Computational Intelligence Applications to PHM

How The Industrial Internet has Transformed my Job

PIERO P. BONISSONE, SSA CHIEF SCIENTIST, GE GLOBAL RESEARCH

bonissone@ge.com

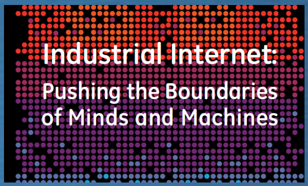
GE imagination at work

Outline

CI Applications to PHM – Industrial Internet

Outline

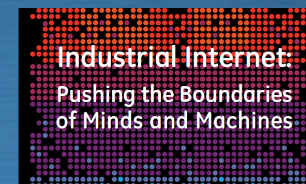
CI Applications to PHM – Industrial Internet



1. The Industrial Internet

Outline

CI Applications to PHM – *Industrial Internet*



1. The Industrial Internet

2. CI in Performance Management: Analytics for PHM

2.1 BC Analytics - (Before the Cloud)

2.2 Cloud Computing

2.3 AC Analytics – (After the Cloud) – Now

2.4 AC Analytics – (After the Cloud) – Future

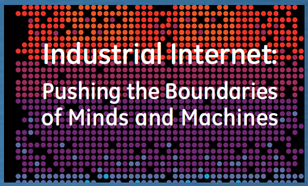
Crowdsourcing, Commoditized Analytics, Model-Agnostic Fusion

… My Job …

GE imagination at work

Outline

CI Applications to PHM — *Industrial Internet*

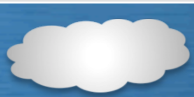


1. The Industrial Internet

2. CI in Performance Management: Analytics for PHM

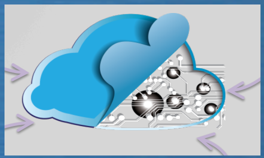
   2.1 BC Analytics - (Before the Cloud)

   2.2 Cloud Computing



   2.3 AC Analytics – (After the Cloud) – Now

   2.4 AC Analytics – (After the Cloud) – Future



   Crowdsourcing, Commoditized Analytics, Model-Agnostic Fusion

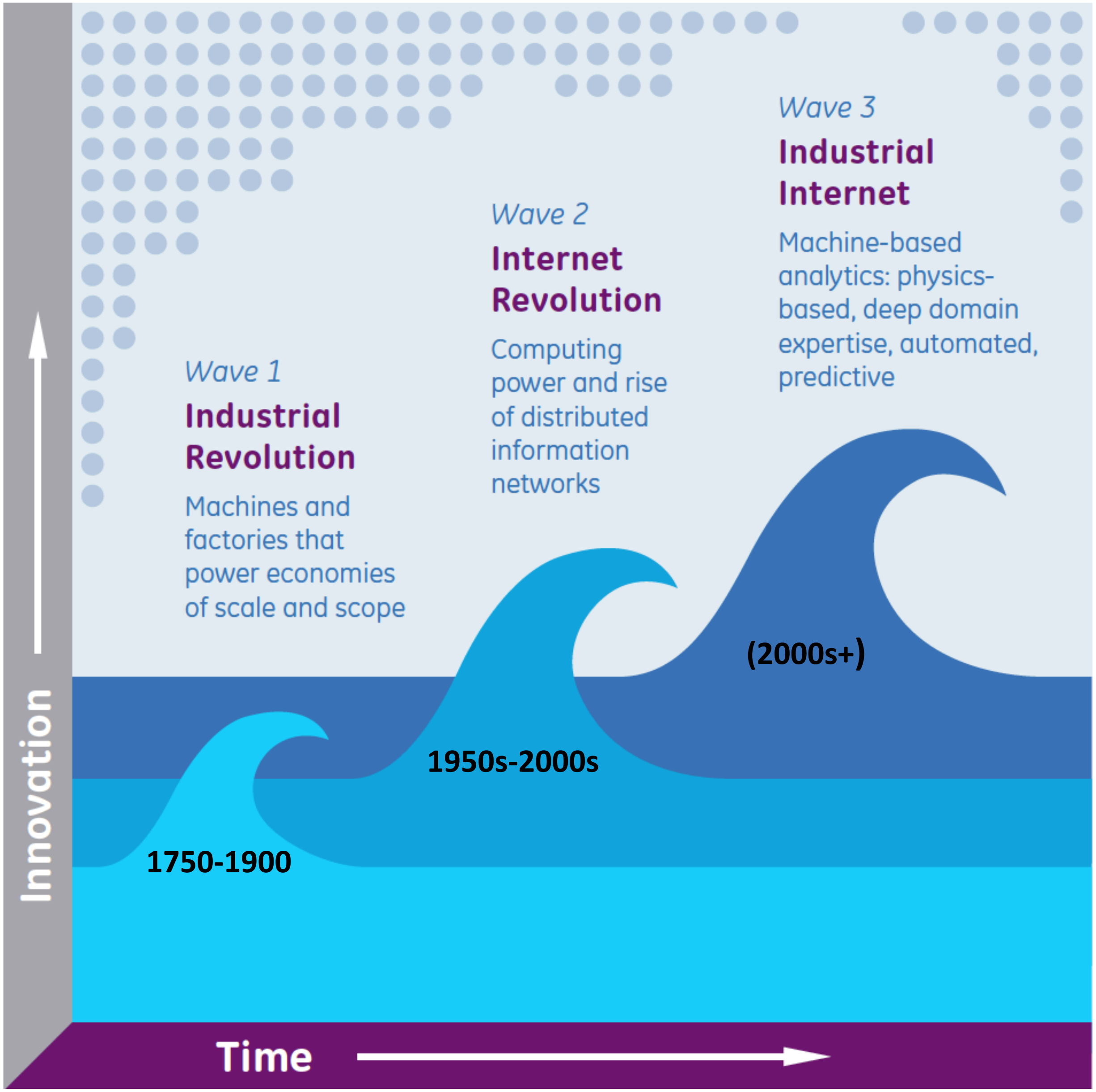
... My Job ...

3. Experiments with Commoditized Analytics

4. Conclusions and Discussion

GE imagination at work

The Three Waves



imagination at work

The Three Waves

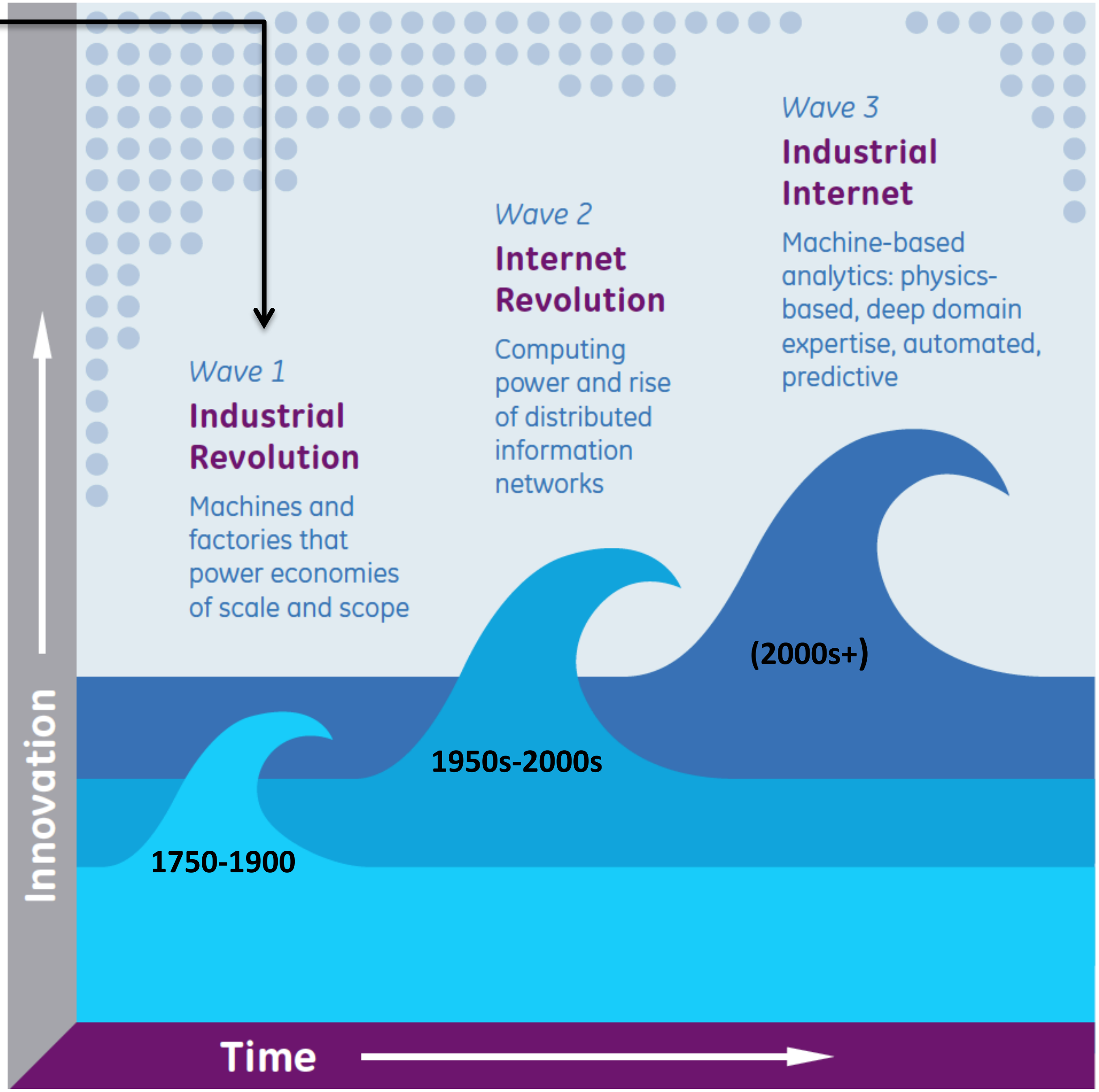
Industrial Revolution - 150 yrs.

What: commercialization of the steam engine, internal combustion engine, electricity.

Impact: Transformed *transportation*, *communication*, and *urban centers.*

Benefits*:* economy of scale + improved efficiency

Side-effects: Resource-intensive, environ. impact

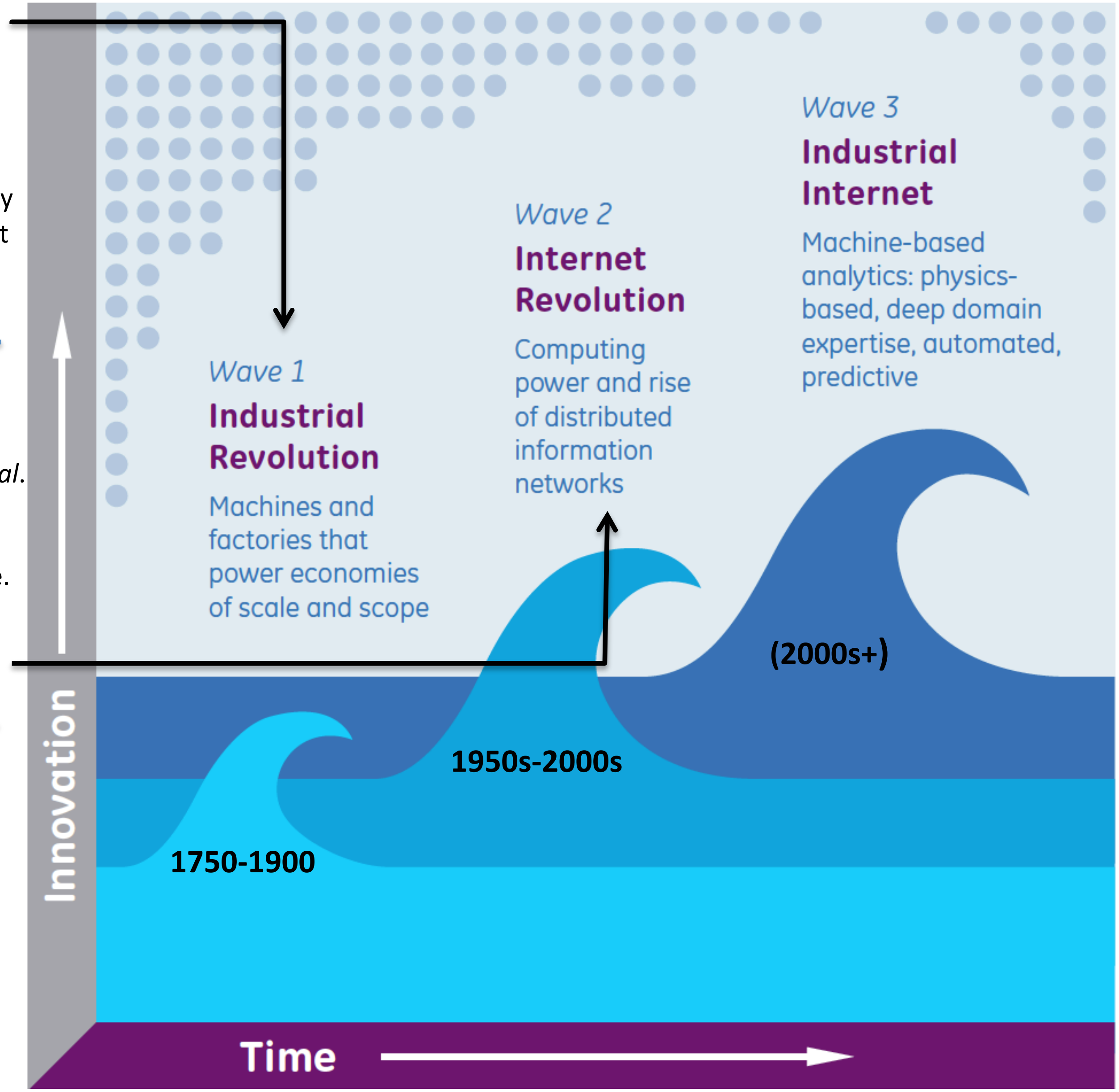


imagination at work

The Three Waves

Industrial Revolution - 150 yrs.
What: commercialization of the steam engine, internal combustion engine, electricity.
Impact: Transformed *transportation*, *communication*, and *urban centers.*
Benefits: economy of scale + improved efficiency
Side-effects: Resource-intensive, environ. impact

Internet Revolution - 50 yrs.
What: mainframes and information packets
Impact: Transformed *analog enterprises to digital*.
Benefits: open networks, horizontal structures, distributed intelligence, e-commerce
Side-effects: Information & knowledge intensive.



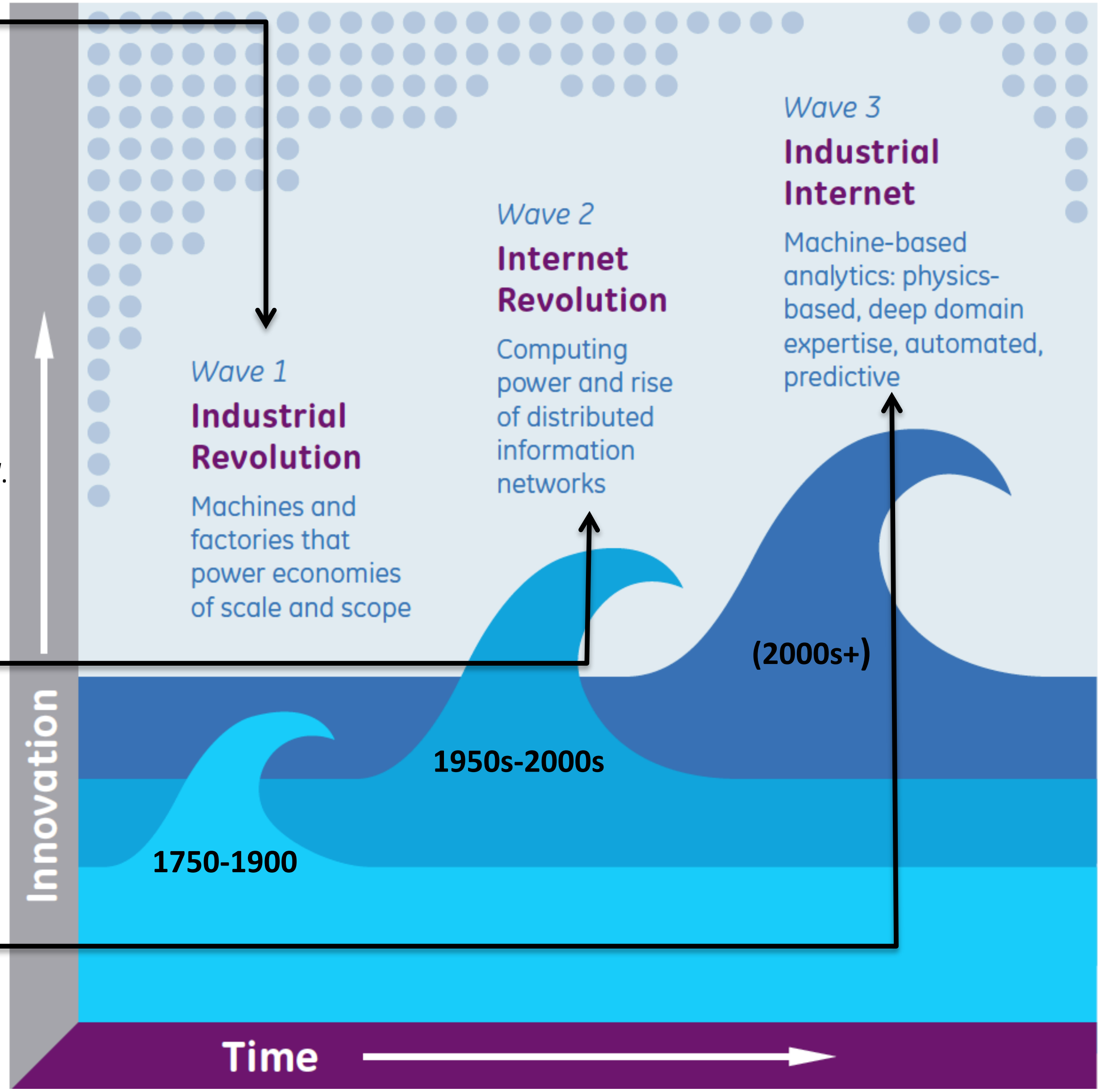
imagination at work

The Three Waves

Industrial Revolution - 150 yrs.
What: commercialization of the steam engine, internal combustion engine, electricity.
Impact: Transformed *transportation*, *communication*, and *urban centers.*
Benefits*:* economy of scale + improved efficiency
Side-effects: Resource-intensive, environ. impact

Internet Revolution - 50 yrs.
What: mainframes and information packets
Impact: Transformed *analog enterprises to digital*.
Benefits: open networks, horizontal structures, distributed intelligence, e-commerce
Side-effects: Information & knowledge intensive.

Industrial Internet - ?? yrs.
What: *Intelligent Devices Intelligent Systems Intelligent Decisioning*
Impact: interconnected intelligent machines *generate big data & drive predictive models.*
Benefits: New service offerings and improved decisions though Integration



imagination at work

Disruption in Industrial Market

Digital industry transformation trends

	INDUSTRY	TRANSFORMATION	ANALOG INDUSTRY	DIGITAL INDUSTRY
2000	**Communications:** Telco's and Cable	Data Transmission	Landline POTS & MCI	Mobile Internet
2005	**Consumer:** Retail Media Gaming, Advertising	Transactions & Interactions	Stores – Music, Book, DVD & Tower Records, Borders, Blockbusters	iTunes Kindle Online Media
NOW	**Industrials:** Energy Healthcare Aviation Mining Transportation …	Sensing Analytics Control & collaboration	Analog products Manual processes Limited use of sensors & software	HC – telemedicine, digital health records, medical devices Aviation – integrated modular avionics Energy – smart grid, smart buildings

First movers & fast followers win

New opportunities emerging…enabled by technology, and driven by mega trends

Rising customer expectations in both cost & complexity reductions

Accelerating pace of Software innovation…real-time capabilities

New competitive threats and challenges…and new business models

imagination at work

The Industrial Internet

A convergence of enabling technologies is setting the stage for the Industrial Internet.

1. Internet of Things

Hyper-connectivity: a living network of machines, data & people

More devices tap into the Internet than people on Earth to use them

2. Intelligent Machines

Increasing system **intelligence** through embedded software

Rise of machines: networked devices overtook the global population in 2011

3. Big Data

Democratization of data

Data **overload**: 2.5 exa-bytes (quintillion bytes) of data created every day

4. Analytics

Generating **data-driven insights**

Enhancing asset performance by detecting & predicting

Algorithms on installed base operational data

INVESTING | 11/30/2011 @ 1:38PM | 1,666 views

GE Goes To California To Surf Industrial Internet

Trefis Team, Contributor

Presence in Silicon Valley
Focused on software & analytics
200+ employees hired in 12 months
Targeting 800+ staff

Shared services
GE digital architecture for industrial solutions
Expertise: user experience, cloud, analytics, platforms, security, collaboration, mobility

Value of Data & Analytics

Monitor fleet of ~25,000* engines … 3.6MM flight records/month

B777

Prognostics

- ✓ Dispatch reliability
- ✓ Preventive maintenance
- ✓ Asset utilization

\+

GE90

Asset Productivity

- ✓ Enhanced service offerings
- ✓ Airline cost structure
- ✓ Fuel performance

=

DATA

90,000 flight records analyzed

~200 parameters per flight record

~18MM parameters per month

System & Optimization

- ✓ Time & space management
- ✓ Fuel efficiency
- ✓ Airspace capacity

Prevent failures = customer efficiency

Streamline operations = increased airline productivity

Integrated systems = value-added services

Drives strong alignment with customers

Creates productivity in long-term service agreements

Value-added services fuels growth

imagination at work

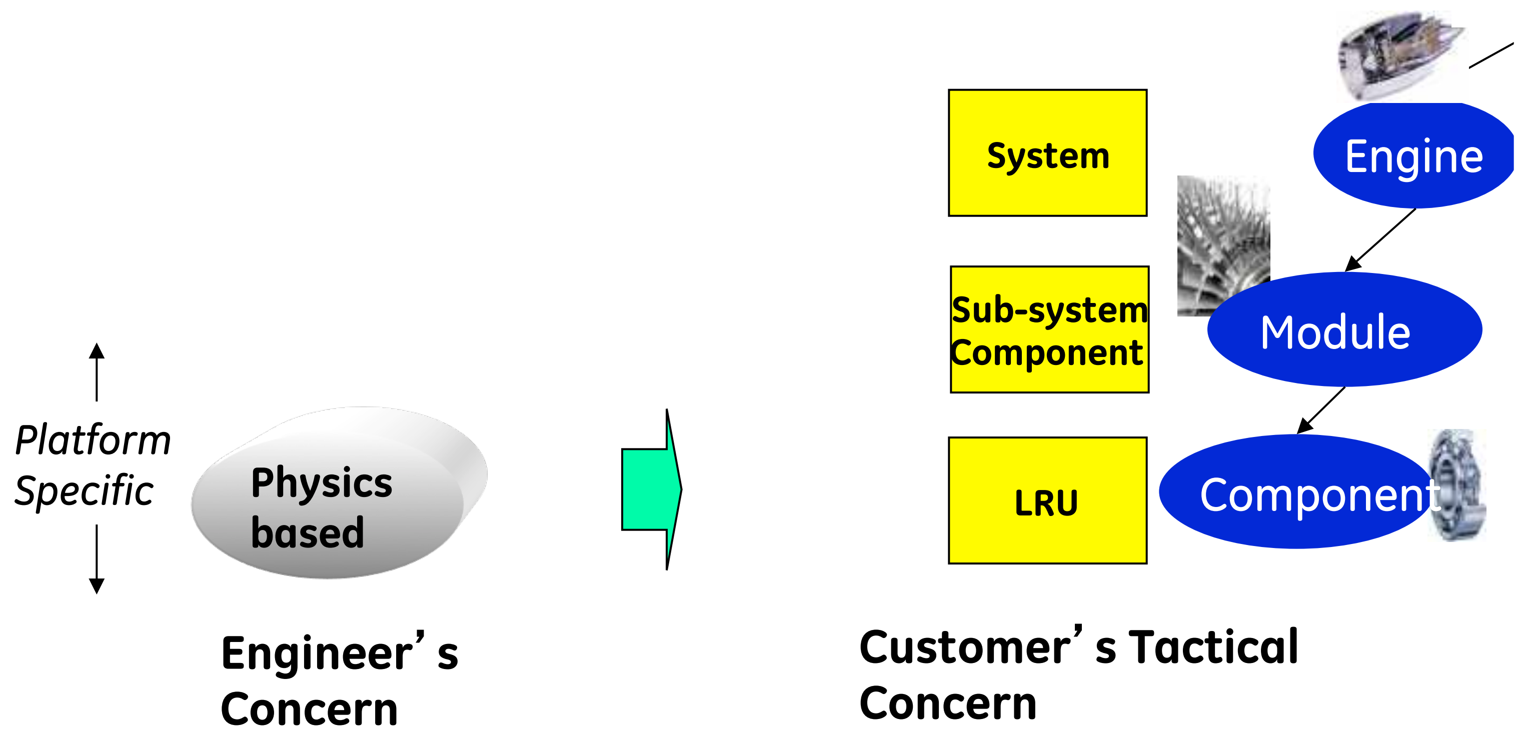
* Includes GE & joint-venture engines with CFM and Engine Alliance. CFM is 50/50 JV with SNECMA. Engine Alliance is 50/50 JV with Pratt & Whitney

2. CI in Performance Management: Analytics for PHM

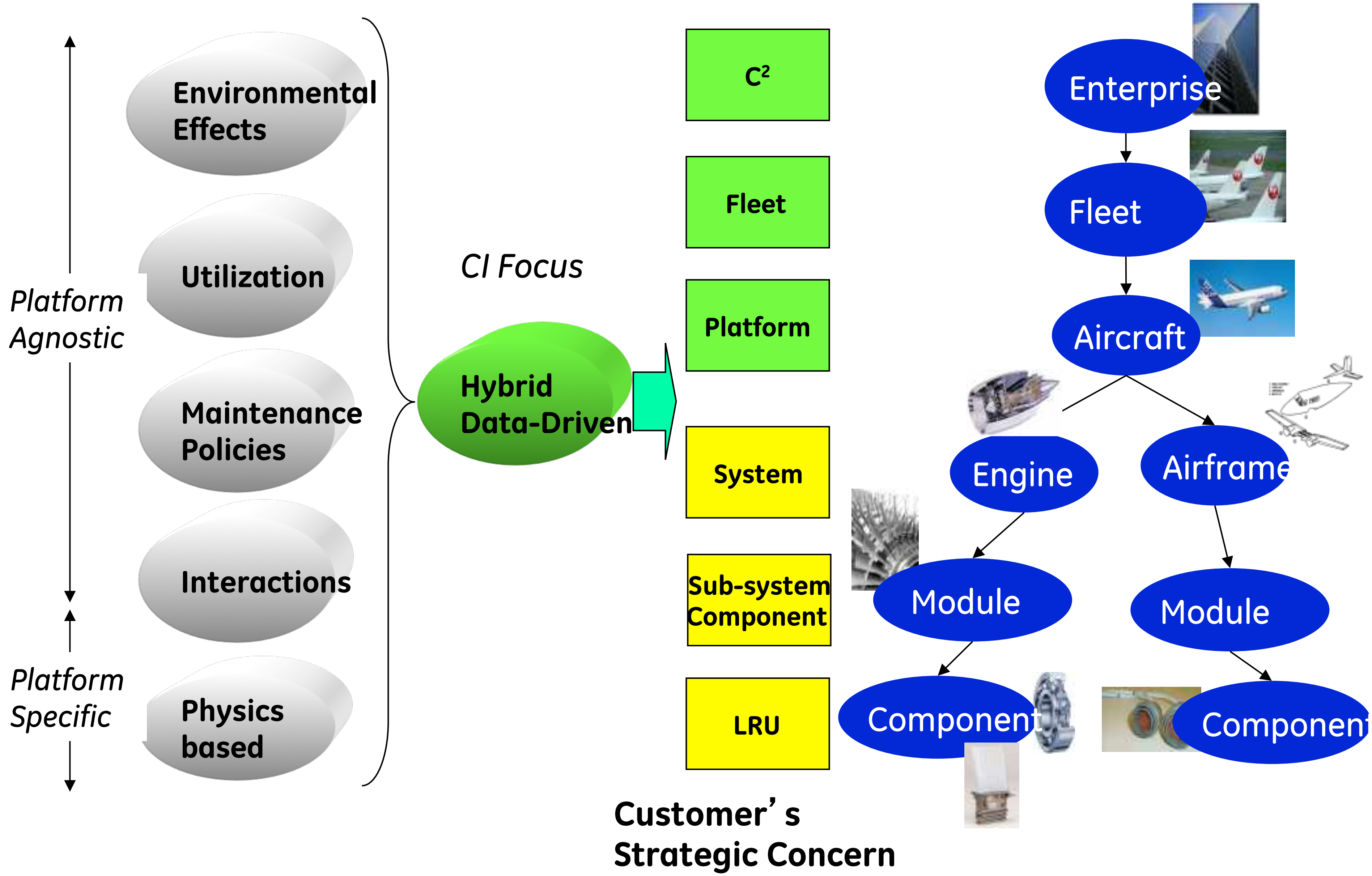
Computational Intelligence

PM Functions

Classical Engineering Approach



Computational Intelligence (CI) Approach

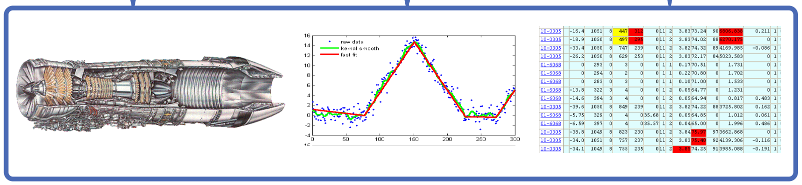


Integrating physics-based with data-driven approaches into hybrid systems

Performance Management Functions

Stage 1: AD + RM&D

Data Acquisition | **Anomaly Detection** | **Diagnostics**



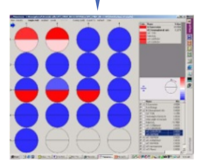
Data Acquisition: Remote monitoring of a fleet of assets

Anomaly Detection: Identification of deviating assets

Diagnostics: Root cause Isolation

Stage 2: Prognostics

Prognostics

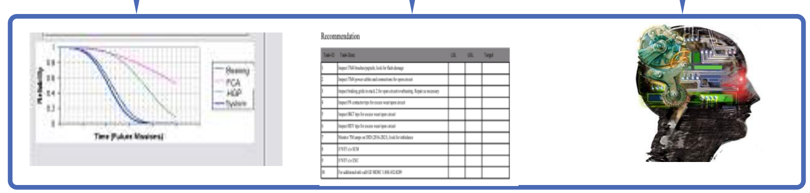


Prognostics: Life prediction.

Unplanned to planned maintenance

Stage 3: Control & Optimization

Control / Advisory Generation | **Optimization** | **Feedback & Learning**



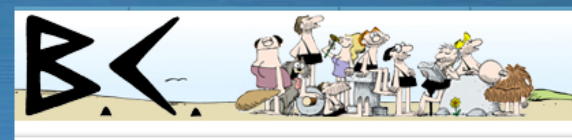
Control/Advisory Real-time safety decisions

Optimization: Offline decisions

Feedback Validation of maintenance cases

imagination at work

2.1 BC Analytics (Before the Cloud)



Manual Process

- Model Building
- Tuning of Legacy Models
- Static Fusion

My Old Jobs:

Building Modeling for Performance System Management (PSM):

- *High Priest of AI & Knowledge Engineering*
- *Machine Learning (ML) Magician*
- *Lazy ML Magician*

imagination at work

My <u>REALLY</u> Old Job (1980's)

My Oldest Job at GE

Job description

High priest of AI and Automated Reasoning,

aka the "<u>Knowledge Engineer</u>"

Main goal

Manually translate experiential knowledge into reasoning (expert) systems

Main outcome

Handcrafted AI-based intelligent systems (Rule-based, Case-Based, Fuzzy, etc.)

