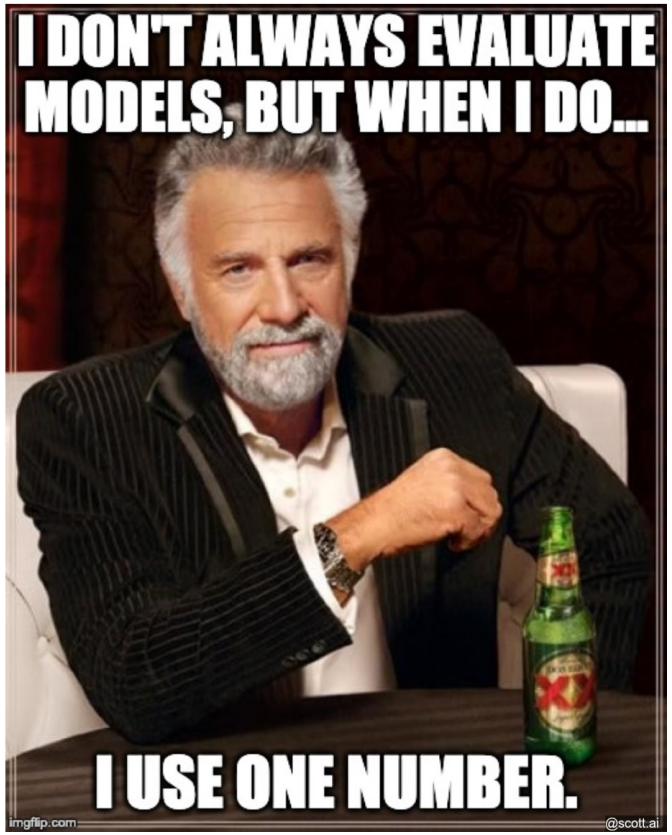
Communicating PHM Model Performance to the Customer



PHM 2023

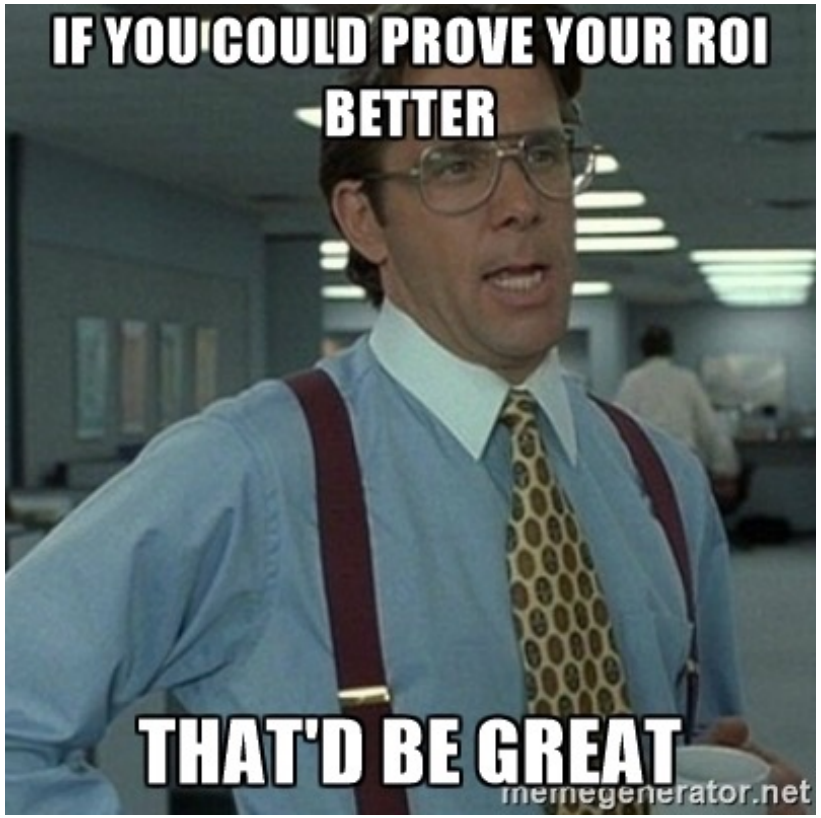
Salt Lake City, Utah

October 30, 2023

Moderator:

Sarah Lukens, Ph.D.

LMI

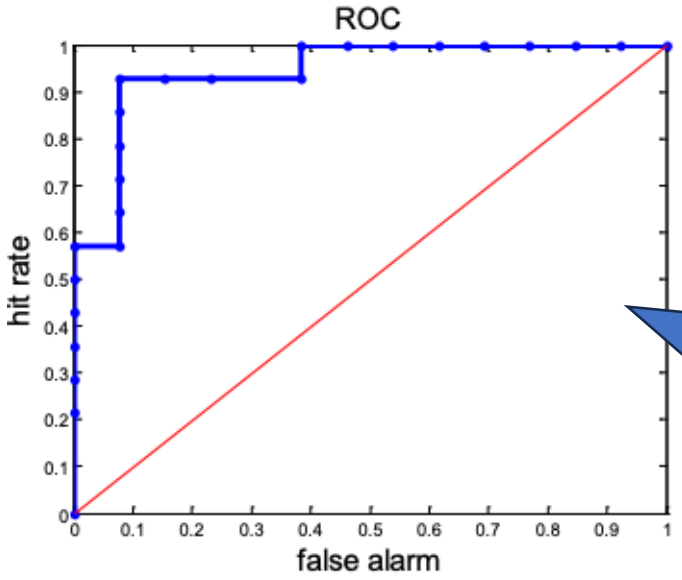


Panel Guidelines

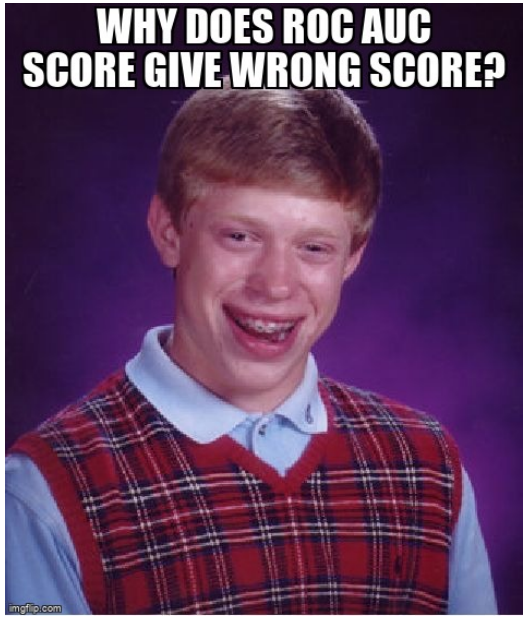
- This panel is not being recorded.
- Questions within the scope of the panel are encouraged. You can submit questions via:
  - The Whova app. Please include “@Panelist” if your question is directed at a specific panelist
  - Raise your hand!
- Volunteer note takers? Send to: sarah.lukens@lmi.org

Panel Background – ROC curves & PHM 2022

- Motivating example in aviation:
  - 12 Engines, 14 faults
  - ~26k negative observations; 14 positive
  - AUC is great!



Different alerting thresholds on the model output generate different fault prediction performance. These are reflected as different points on the ROC curve.



Panel Background

Very long audience discussion after PHM tutorial last year

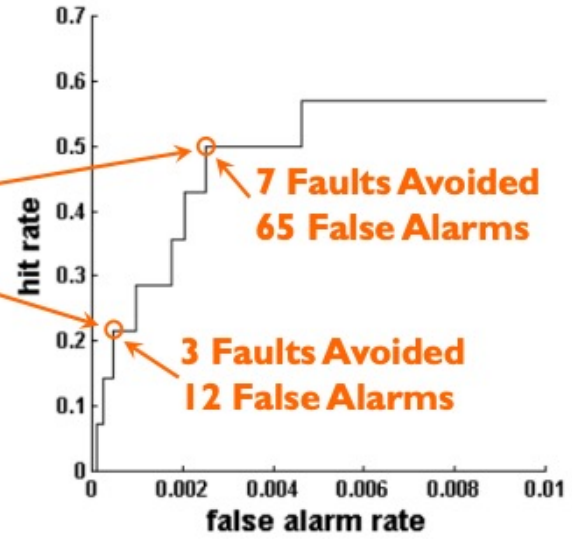
Here was the trigger:

Economic Analysis

Recall: $75k/incident

- At most $8076/false alarm @ 0.5 hit rate
- At most $18,750/false alarm @ 0.21

$75k/incident means avoiding 7 out of the 14 faults has savings of $525k

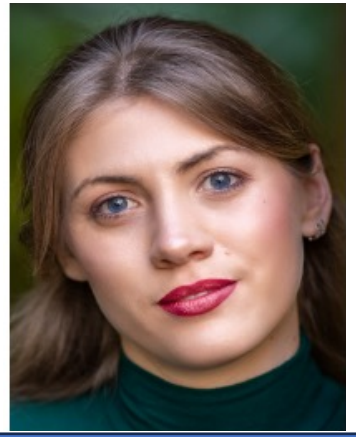


Model threshold selection based on balancing the cost of avoiding a fault with the number of false alarms that must be tolerated and their costs

Panelists

Communicating PHM Model Performance to the Customer

Peng Liu. Ph.D.
JMP



Katarina Vuckovic
Collins Aerospace

Dmitry Belov, Ph.D.
SLB

Jesse , Ph.D.
Global Technology Connection, Inc.

ROC & PR

Observations & Opinions

Peng Liu, PhD

Principal Research Statistician Developer

JMP Statistical Discovery LLC

PHM 2023

About me

Peng Liu

Build smarter algorithms and better GUI for analyzing data in JMP, a beautiful statistical discovery software.

JMP Division, SAS Institute Inc. ·

North Carolina State University

Greater Chicago Area · Contact info

Build and maintain JMP's reliability and time series products since 2007.
Specialized in reliability data analysis, reliability engineering, time series analysis, time series forecasting, machine learning techniques, discrete event simulation, software engineering, parallel computing, graphical user interface design and implementation, software architecture, etc.

