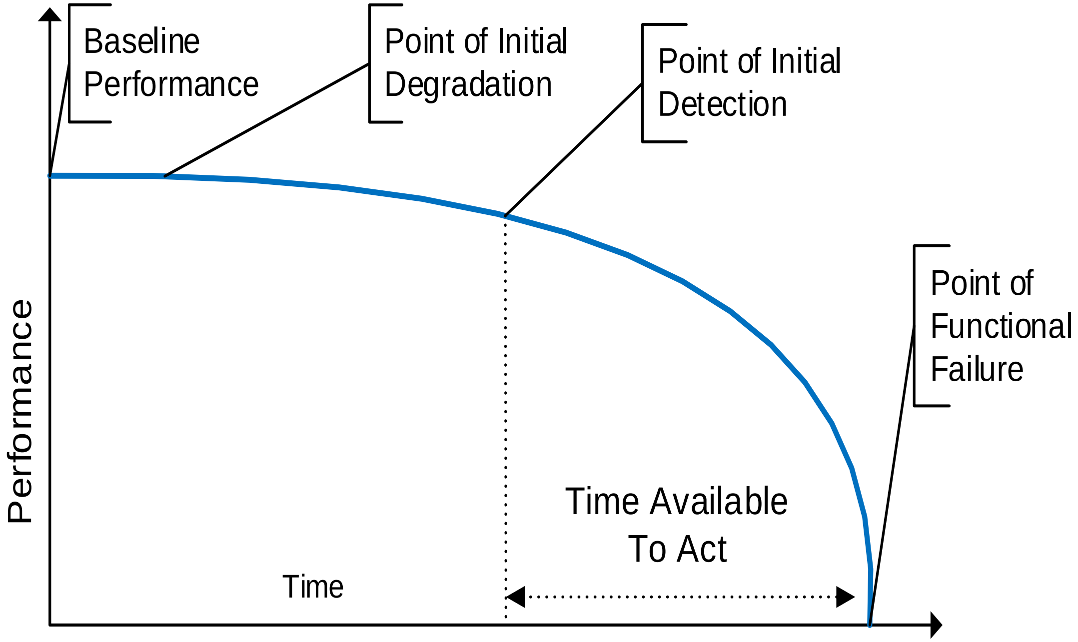


Generator Bearing - *Aging impact*

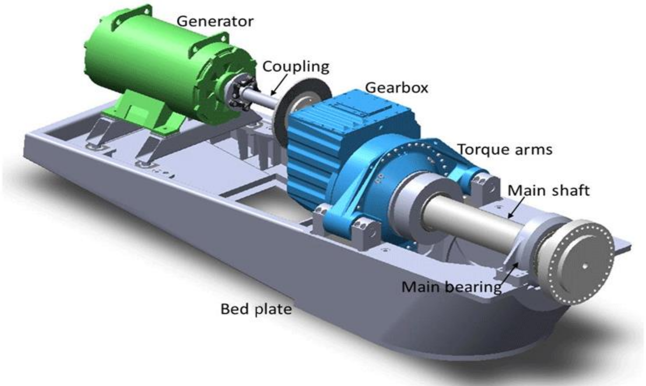
Performance Based (Data-Driven)	• Temp. Monitoring • Vibration/Noise Monitoring
Performance Based (Model-Based)	• Comparing against laws of physics • Manufacturer Specifications
Direct Monitoring	• X-Ray Gearbox • Corrosion/Erosion Measurements

Steps:

- Built baseline model of healthy conditions
- Preform monitoring against the baseline
- Detecting performance/temperature anomalies



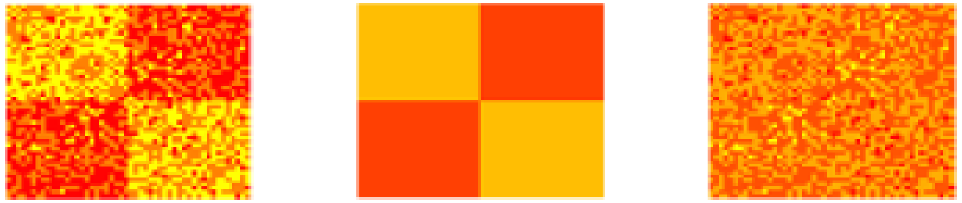
Based on John Moubray P-F Curve
In Reliability Centered Maintenance