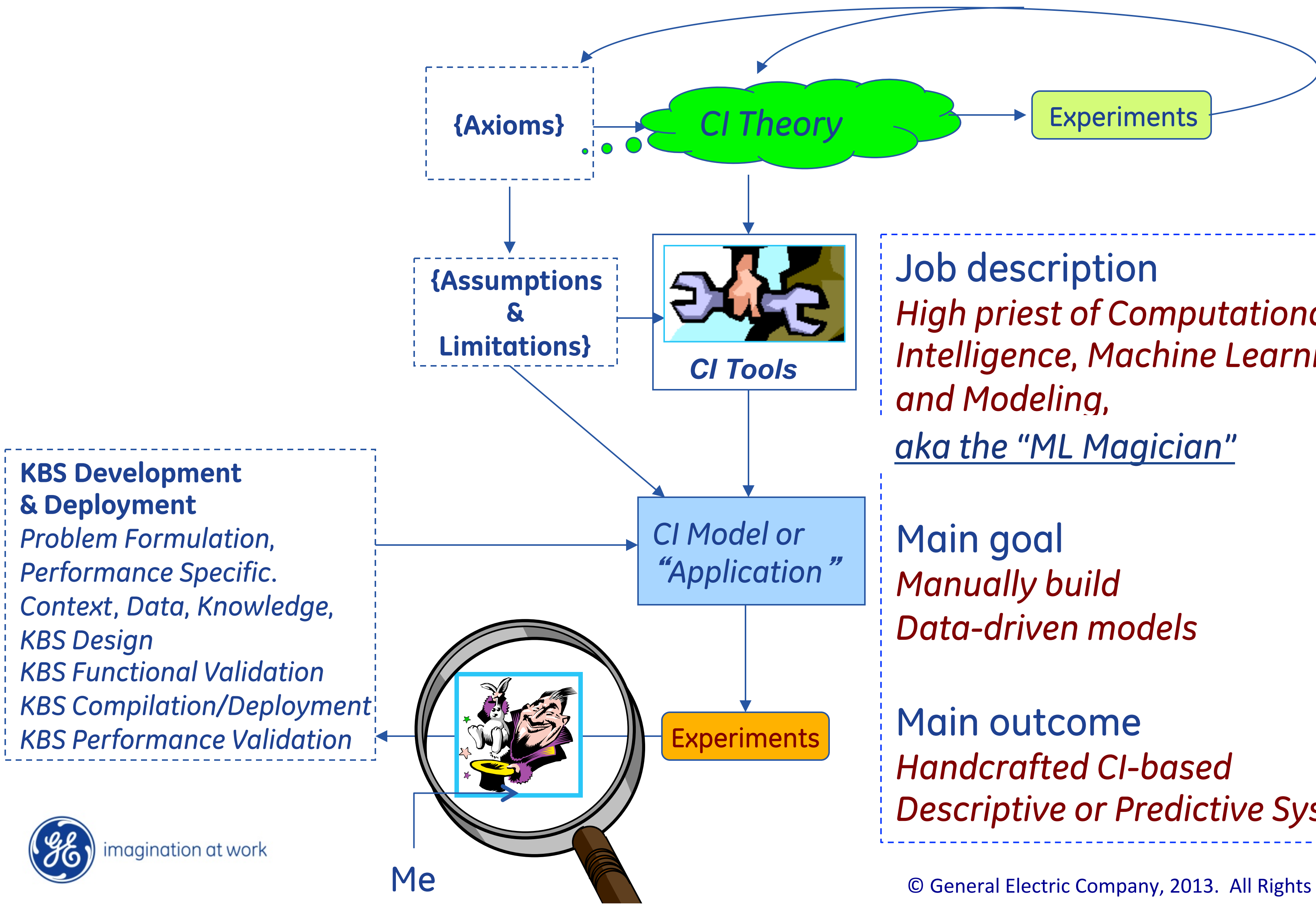
Me

. . . and diesels are serviced with artificial intelligence.

GE applies artificial intelligence to the repair of diesel-electric locomotives enabling the serviceman to diagnose the exact cause of difficulty while ensuring that he doesn't overlook unreported problems which might cause the locomotive to fail soon after "servicing."

imagination at work

My job two decades ago (1990's): ML Magician



Job description

High priest of Computational Intelligence, Machine Learning, and Modeling,

aka the "ML Magician"

Main goal

Manually build

Data-driven models

Main outcome

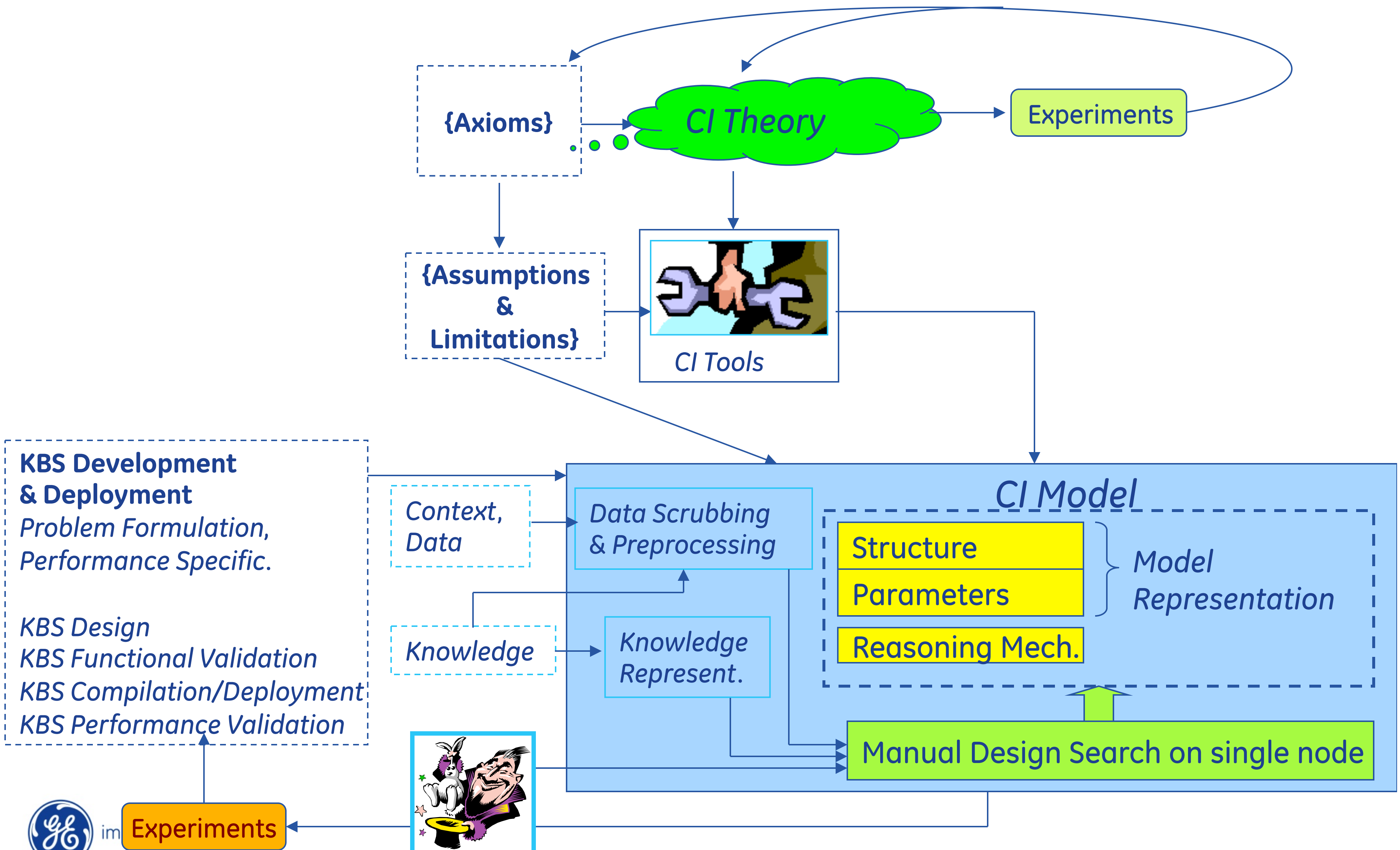
Handcrafted CI-based

Descriptive or Predictive Systems

imagination at work

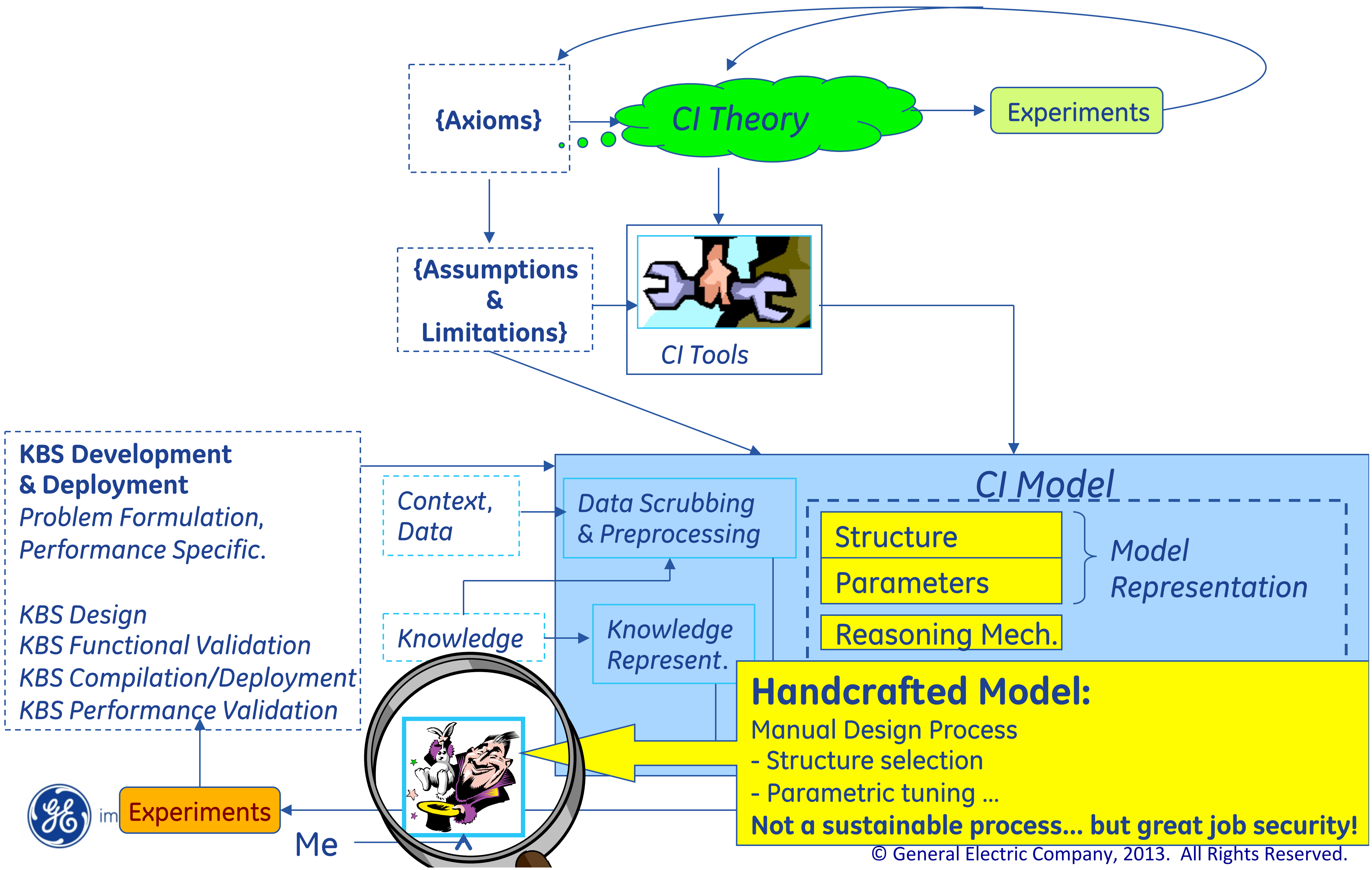
… ➔ Job Security

2.1 BC Analytics:
Manual Process

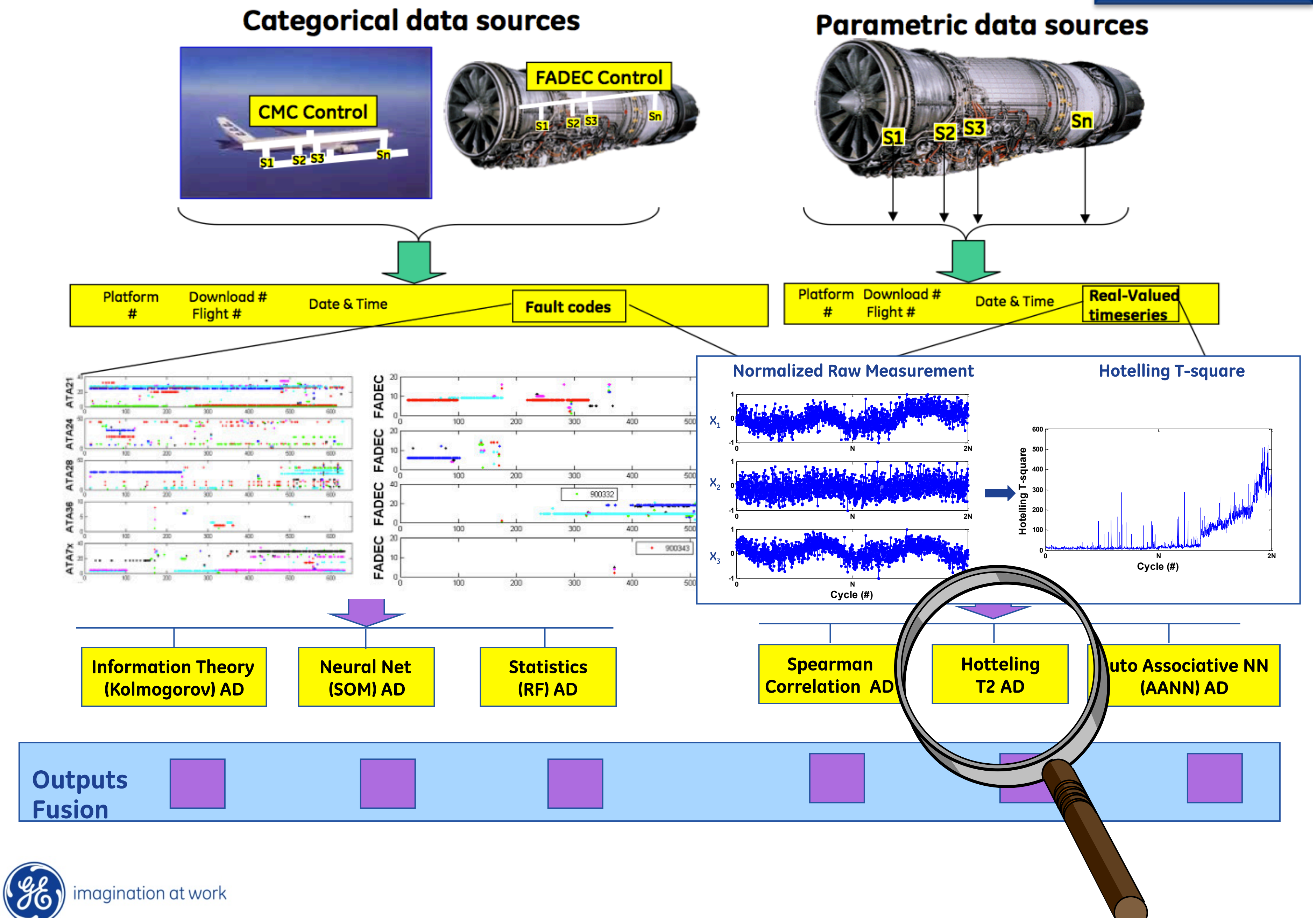


… → Job Security

2.1 BC Analytics:
Manual Process



Example: Data Driven Model from Time Series - Anomaly Detection



imagination at work

My job a decade ago (2000's) → Still employed

Job description

A more efficient "Magician" relying on Meta-heuristics (MH), such as Evolutionary Algorithms, to search in the model design space (on single node)

aka "the Lazy ML Magician"

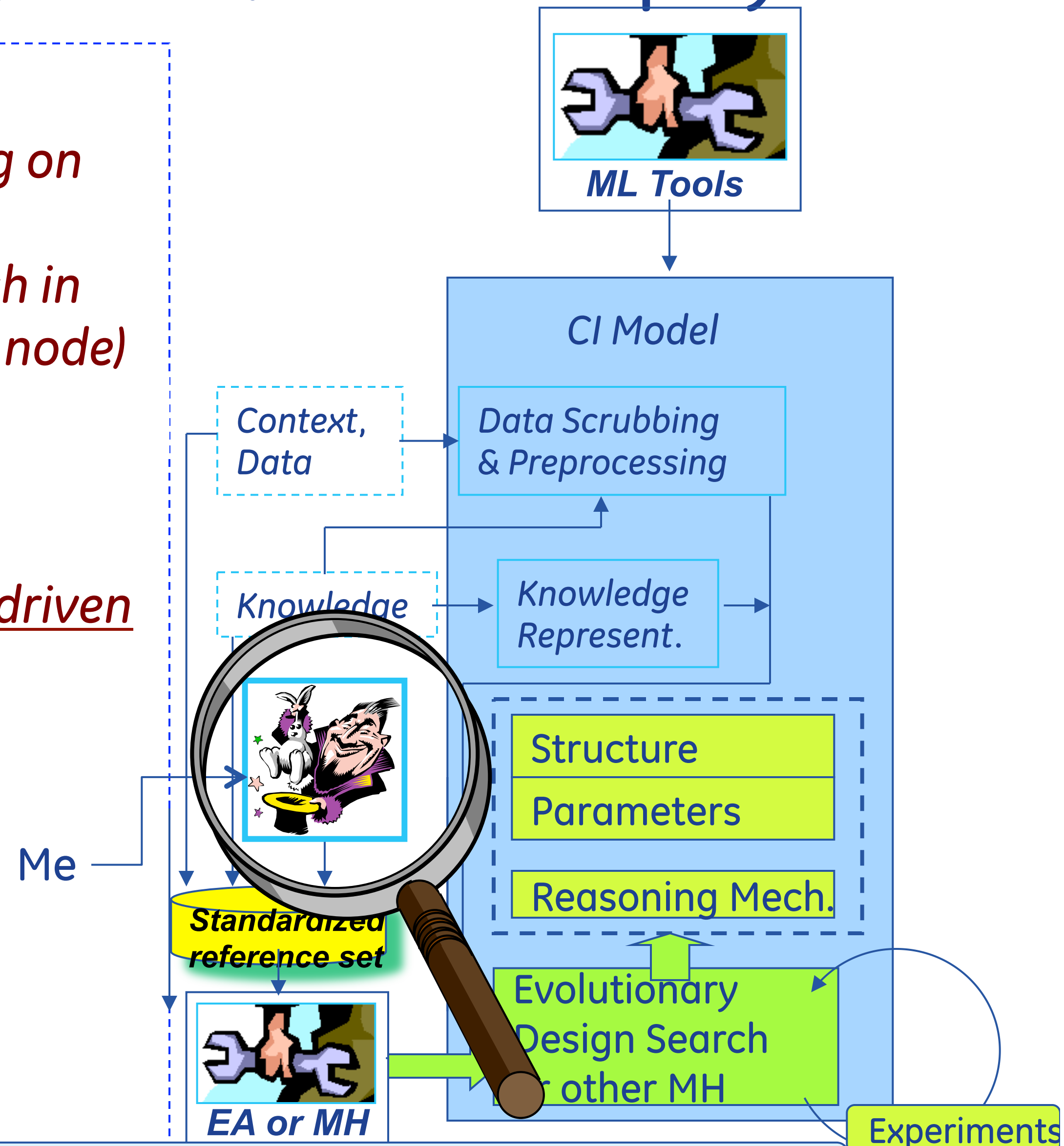
Main goals

(1) Create a process to build data-driven models semi-automatically

(2) Manually create static model ensembles and Fusion

Main outcome

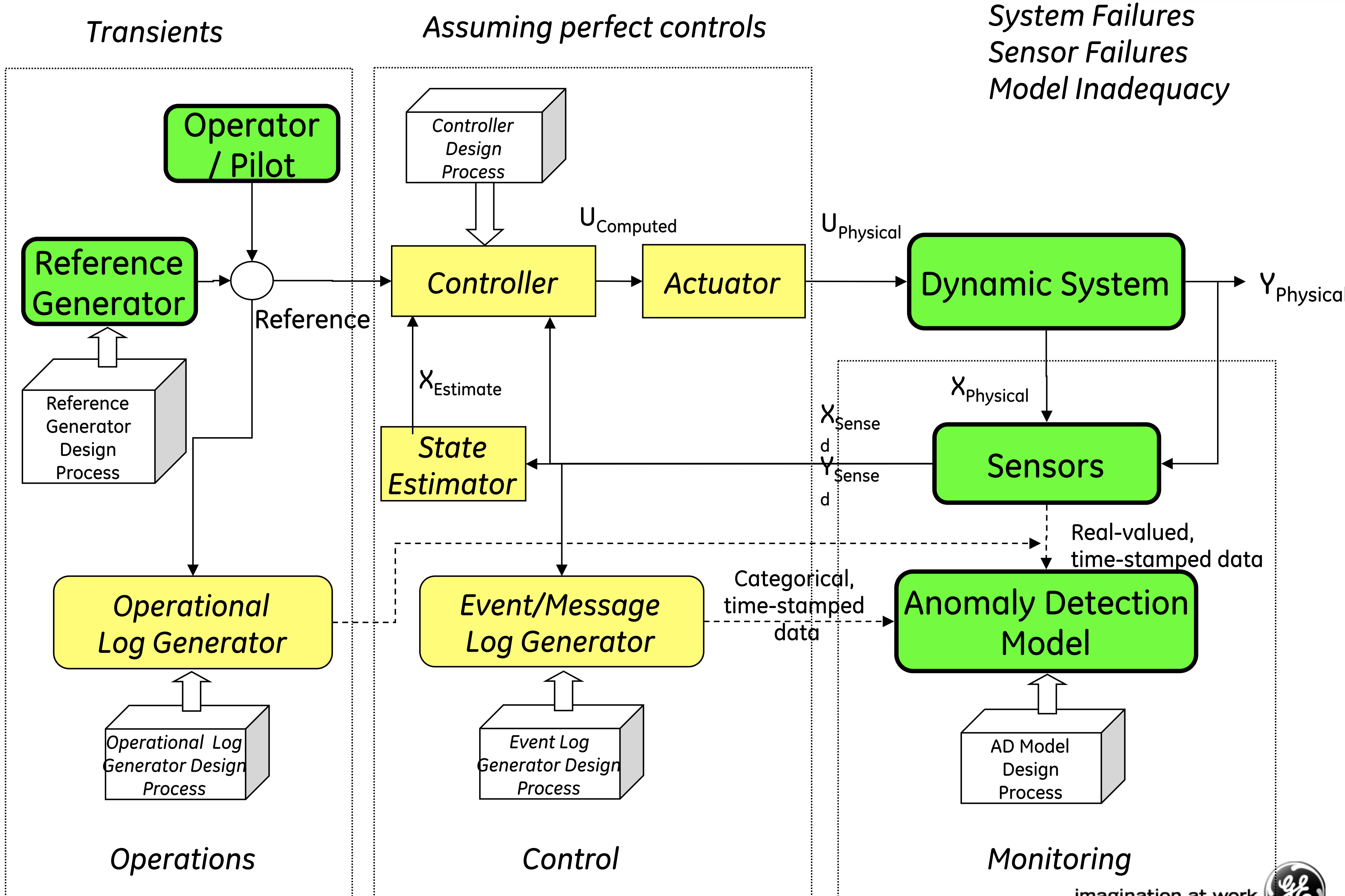
Process based CI-based Descriptive or Predictive Models



Still Issues: time-to-market, provisioning, lifecycle maintenance,...

Example 1: Semi-Automate Anomaly Detection - Sources of Anomalies

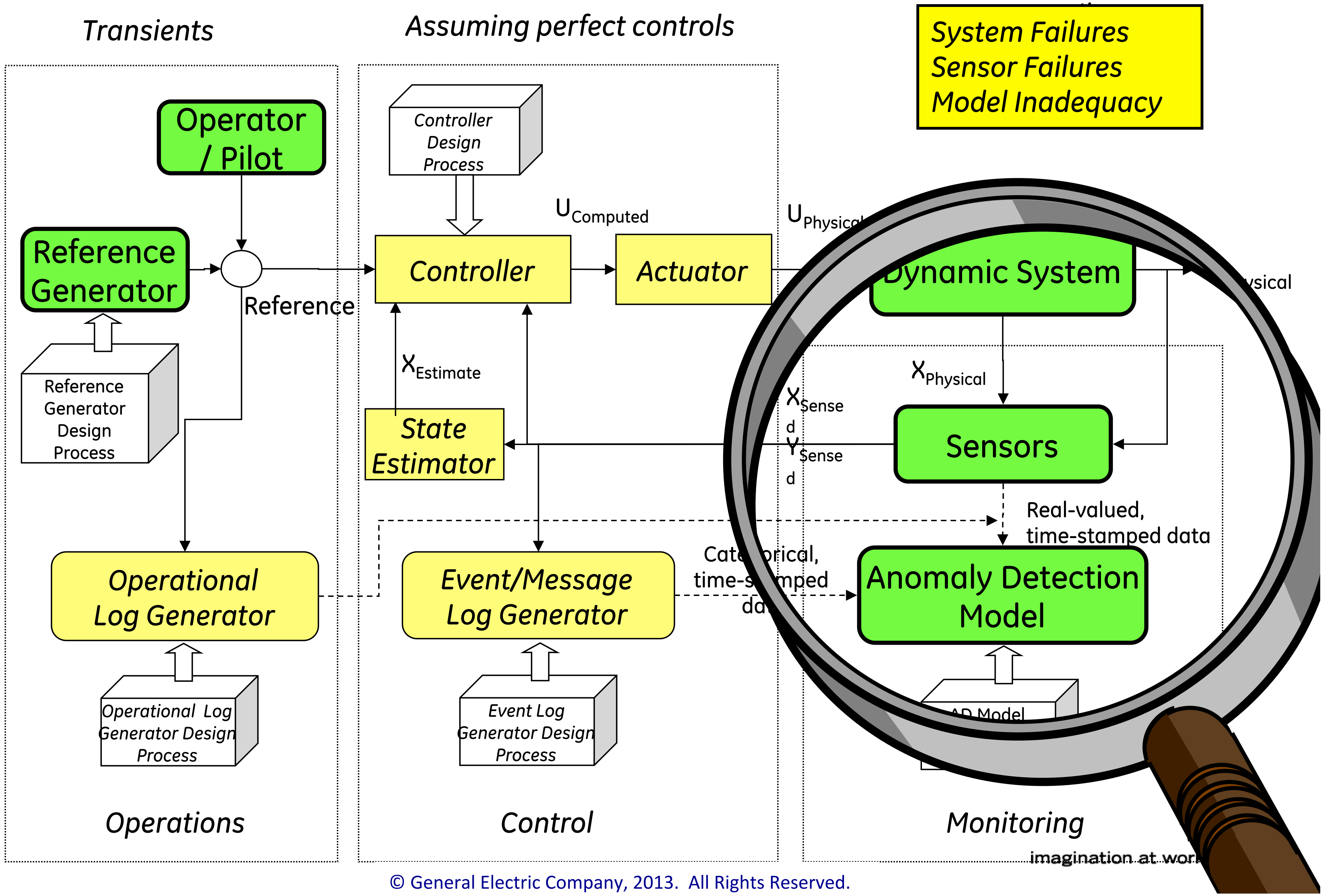
2.1 BC Analytics:
GE-90 Anomaly Detection
Problem Formulation



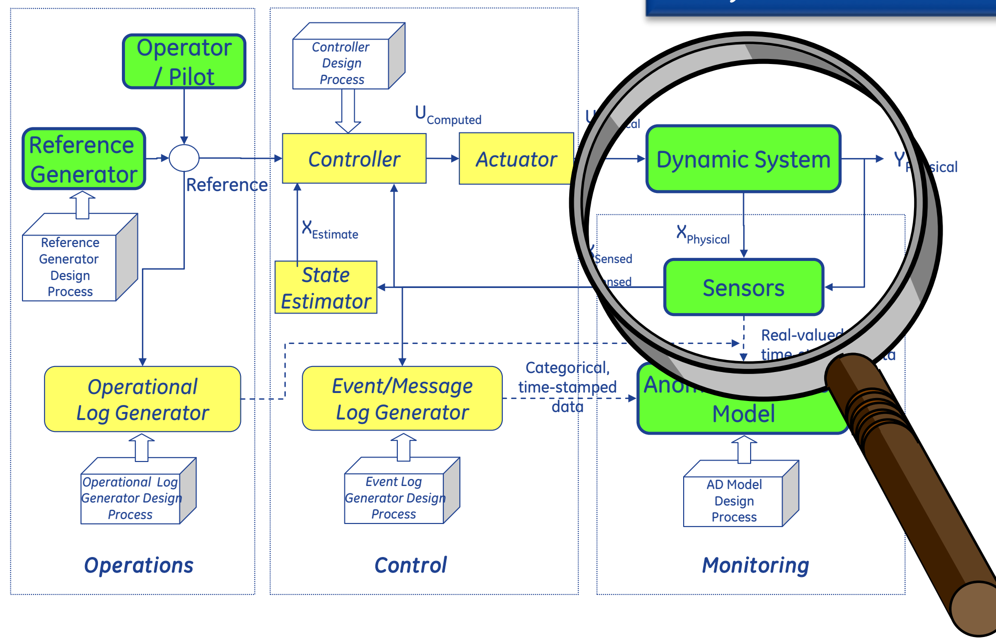
imagination at work

Example 1: Semi-Automate Anomaly Detection - Sources of Anomalies

2.1 BC Analytics:
GE-90 Anomaly Detection
Problem Formulation



Dynamic System: Simulated Aircraft Engine [GE 90]



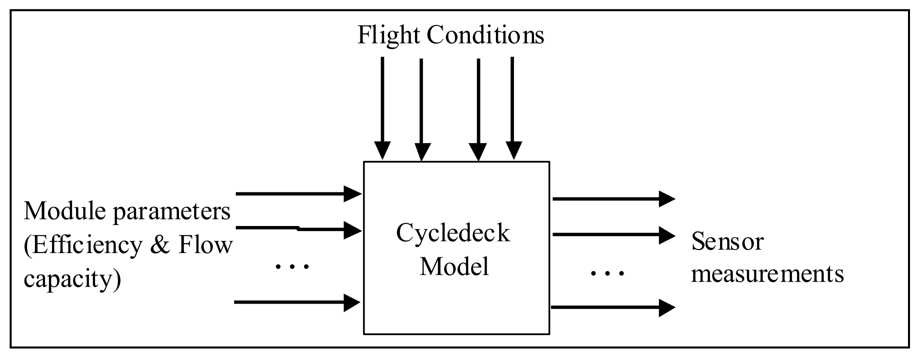
Physics-Based Simulation

–**CLM:** Component Level Model is a physics-based thermodynamic model widely used to simulate the performance of a commercial aircraft engines.

–**Flight Regime:** Flight conditions, such as altitude, Mach number, ambient temperature, and engine fan speed, and a large variety of model parameters, such as module efficiency and flow capacity are inputs to the CLM

–**Outputs:** CLM's outputs are the values for **pressures, core speed and temperatures** at various locations of engine, which simulate sensor measurements.

–**Noise**: Realistic values of sensor noise can be added after the CLM calculation.