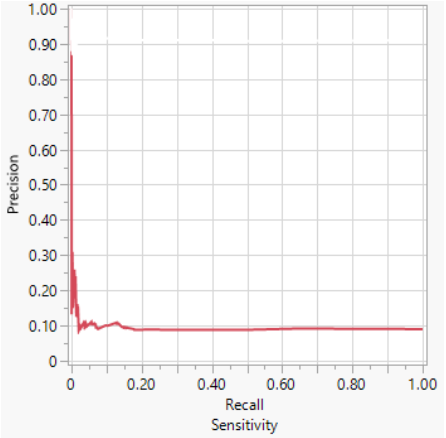
Outline

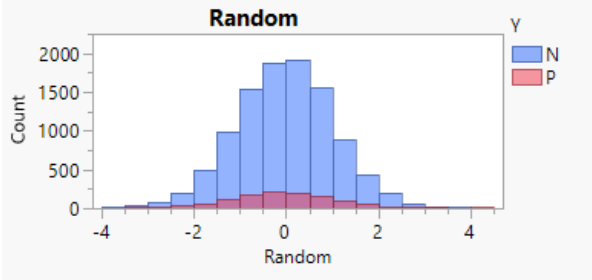
- The purpose of PR according to the PLOS paper
- How I use ROC & PR
- Additional ROC & PR simulations
- Personal thoughts

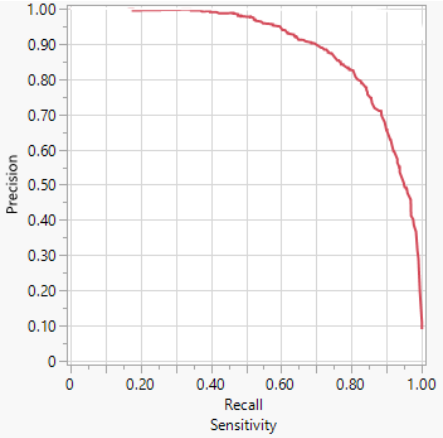
Comments on Fig. 5 in the PLOS Paper

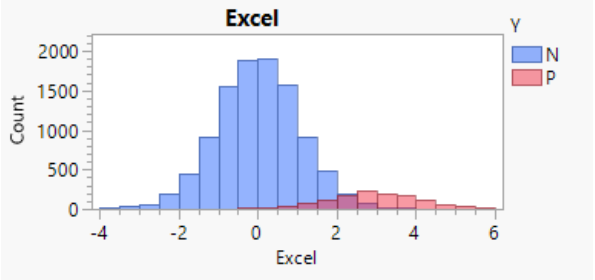
Summary of the simulation. The overall results of the simulations suggest that PRC is the most informative and powerful plot for imbalanced cases and is able to explicitly reveal differences in early-retrieval performance.

- How about a histogram?







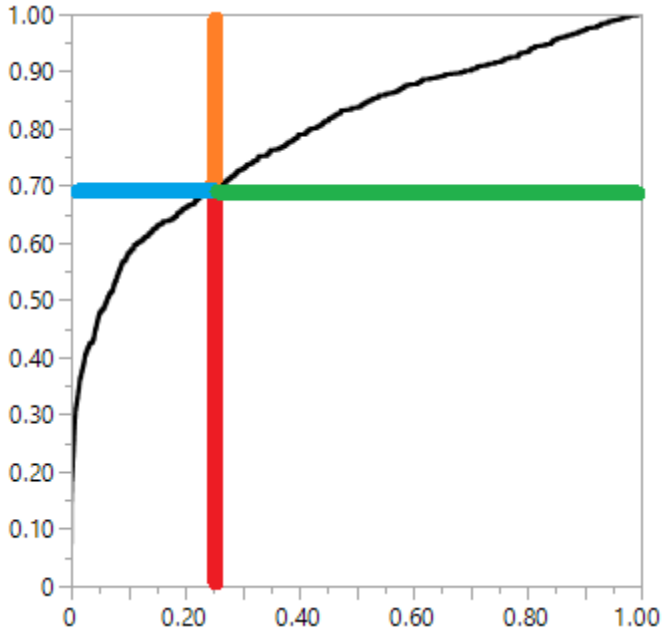


ROC

- Y-axis: $TPR(\tau) = \frac{N_{\blacksquare}(\tau)}{N_{\blacksquare}} = F_{\blacksquare}(\tau)$
- X-axis: $FPR(\tau) = \frac{N_{\blacksquare}(\tau)}{N_{\blacksquare}} = F_{\blacksquare}(\tau)$
- $\frac{N_{\blacksquare}}{N_{\blacksquare}}$ is irrelevant.
- Always monotone
- How I use the entire curve
- How I would use a point on the curve
- Drawback: disrespect $\frac{\blacksquare}{\blacksquare}$ proportion

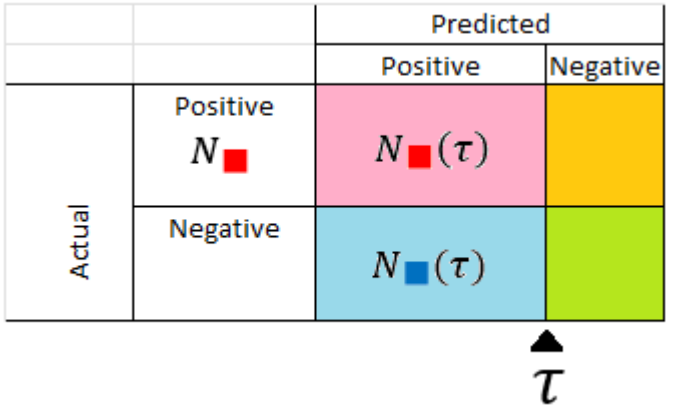
		Predicted	
		Positive	Negative
Actual	Positive $N_{\blacksquare}$	TP $N_{\blacksquare}(\tau)$	FN
	Negative $N_{\blacksquare}$	FP $N_{\blacksquare}(\tau)$	TN

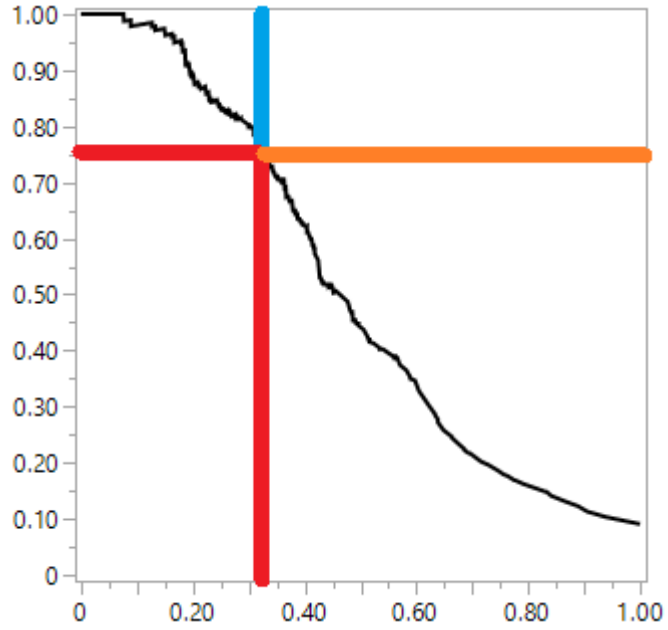
τ

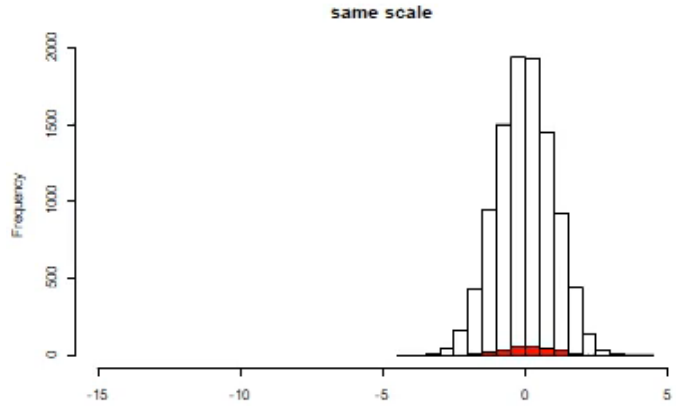


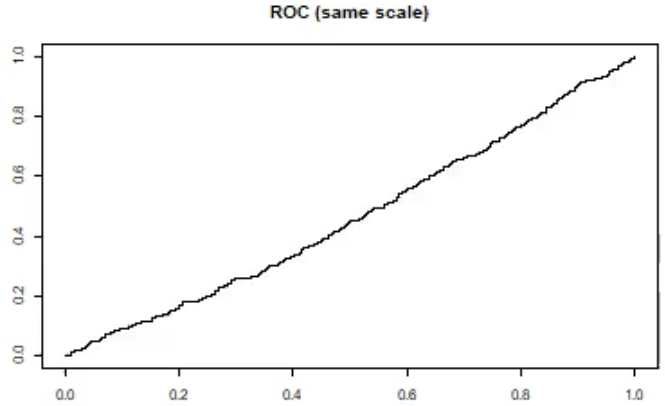
PR

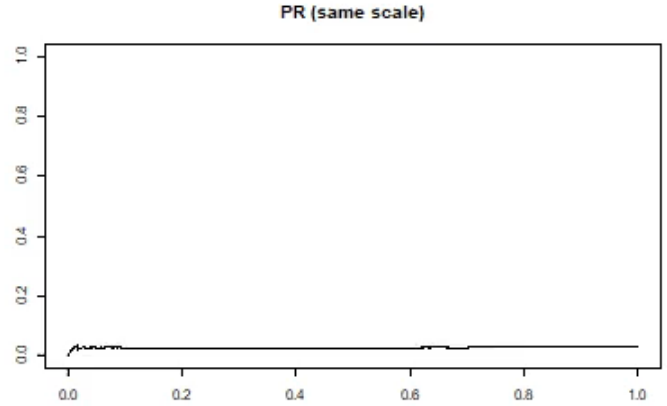
- Y-axis: $P(\tau) = \frac{N_{\blacksquare}(\tau)}{N_{\blacksquare}(\tau)+N_{\blacksquare}(\tau)}$
- X-axis: $R(\tau) = TPR(\tau) = \frac{N_{\blacksquare}(\tau)}{N_{\blacksquare}} = F_{\blacksquare}(\tau)$
- It is not monotone
- How I use the entire curve
- How I would use a point on the curve
- $\frac{N_{\blacksquare}(\tau)}{N_{\blacksquare}}$ could be a misleading factor, disproportional scales