Tidriri, K., Chatti, N., Verron, S., & Tiplica, T. (2016). Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. Annual Reviews in Control, 42, 63–81. https://doi.org/10.1016/j.arcontrol.2016.09.008

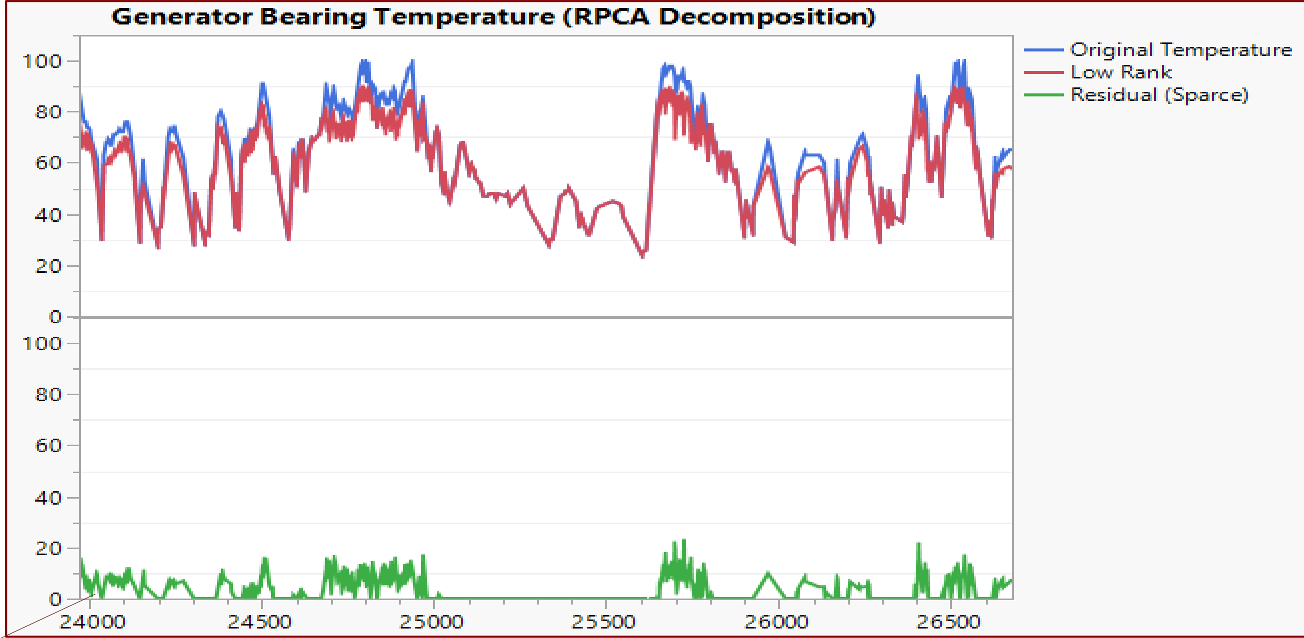
Robust Principal Component Analysis (RPCA)

- RPCA is statistical procedure to isolate "noise"
  - Fast and efficient
- Splitting signals to:
  - Low-Rank
  - Residuals/Sparse (noise)
- Low-Rank: defines the major process changes
  - Power Production
  - Ambient Conditions
- Residuals:
  - Fluctuations beyond expected process behavior



Original Low Rank Residual (Sparse)

Yuxin Chen, Princeton University, Fall 2020, ELE 520: Mathematics of Data Science