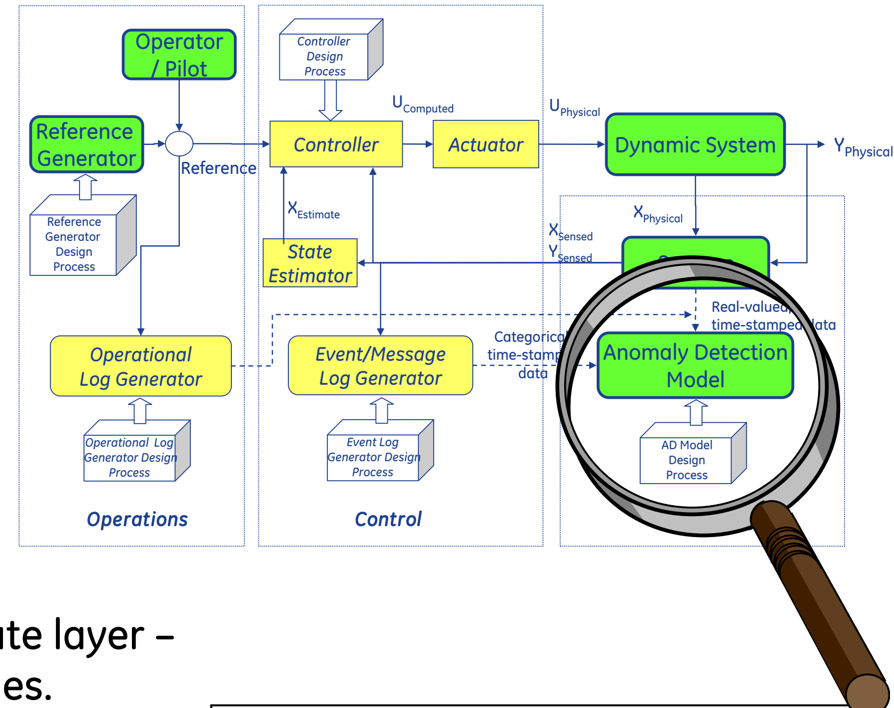
imagination at work

Basic AD Model: Auto-Associative Neural Network

2.1 BC Analytics:
GE-90 Anomaly Detection
AD Model

Rationale

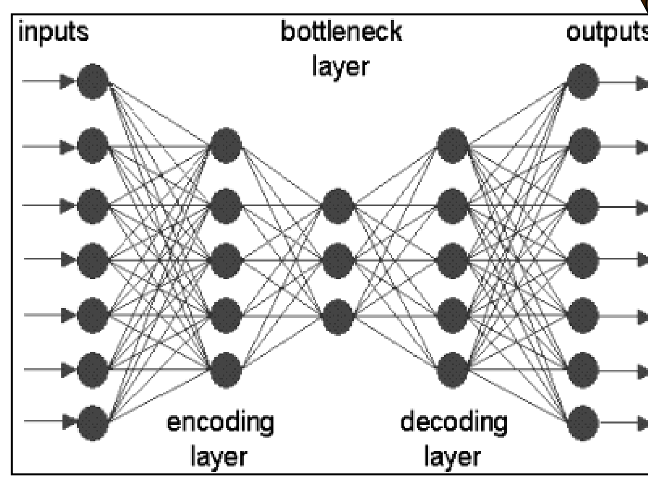
The Auto-Associative Neural Network (AANN) leverages covariance information like other approaches (SRC and T2). The AANN also produces sensor estimated values to replace the ones generated by faulty sensors. This approach provides a better discrimination between sensor faults and system component faults.



Definition/Properties

AANN computes the largest Non-Linear Principal components (NLPCA) - the nodes in the intermediate layer – to identify and remove correlations among variables.

NLPCA uncover both linear and nonlinear correlations, without restriction on the type of the nonlinearities present in the data.



Computation

Traditional NN training with back-propagation

Variable Contribution

Residuals magnitude/distribution

imagination at work

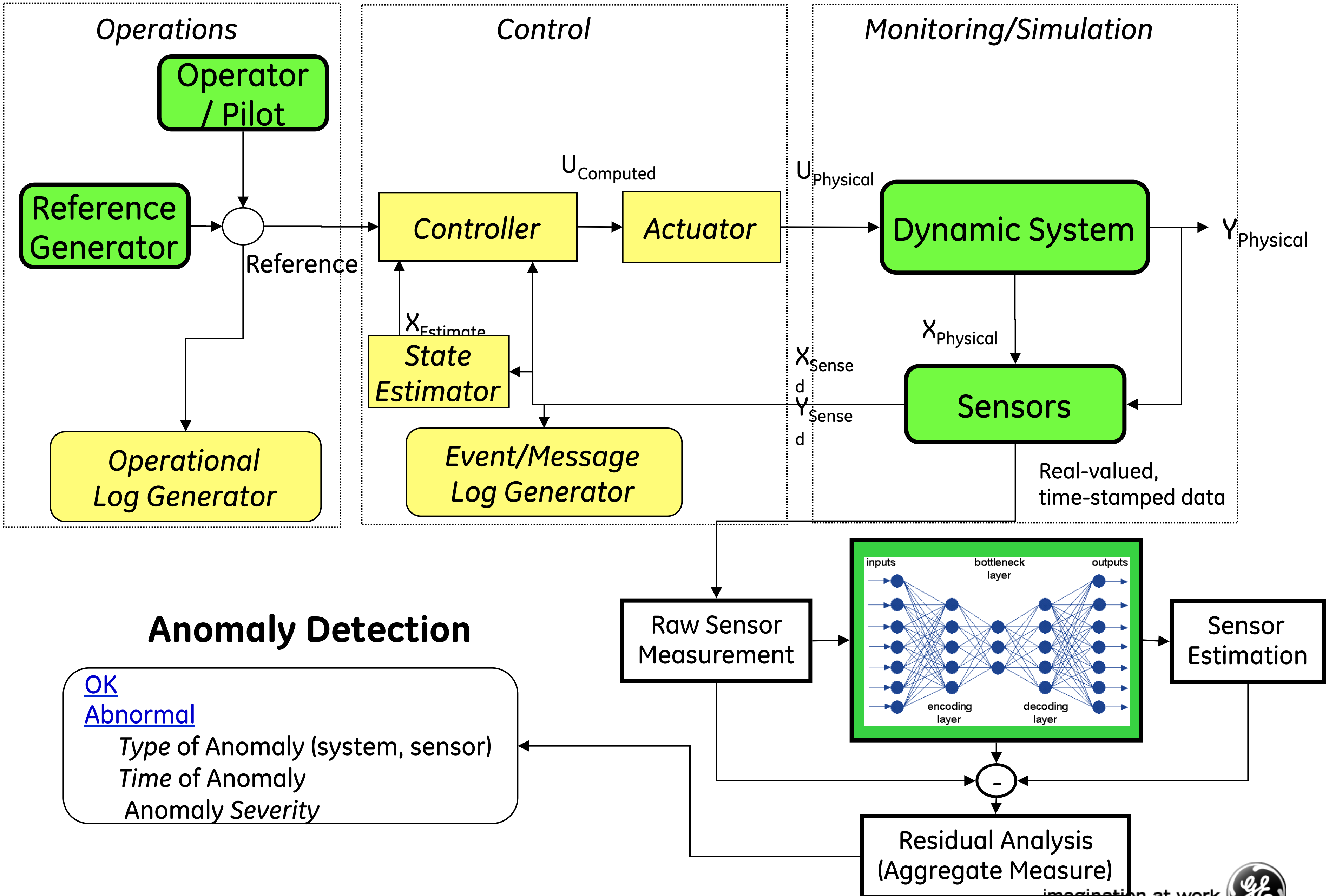
Experiments with Simulated GE90 Aircraft Engines

- Experiment Setup
- Segmentation of the Operating Space
- Experiments

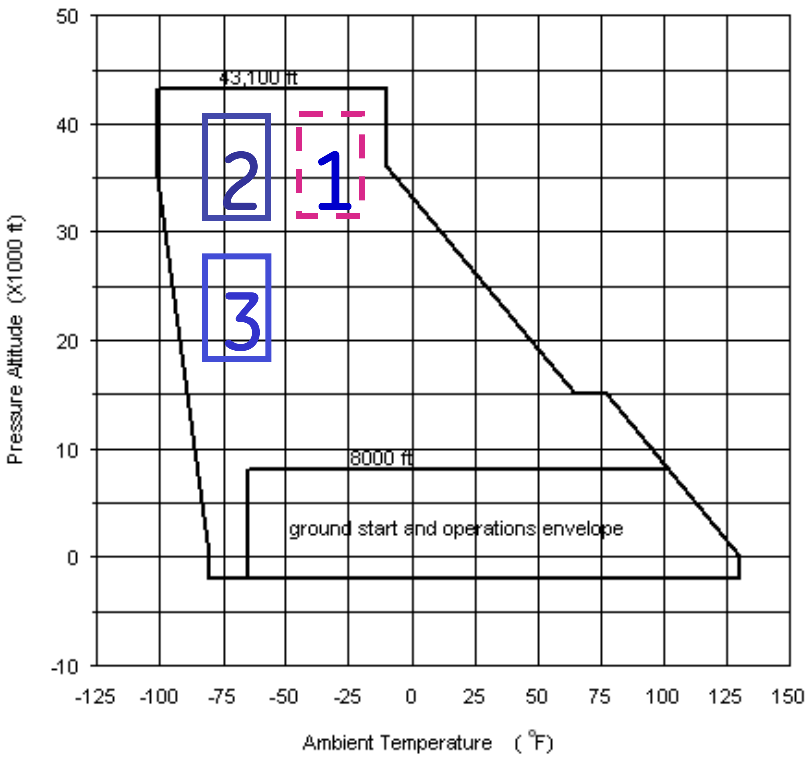
1st	- 3 local models
2nd	- 1 Global Model
3rd	- 3 local Models + Supervisory Model

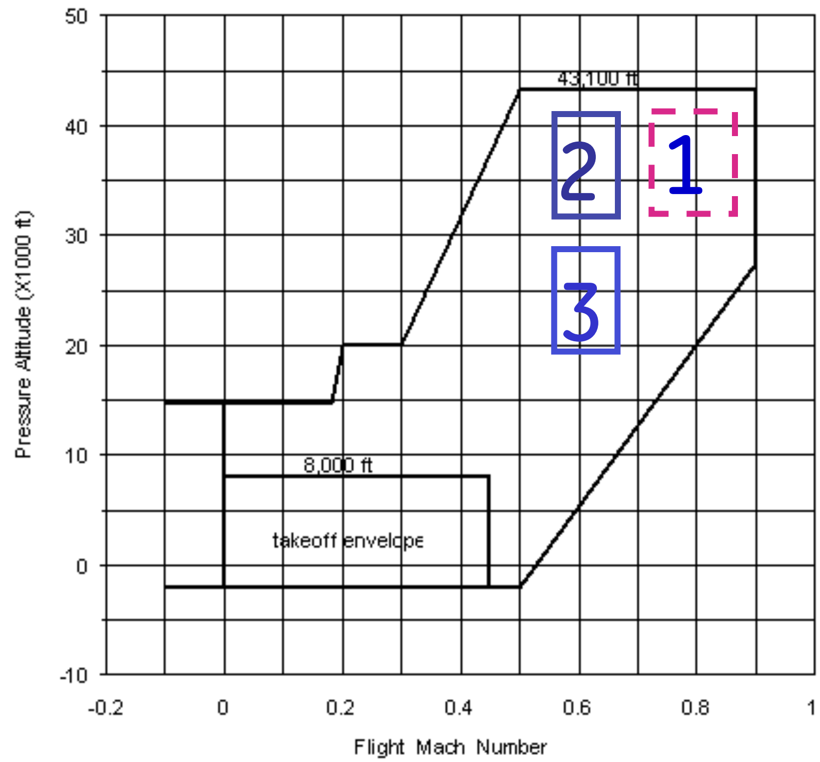
imagination at work

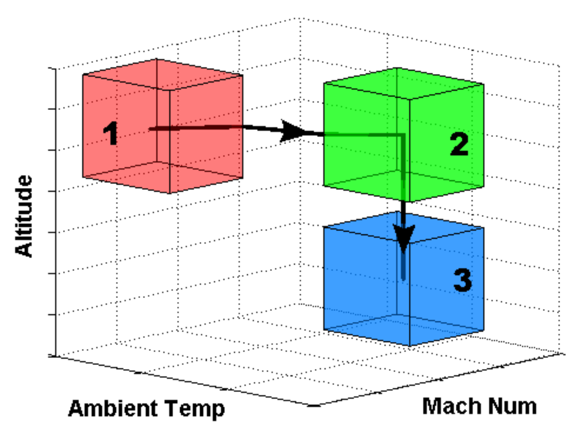
Experiment Setup



Segmentation of the Operating Space
Three regions in the Flight Envelops





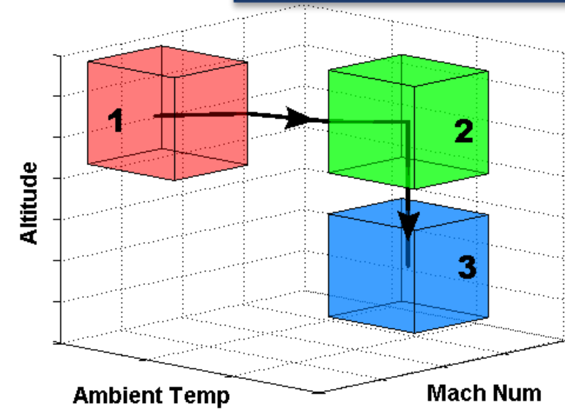


Experiments

Experiments Settings

– We used a **steady state CLM** model for a commercial, **high-bypass, twin-spool, turbofan engine**.

– We can manipulate flight conditions to simulate different operation regimes (i.e. flight envelops of aircraft) and generate data corresponding to them



1st Experiment

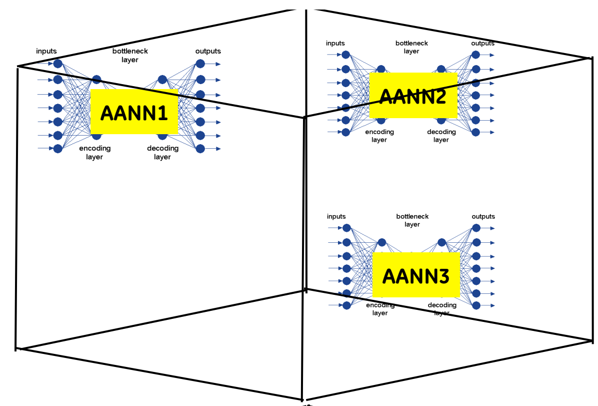
Three AANN's: One for each region in the flight envelop (region)

Vary ALT, Mach and Tamb ->1000 normal operating pts for each region

Run each operation point through CLM to generate a 9x1 sensor vector

900 points for training (200 for validation); 100 points reserved for test

Each local model performs very well (better than global model) in region of competence, but performs poorly outside of its scope



2nd Experiment

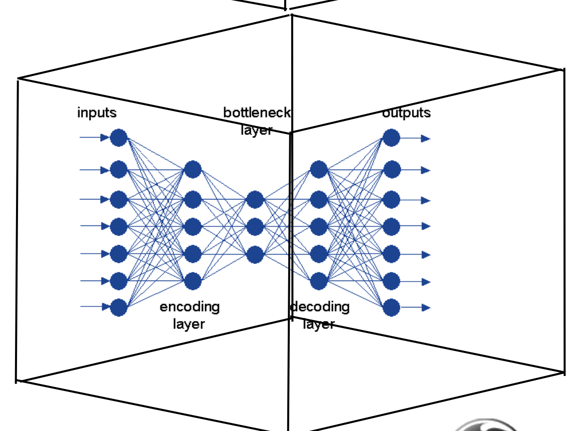
One Global AANN

Train on same 2700 training data points from experiment 1

Run each operation point through CLM to generate a 9x1 sensor vector

Test on the left 300 points

Global model performs fair across all three regions - shows higher variance than each local AANN operating within its scope



Experiment 3

3rd Experiment

- **Three AANN's:** One for each flight envelop region
- **Fuzzy Supervisory Model (FSM)** to interpolate among local AANN's

Simulate the change of flight conditions

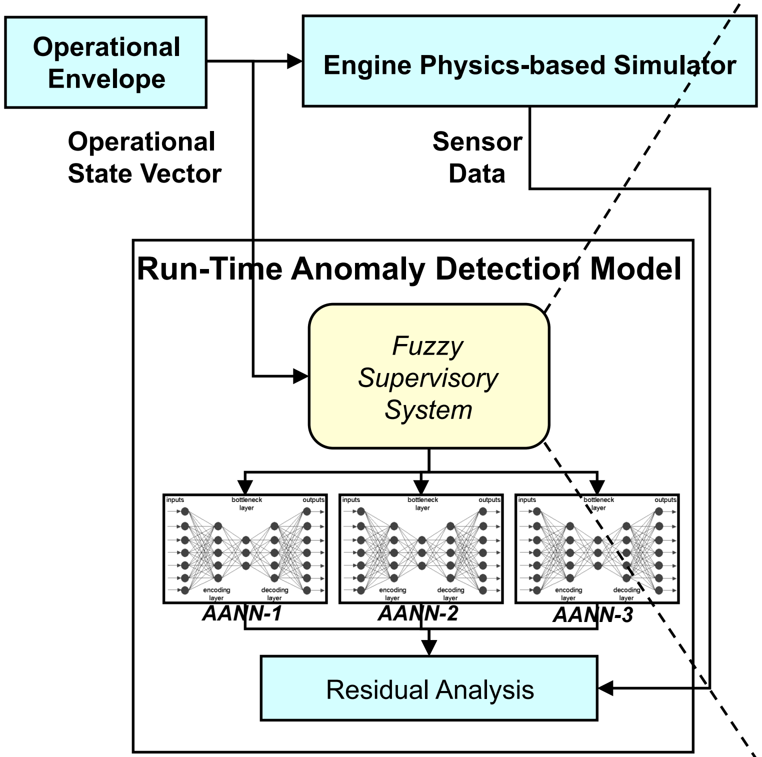
FE1: 200 pts

FE1 → FE2: 200 pts

FE2: 200 pts

FE2 → FE3: 200 pts

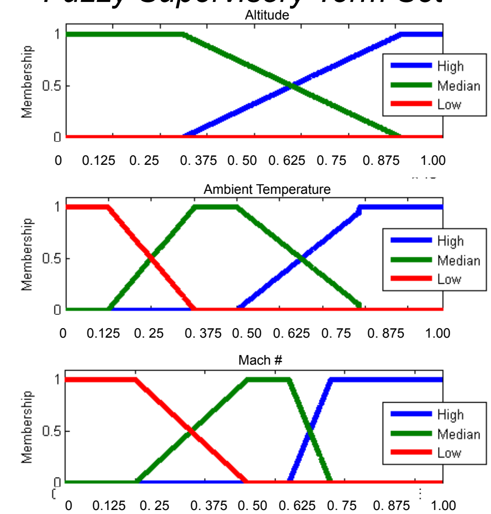
FE3: 200 pts



Fuzzy Supervisory Rule Set

State Variables		Altitude	Amb. Temp.	Mach #	Model #
RULES	R1	High $V_{1,1}(Input_1)$	High	High	AANN-1
	R2	High $V_{2,1}(Input_1)$	Medium	Medium	AANN-2
	R3	Medium $V_{3,1}(Input_1)$	Low	Low $V_{3,3}(Input_3)$	AANN-3

Fuzzy Supervisory Term Set



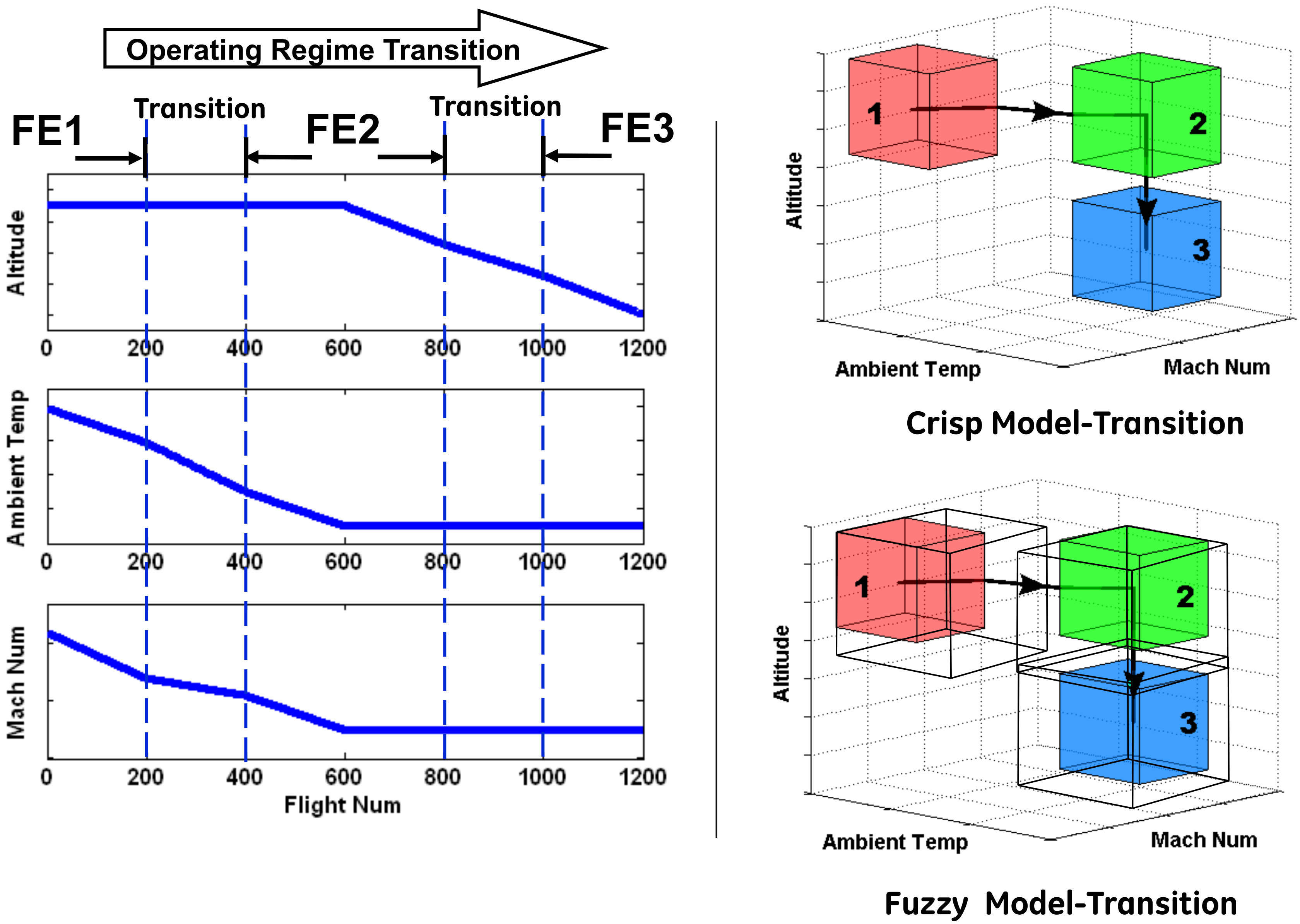
Tested the simulated data on the Fuzzy Supervisory Model + AANN1, AANN2, AANN3

Intentionally making transitions in spaces not covered by any pre-trained flight envelope

Hierarchical structure performs very well across all regions – including transitions

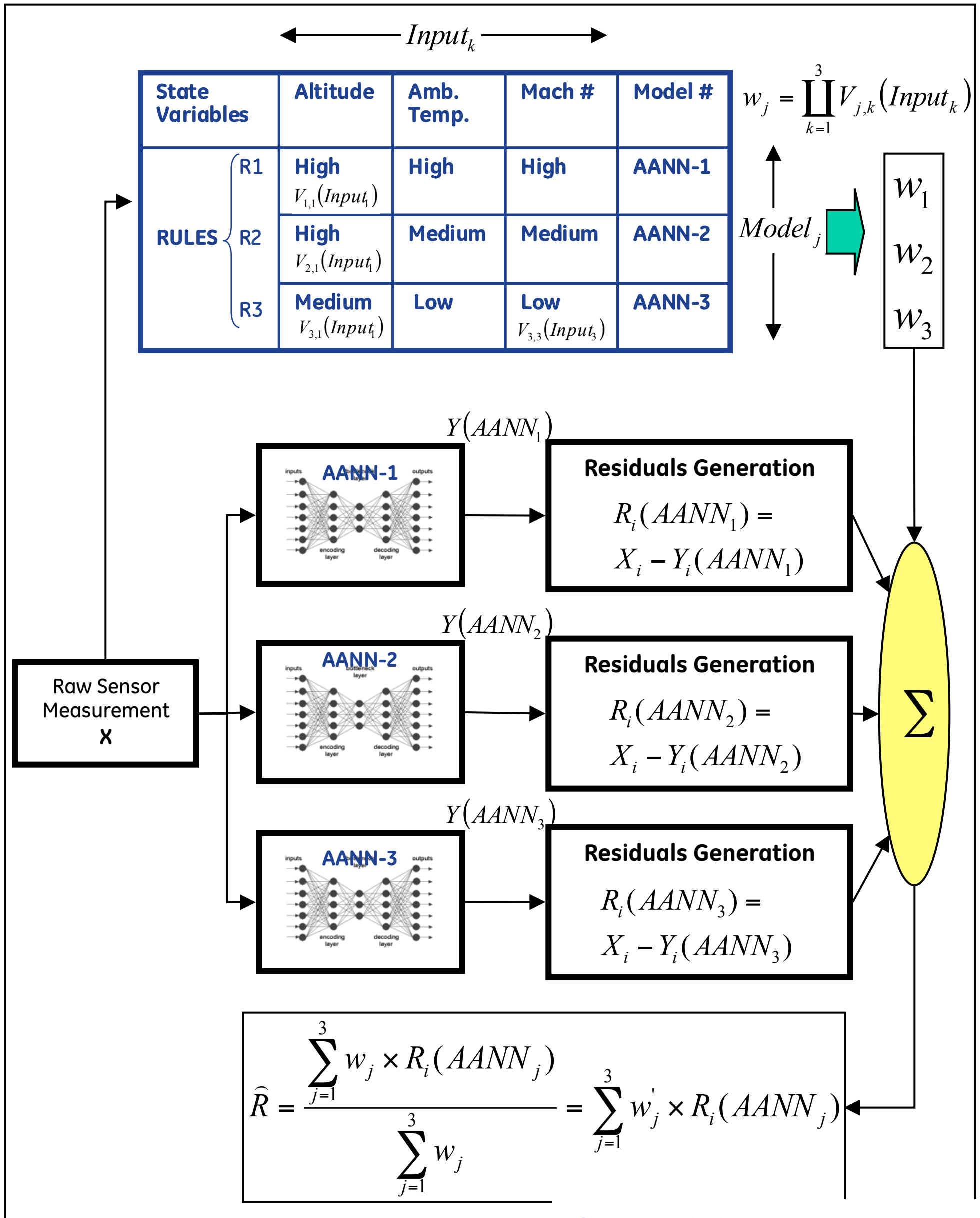
Created a **Fuzzy Supervisory for three local models**
➔ Higher performance across all regions

Flight Envelop Transitions

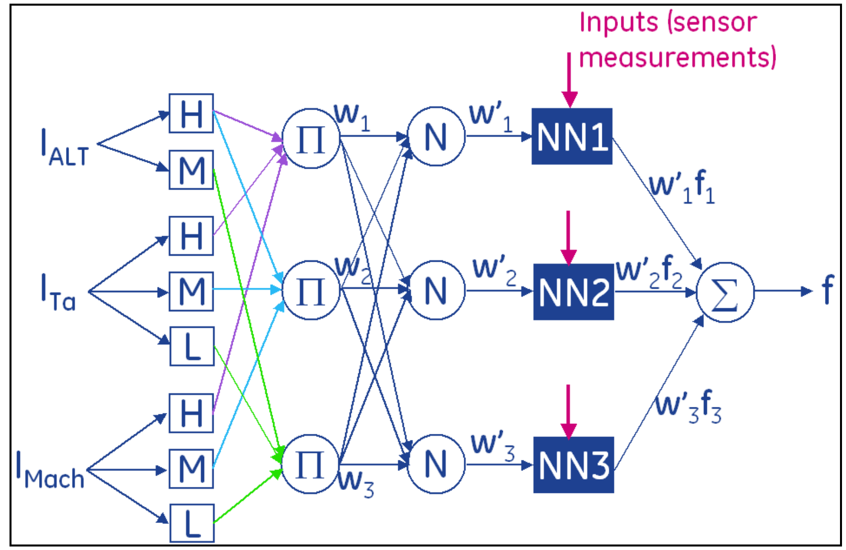


Transition Management Using Fuzzy Supervisory

AANN Interpolation by Fuzzy Supervisory



Network Implementation



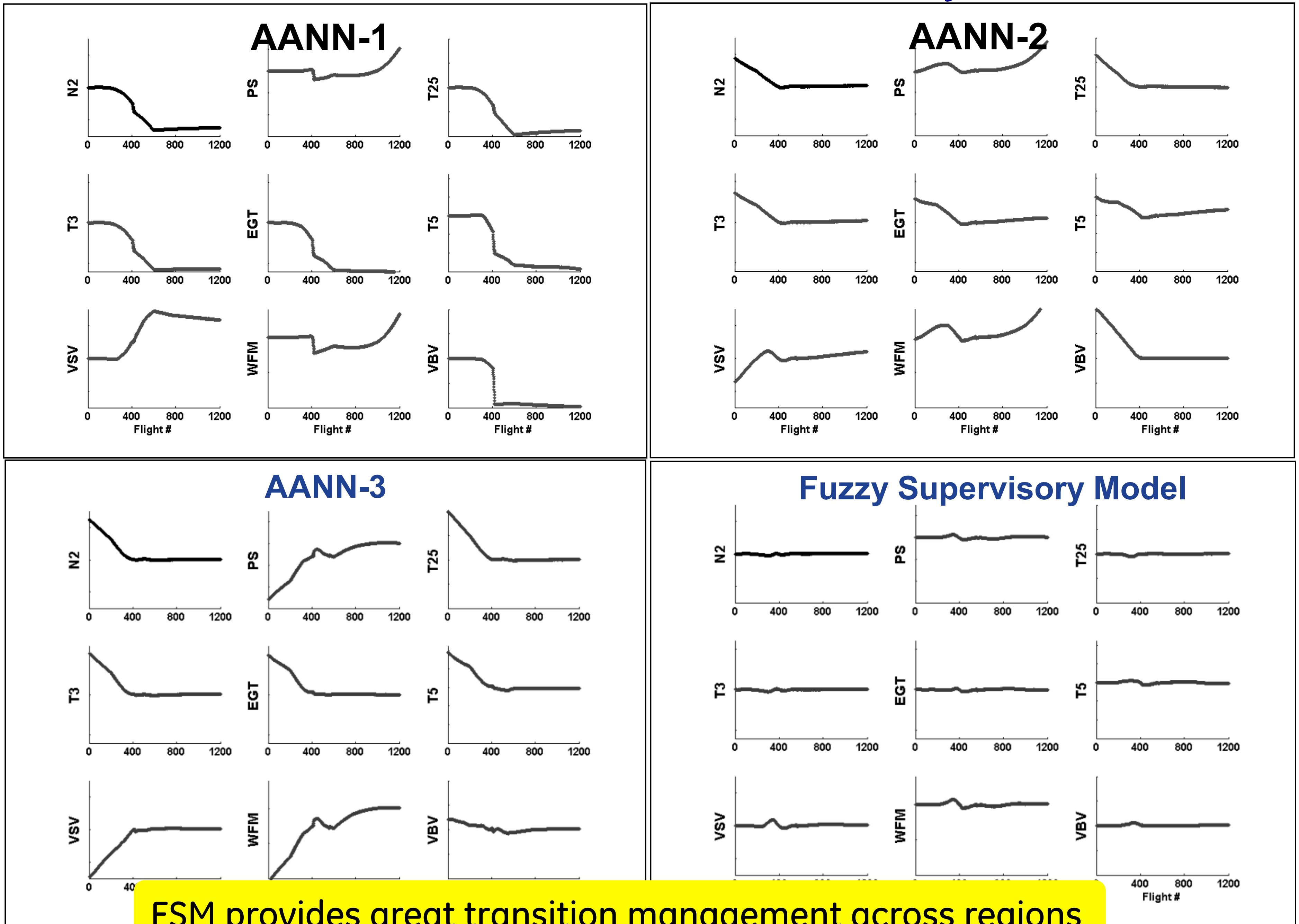
$w_i = \prod_{j=1..3} x_{ij}(I_i)$ $w'_i = w_i/(w_1+w_2+w_3)$