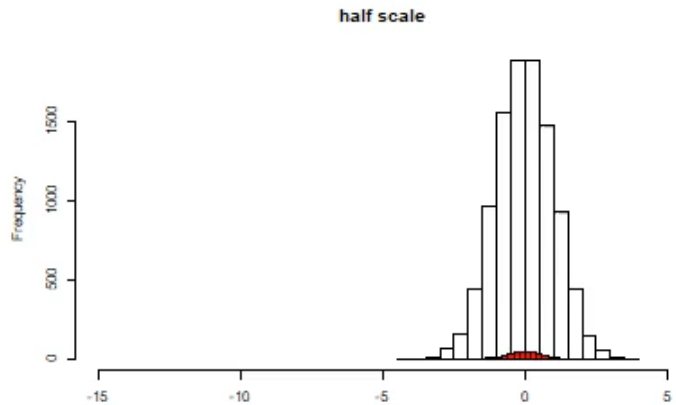


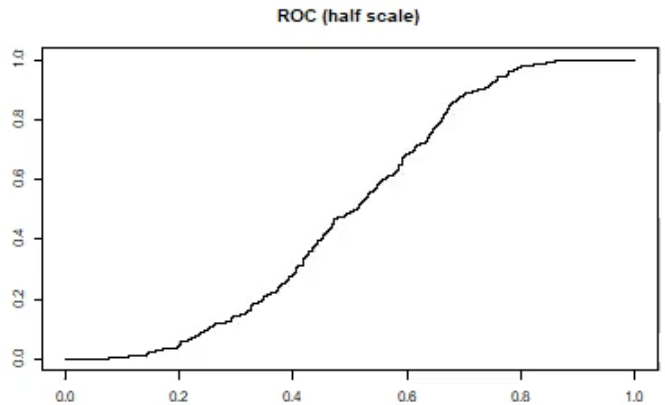


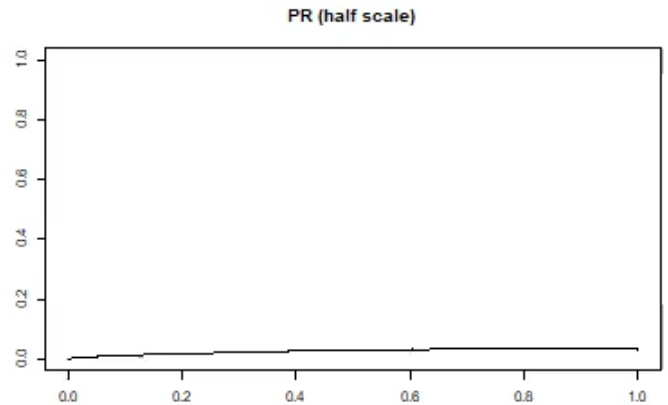

same scale
Frequency
0
500
1000
1500
2000
-15
-10
-5
0
5

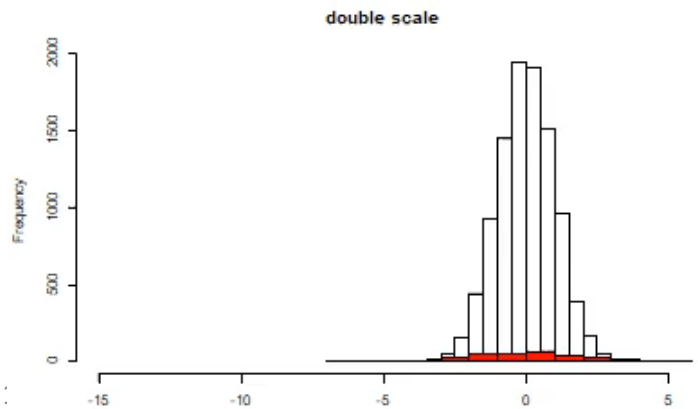

ROC (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

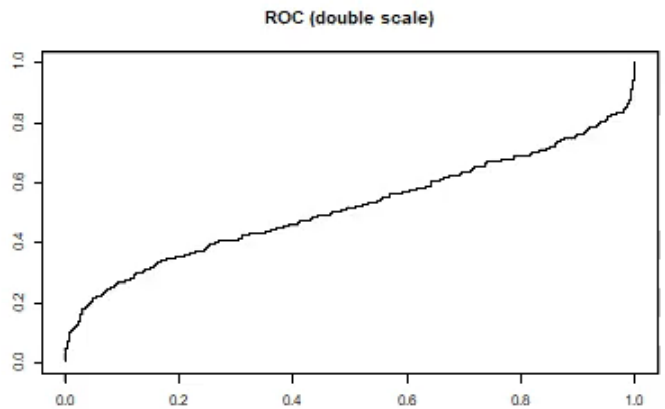

PR (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

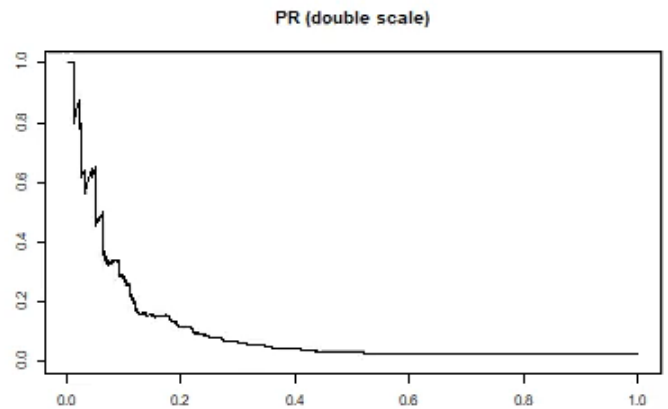

half scale
Frequency
0
500
1000
1500
-15
-10
-5
0
5

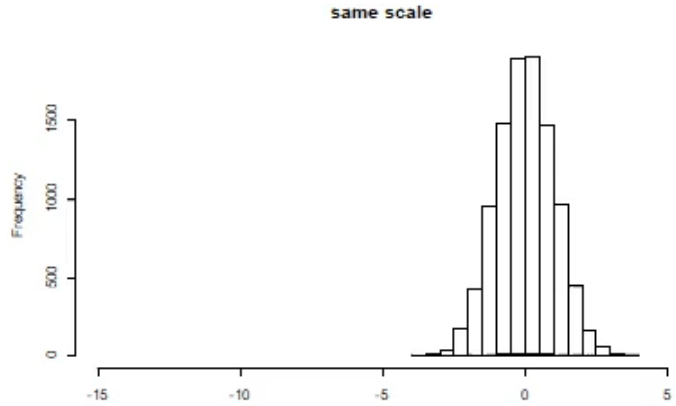

ROC (half scale)
0.0
0.2
0.4
0.6
0.8
1.0

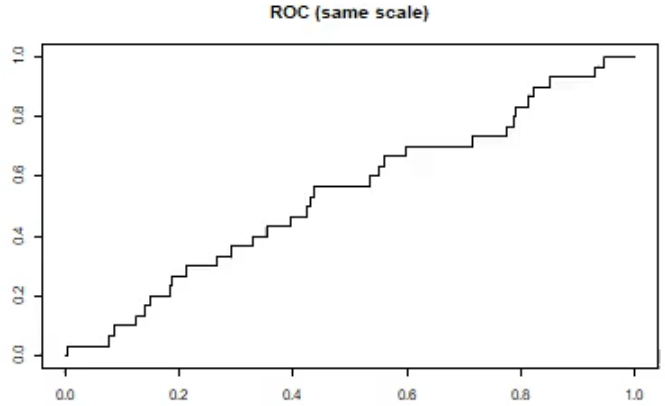

PR (half scale)
0.0
0.2
0.4
0.6
0.8
1.0

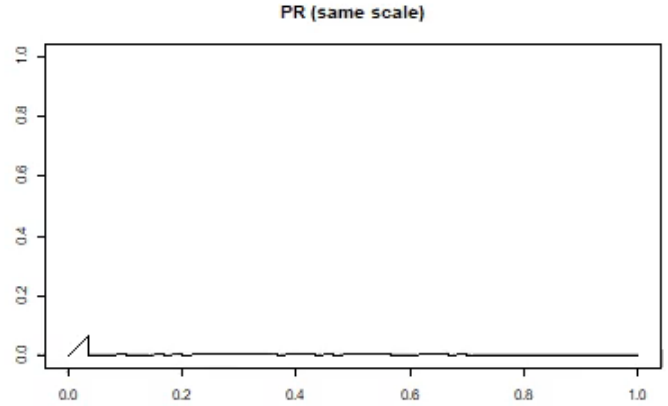

double scale
Frequency
0
500
1000
1500
2000
-15
-10
-5
0
5

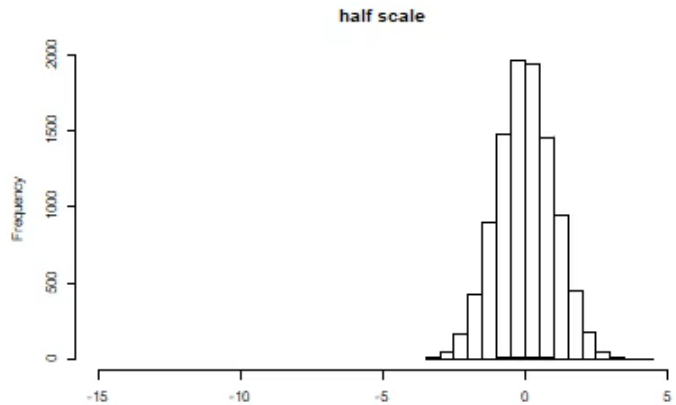

ROC (double scale)
0.0
0.2
0.4
0.6
0.8
1.0

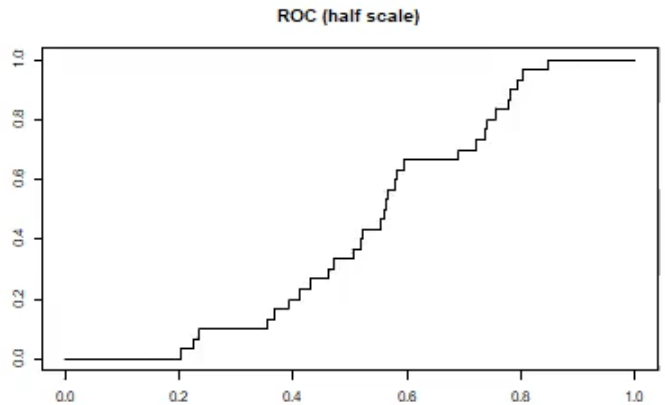

PR (double scale)
0.0
0.2
0.4
0.6
0.8
1.0

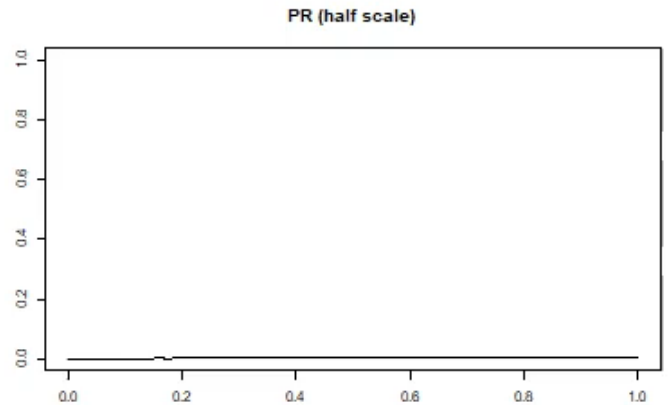

same scale
Frequency
0
500
1000
1500
-15
-10
-5
0
5

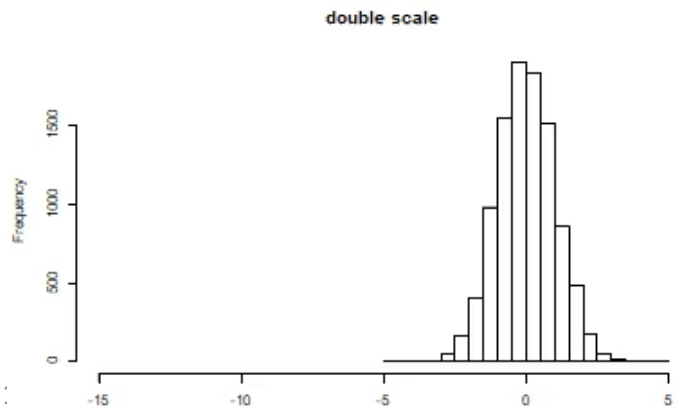

ROC (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

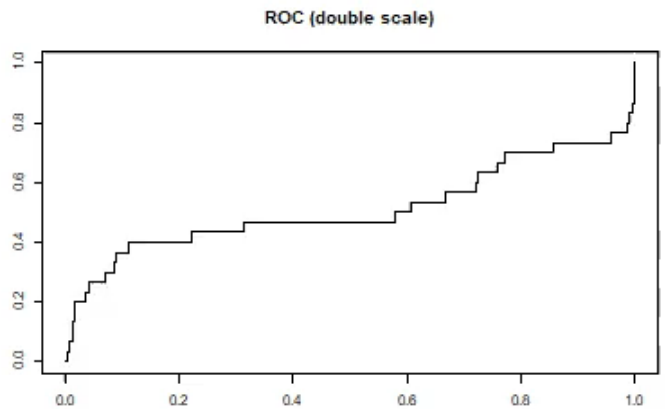

PR (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

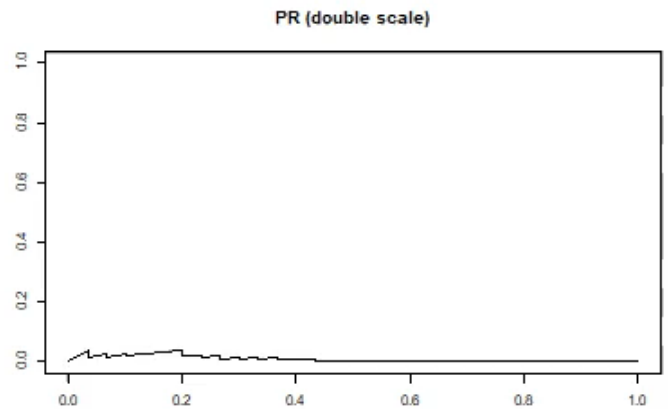

half scale
Frequency
0
500
1000
1500
2000
-15
-10
-5
0
5

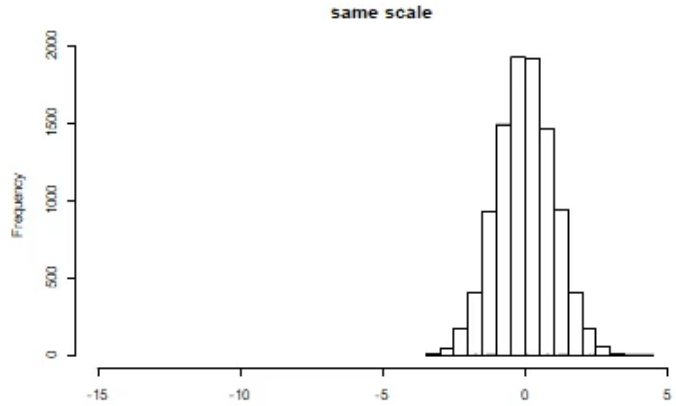

ROC (half scale)
0.0
0.2
0.4
0.6
0.8
1.0

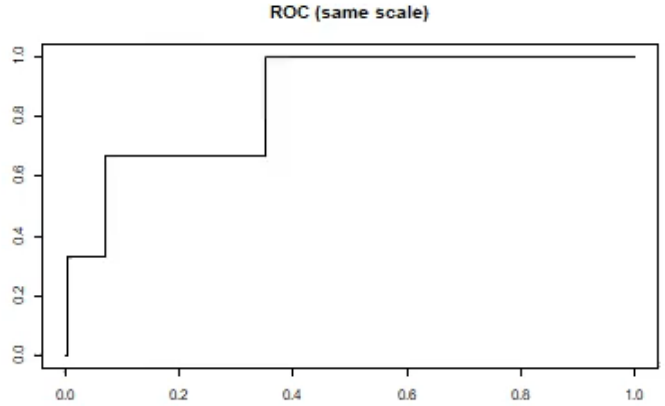