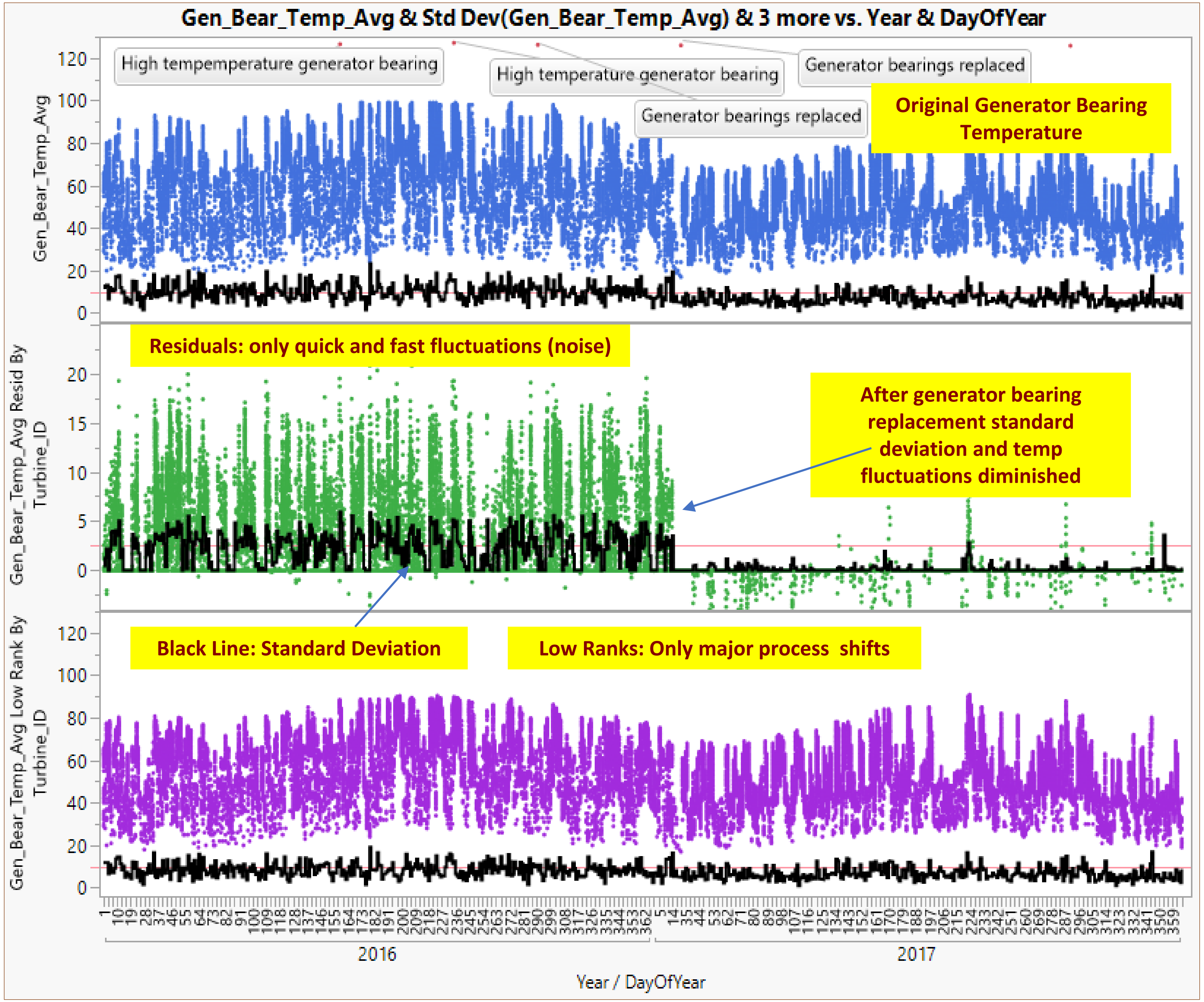
Delta

Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. 2011. Robust principal component analysis? J. ACM 58, 3, Article 11 (May 2011), 37 pages. DOI: https://doi.org/10.1145/1970392.1970395

Robust Principal Component Analysis

Splitting inputs based on major shifts in the process:

- Residual/Sparse:
  - high frequencies
  - fast rapid movements
  - Short term changes
- Low Rank
  - major shifts in process (steady state)
  - seasonal/major changes



BCNET CONNECT

The Methodology - *Simplified*

No Silver Bullet

- …access data & generate data: → assets of the business.