Figure Of Merit (FOM)

$$FOM = \sqrt{\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{R_{ij}}{\overline{X}_i}\right)^2}$$

n is the number of the variables (sensors)
m is the number of data points (measurement)
R_{ij} is the residual between true measurement and AANN estimation,
$\overline{X}_i$ is the mean of the true measurement

Residuals for each AANN and for hierarchical system (with FSM)



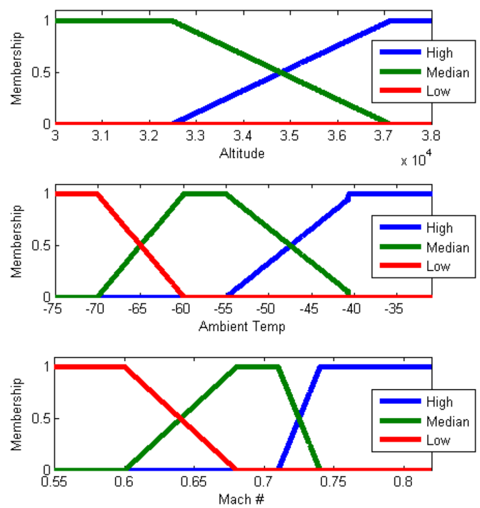
FSM provides great transition management across regions

Design Tuning

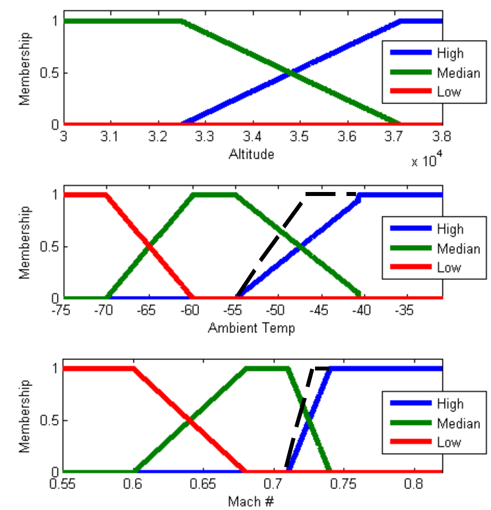
- Tuning the Fuzzy Supervisory Model
  - Manual tuning of FSM State Partitions
  - Automated tuning of FSM State Partitions

Manual FLS Tuning: Membership function parameters

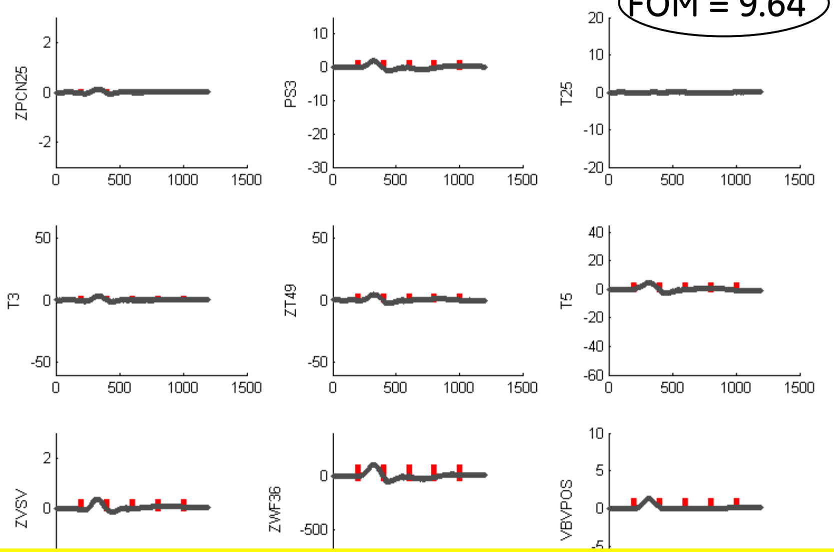
2.1 BC Analytics:
GE-90 Anomaly Detection



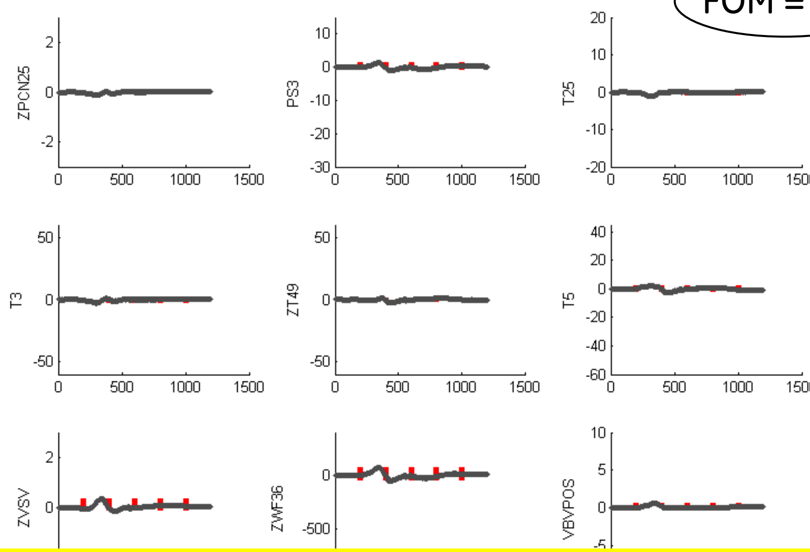
Manual
Tuning
(extending AANN1 scope)



FOM = 9.64



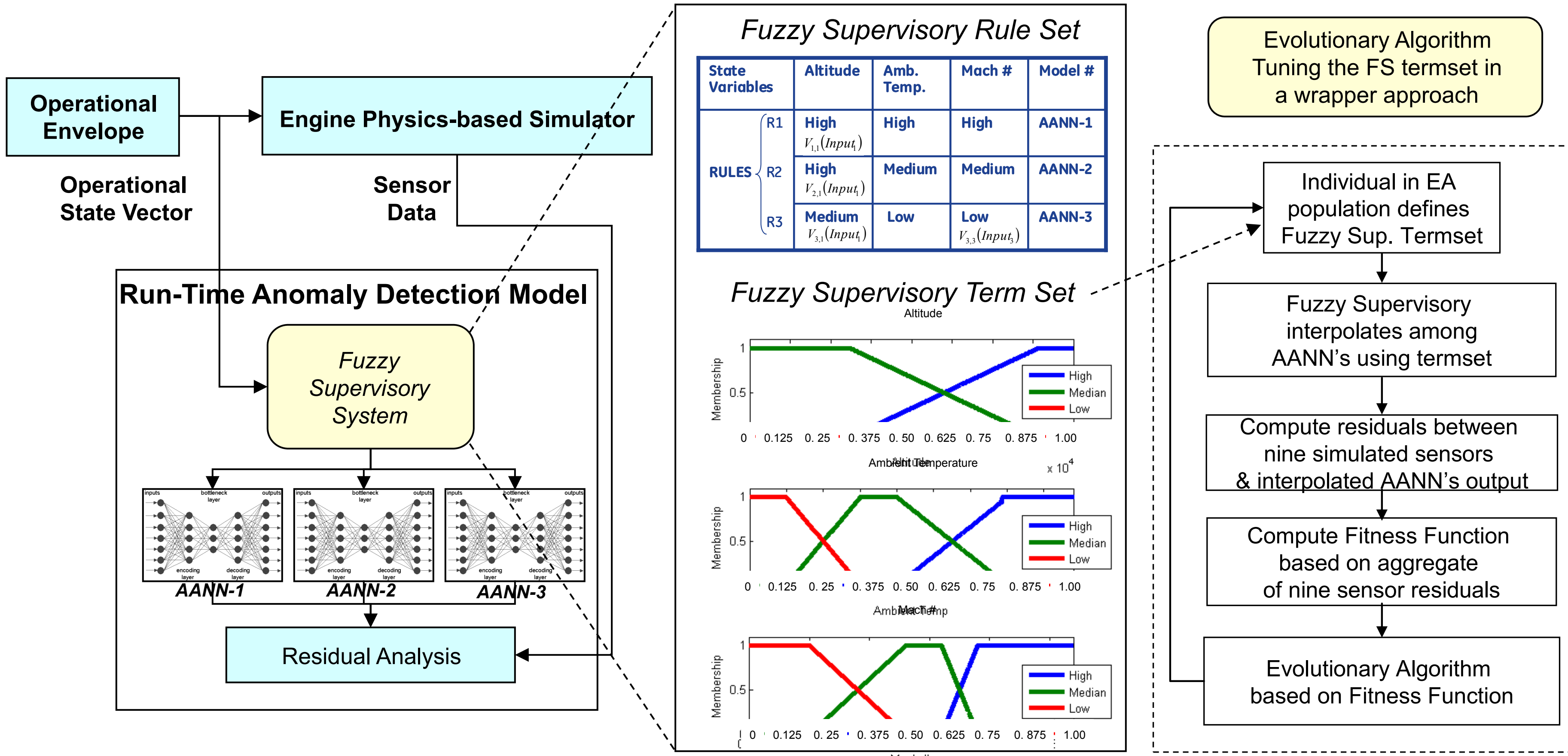
FOM = 7.25



Manual tuning, extending AANN1's scope, lead to a 25% FOM improvement
We could use FOM for gradient or evolutionary parametric tuning

Automated FLS Tuning with an EA (Wrapper Approach)

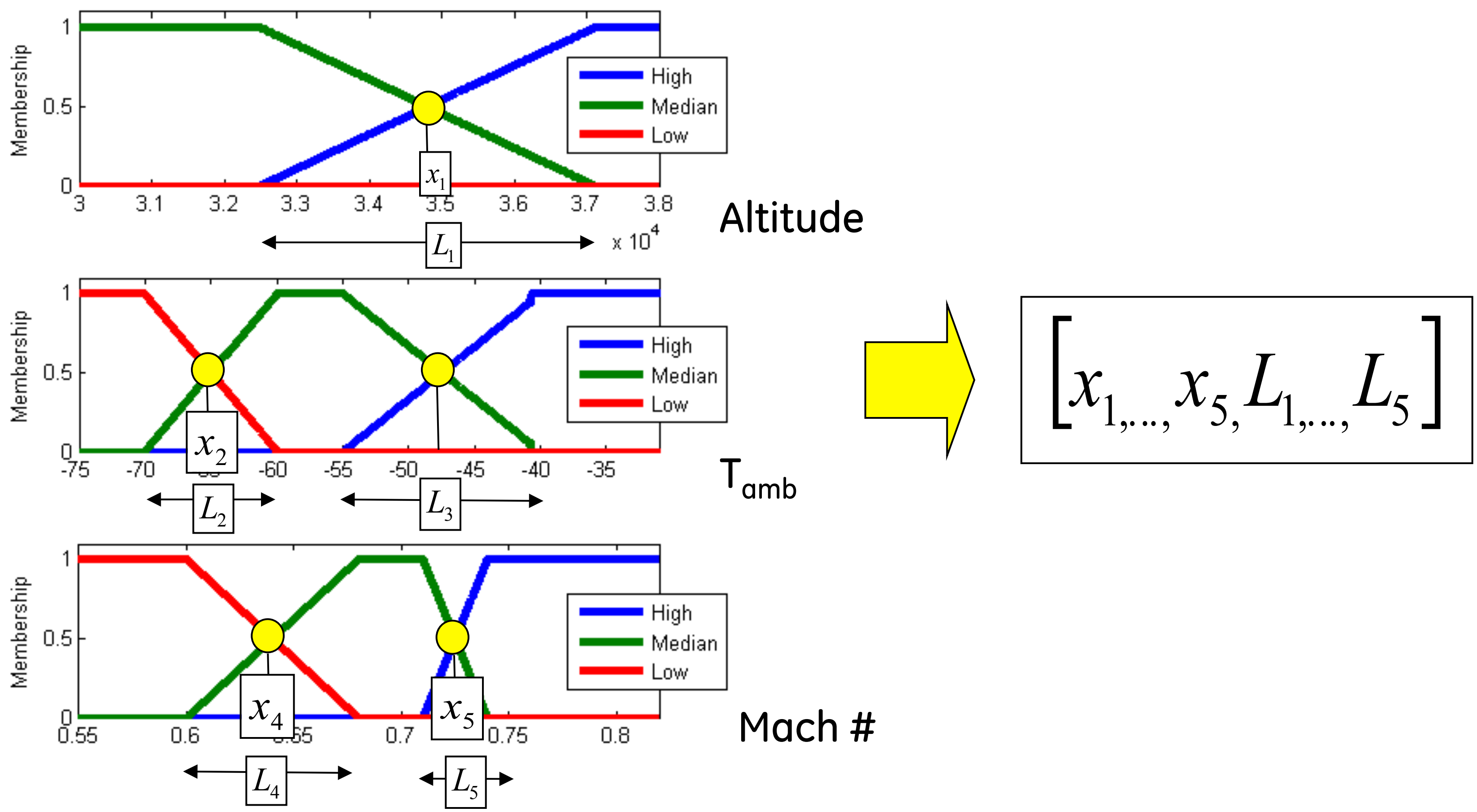
2.1 BC Analytics:
GE-90 Anomaly Detection



Problem Instance	Problem Type	Model Design (Offline MH's)	Model Controller (*Online MH's*)	Object-level models
Anomaly Detection	1-class Classification	**EA tuning of fuzzy supervisory termset**	**Fuzzy Supervisory**	***Multiple Models:*** **Ensemble of AANN's**

Reference: "A Systematic PHM Approach for Anomaly Resolution: A Hybrid Neural Fuzzy System for Model Construction", Proc. PHM 2009, San Diego, CA, Sept 27-Oct 1, 2009. -

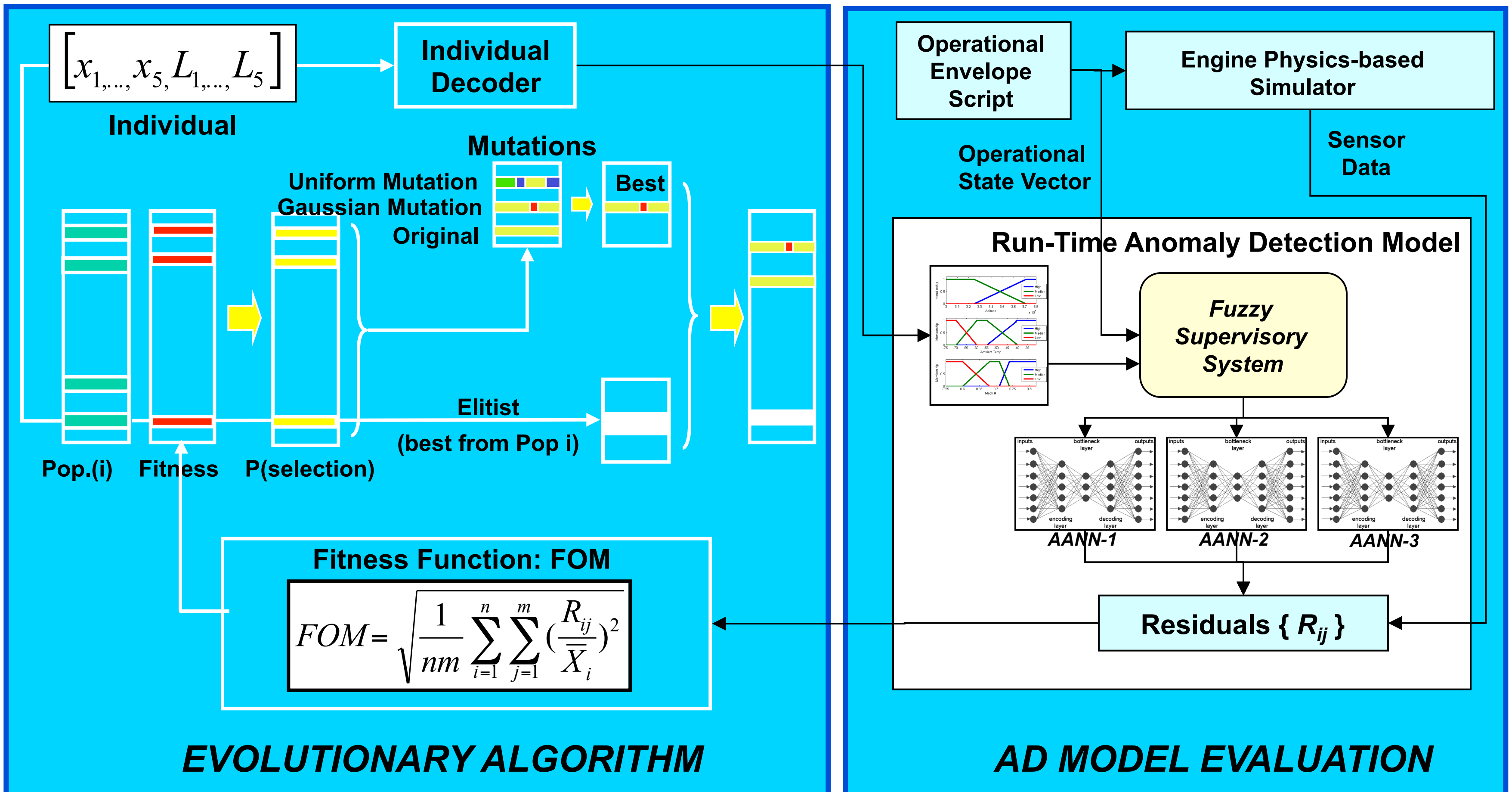
Automated FLS Tuning: Encoding Trapezoidal Membership functions



Encoding the abscissa of the slope intersections (x_i) and the lengths of the bases of each triangle (L_i) as an individual in the Evolutionary Algorithm population

imagination at work

Evolutionary Search for Tuning a Fuzzy Supervisory System using a Wrapper Approach

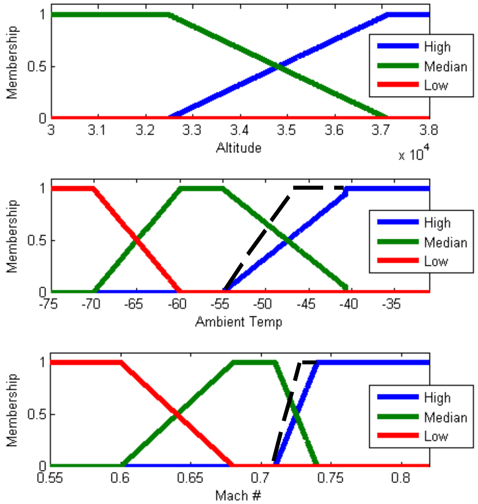


Pop Size = 500 individuals
GenMax = 1,000 generations

imagination at work

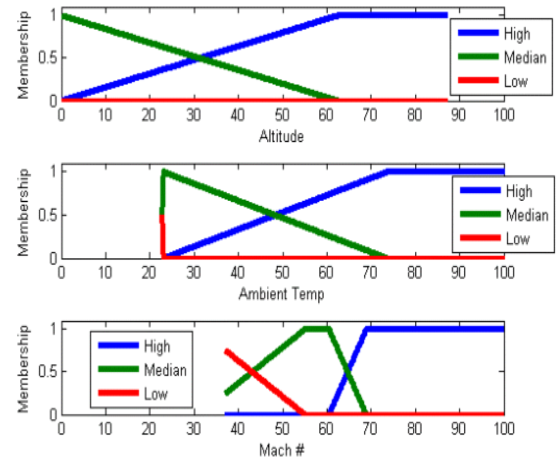
Anomaly Detection - Results

Automated FLS Tuning: Membership function parameters



Meta-Heuristic
Tuning

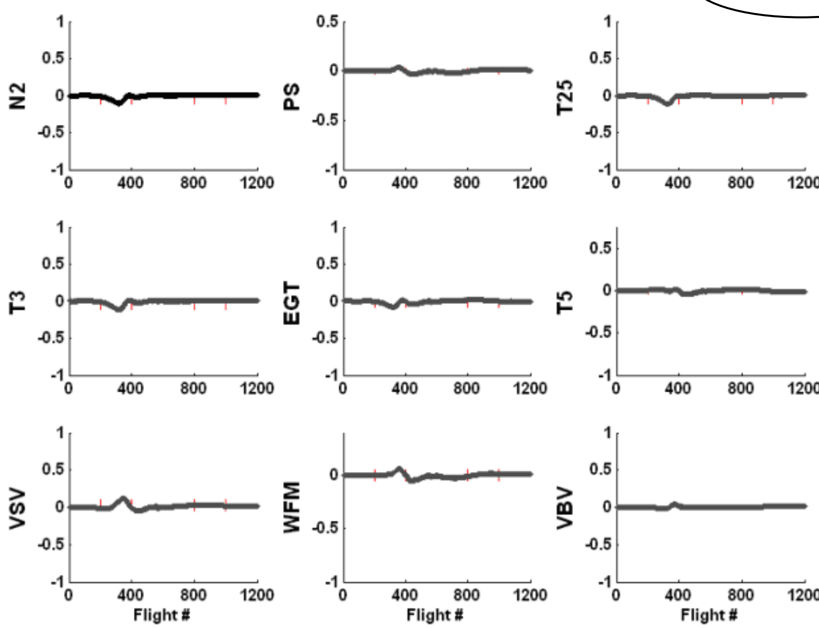
Use **Global Tuning**
(based on FOM fitness function
and Genetic Algorithm)
to further improve results



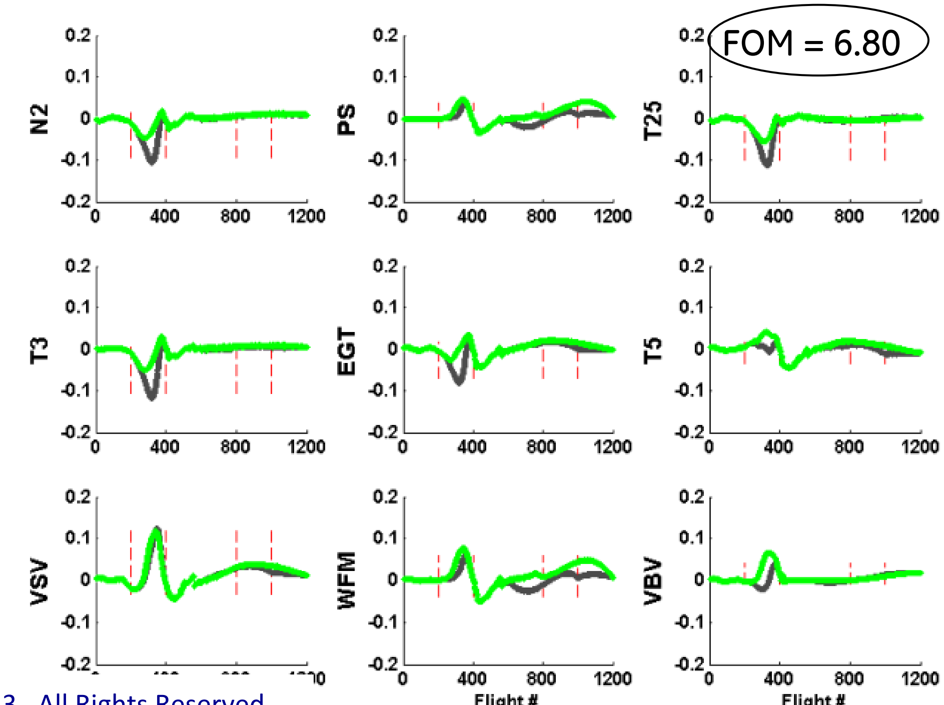
Note: Magnified scale to
enhance comparison

Before GA   After GA

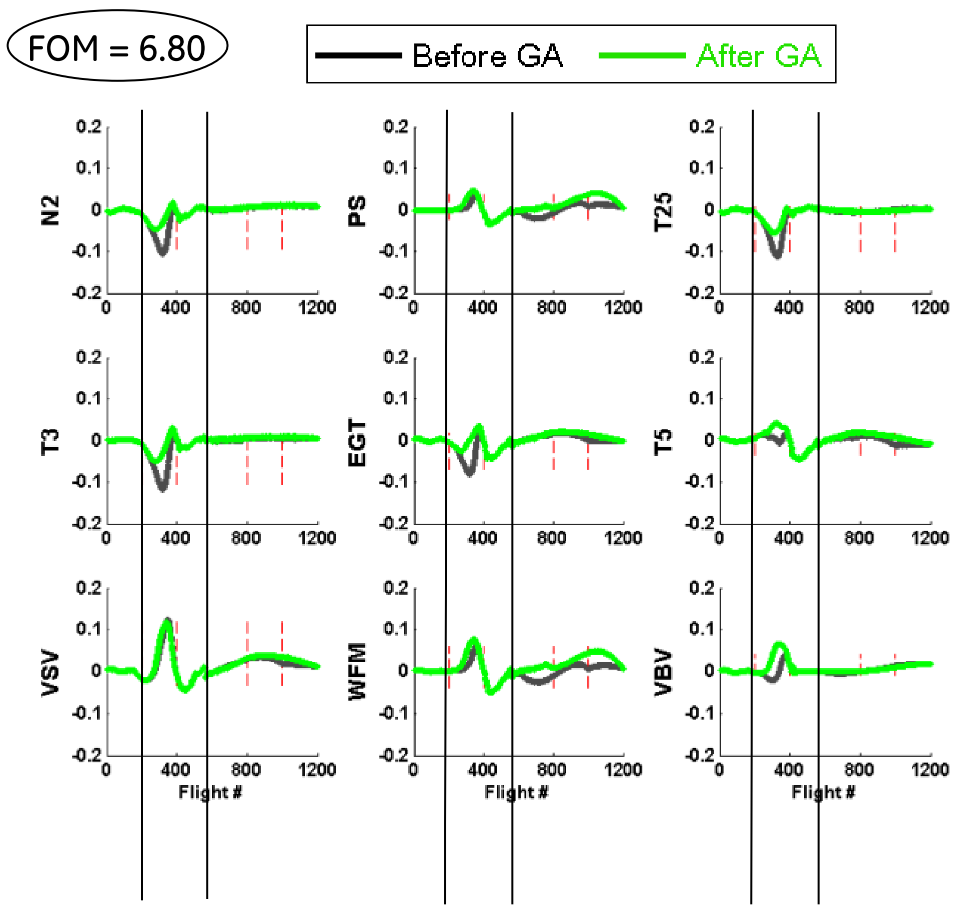
FOM = 7.25

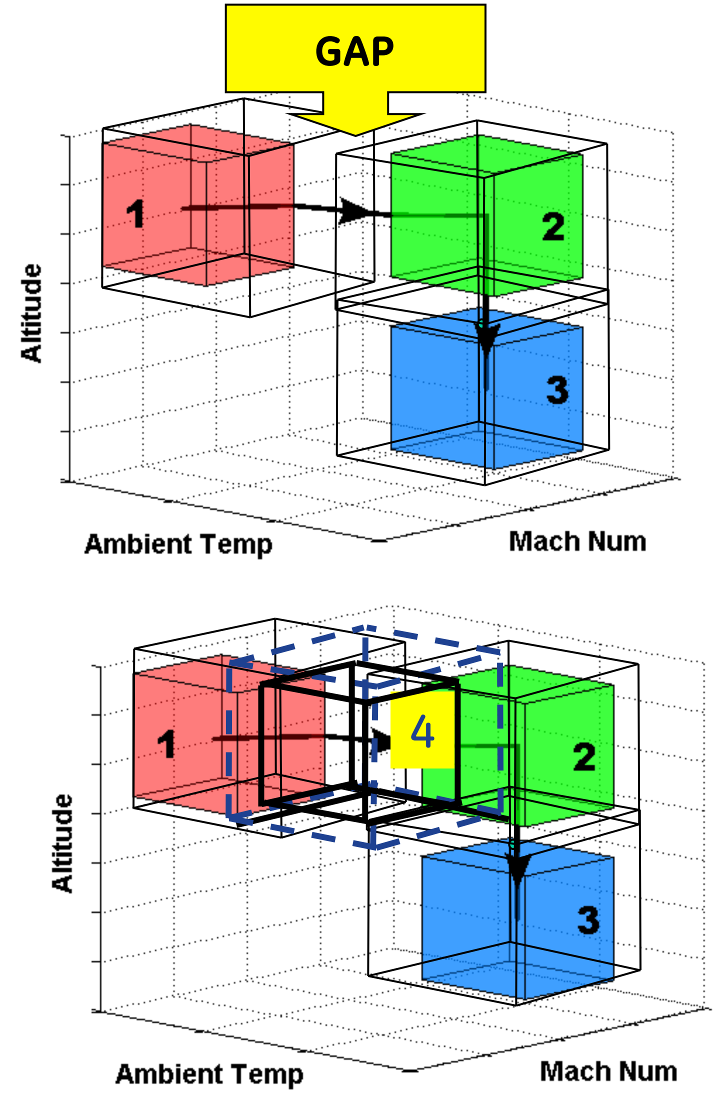


FOM = 6.80



Improved AD Design : Add AANN4 +retune FLS





Most residual errors occur in the [200, 600] interval, indicating a performance limit that cannot be addressed only by tuning the FLS. Rather it suggests the need for an additional AANN-4 to provide better coverage in that region