PR (half scale)
0.0
0.2
0.4
0.6
0.8
1.0

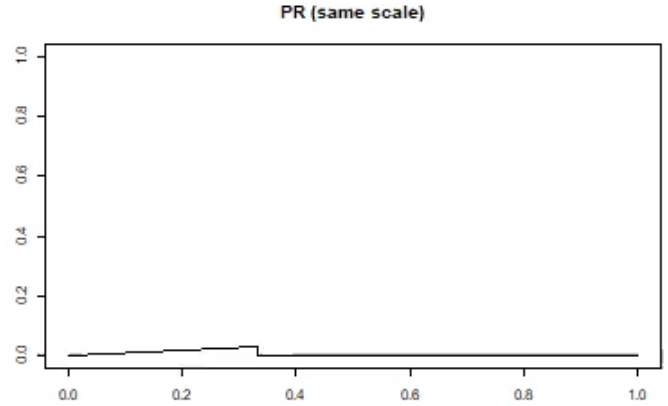

double scale
Frequency
0
500
1000
1500
-15
-10
-5
0
5

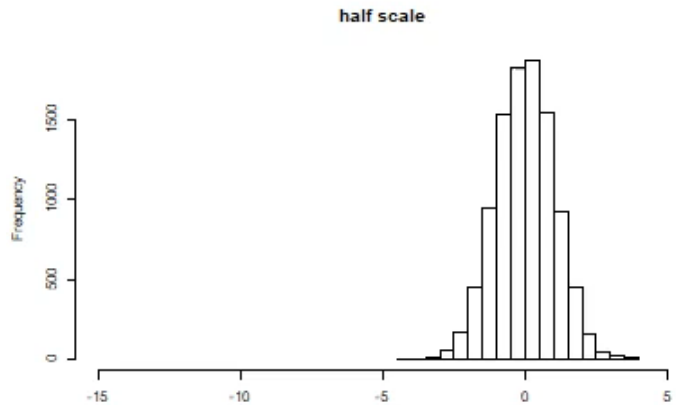

ROC (double scale)
0.0
0.2
0.4
0.6
0.8
1.0

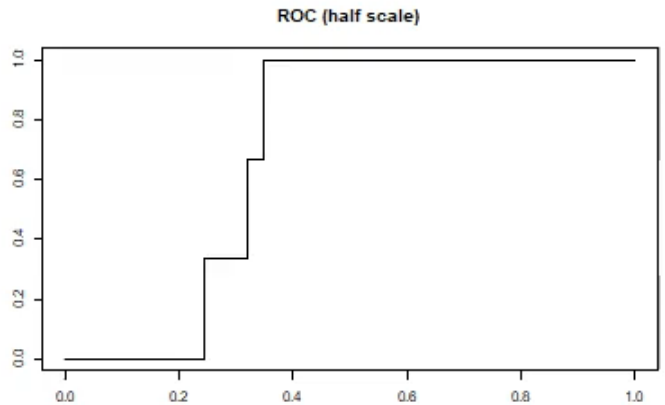

PR (double scale)
0.0
0.2
0.4
0.6
0.8
1.0

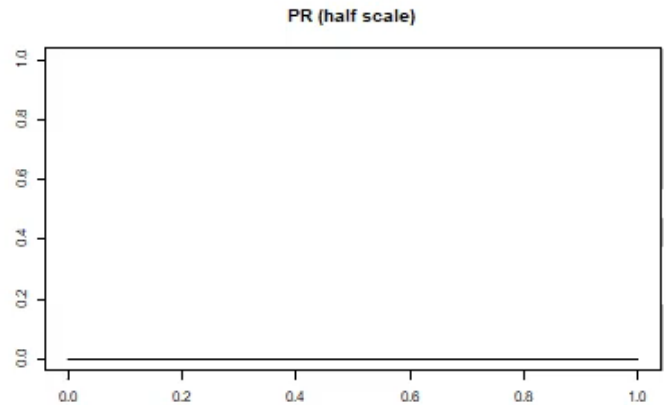

same scale
Frequency
0
500
1000
1500
2000
-15
-10
-5
0
5

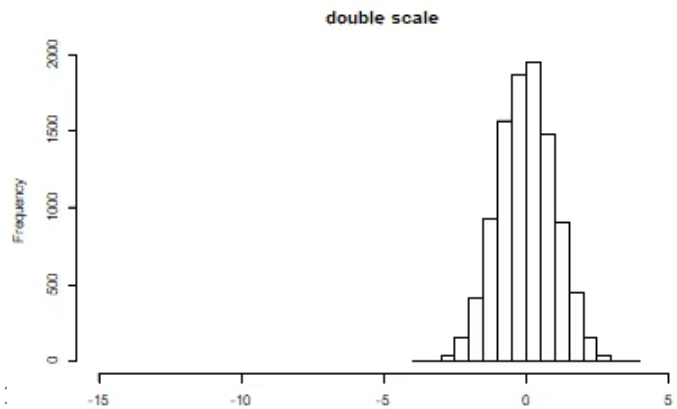

ROC (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

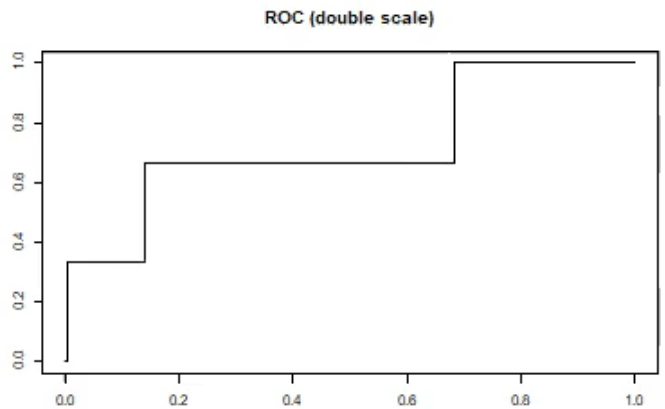

PR (same scale)
0.0
0.2
0.4
0.6
0.8
1.0

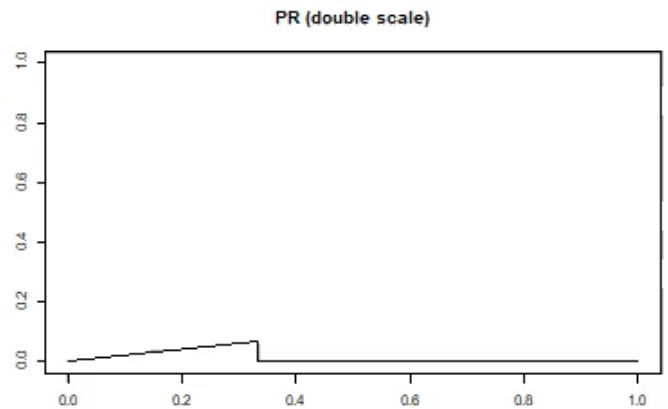

half scale
Frequency
0
500
1000
1500
-15
-10
-5
0
5


ROC (half scale)
0.0
0.2
0.4
0.6
0.8
1.0


PR (half scale)
0.0
0.2
0.4
0.6
0.8
1.0

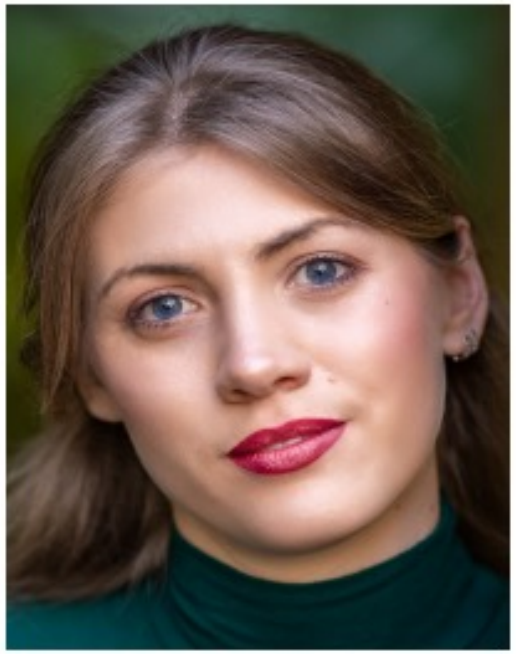

double scale
Frequency
0
500
1000
1500
2000
-15
-10
-5
0
5


ROC (double scale)
0.0
0.2
0.4
0.6
0.8
1.0


PR (double scale)
0.0
0.2
0.4
0.6
0.8
1.0

Observations & Opinions

- If you have a verified model which makes complete separable classes with 100% confidence, none of the metrics matter.
- Only clear when there is a dominant curve. For model building.
- Only clear when ROC or PR are monotone. For trade-off decision?
- ROC and PR are equivalent for model comparison if there is a dominant curve. And the AUC argument is legit.
- All other situations …
- Additional consideration: uncertainty in ROC, PR, AUC, …