The Hidden Requirement :
Data and the solution environment itself needs to be regulated, standardized and evolved in a controlled manner.

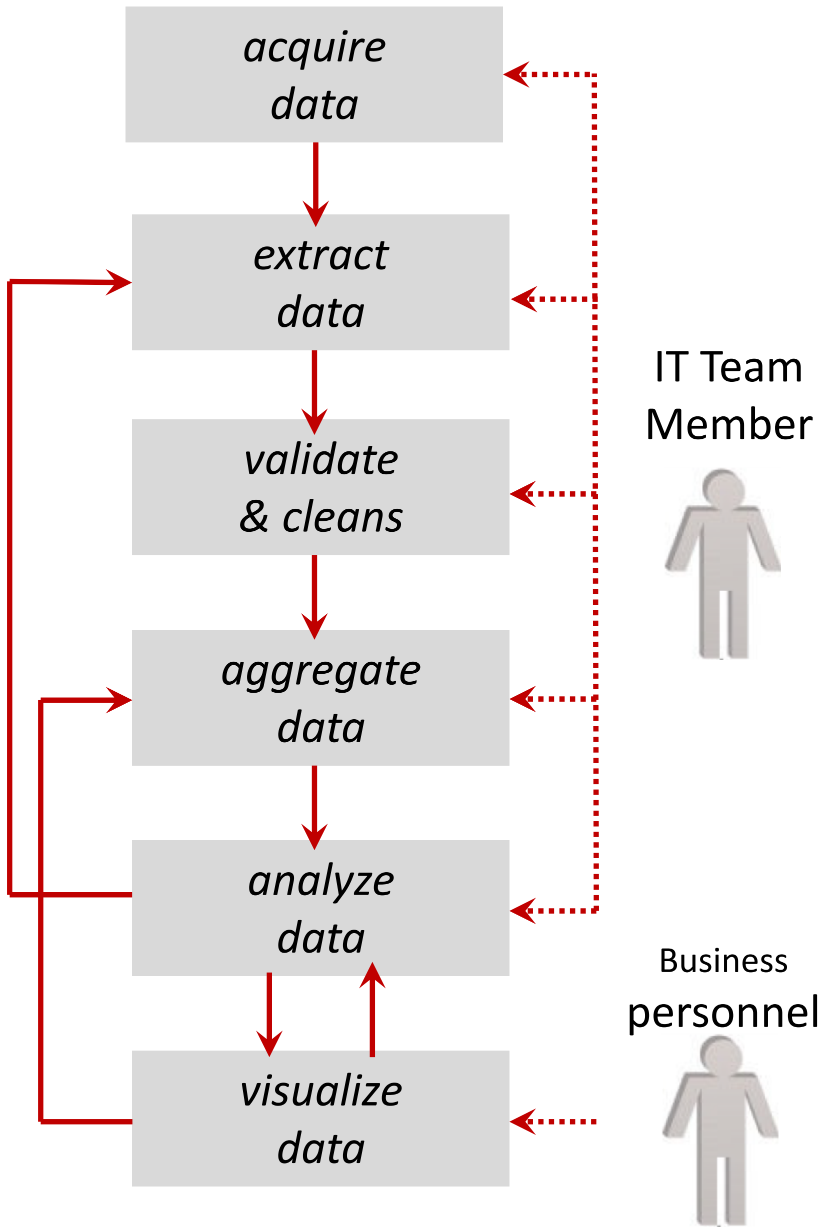
<u>Typical Capabilities include:</u>

Standardizing:

- ❑ Data tagging process
- ❑ Metadata design

Policies:

- ❑ regulate the kind of ***external data*** that can be acquired
- ❑ policies for data privacy and data anonymization
- ❑ data archiving data sources and analysis results
- ❑ data cleansing and filtering



Considerations: Governance implications!