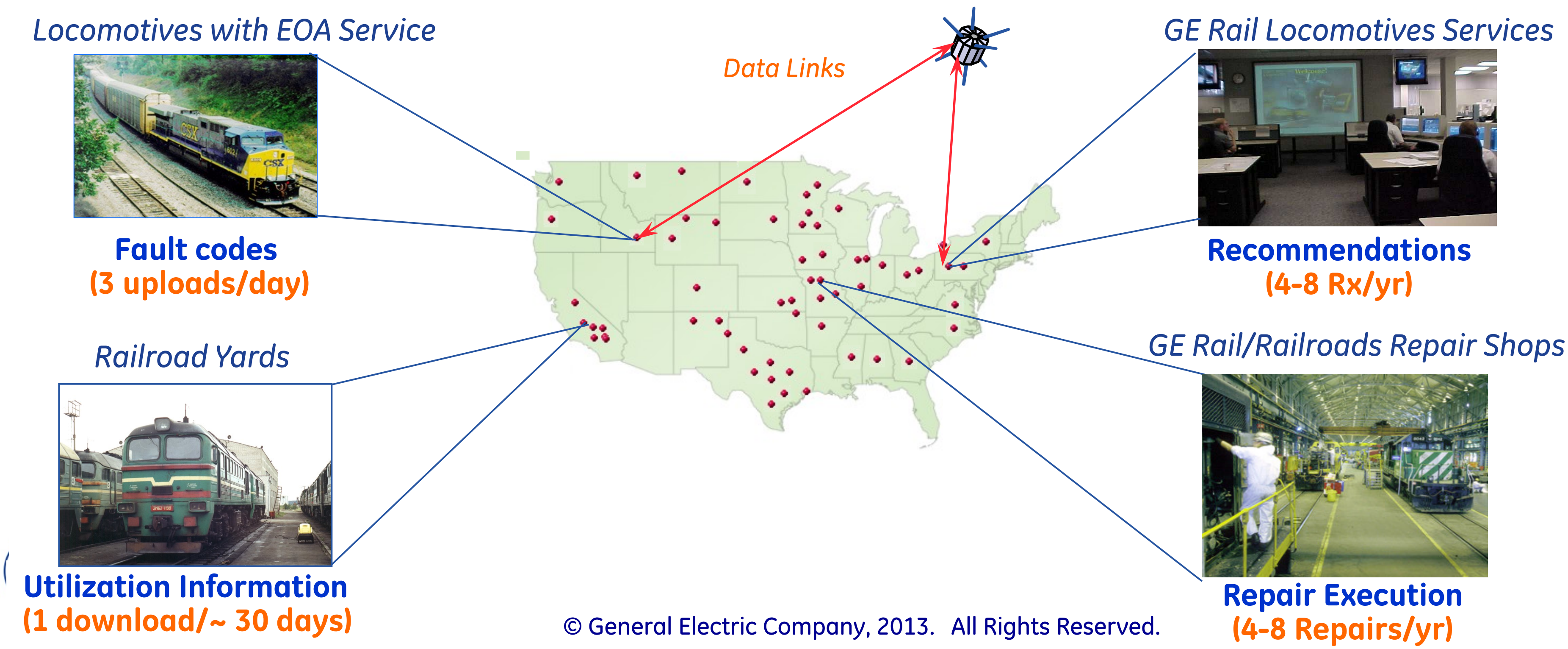
Cloud

2.1 BC Analytics:
Fleet Health Ranking

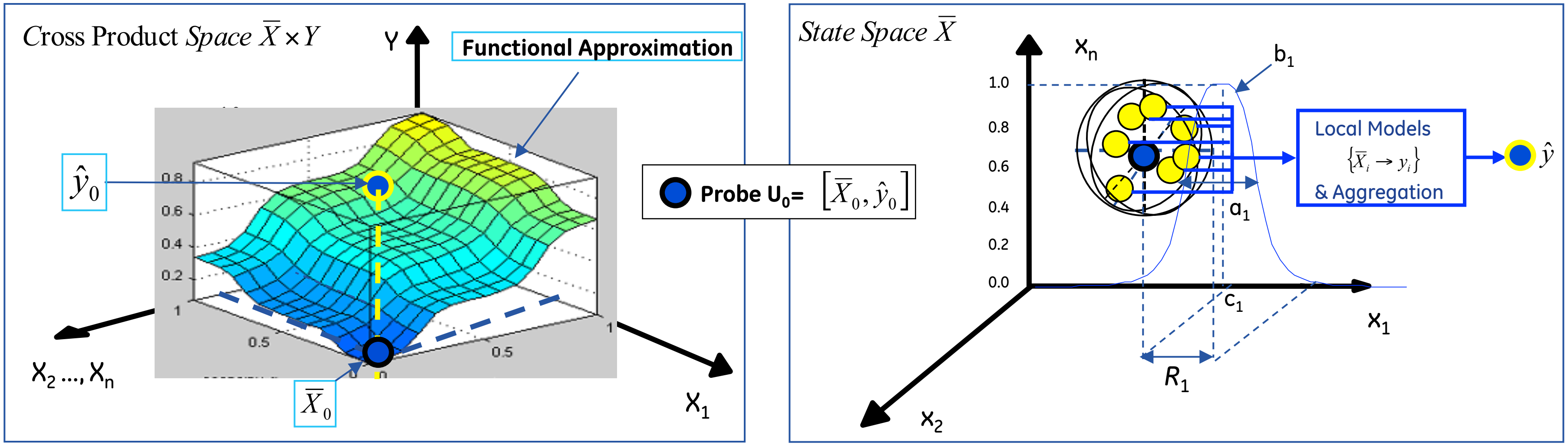
Example 2 : Rank a Fleet of Locomotives based on the units' Remaining Useful Life (RUL)

Data Categories:

- Configuration Information	*(Source: GE Rail)*
- Maintenance & Repair Information	
Fault Codes	*(Source: Locomotive's EOA)*
Recommendations	*(Source: GE Rail)*
Repairs	*(Source: GE Rail / Railroads)*
- Utilization Information	*(Source: Railroads)*



Functional Approximation vs. Instance-Based



Neural Networks

Radial Basis Functions — *Feed-forward Networks*

Fuzzy Systems

Takagi Sugeno Systems — *Mamdani-type Systems*

Neural Fuzzy Systems (ANFIS)

AI

Reasoning — *ML*

(Continuous) CBR — *Lazy Learning*

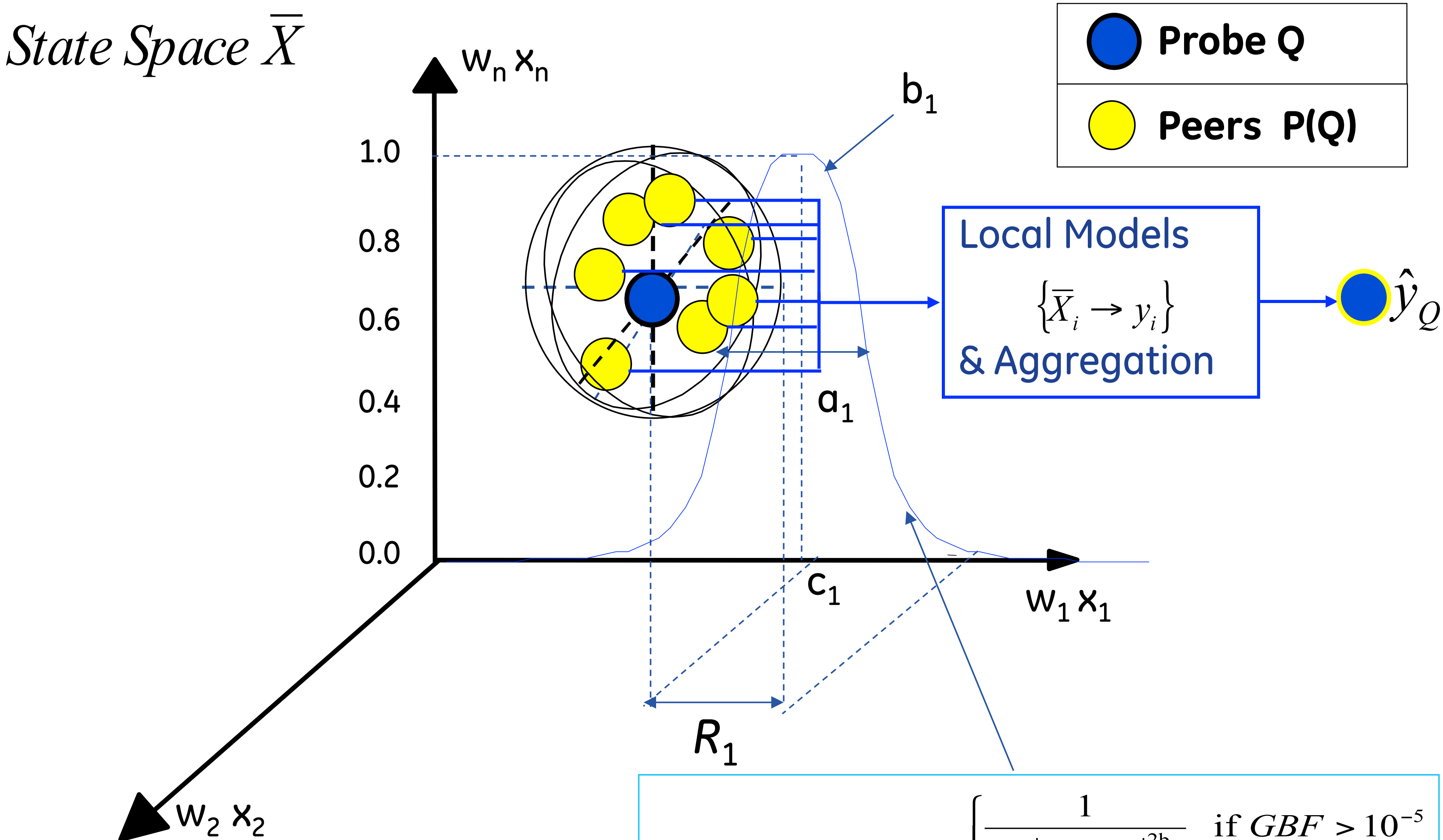
Statistics

Kernel-Based Models

Fuzzy Instance Based Models (F-IBM)

imagination at work

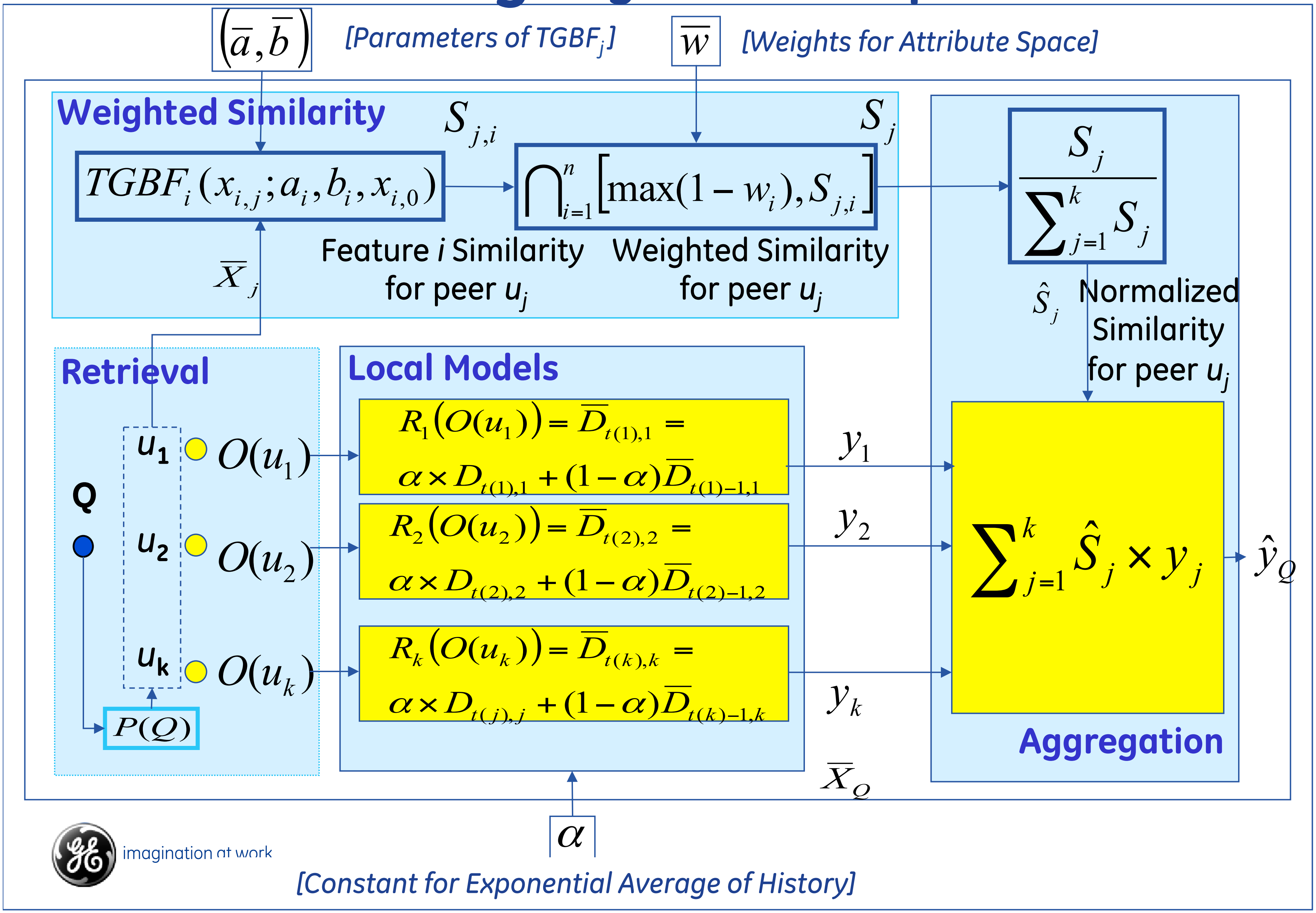
F-IBM Reasoning: Geometrical Interpretation



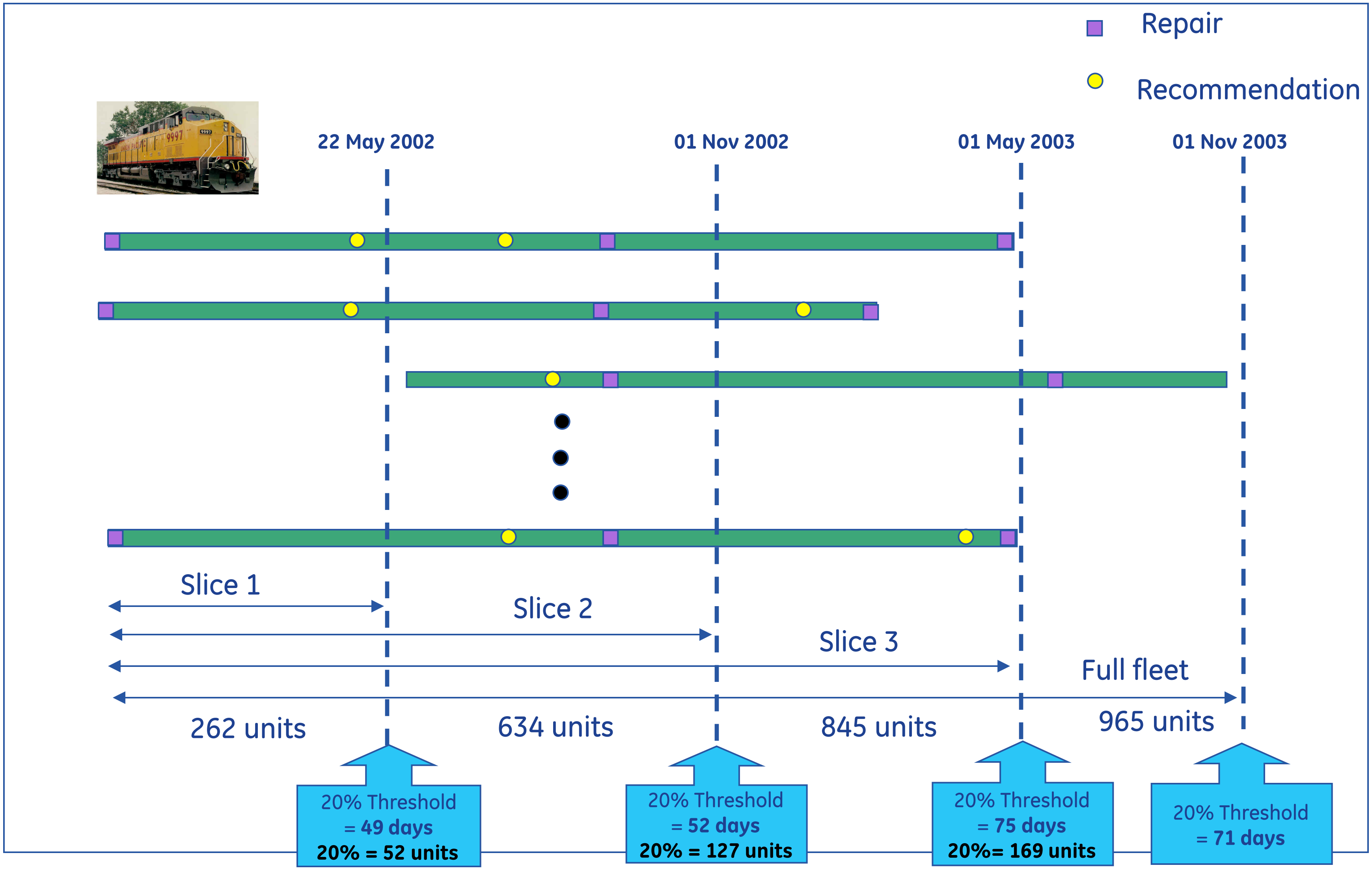
$$TGBF_i(x_i; a_i, b_i, c_i) = \begin{cases} \dfrac{1}{1 + \left| \dfrac{x_i - c_i}{a_i} \right|^{2b_i}} & \text{if } GBF > 10^{-5} \\ 0 & \text{otherwise} \end{cases}$$

imagination at work

F-IBM Reasoning: Algebraic Interpretation



Data Slices



imagination at work

Evolutionary Search

EA to Search in Design Space

The EA is composed of a population of individuals ("chromosomes"), each of which contains a vector of elements representing distinct tuneable parameters within the FIM configuration.

The EA used two types of mutation operators (Gaussian and uniform), and no crossover. Its population (with 100 individuals) was evolved over 200 generations

Each chromosome defines an instance of the attribute space used by the associated model by specifying a vector of weights $[w_1, w_2, \ldots, w_n]$.

If $w_i \in \{0,1\}$ we perform *attribute selection*, i.e., we select a crisp subset from the universe of potential attributes.

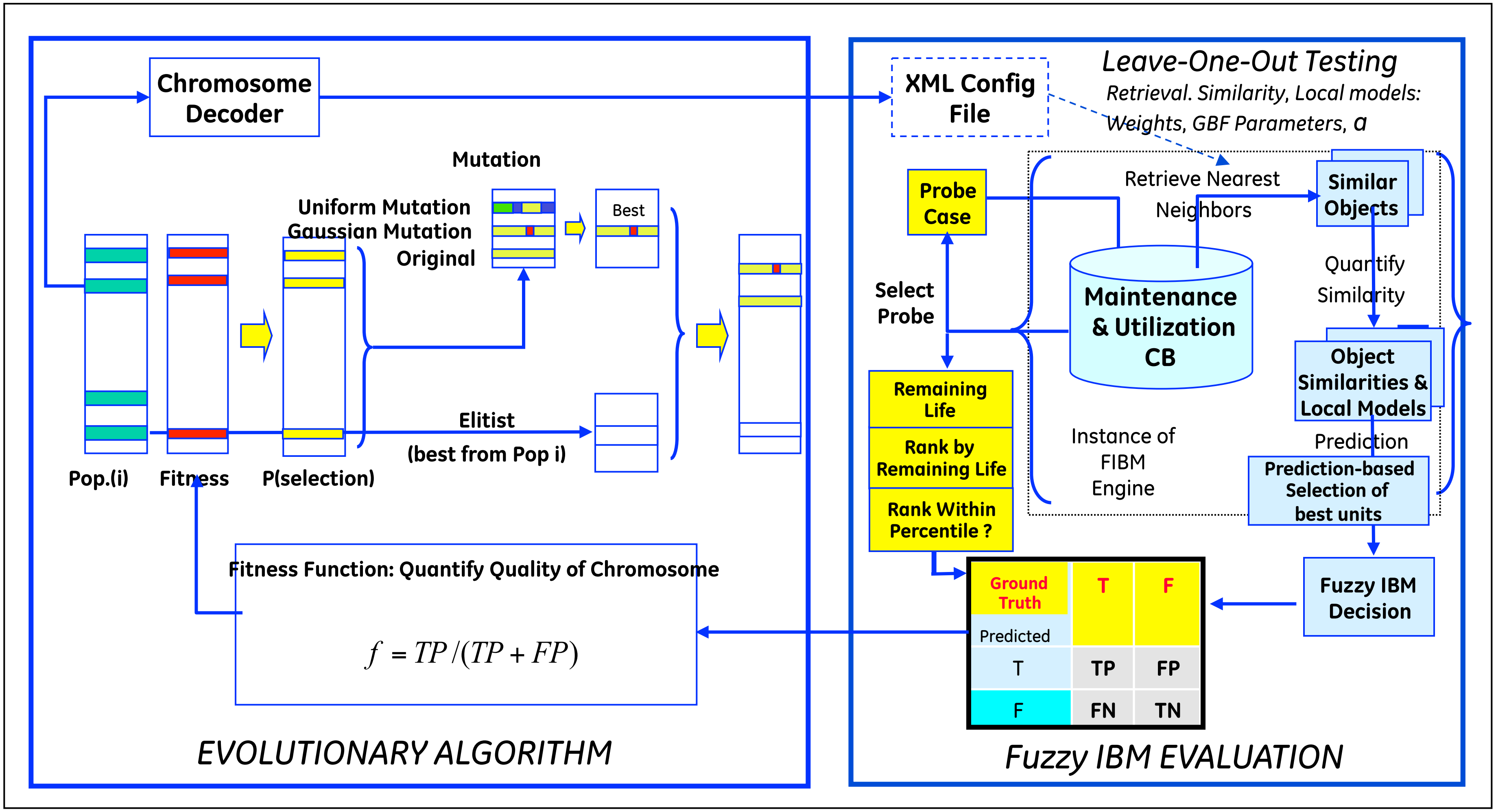
If $w_i \in [0,1]$ we perform *attribute weighting,* i.e., we define a fuzzy subset from the universe of potential attributes

Chromosome representation:

$$\boxed{[w_1\ w_2\ \ldots\ w_n][(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)][\alpha]}$$

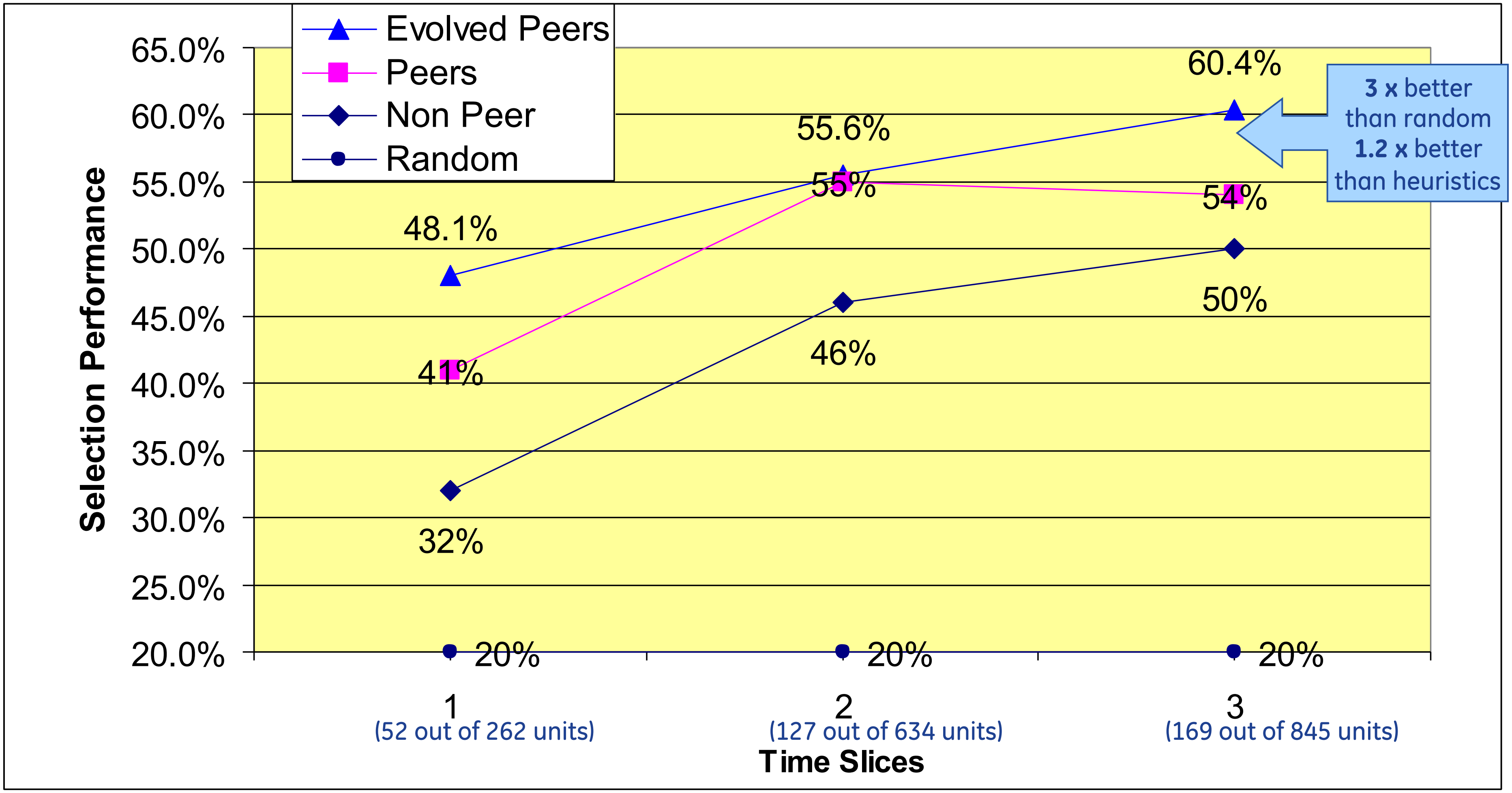
imagination at work

Evolutionary Search for Designing A Fuzzy Instance-Based Model using a Wrapper Approach



imagination at work

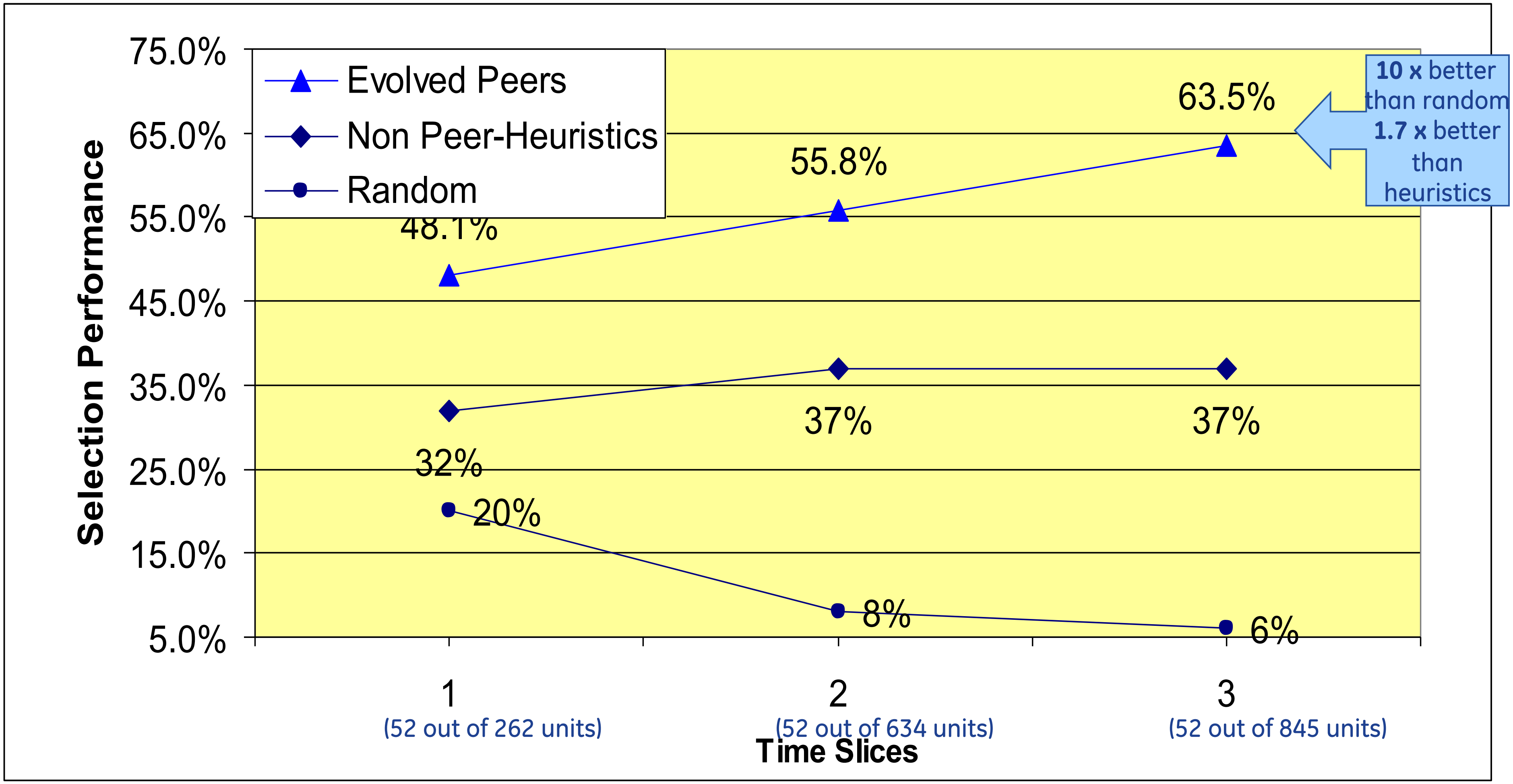
Peers Evolution over Time: Estimating Best Current performers



(Targeting top 20% of units for each Time Slice)

imagination at work

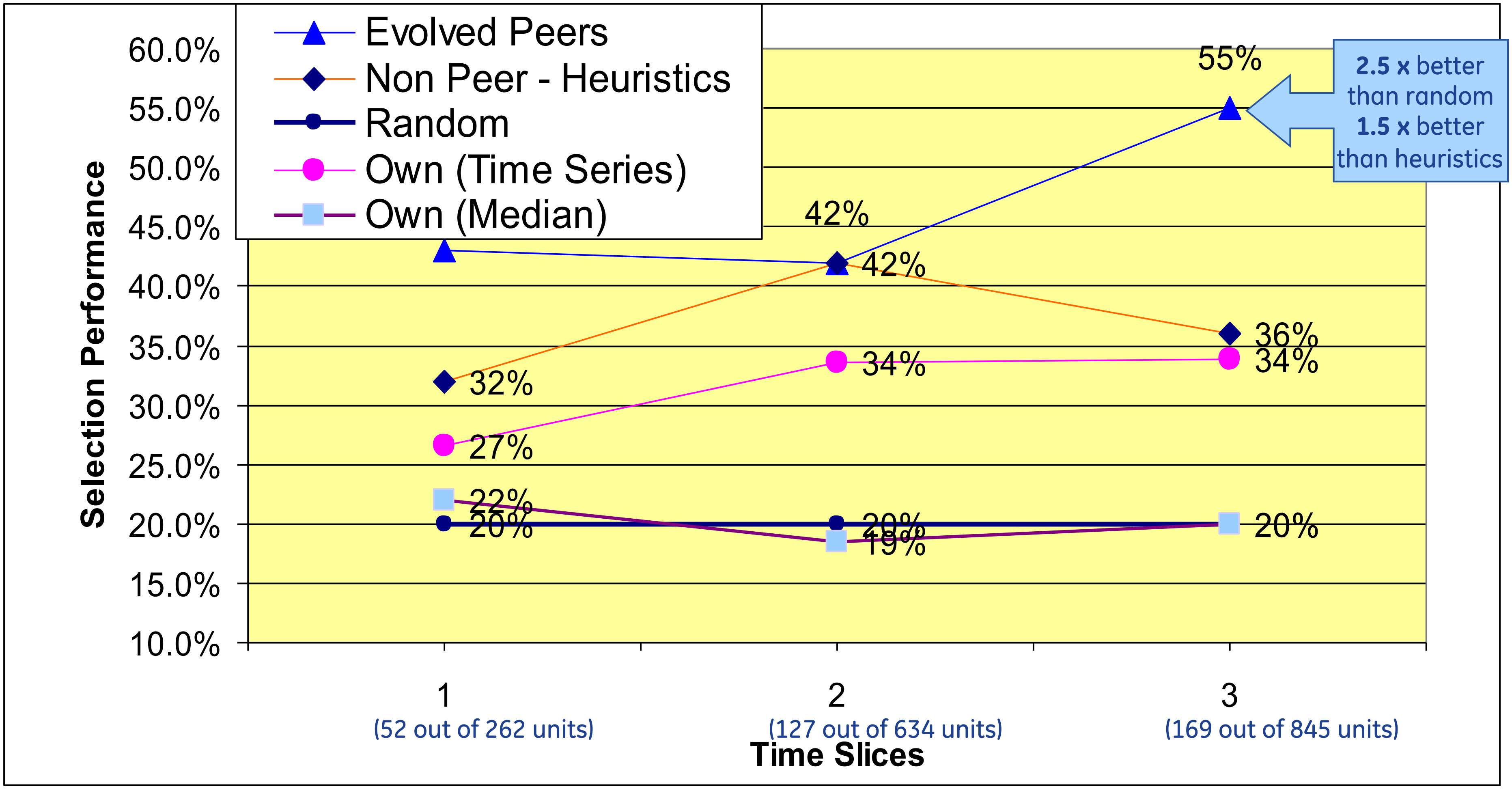
Peers Evolution over Time:
Estimating best Current performers



(Targeting Top 52 units for each Time Slice)

imagination at work

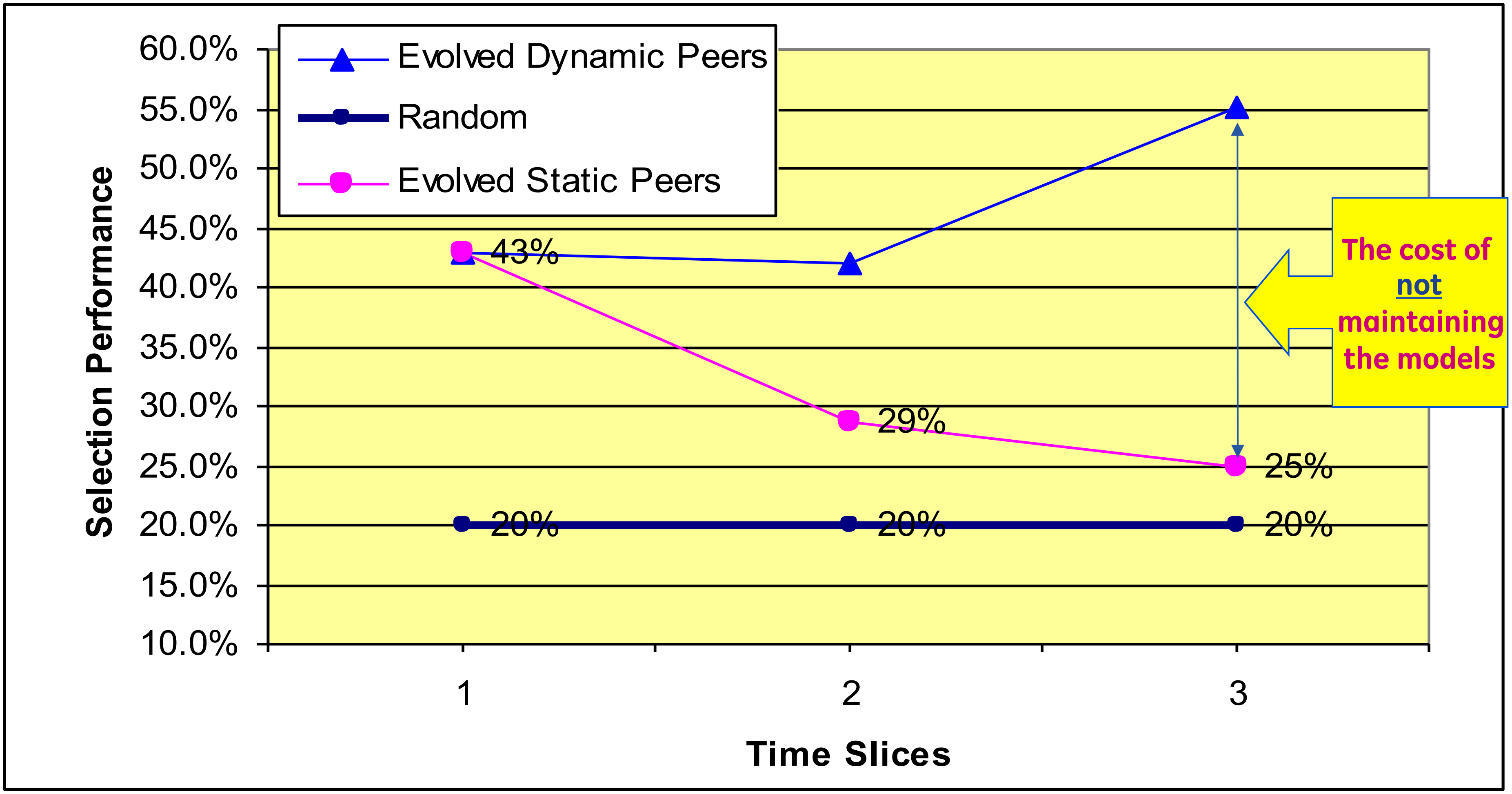
Peers Evolution over Time: Estimating best future performers



(Targeting top 20% of units for each Time Slice)

imagination at work

Experiment 3 bis – Comparing Evolved, Dynamic (updated) Peers with Evolved Static (Non-updated) Peers: Estimating Best Future Performers



The Benefit of Automated Peer Redesign/Update

2.2 Cloud Computing: The Enabler

GE imagination at work

Cloud Computing

" … sometimes an emerging technology is neither awesome nor suspicious, but is **so obviously the right answer that it is only a matter of time before it inevitably permeates our societ**y.