Katarina Vuckovic
Collins Aerospace
Data Science: Predictive Analytics Group

Discussion Topic:
Operational Prognostic Model Evaluation

ABOUT ME

Education:

- *University of Central Florida*
  - Ph.D. Electrical Engineering (Expected 2024)
- *Florida Institute of Technology*
  - M.S. Electrical Engineering (2019)
  - B.S. Electrical Engineering (2017)
  - B.S. Aerospace Engineering (2016)

Experience:

- *Colins Aerospace*
  - Data Scientist (2022- Present)
  - Avionics System Engineer (2017-2021)

Areas of Interest:

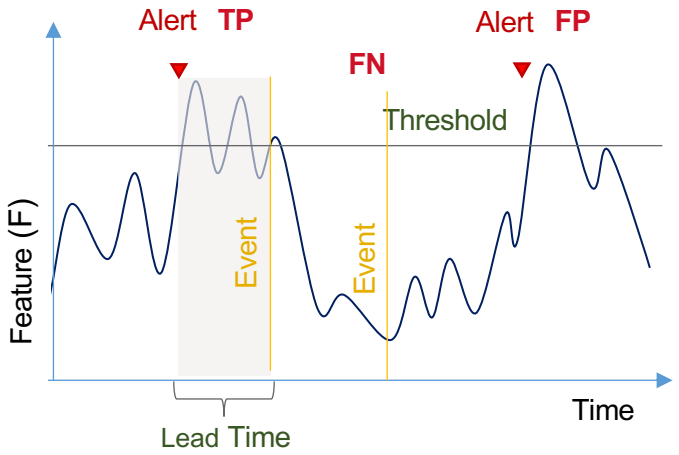
- Data Science
- Modeling and Simulations
- Machine Learning
- Aviation

Collins Aerospace
An RTX Business

PROGNOSTIC PERFORMANCE METRICS

BINARY CLASSIFIER

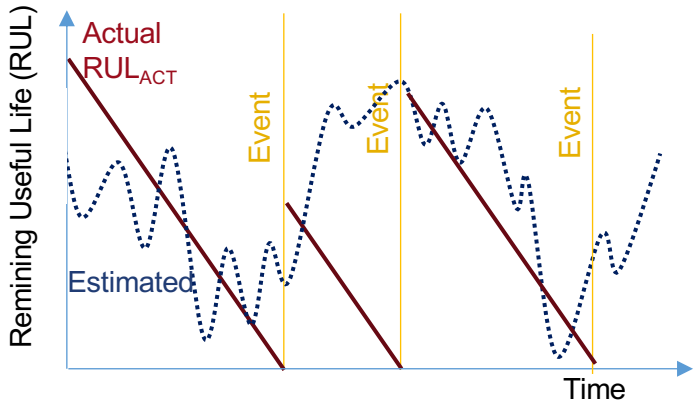
Predicted \ Reality	-No Event	Event	
Alert +	False Positive FP	True Positive TP	Precision $\frac{TP}{TP + FP}$
No Alert -	True Negative TN	False Negative FN	
		Recall $\frac{TP}{TP + FN}$	



Arbitrary distinction boundaries: max Lead time?

Discrete states not efficient as an optimizer

REMAINING USEFUL LIFE ERROR



Prediction Error : $\varepsilon = RUL_{EST} - RUL_{ACT}$

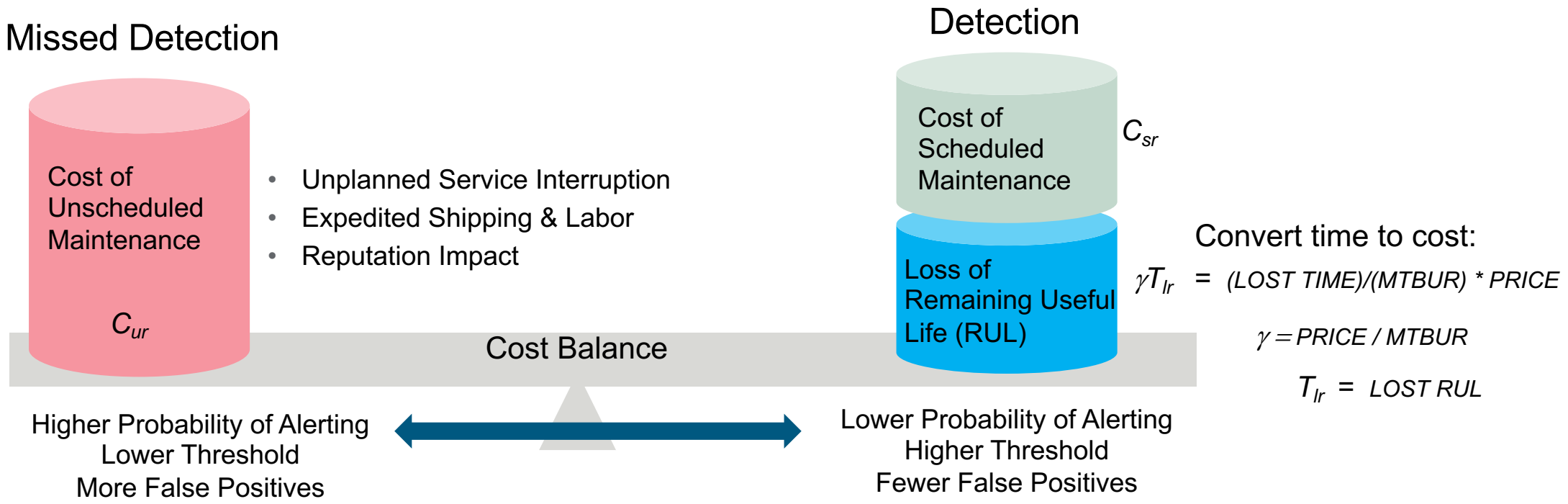
Mean Absolute Error (MAE)	Mean Absolute Percentage Error (MAPE)	Root Mean Squared Error (RMSE)
$\frac{1}{n}\sum_{i=1}^{n}\lvert\varepsilon_i\rvert$	$\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{\varepsilon_i}{RUL_{ACT,i}}\right\rvert * 100\%$	$\sqrt{\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n}}$

MAE/RMSE: Error equally important throughout life

MAPE: Error singularity at end of life

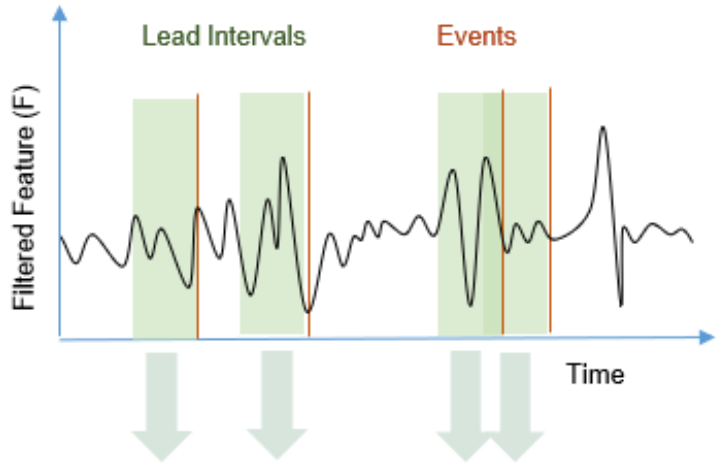
No accounting for unobserved failure modes

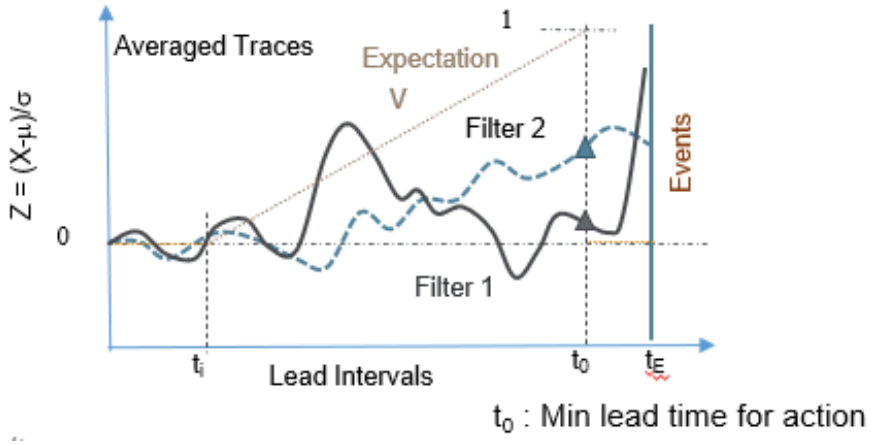
OPERATIONAL CONSIDERATIONS



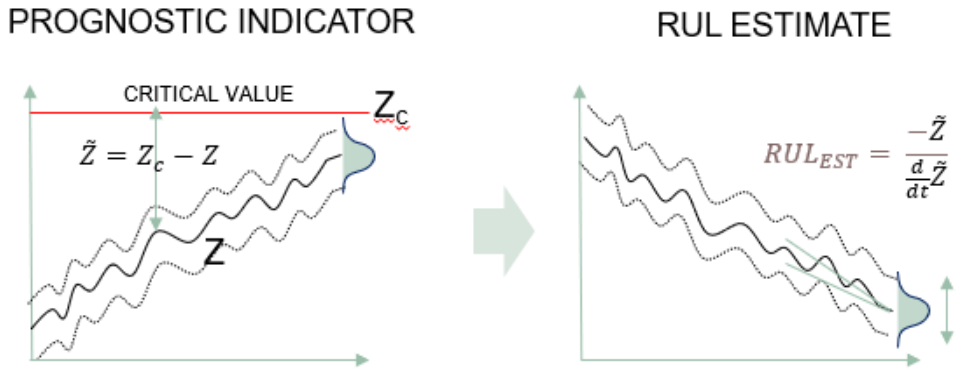
PROGNOSTIC OBJECTIVES

FEATURE CORRELATION TO RUL





PROGNOSTIC FEATURE UNCERTAINTY



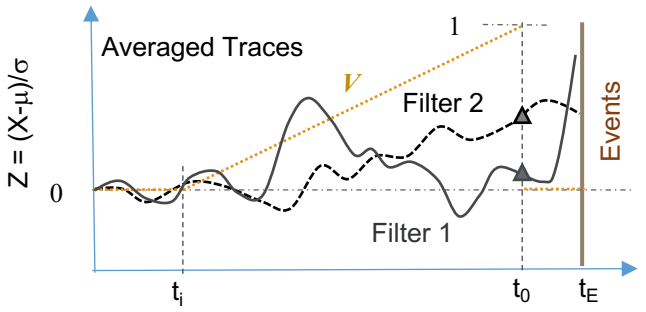
- The wider the uncertainty the more RUL we lose

- Filter signal
- Average the lead time interval behavior
- Adjust filter coefficient to improve filtering
  - Optimize for correlation to RUL and low scatter.