Machine Learning Design Patterns

Data Exploration Patterns:
Central Tendency Computation
Variability Computation
Associativity Computation
Graphical Summary Evaluation

Model Optimization Patterns :
Ensemble Learning
Frequent Model Refactoring
Lightweight Model Implementation
Incremental Model Learning

Data Reduction Patterns:
Feature Selection
Feature Extraction

Model Evaluation Patterns :
Training Performance Evaluation
Prediction Performance Evaluation
Baseline Modeling

Data Wrangling Patterns:
Feature Imputation
Feature Encoding
Feature Discretization
Feature Standardization

Unsupervised Learning:
Category Discovery
Pattern Discovery

Supervised Learning:
Numerical Prediction
Category Prediction

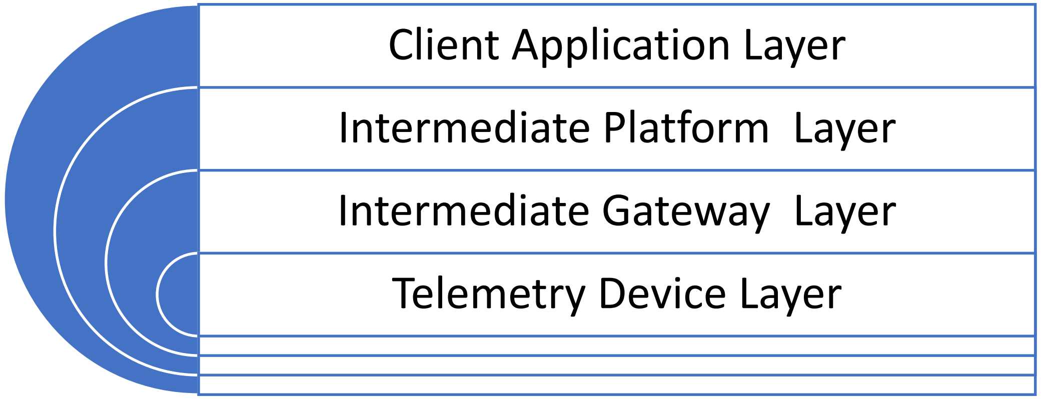
https://patterns.arcitura.com/machine-learning-patterns

Challenges & Opportunities

The Industrial IoT platform:

- Monitors IoT endpoints and event streams
- Supports and translates a variety of manufacturer and industry proprietary protocols
- Analyzes data at the edge and in the cloud
- Integrates and engages IT and OT systems in data sharing and consumption
- Enables application development and deployment
- Can enrich and supplement OT functions for improved asset management life cycle strategies and processes
- In some emerging use cases, it may obviate some OT functions.

Gartner Magic Quadrant Oct 2021



The layering architecture provides flexibility and extensibility.