Cloud computing is in this latter class—its characteristics of economic value, technological elegance, and business empowerment all combine to **make it the clear architecture of choice for the bulk of information technology needs of the 21st century**…."

Source: Cloud Computing: The Inevitable Answer, A. Pasik, June 12, 2012, The Institute, IEEE
http://theinstitute.ieee.org/ieee-roundup/opinions/ieee-roundup/cloud-computing-the-inevitable-answer

imagination at work

The inevitability of the Cloud

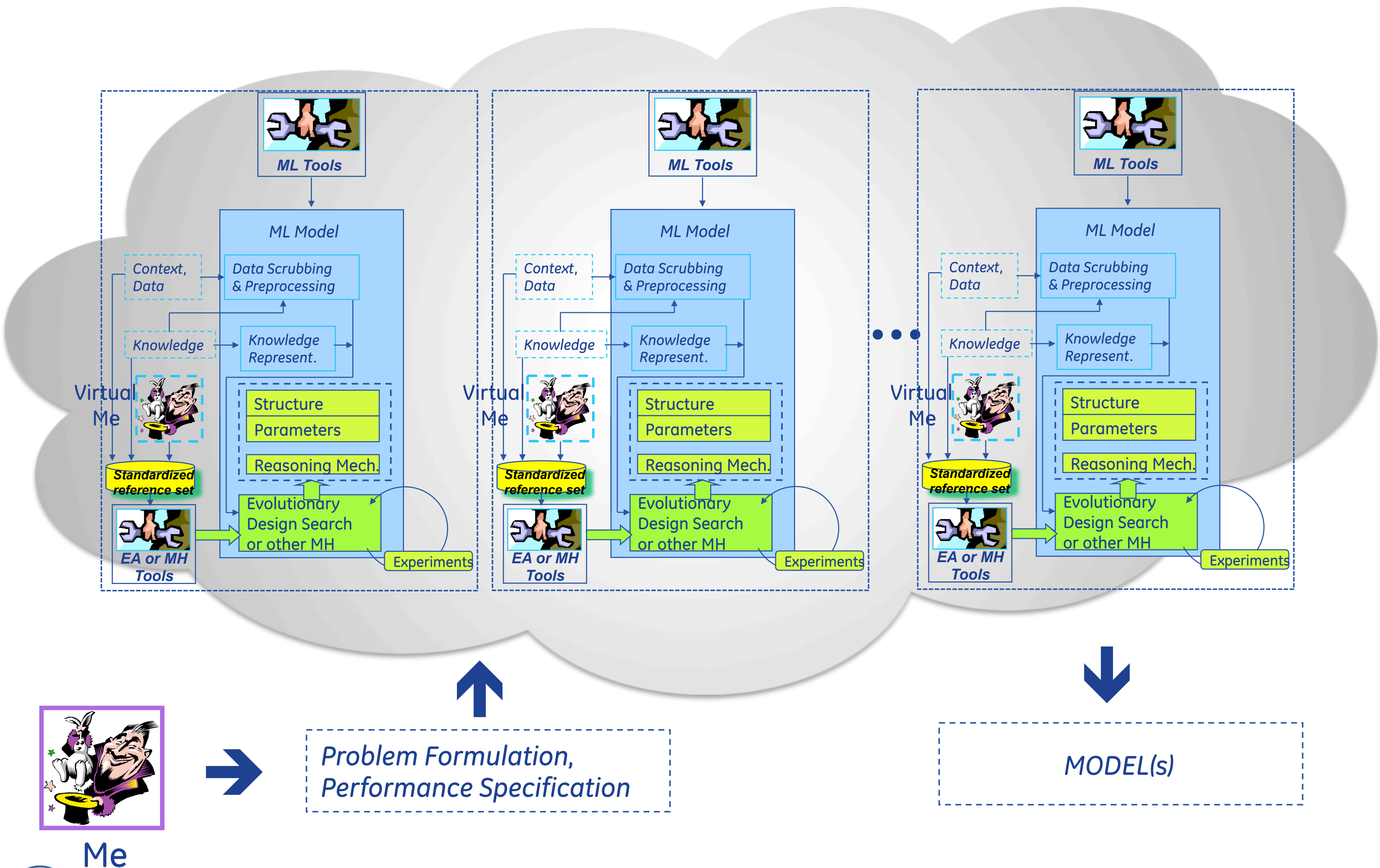
2.3 AC Analytics (After the Cloud) - NOW

Analytics Cloud

- Tuning Legacy Models
- Automating new model building
- Fusion as a meta-model

GE imagination at work

My New Job → Still employed ?



imagination at work

GE Analytics Cloud

Ecosystem for collaborative analytics development

Data Scientist

Develops Algorithms and models to crunch data

Subject Matter Expert

Engineer with technical expertise & domain knowledge

Application Developer

Developer integrates a method into a production environment or builds a specific using a method

Composer

Expert at combining algorithms and analytics, machine requirements and data feeds into "methods"

- Make analytics scalable & repeatable
- Accessible & amenable: empower different users
- Speed up learning process, discover what we don't know
- Deploy analytics into applications & processes

iScale – an internal project similar to MLBase

imagination at work

2.4 AC Analytics (After the Cloud) - Future

- Crowdsourcing, Commoditized Analytics
- Model Agnostic Fusion

CI Genetable

THE COLLECTIVE INTELLIGENCE GENETABLE

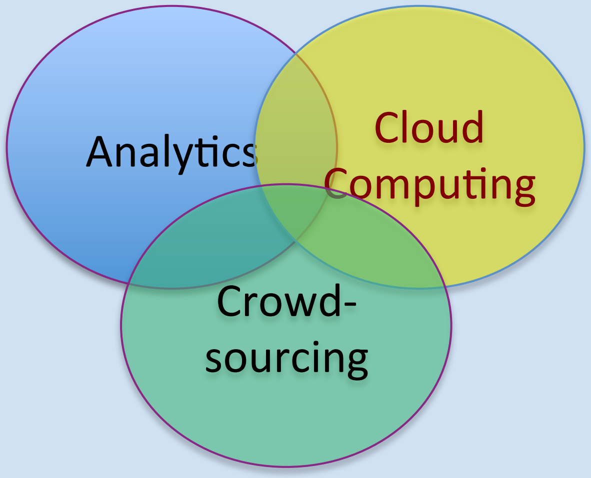
Each of the principal genes — or building blocks — of collective intelligence can be combined with other genes to create the appropriate system for accomplishing a task. This table lists the principal genes and describes the conditions under which they're most useful.

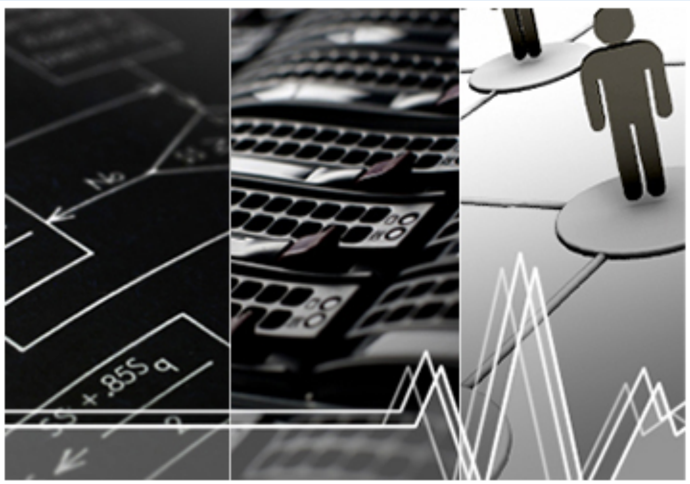
QUESTION	GENE	USEFUL WHEN ...
Who	**Crowd**	•Resources useful in performing activities are distributed widely or in places not known in advance. •Activities can be divided into pieces satisfactorily (necessary information can be shared; gaming and sabotage can be managed). •Crowds can do things cheaper, faster, with higher quality or with higher motivation.
	Hierarchy (or, Management)	•Conditions for Crowd aren't met.
Why	**Money**	•Many factors apply, too complex to list here. But there are two rules of thumb: – Appealing to Love and Glory, rather than Money, can often (but not always) reduce costs. – Providing Money and Glory can often (but not always) influence a group's direction and speed.
	Love	
	Glory	
How — Create	**Collection**	•Conditions for Crowd, *plus* ... •Activity can be divided into small pieces that can be done (mostly) independently of each other.
	Contest	•Conditions for Collection, *plus* ... •Only one (or a few) good solutions are needed.
	Collaboration	•Activity cannot be divided into small independent pieces (otherwise Collection would be better). •There are satisfactory ways of managing the dependencies among the pieces.

Source*:* T. Malone, R. Laubacher, C. Dellarocas, The Collective Intelligence Genome, *MIT Sloan Management Review*, 51(3) :21-31 (2010) http://www.lhstech.com/chair/Articles/malone.pdf

The Vision – Already shaped for us...

UC Berkeley AMP Lab





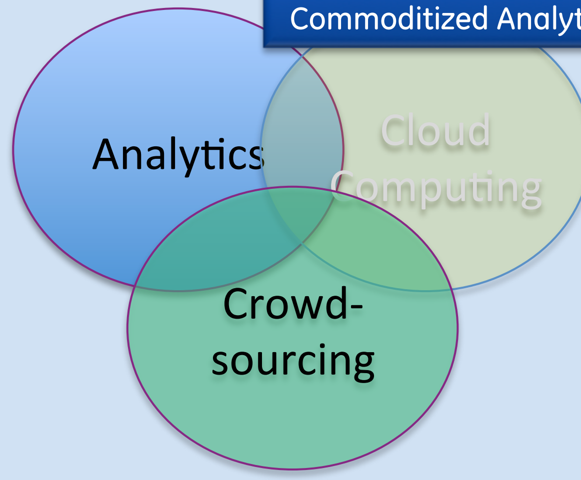
AMP: ALGORITHMS MACHINES PEOPLE

TURNING UP THE VOLUME ON BIG DATA

Working at the intersection of three massive trends: powerful machine learning, cloud computing, and crowdsourcing, the AMPLab is integrating Algorithms, Machines, and People to make sense of Big Data. We are creating a new generation of analytics tools to answer deep questions over dirty and heterogeneous data by extending and fusing machine learning, warehouse-scale computing and human computation. We validate these ideas on real-world problems including participatory sensing, urban planning, and personalized medicine with our application and industrial partners.

Source: http://amplab.cs.berkeley.edu/

amplab



Kaggle.com

The Fight Club for Geeks

Kaggle's Contests: Crunching Numbers for Fame and Glory

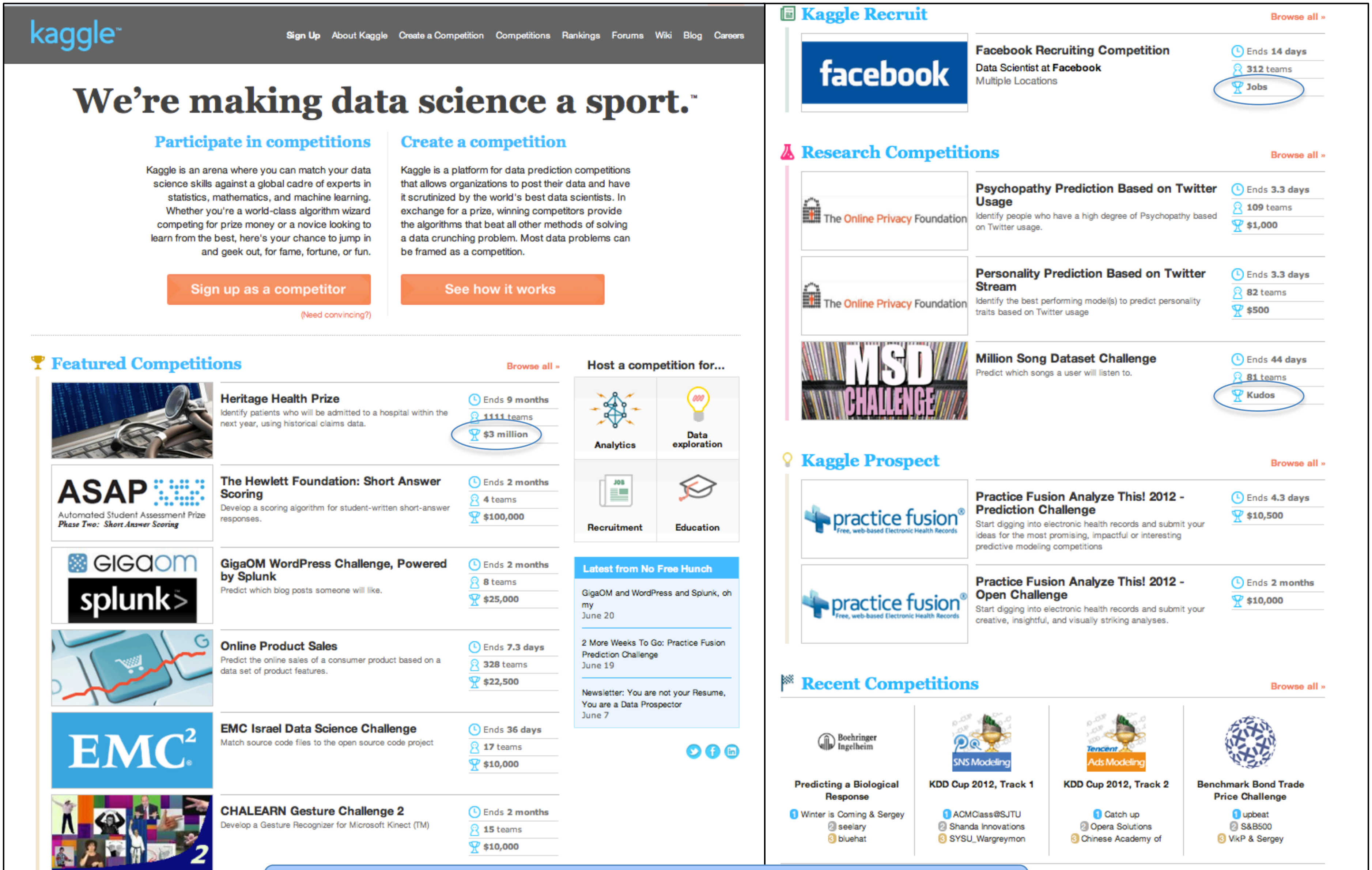
The startup's competitions lure PhDs and whiz kids to solve companies' data problems

Photograph by Matthew Scott for Bloomberg Businessweek

Source: A. Vance, Kaggle's Contests: Crunching Numbers for Fame and Glory, Vance, A., *Businessweek*, Jan. 04, (2012)

http://www.businessweek.com/magazine/kaggles-contests-crunching-numbers-for-fame-and-glory-01042012.html

Kaggle.com



Motivation: Money, Jobs, Kudos (fame and glory)

Source: http://www.kaggle.com/

My Future Job

Building Modeling for Performance Management using Lazy Meta-Learning:

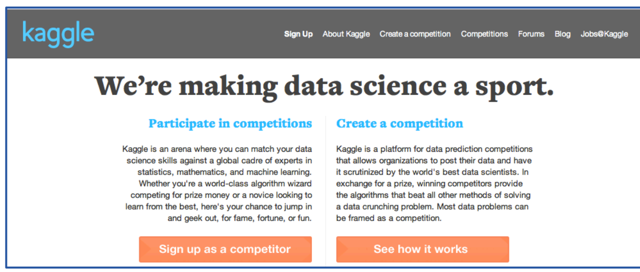
- Part-time Magician ?
(... or the Curse of the Cloud)

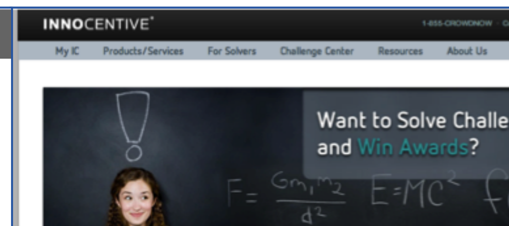
imagination at work

Potential commoditization of analytical models

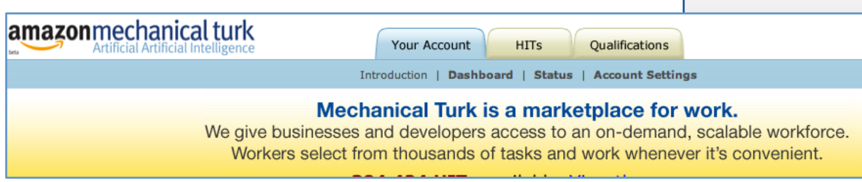
Source Type	Mechanism / Incentive	Examples	URL
Crowdsourcing Crowdsourced models	*Model Competition; R&D Competition/ Prizes*	***Kaggle*** ***TunedIT*** **CrowdANALYTICX** ***INNOcentive***	www.Kaggle.com www.TunedIT.org www.Crowdanalytix.com www.innocentive.com/
	HITS / Direct payments	***Amazon Mechanical Turk***	www.mturk.com/mturk
	Game / credentials (fame & glory)	***Foldit***	http://fold.it/portal/

Hosts for ML or R&D competitions

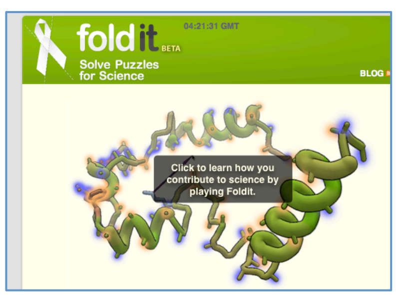




Crowdsourcing Market
(HITS = Models)



Example of Specialized Molecular biology problem solving



Potential commoditization of analytical models

Source Type	Mechanism / Incentive	Examples	URL
Crowdsourcing Crowdsourced models	*Model Competition;* *R&D Competition/ Prizes*	***Kaggle*** ***TunedIT*** **CrowdANALYTICX** ***INNOcentive***	www.Kaggle.com www.TunedIT.org www.Crowdanalytix.com www.innocentive.com/
	HITS / Direct payments	***Amazon Mechanical Turk***	www.mturk.com/mturk
	Game / credentials (fame & glory)	***Foldit***	http://fold.it/portal/
Outsourcing to Academia or Vendors Outsourced models	*University Grants; Contracts / Money & Contacts*	***AMP:*** Algorithms, Machines, and People [UC Berkeley]; ***CMU*** ML Dept; ***MIT*** CSAIL; …	wwwamplab.cs.berkeley.edu

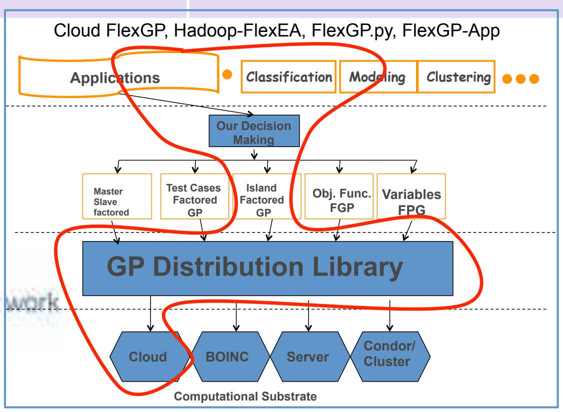
Major Universities with top ML capabilities

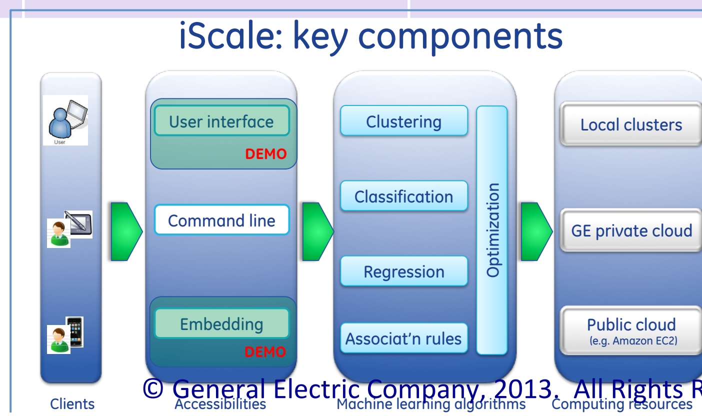
Potential commoditization of analytical models

2.4. AC Analytics (Future)
Commoditized Analytics

Source Type	Mechanism / Incentive	Examples	URL
Crowdsourcing Crowdsourced models	*Model Competition; R&D Competition/ Prizes*	***Kaggle*** ***TunedIT*** **CrowdANALYTICX** ***INNOcentive***	www.Kaggle.com www.TunedIT.org www.Crowdanalytix.com www.innocentive.com/
	HITS / Direct payments	***Amazon Mechanical Turk***	www.mturk.com/mturk
	Game / credentials (fame & glory)	***Foldit***	http://fold.it/portal/
Outsourcing to Academia or Vendors Outsourced models	*University Grants; Contracts / Money & Contacts*	***AMP:*** Algorithms, Machines, and People [UC Berkeley]	wwwamplab.cs.berkeley.edu
Cloud-based automation Meta-Heuristics generated models	*Cloud-based Meta-heuristics for model design / Bleeding Edge Research; Innovation; Papers*	***FLexEA*** [MIT CSAIL] ***iScale*** [GE GR]	http://groups.csail.mit.edu/EVO-DesignOpt/evo.php?n=Site.FlexGP

EvoDesignOpt
FlexGP
Flexibly Factored, Flexibly Scaled Genetic Programming



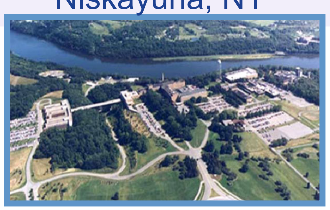


Potential commoditization of analytical models

2.4. AC Analytics (Future)
Commoditized Analytics

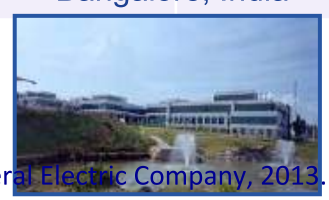
Source Type	Mechanism / Incentive	Examples	URL
Crowdsourcing Crowdsourced models	*Model Competition; R&D Competition/ Prizes*	***Kaggle*** ***TunedIT*** **CrowdANALYTICX** ***INNOcentive***	www.Kaggle.com www.TunedIT.org www.Crowdanalytix.com www.innocentive.com/
	HITS / Direct payments	***Amazon Mechanical Turk***	www.mturk.com/mturk
	Game / credentials (fame & glory)	***Foldit***	http://fold.it/portal/
Outsourcing to Academia or Vendors Outsourced models	*University Grants; Contracts / Money & Contacts*	***AMP:*** Algorithms, Machines, and People [UC Berkeley]	wwwamplab.cs.berkeley.edu
Cloud-based automation Meta-Heuristics generated models	*Cloud-based Meta-heuristics for model design / Bleeding Edge Research; Innovation; Papers*	***FLexEA*** [MIT CSAIL] ***iScale*** [GE GR]	http://groups.csail.mit.edu/EVO-DesignOpt/evo.php?n=Site.FlexGP
Industrial R&D Traditional GE GRC models	*R&D ML Projects (manual dev. on single node or cluster) / Research; Innovation; Patents;*	Tree Ensemble Classifiers [TEC] Prognostics,	www.ge.geglobalresearch.com

Niskayuna, NY



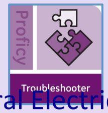
San Ramon, CA

Bangalore, India



imagination at work

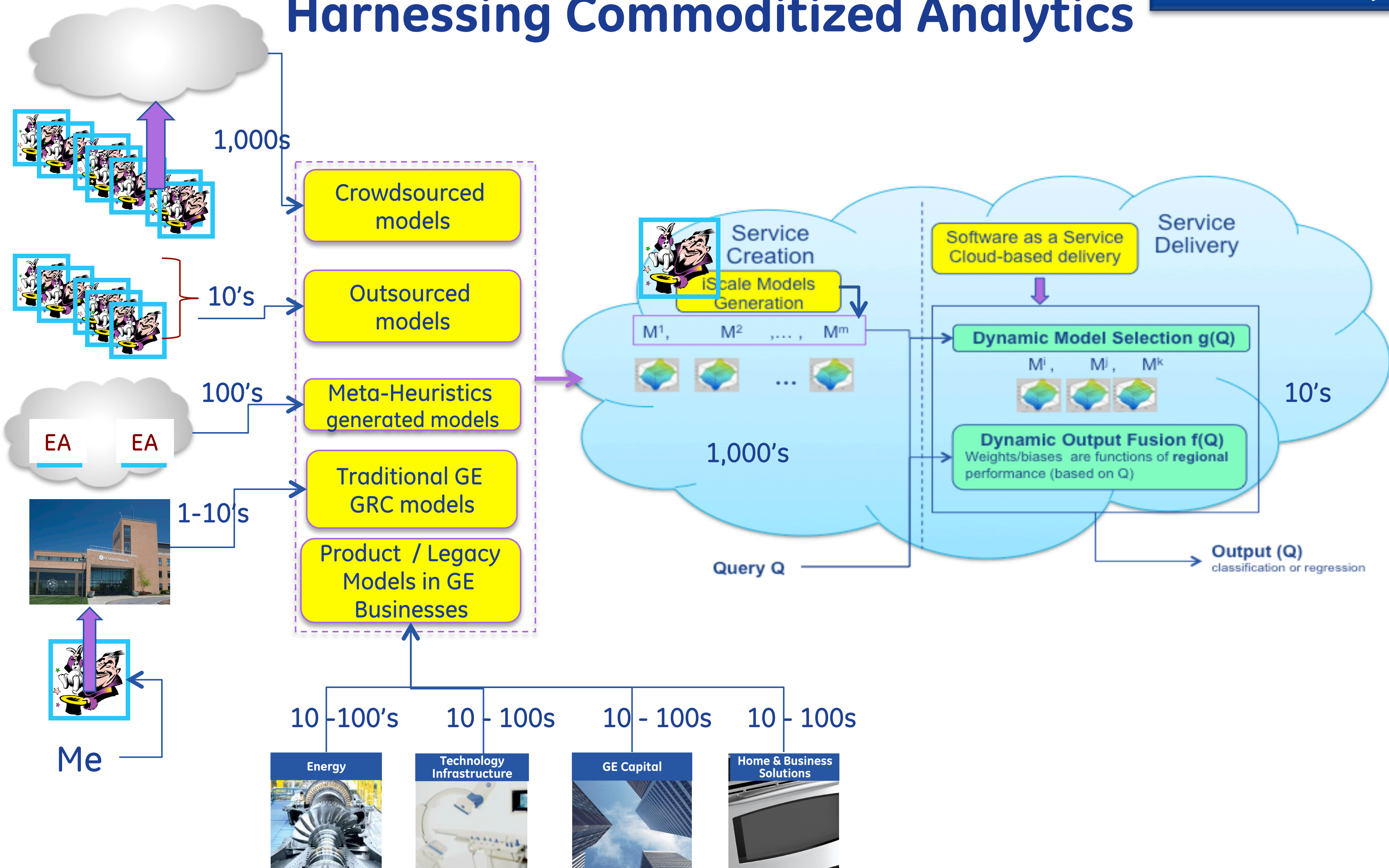
Potential commoditization of analytical models

Source Type	Mechanis m / Incentive	Examples	URL
Crowdsourcing Crowdsourced models	*Model Competition; R&D Competition/ Prizes*	***Kaggle*** ***TunedIT*** **CrowdANALYTICX** ***INNOcentive***	www.Kaggle.com www.TunedIT.org www.Crowdanalytix.com www.innocentive.com/
	HITS / Direct payments	***Amazon Mechanical***	www.mturk.com/mturk
			portal/
Outsourc Academia or Outsour mode			.cs.berkeley.edu
Cloud-b automatio Meta-Heuristics generated models	*model design / Bleeding Edge Research; Innovation; Papers*	***iScale*** [GE GR]	s.csail.mit.edu/ signOpt/evo.php?n=Site.FlexGP
Industrial R&D Traditional GE GRC models	*R&D ML Projects (manual dev. on single node or cluster) / Research; Innovation; Patents;*	Tree Ensemble Classifiers [TEC] Prognostics, …	www.ge.geglobalresearch.com
Commercial Products/ Product / Legacy Models in GE Businesses	*Existing commercial product offerings; Deployed models / Growth, Margins*	**SmartSignal** **Proficy** **System1** **ProDaps**  	www.smartsignal.com