IMPLEMENTATION

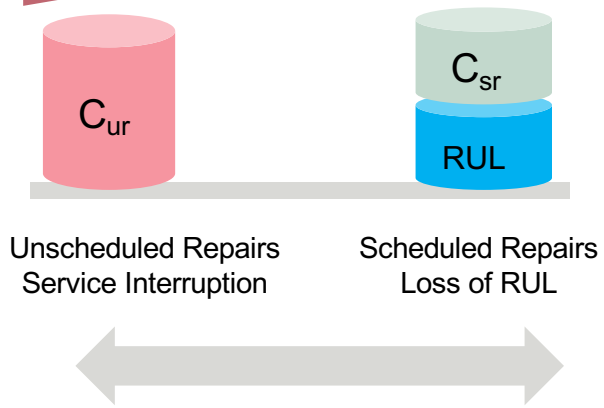
Optimize Detection

- Continuous signal
- Correlation to events



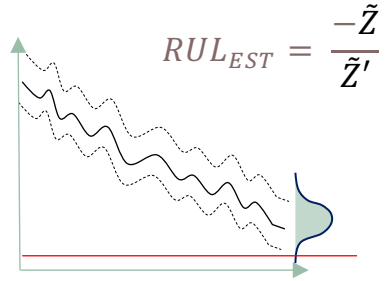
Operational Impact

- Unscheduled Costs
- Life Impact



Address Variation

- Cost associated with estimation error
- Re-optimize filter with weight on scatter



- Cost savings from prognostic removal is difference between unscheduled (C_{ur}) and scheduled removal (C_{sr})
- Savings must exceed the value of lost remaining life and associated maintenance overhead
- Lost life is proportional to the RUL estimate uncertainty

Dmitry Belov, PhD – BIO

Apr 2008 – Nov 2014 (Moscow, Russia)

Senior Research Scientist

Nov 2014 – Mar 2017 (Fuchinobe, Japan)

Senior Physicist

Apr 2017 – May 2020 (Houston, USA)

Senior M&S Engineer

June 2020 – Dec 2021 (Houston, USA)

Data Scientist

June 2020 – Dec 2021 (Cambridge, UK)

Principal Data Scientist

Main expertise:

- Modeling & Simulation / FEA (wave propagation, dynamic simulations, strength analysis, etc.)
- Fatigue & PHM
- CAD/CAE systems (ANSYS, ABAQUS, LS-DYNA)
- Data processing (data scientist domain)
- Digital Twins
- Composite materials
- Wireline domain (acoustic, ultrasonic, seismic) & Drilling domain

Featured PHM articles:

Belov, D. et al. (2021) "Hybrid Approach for Health Monitoring of Mud Motor Fleet", PHM Society European Conference, 6(1), 10.

Belov, D. et al. (2021) "Data-Driven PHM Solution for Health Monitoring of Mud Motor Power Sections While Drilling", SPE-205219-MS, SPE Europec featured at 82nd EAGE Conference and Exhibition, Amsterdam, The Netherlands

Belov, D. et al. (2023) "Mud Motor Digital Maintenance with an Industry-Unique PHM Solution", SPE-212505-MS, 2023 SPE/IADC International Drilling Conference and Exhibition, Norway, Stavanger

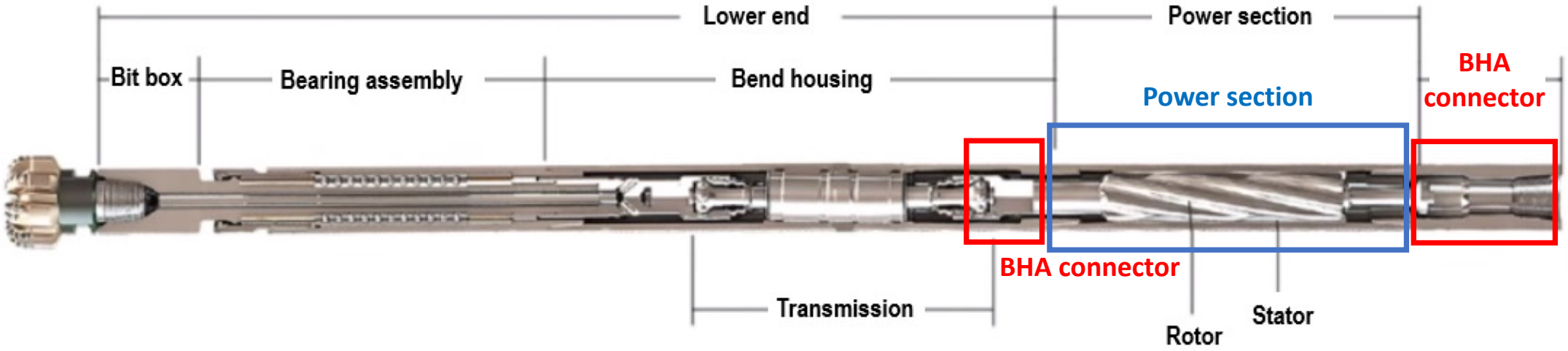
Belov, D. et al., (2023) "BHA Critical Connection PHM Model for Drilling Industry",ISBN:979-8-3503-0014-7, 2023 PHM IEEE Conference, Paris, France

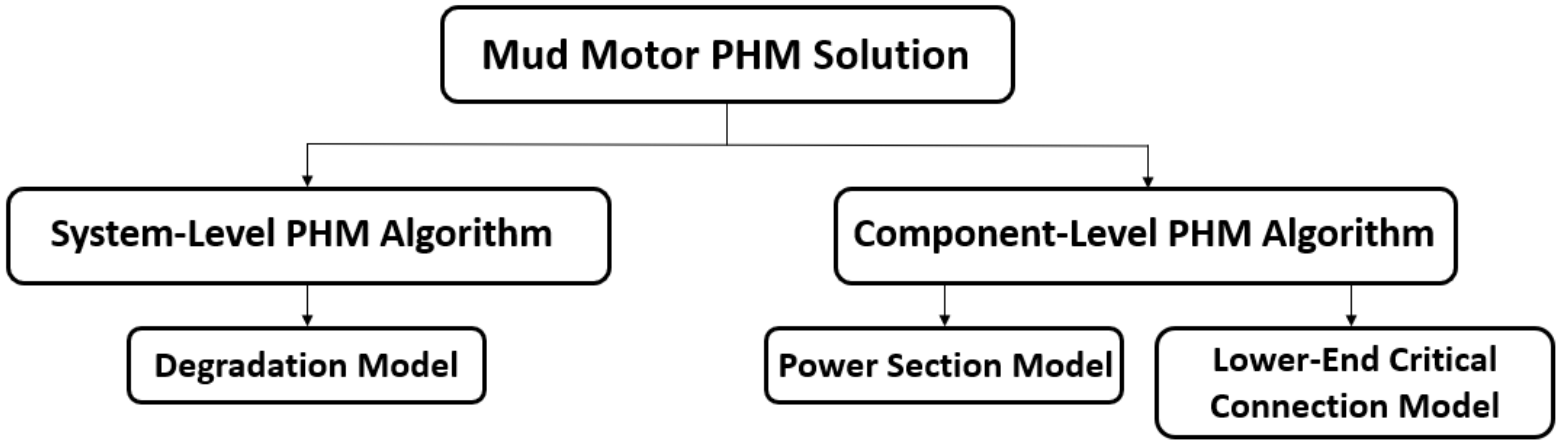
Belov, D. et al., (2023) "The Digitization of Mud Motor Power Section Life Cycle: From Concept to Operation", SPE-216138-MS, ADIPEC, Abu Dhabi, UAE

Belov, D. et al., (2023) "Revolutionizing Drilling Operations: An Automated Approach for Drilling Analysis and Optimization Using Real-Time Data", OTC-32892-MS, Offshore Technology Conference Brazil, Rio de Janeiro, Brazil