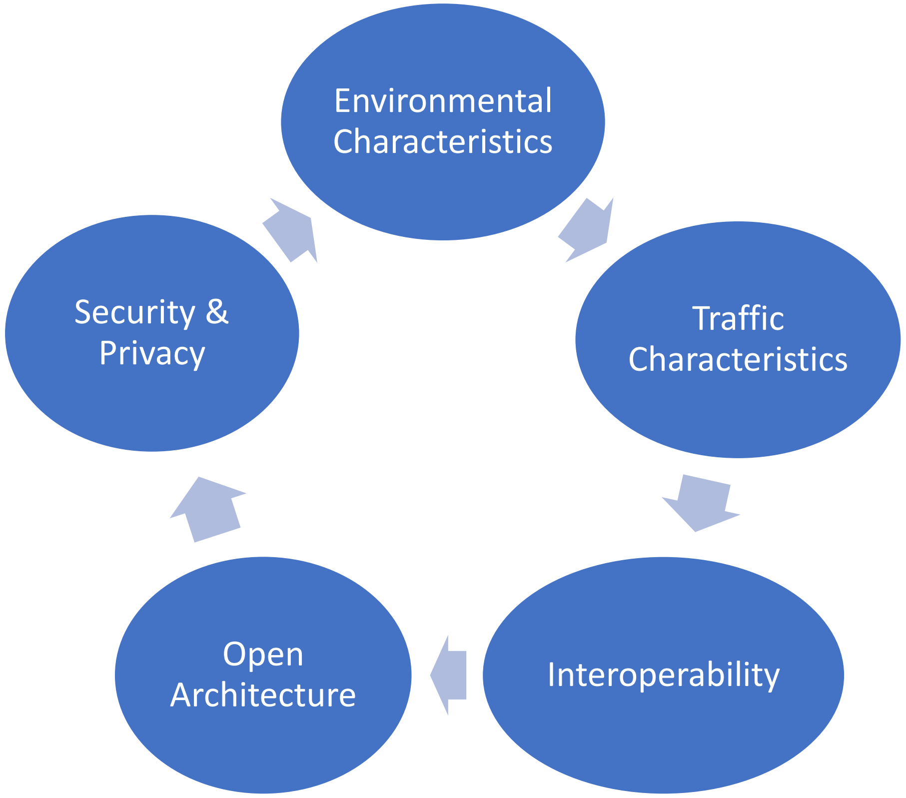
BCNET CONNECT

IoT – *Forces and Characteristics*

There are several underpinning issues that come into play in the architectural design as well as field deployment of IoT Applications.

It is necessary to map out different possible interaction scenarios that the underlying architecture needs to support, examples include:

- Flow of telemetry data
- Flow of command data
- Connectivity management
- Device Registry & Discovery
- Location & distribution of solution logic
- Transport protocols & Hardware components
- Nonfunctional capabilities such as scalability and security

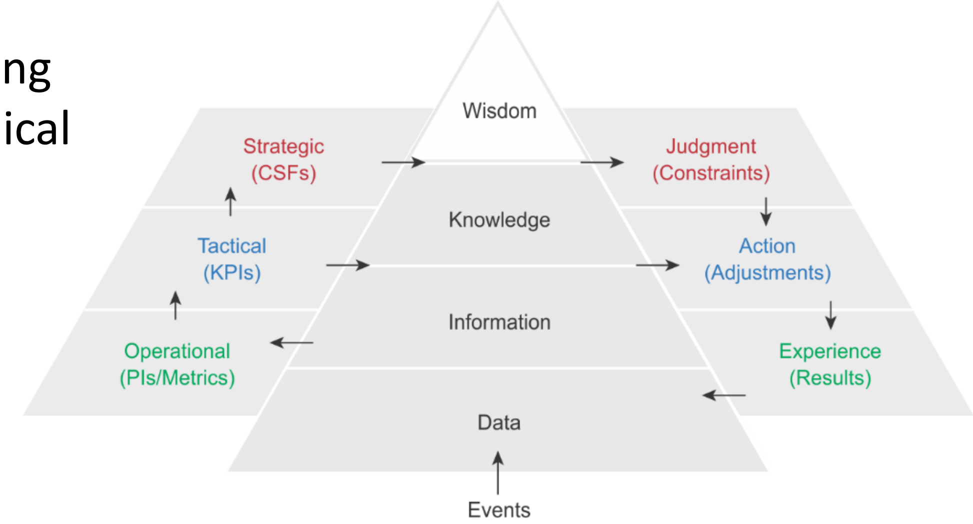


BCNET CONNECT

https://patterns.arcitura.com/internet-of-things-iot-patterns

Lessons Learned

- Data is never perfect, do not look for one.
- Preprocessing is an iterative task and may include several cycles.
- Using big data offers opportunities to expand the conventional knowledge.
- Problem space will include many moving parts, expect complexity explosion!
- Big Data solutions access data and generate data, all of which becomes assets of the business. Regulatory aspects should not be ignored.
- Leveraging frameworks and design patterns is a healthy option to move forward with the rise of demands and complexities.
- Understanding the domain knowledge is key; so is the power of modeling and establishing a business architecture, prior to implementing a technical solution.



Big Data Fundamentals: Concepts, Drivers & Techniques (pearsoncmg.com)

Discussions:

This Photo by Unknown Author is licensed under CC BY-SA-NC

Alan Rezazadeh
https://www.linkedin.com/in/alanrezazadeh/
Alan.Rezazadeh@sait.ca

Samuel Rostam
https://www.linkedin.com/in/samuelrostam/
srostam@nyit.edu

BCNET CONNECT