We need to harness value from all available analytic models, **regardless of their sources**, including crowdsourced, outsourced, traditional, and proprietary OEM models

What: Lazy Meta Learning - Harnessing Commoditized Analytics

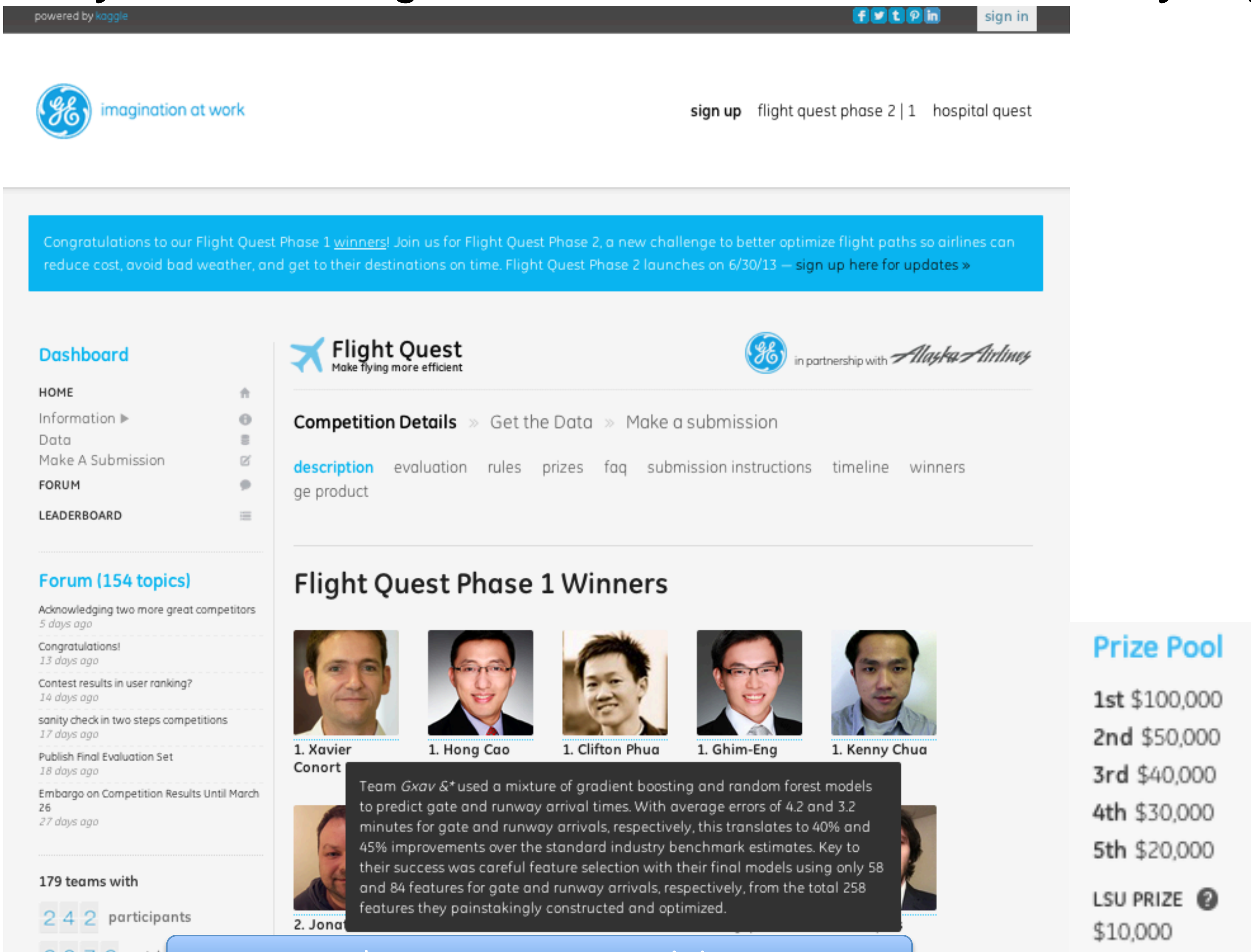
2.4. AC Analytics (Future) Agnostics Fusion for Commoditized Analytics



Based on run-time queries, a cloud-based service automatically selects and fuses the best models, intelligently leveraging model meta-information

GE Flight Quest at Kaggle

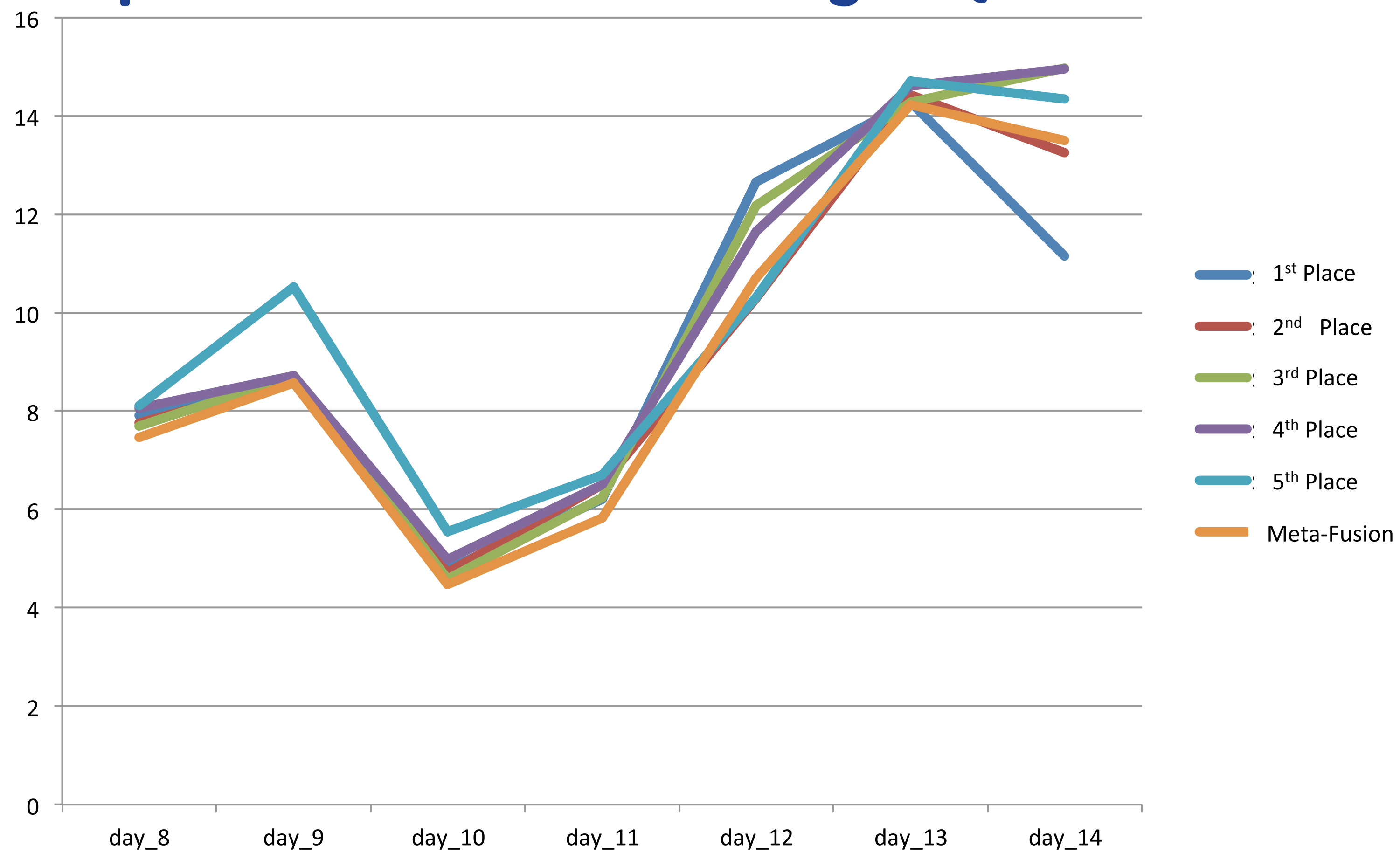
Predict runway arrival time & gate arrival time for each Alaskan Airways flight



$250K, 1,000's models

Source: https://www.gequest.com/

Fusion of top five models for GE Flight Quest



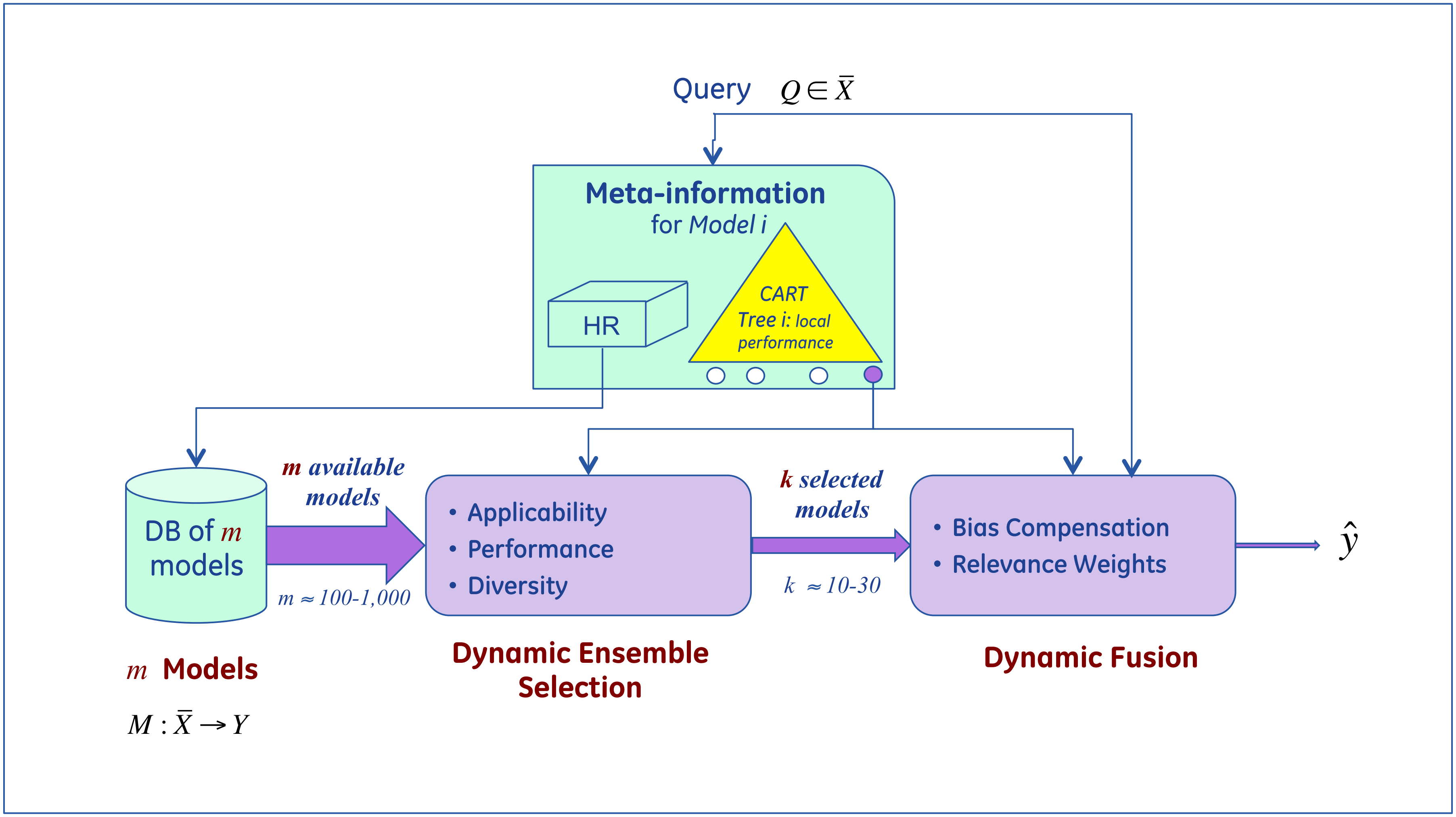
Team	day_1	day_2	day_3	day_4	day_5	day_6	day_7	day_8	day_9	day_10	day_11	day_12	day_13	day_14
1st place	5.8449	5.0991	5.0606	4.9187	5.7752	4.9718	5.4696	7.9038	8.5997	4.7784	6.2098	12.6572	14.2715	11.1484
2nd Place	5.9595	5.2351	5.3171	4.6769	5.6437	4.9331	5.3707	7.7655	8.5957	4.7160	6.5076	10.2944	14.4285	13.2421
3rd Place	5.7560	5.1777	4.8639	4.6531	5.8765	5.0068	5.3799	7.6954	8.7252	4.5796	6.2310	12.1822	14.2826	14.9680
4th Place	6.0796	5.4643	5.3003	4.9229	6.2737	5.2565	6.0158	8.0536	8.7253	4.9750	6.4990	11.6501	14.6106	14.9516
5th Place	6.5626	5.7937	5.5331	5.4074	6.2780	5.8030	6.3246	8.1123	10.5216	5.5470	6.7076	10.3074	14.7022	14.3382
Meta_fusion	← Meta-data Training →							7.4690	8.5635	4.4738	5.8317	10.7100	14.2252	13.4958
Minimum								7.6954	8.5957	4.5796	6.2098	10.2944	14.2715	11.1484

3. Experiments with Commoditized Analytics

Lazy Meta-Learning

Experiments with a Regression Problem

Lazy Meta Learning
Dynamic Ensemble Selection + Dynamic Fusion

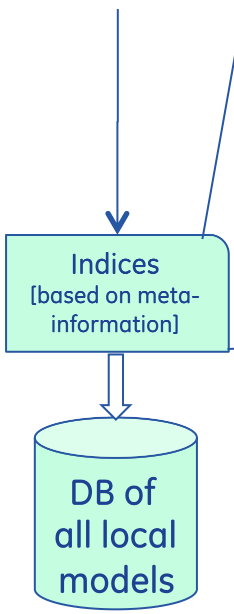


imagination at work

Experiments

How (Details) – Dynamic Ensemble: Meta Information

3. Lazy Meta-Learning: Dynamic Ensemble



m Models

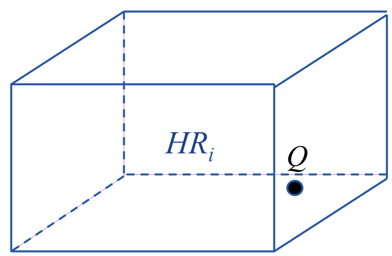
$m \approx 100\text{-}1{,}000$

Model Creation & Meta- Information

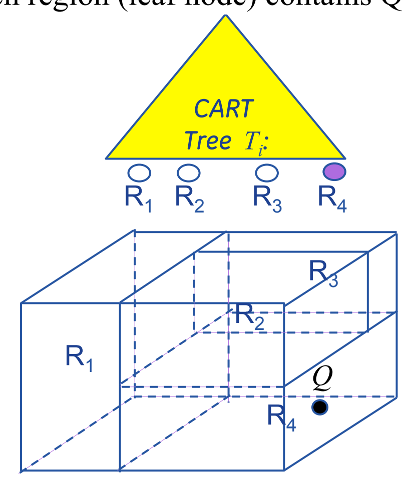
Meta-Information

Model	Training Set	Training Set (updated over time)
Label: M_i	Label: TS_i	Label: VS_i
Model Mapping:	Model Applicability:	Local Model Performance:
$M_i: X_i \rightarrow Y_i$	Hyper-rectangle: HR_i	CART Tree: $T_i: X_i \rightarrow e_i$
$X_i \in R^n;\ Y_i \in R\ (regressions)$	$HR_i \in R^{2,n}$ is the range	$X_i \in R^n;\ e_i \in R\ (regressions)$
$Y_i \in [0,1]^{(C+1)} (classif.)$	of values of each of the n features over TS_i	$e_i \in [0,1]^{(C+1)} (classif.)$

Applicability of Model M_i :
"Is Q in HR_i ?"



Local Performance of Model M_i :
"Which region (leaf node) contains Q?"

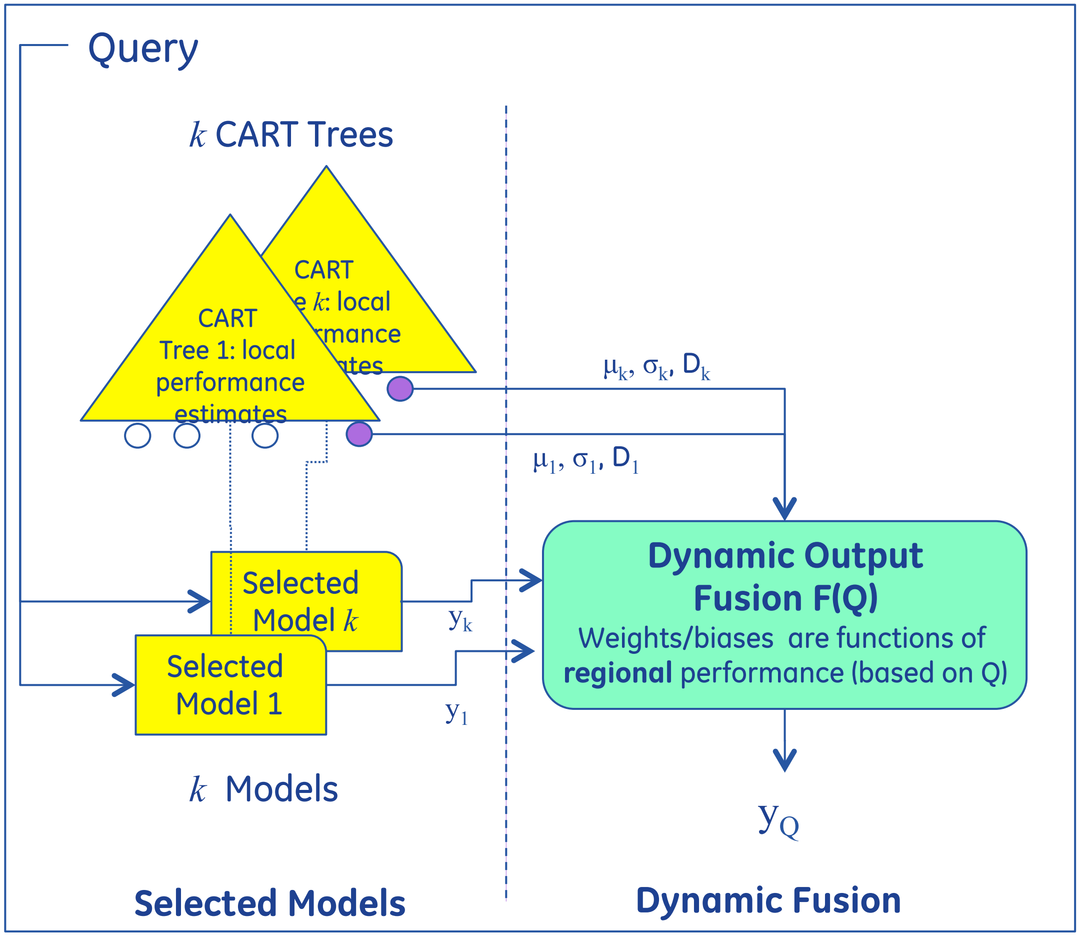


Leverage DB efficiency (indices) to retrieve applicable models in *log m*

Built the **indices** for the DB

Experiments

How (Details) – Dynamic Fusion

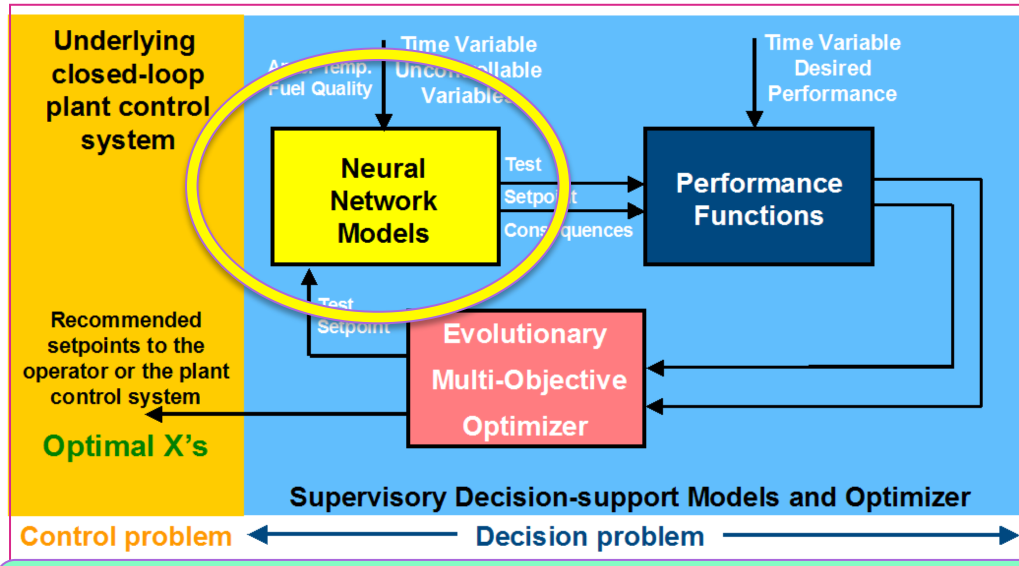


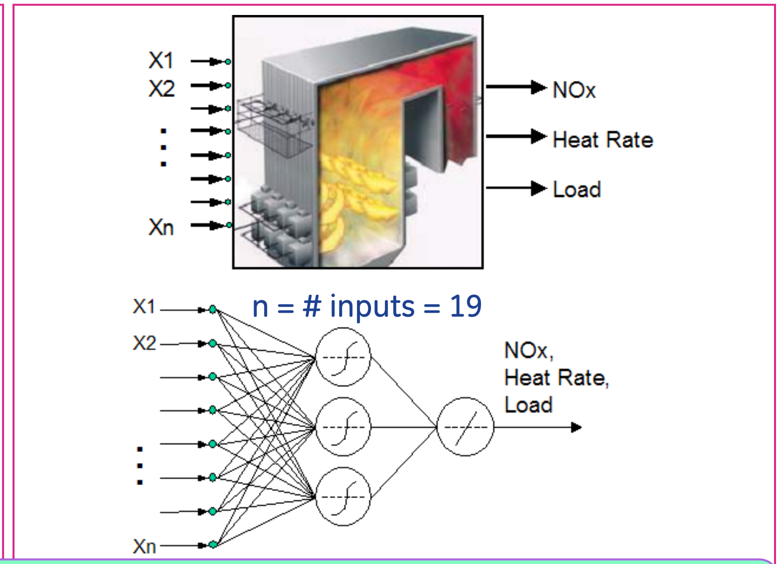
Dynamic bias compensation for each model based on bias in CART leaf node

Dynamic weighted fusion based on each model prediction error (Std. Dev. of error) in CART leaf node

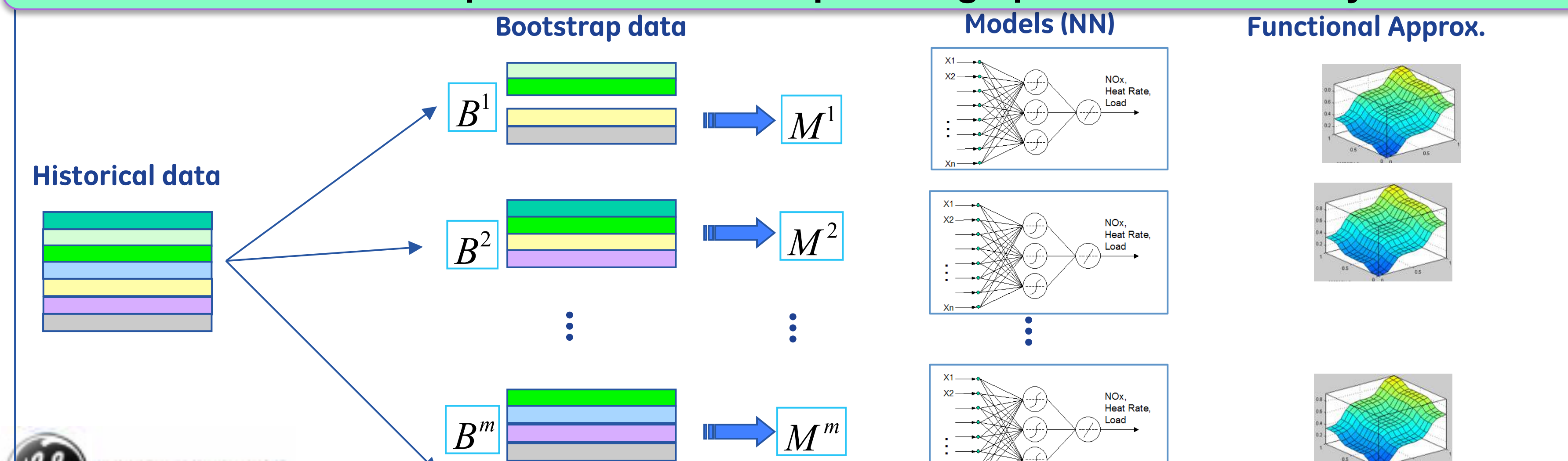
imagination at work

Lazy Meta-Learning Experiments



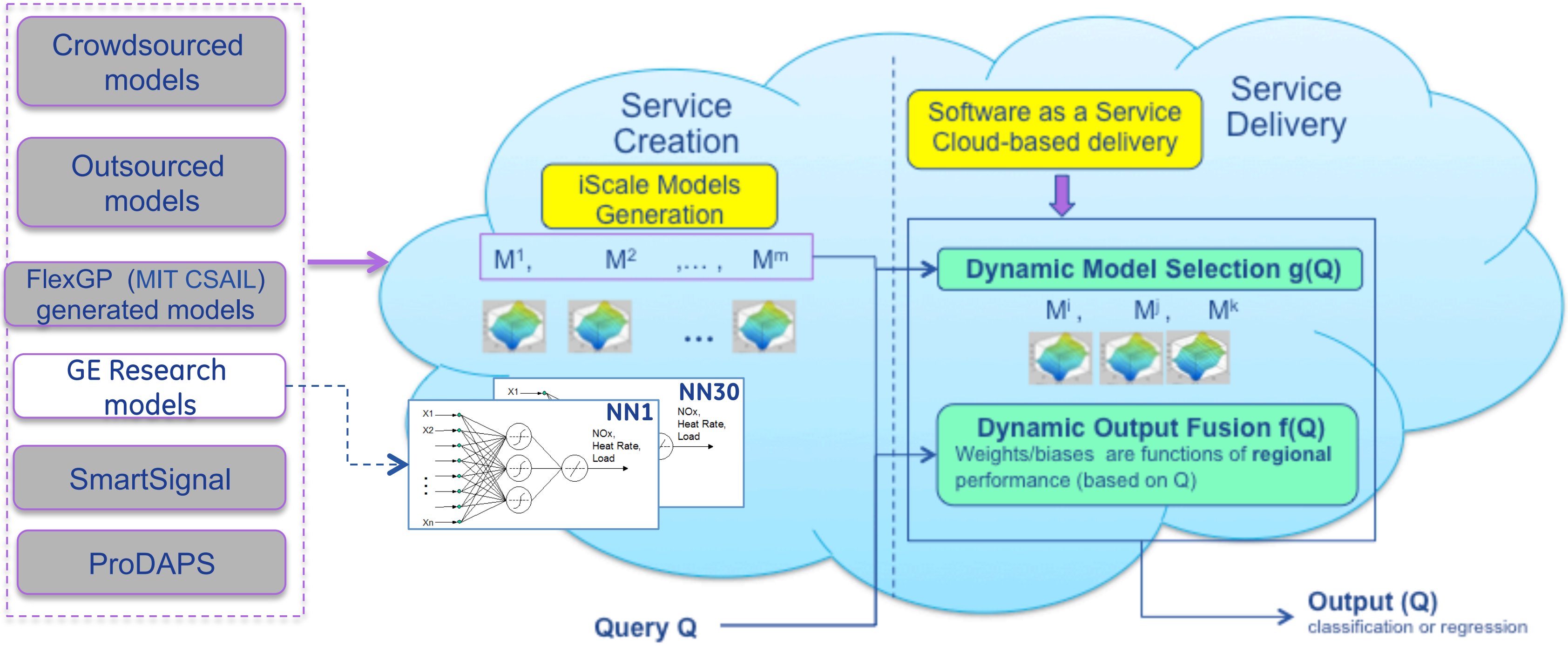


Find the "best" set points X to meet the load, minimize Heat Rate, and minimize NOx based on NN prediction. This requires high prediction accuracy



Reduced NN prediction uncertainty with an ensemble of 30 diverse NN's (trained by using bootstrapping)

Lazy Meta-Learning Exp.1 (Regression)

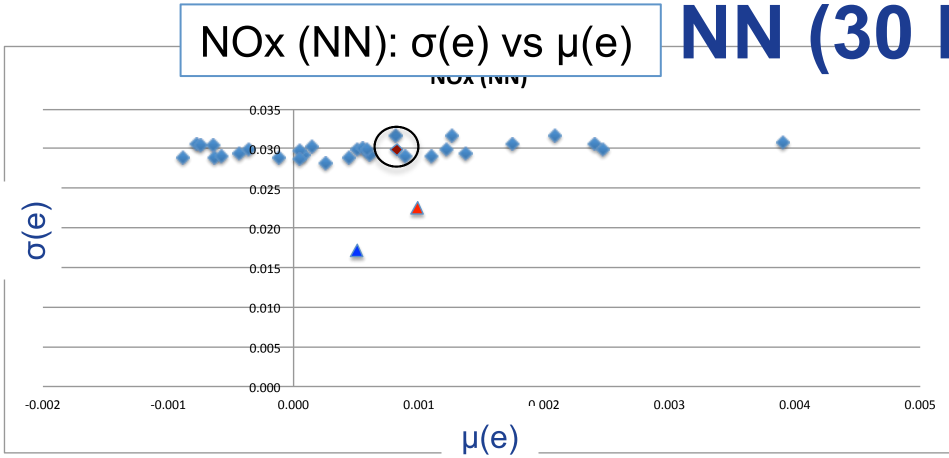


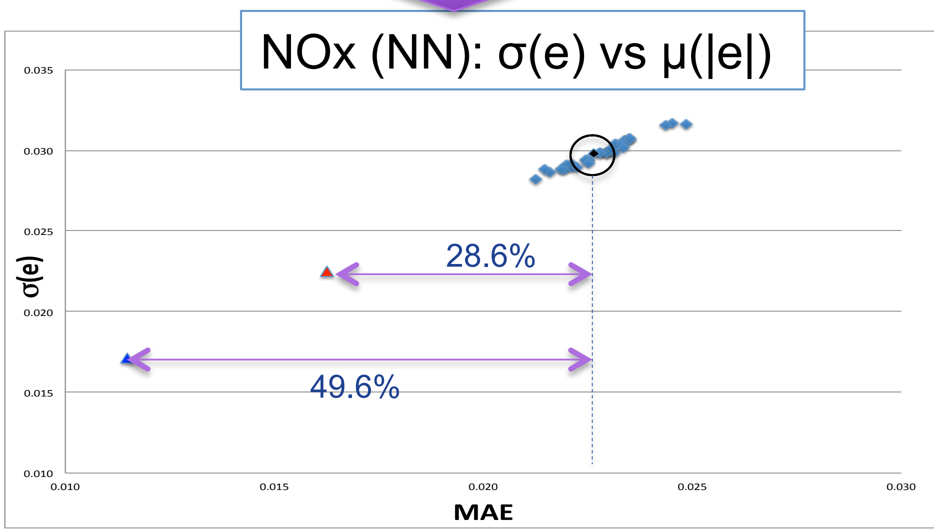
Used 30 Feed-forward NN's trained with bootstrapping, and performed design experiments in Matlab (on single node)