Dmitry Belov, PhD – The Importance of Effective Communication

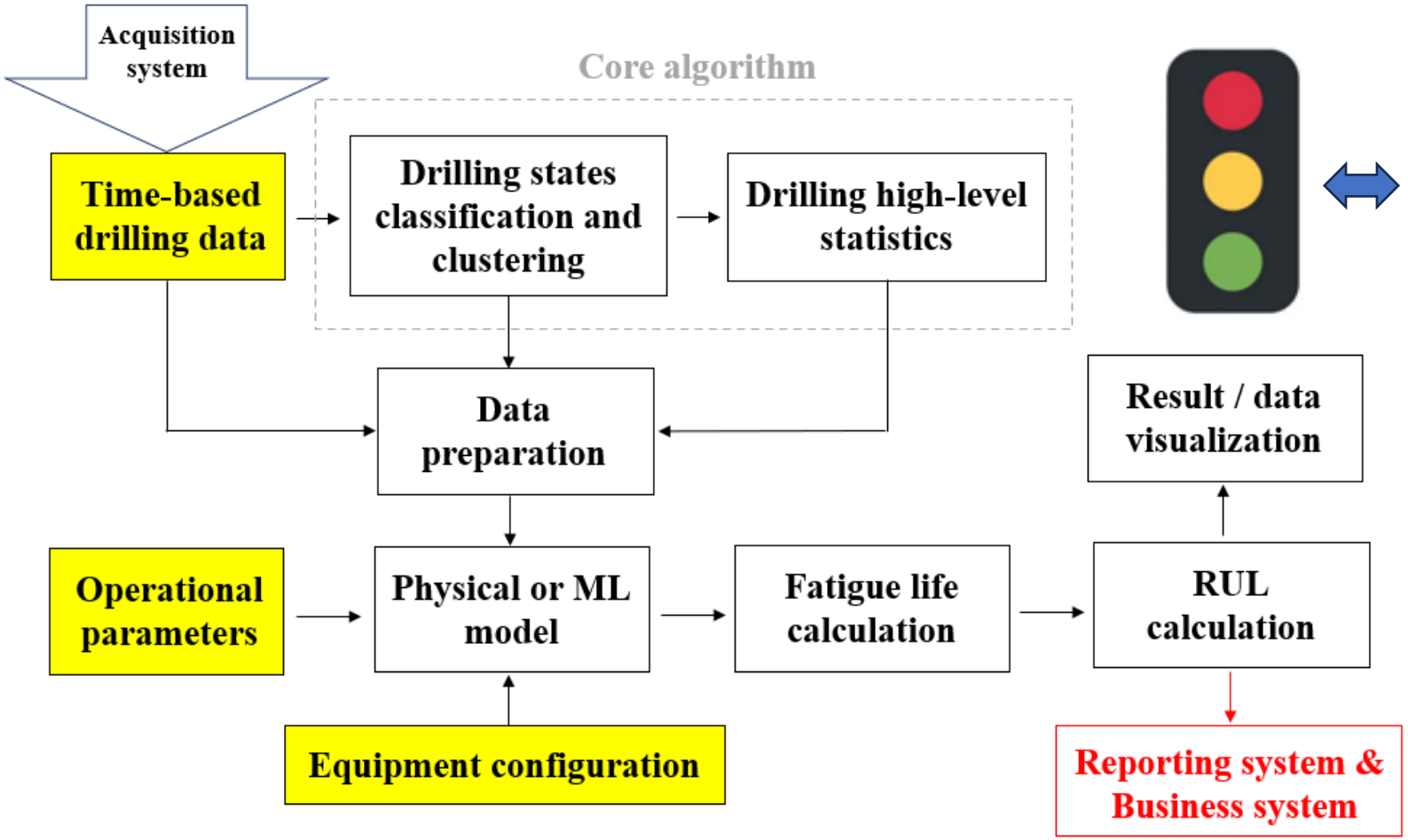
Example of the Relevant PHM Algorithms: Mud Motor PHM*

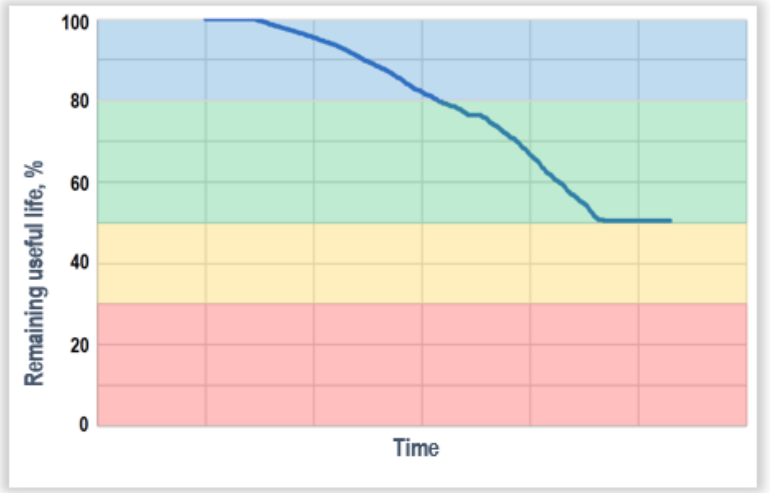




*The materials are sourced from open literature (refer to the list in the BIO section): Belov et al. (2021, b) and Belov et al. (2023, a)

Example of the Relevant PHM Algorithms: Component-Level Models*





Primary User Concerns:

Why should I trust your algorithm?

How reliable are your thresholds?

Will it truly offer me a benefit?

How can I market it to my customers?

What is the margin of error?

*The materials are sourced from open literature (refer to the list in the BIO section): Belov et al. (2021, a) and Belov et al. (2023, d)

Dmitry Belov, PhD – The Importance of Effective Communication

Conclusion

How to prove your model

- Explain the chosen approach/model clearly
- Provide a comprehensive overview of the covered failure mechanisms
- Present positive validation results based on historical data
- Conduct field trials for new clients

How to prove the value to the customer

- Provide the success story from the existing customers
  - Red money
  - NPT reduction
- Measure the impact your algorithm delivers to the customer's saving
  - Accumulated saving algorithm
  - Cloud-based storage with results
  - Real-time dashboard (Dataiku)

How to inspire commencement of use

- Offer a user-friendly interface with minimal input requirements
- Maximize the level of automation for the algorithm
- Include pre-developed integration with the customer's acquisition systems and support various input data formats
- Ensure clear results with actionable recommendations
- Provide statistical insights and easily understandable reports

Typical frequent objections from potential users?

I have 20 years of experience and can determine the right time to replace equipment…

I need an exact RUL percentage; I want to eliminate any probability and uncertainty...

Jesse Williams, PhD – Bio Sketch

Experience

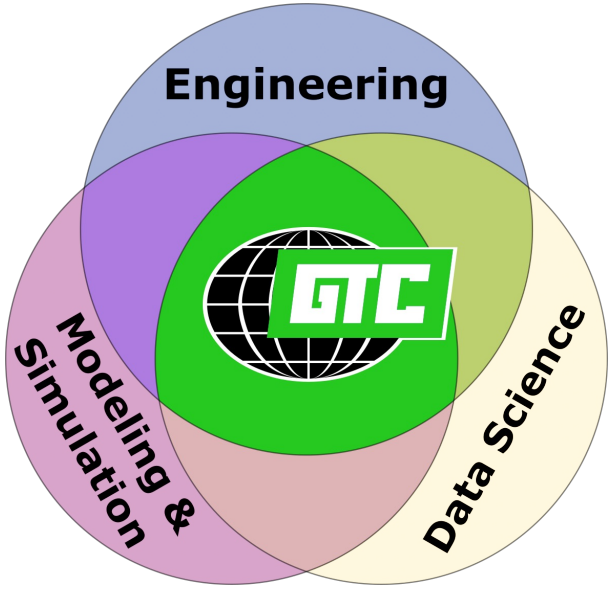
- **Global Technology Connection**, Inc., Atlanta
  - *CTO*
- **LIM Innovations**, San Francisco
  - *CTO/VP Engineering*
- **Nation Institute of Material Science** (Japan)
  - *Research Scientist*
- **University of California, Santa Barbara**
  - *Ph.D.* (2008)
  - Materials Science and Engineering
- **University of Pittsburgh**
  - *B.S.E.*
  - Materials Science and Engineering

Publications

- Peer-reviewed publication: 26
- Patents: 18 (full), 1 (provisional)

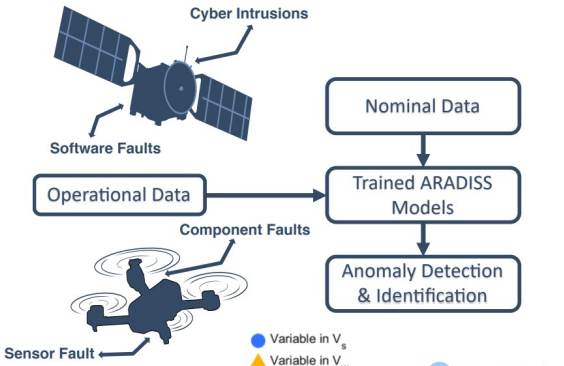
Areas of expertise

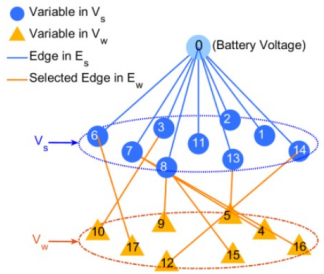
- Data science
- Modeling and simulation
- Python development
- Hardware development
- Networking (IT)
- Dashboarding and visualization
- Engineering

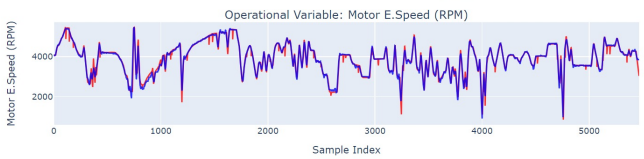


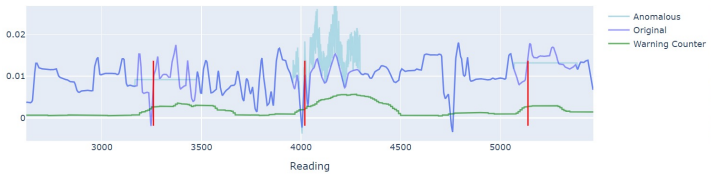
Jesse Williams, PhD – Technologies in development

Platform agnostic anomaly detection

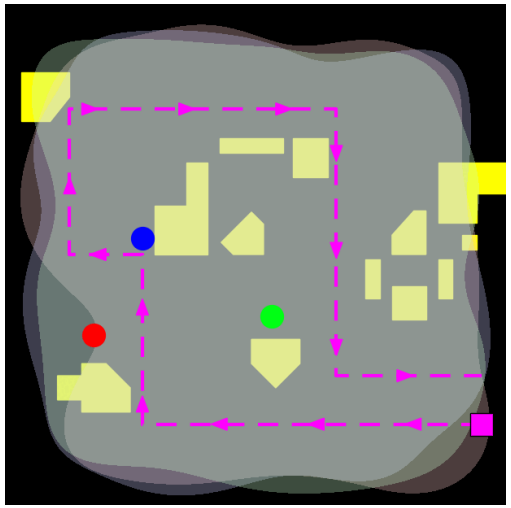


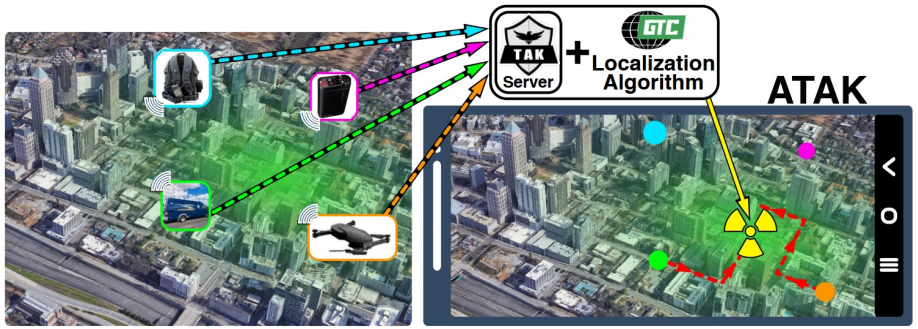




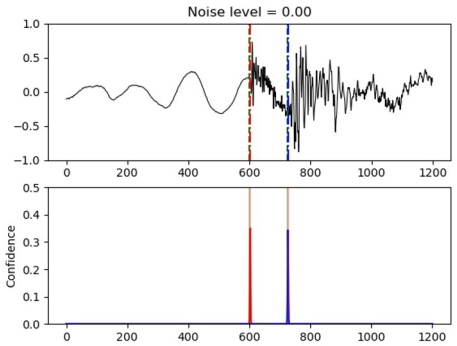


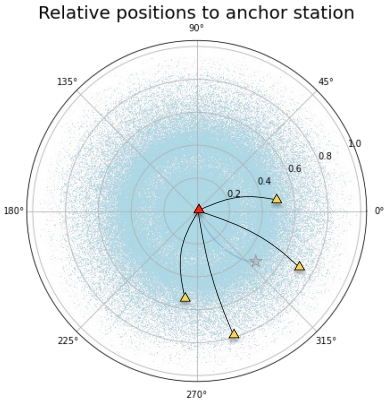
Sensor fusion for Radiation source location

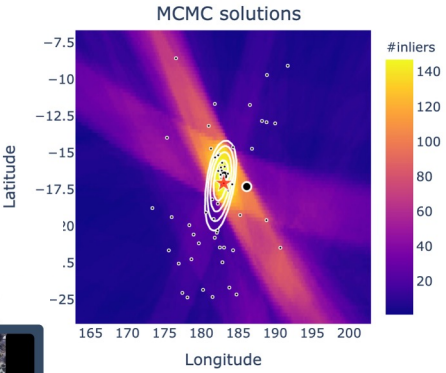


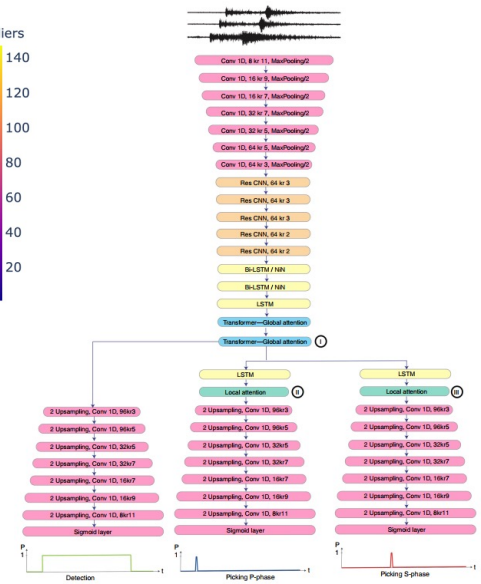


Seismic signal analysis for Earthquake/explosion monitoring









Mousavi, *et al.*,Nature Communications, 2020

Jesse Williams, PhD – Communicating with the Customer

Model assessment

- Make sure that **you** understand your model(s)
- Clearly articulate to the customer any limitations
- What error have you (not) captured in the model
- Are there unintended biases
- Be skeptical

Presentation and style

- Using a reporting medium that resonates with the customer
  - Report
  - Slide deck
  - Interactive notebook (Mercury)
  - Dashboard (Plotly Dash)
- Better visuals! But make sure they are relevant
  - Stats blogs
  - Plotting package galleries (seaborn or ggplot)
  - The Grammar of Graphics

Value to the customer

- Quantify the value your algorithm brings to the customer
  - Reduced man hours
  - Reduced downtime
  - Increased production
  - Mission readiness
  - Increased security
- Quick wins for renewed contracts
  - Low-hanging fruit
  - Rapid integration
    - Microservice/API

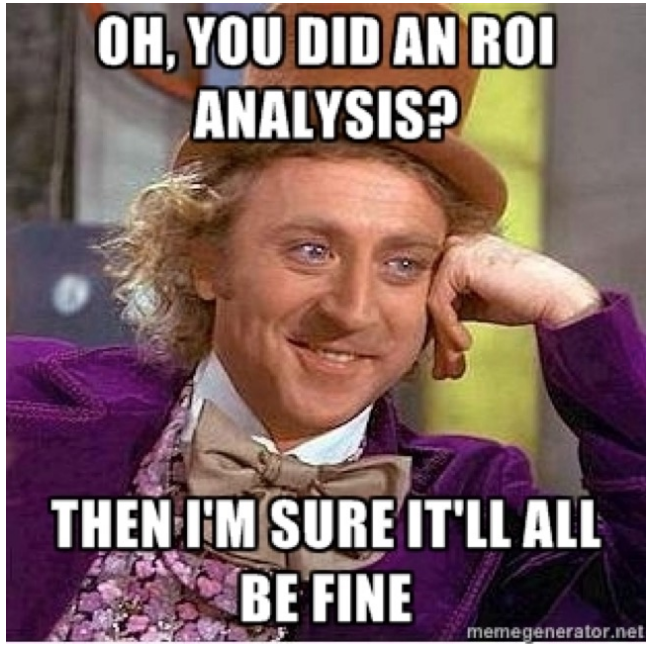
Questions

1. Return on Investment (ROI): Inevitably, customers always want to know the ROI. How do you approach not only measuring ROI but also tracking ROI and communicating with the customer?
2. Often, formal model performance summaries can be highly technical and difficult to read. Do you have tips, tricks or past successes for creating visuals which summarize model performance that someone can quickly understand?
3. For PHM models, are there any requirements from customers that you hear over and over? Or requirements you don’t hear from customers that you think they should be asking?

Panelist Question 1

Return on Investment (ROI): Inevitably, customers always want to know the ROI.

How do you approach not only measuring ROI but also tracking ROI and communicating with the customer?



Panelist Question 2

Often, formal model performance summaries can be **highly technical** and **difficult** to read.

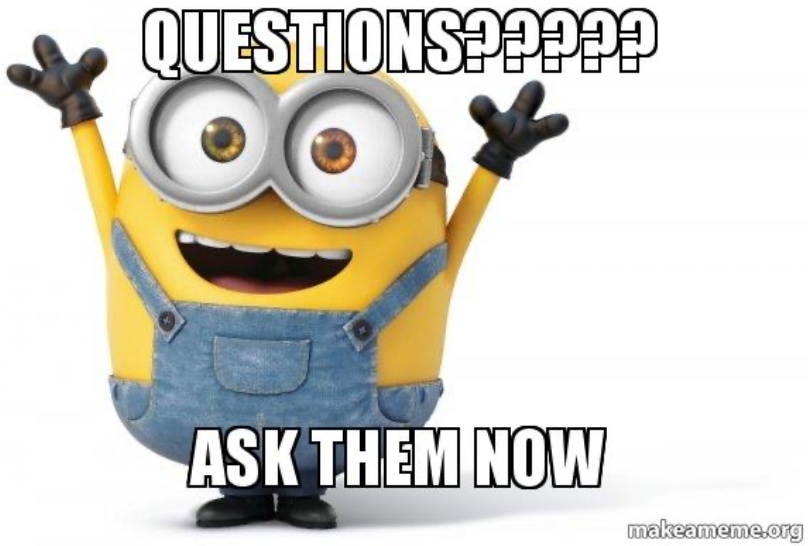
Do you have tips, tricks or past successes for creating visuals which summarize model performance that someone can quickly understand?

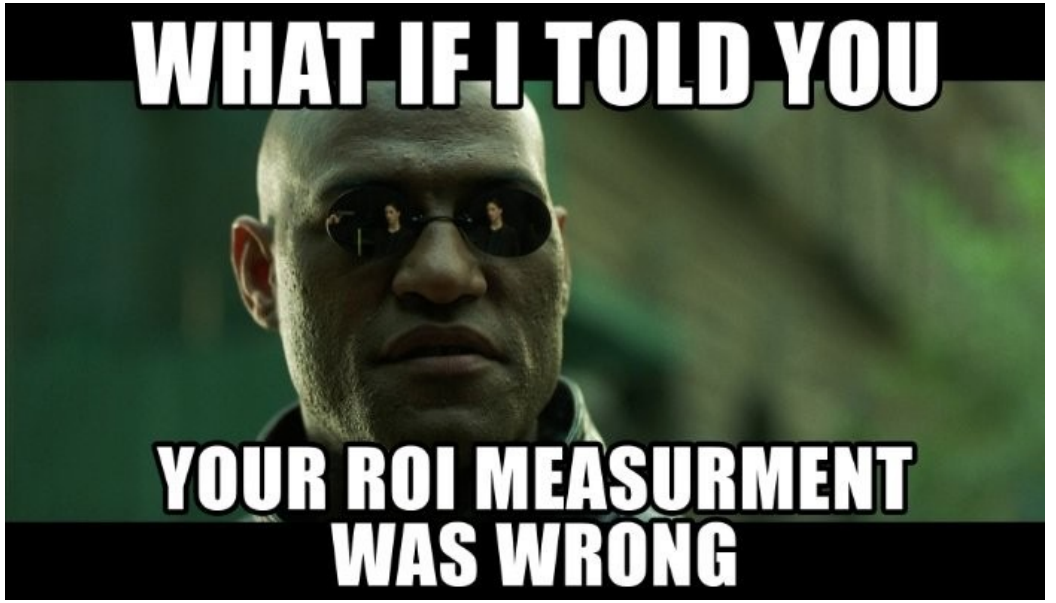
Panelist Question 3

For PHM models, are there any requirements from customers that you hear over and over?

Or are there requirements you don't hear from customers that you think they should be asking?

Audience questions





Notes from the panel session – page 1

- Thank you Fabian Rubilar Jamen!

- Question 1) ROI
- Try to generate a PMV as fast as possible
- Many times even the clients doesn't know the ROI and/or the costs of their failures
- Good trick is to calculate an estimation given the available industry data
  - Some ways to present the estimations:
    - § Cost/risk of the unavailability vs false-positive costs

Notes from the panel session – page 2

Question 2) Visualization tips

- Metrics: do NOT talk about accuracy, especially in highly biased datasets. Find the right metric
- Avoid the “black box” term.
- Try to present uncertainty from the beginning to the client (to avoid unmatched expectations)
- Some “core” plots you might consider:
  - A plot to best represent data
  - A plot to represent the model
  - Maybe an overlay of the data and model
  - If you have several models, it is good to think how they should be presented

Notes from the panel session – page 3

Question 3) For PHM models, are there any requirements from customers that you hear over and over? Or are there requirements you don't hear from customers that you think they should be asking?

- Make the concepts and definitions as part of your project initial presentations, thus ensuring a minimum leverage of knowledge
- Are the results exact/certain? Convince the client about the results by the number of events detected, for example.
- Find a middle point between the languages of both clients and PHM scientists.
- Will the solution be simple? The deployed tool should be simple and easy to integrate to their operations. A good way to do so is creating APIs, clients seems to “love” microservices.
- Underestimation of false positives. Some industries/applications are more “allergic” to false positives.
- Give enough importance to the preliminar data analysis/feature extraction process, since about 75% of the prognostics problem is having good input information.

Audience questions through the Whova App

- To all presenters: performance determined at design time using historical data may not always hold when deployed. How do you track that? How long do you need to wait to track and identify this issue? Is there risk of losing customer trust in the meanwhile?
- Do you have experience applying A/B testing to PHM models, perhaps by holding out assets, to prove effectiveness in cases where you can let failures play out? Back testing works great for communicating results until you deploy and start changing the future.
- @Katarina: if you are working for a customer that does not currently have a value for lost RUL on a platform or component, where would you start for optimizing the loss in early predictions?