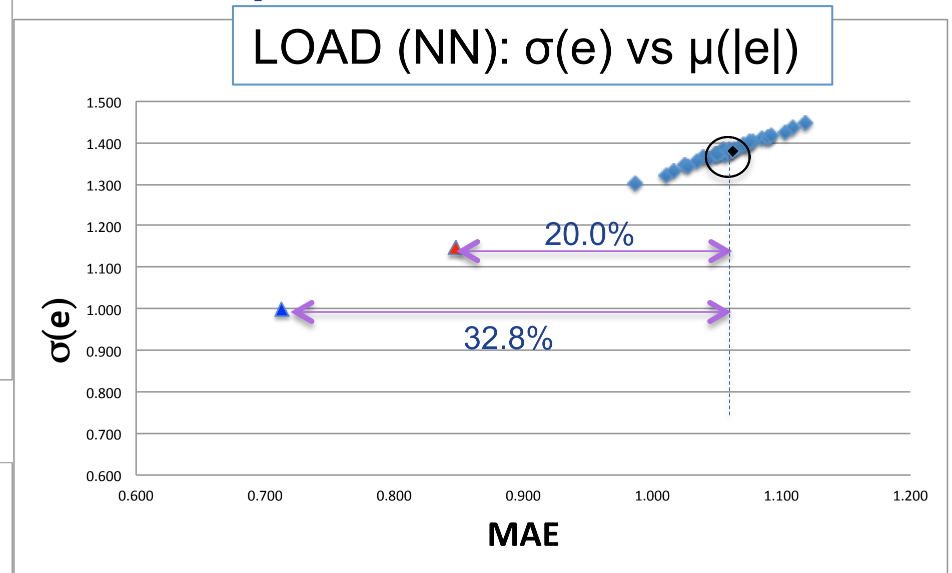
- Defined **design parameters** and automation **processes** to implement Customized Analytics **for regression** problems (typical of *Prognostics* problems)
- Partially validated approach using historical data set from GE Energy (coal-fired boilers)

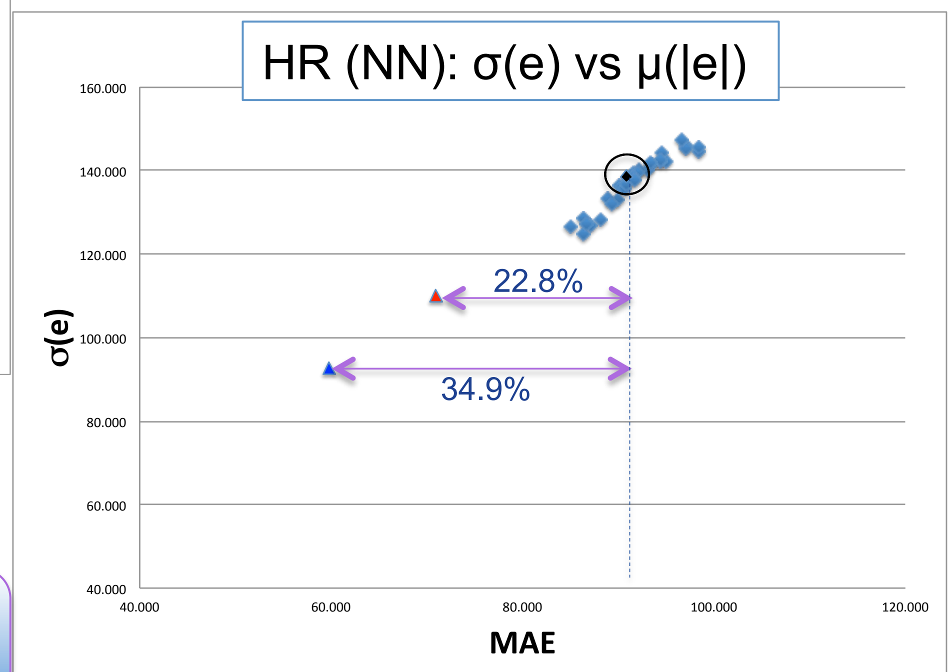
imagination at work

NN (30 Models)





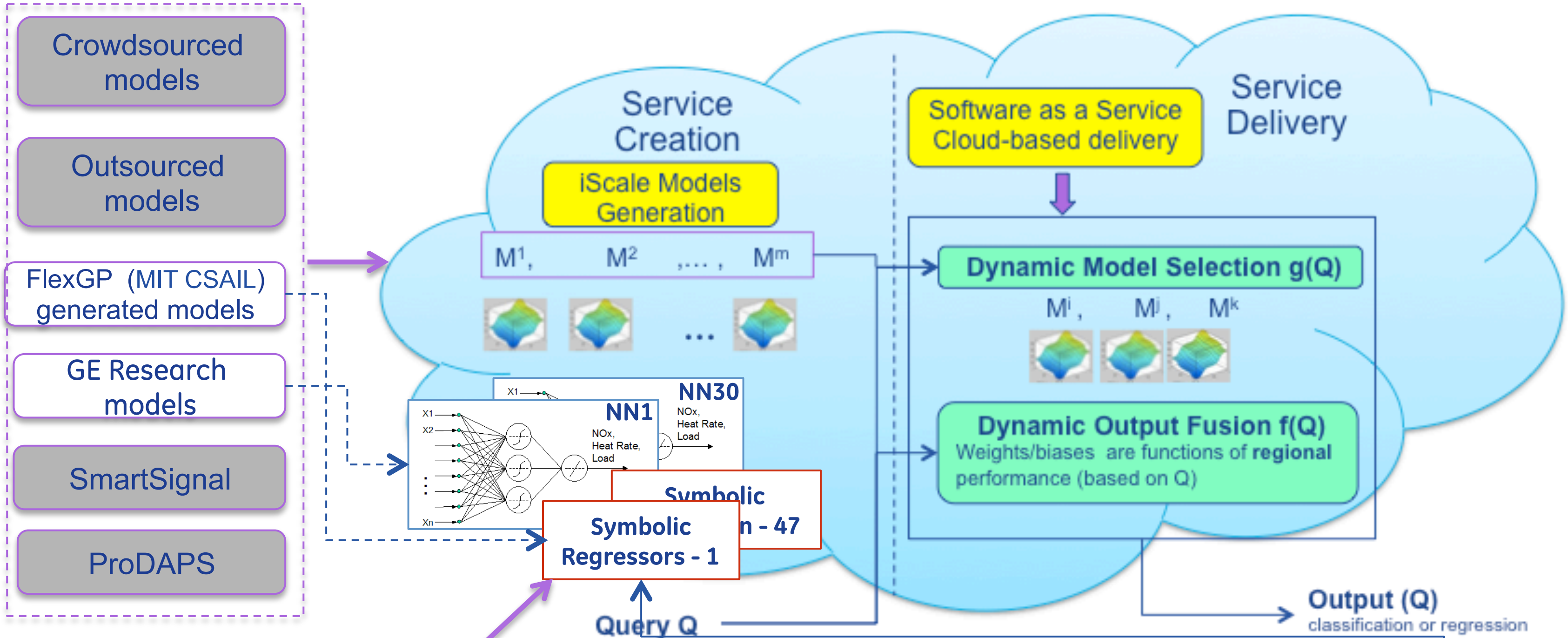


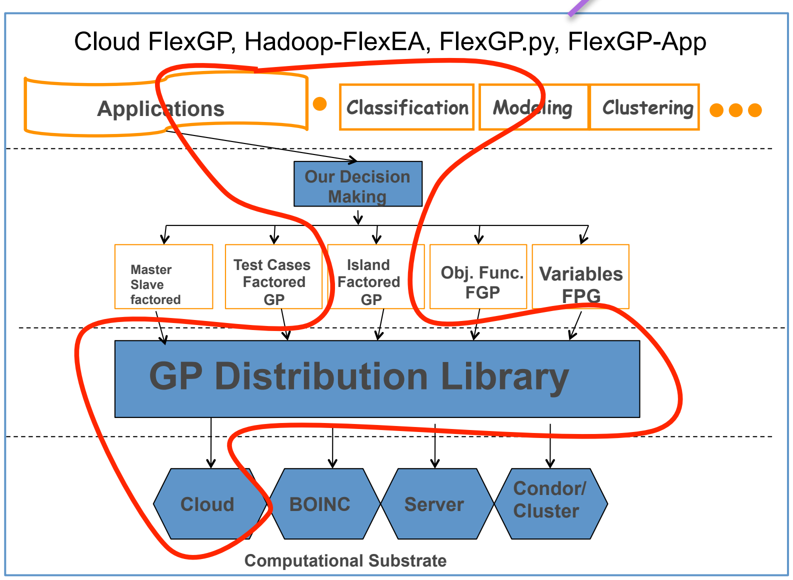


LEGEND:
Baseline Average
Custom Fusion 1 (n≥25)
Custom Fusion 2 (n≈4)

Results on ~2,250 records validation set records (disjoint from ~5,000 training records)
Improvement of 32%-49% with Fusion 2

Lazy Meta-Learning–Exp. 2 (regression)

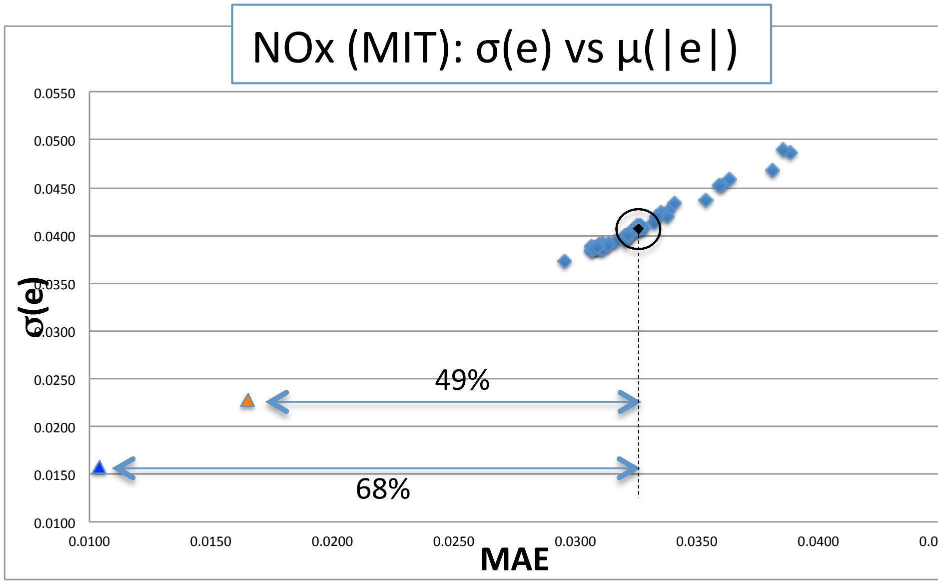


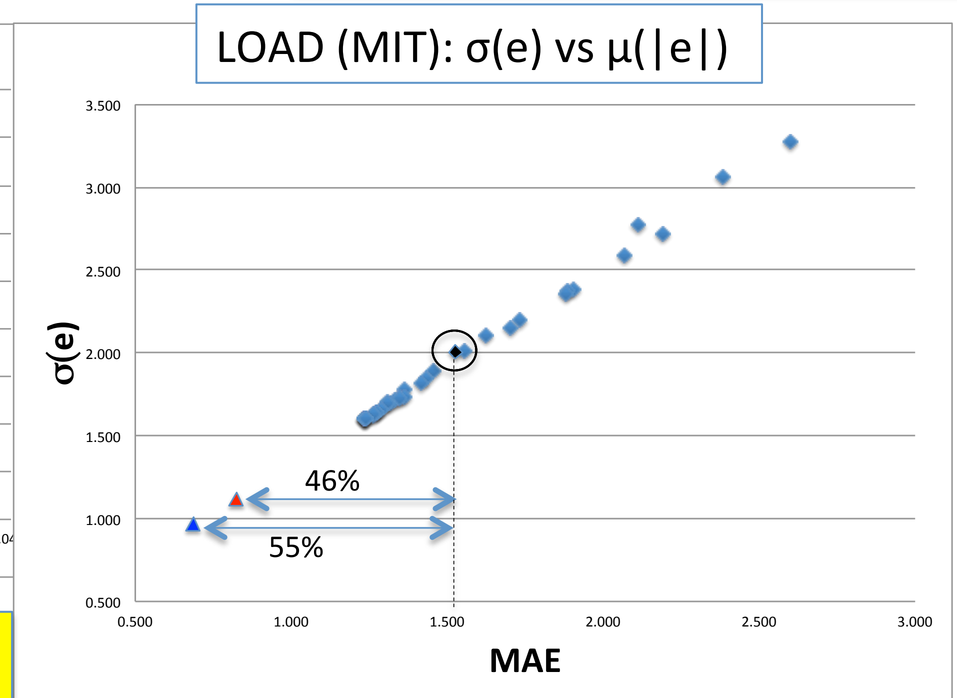


- Used MIT CSAIL's FlexGP to generate 36-47 symbolic regression models, trained on same data set (diversified by bootstrapping, randomized feature subset selection, different function sets, and different grammars)
- Rerun experiments for a more complete test of Customized Analytics for regressions

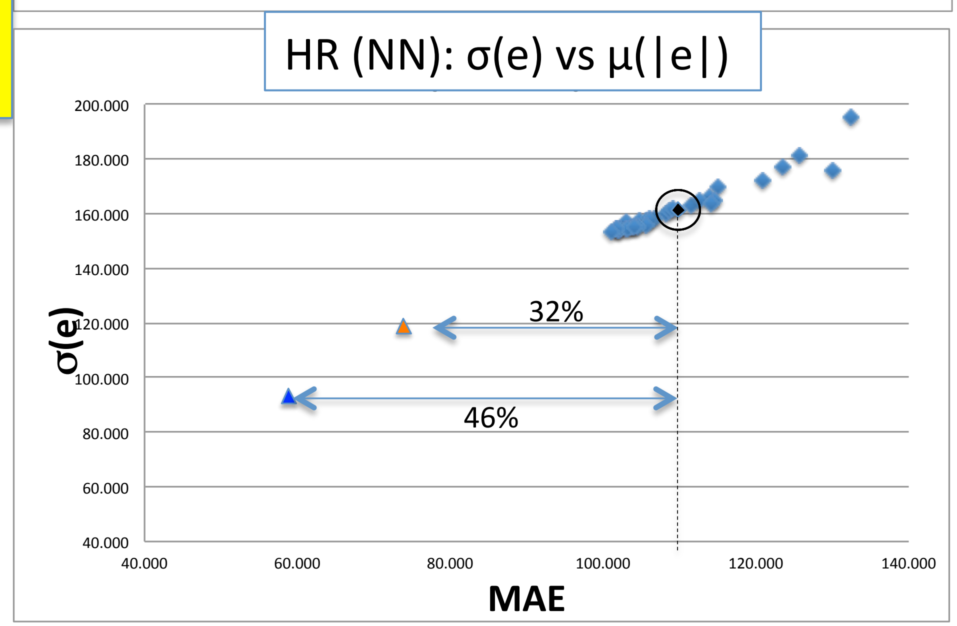
Complete validation of Customized Analytics for Regression

MIT Symbolic Regressors (36 and 46 models respectively)



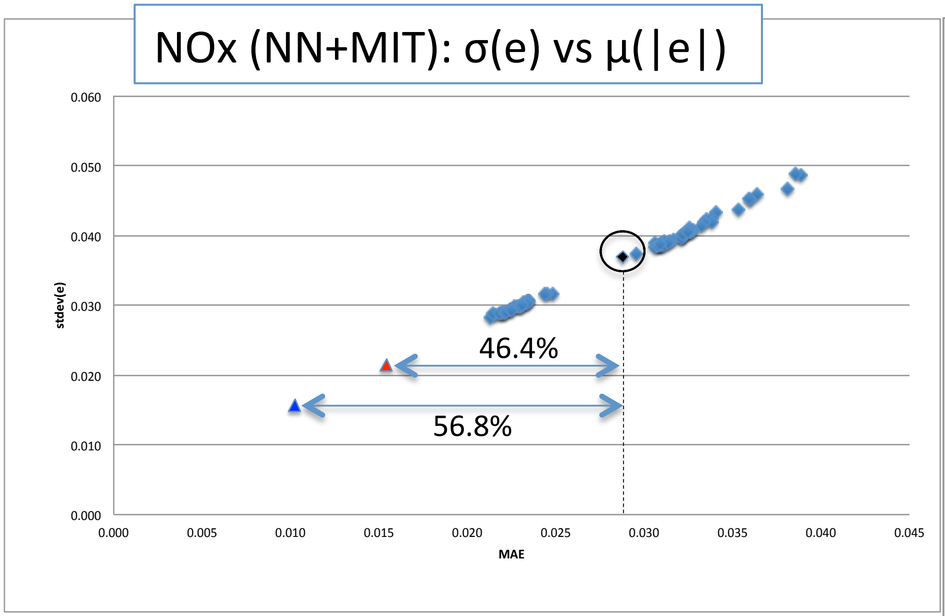


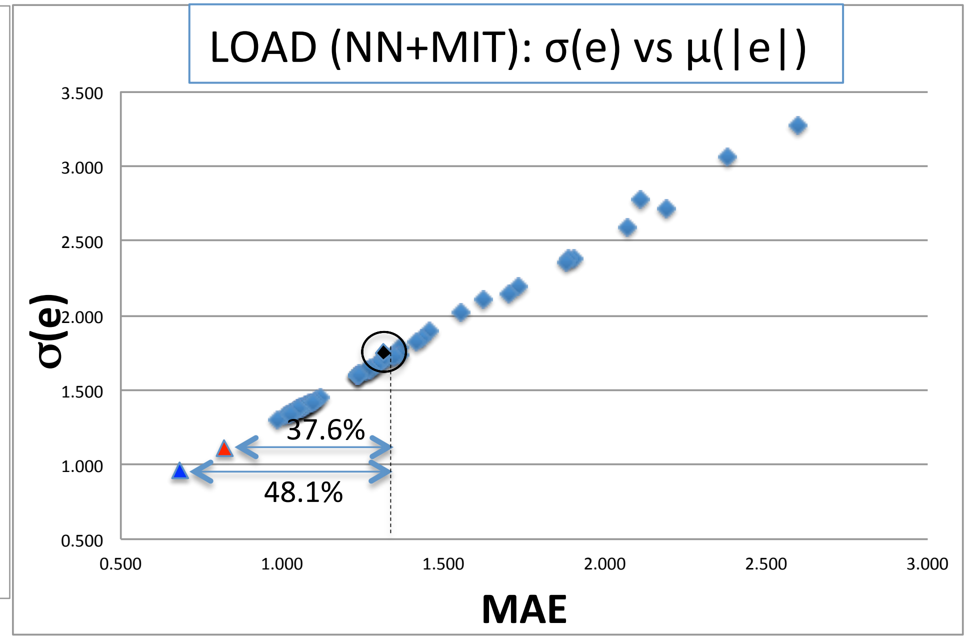
LEGEND:
Baseline Average (~ 40-50% worse than NN Baseline)
Custom Fusion 1 (n≥25)
Custom Fusion 2 (n≈4) (~ 10% better than Fusion2 of NN)



Improvement of 46%-62% with Fusion 2

NN + MIT Symbolic Regressors (66, 76 models respectively)



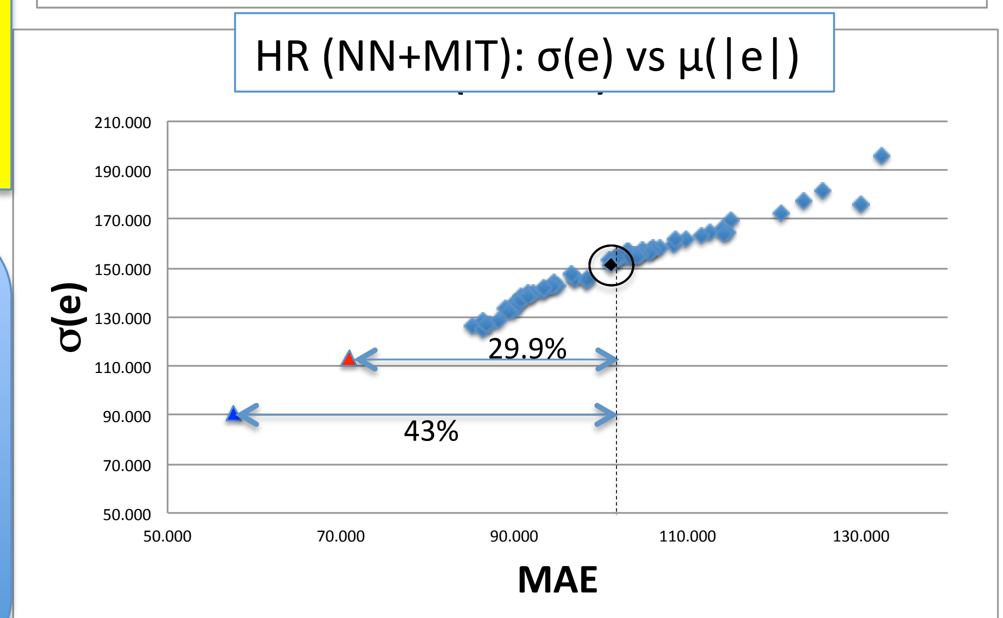


LEGEND:

Baseline Average ◆ (convex sum of NN & MIT baselines)

Custom Fusion 1 (n≥25) ▲

Custom Fusion 2 (n≈4) ▲ (slightly better than MIT fusion)



- The 30 NNs provided an **40-50%** better baseline for all three outputs than the MIT models
- The fusion of the MIT models provided **~10% better** results than the fusion of the NNS
- **By injecting diversity** into low performing (i.e. weak) models **we can still extract high performance from their fusion**
- The fusion of all models was only slightly better than the fusion of MIT modes alone (declining contribution of additional models)

4. Conclusions

GE imagination at work

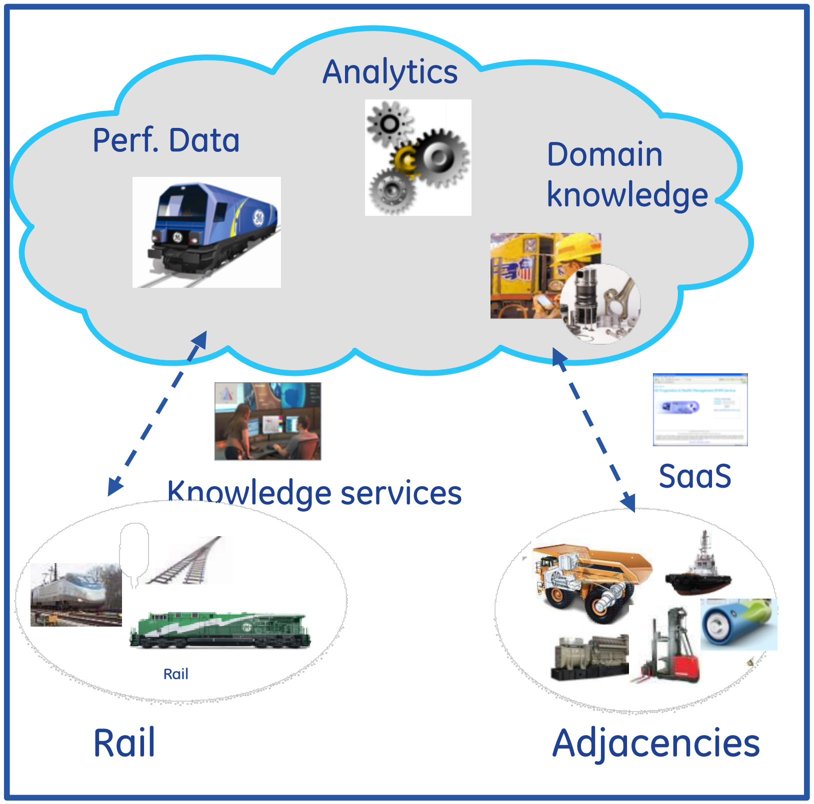
Conclusions

Crowdsourcing: The power of the people (augmented by collaboration tools)

Cloud Computing: The enabler

Analytics: Still the secret sauce…?

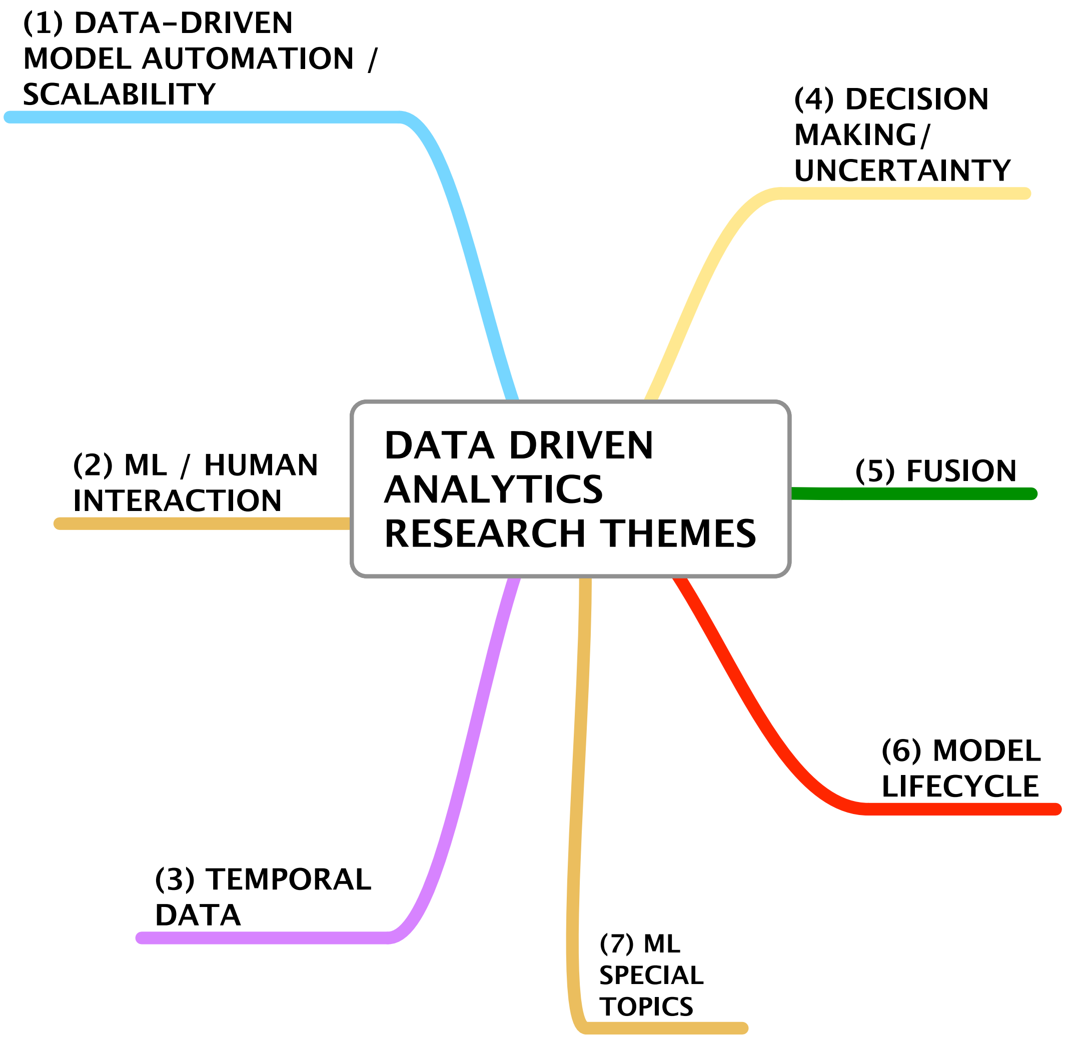
Trend: Intersection of two or three of these components (AMP, Kaggle, DREAM)



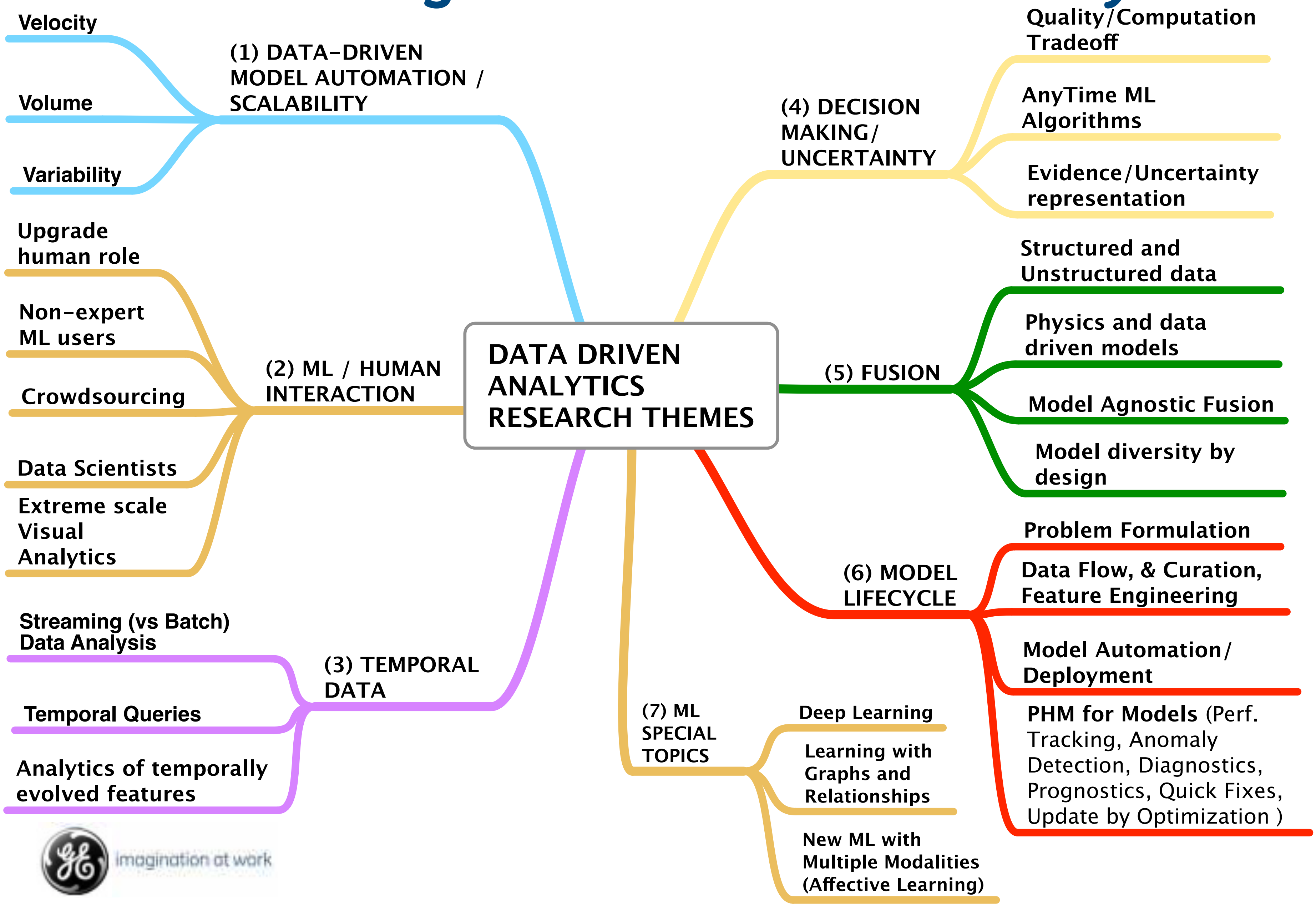
Past Job Responsibilities:
- Handcraft KB and ML models
- Use EA-processes to derive models
- Create static model ensembles

Future Job Responsibilities:
- Harness commoditization trend in analytics to automate **dynamic model ensemble and fusion** using Lazy Meta-Learning
- Leverage Analytics for the Industrial Internet

Future Challenges for Data-Driven Analytics



Future Challenges for Data-Driven Analytics



imagination at work

Thank you.

Piero P. Bonissone
Chief Scientist, GE GR
bonissone@ge.com