Evolving Local Fuzzy Models to Adapt in a Dynamic Environment:

An Example in Asset Management

Piero P. Bonissone

GE Global Research
Schenectady, NY, USA

imagination at work

Outline

1. PHM at GE Global Research
2. Issues in Model Lifecycle
3. Computational Intelligence and Model Generation
3. Fuzzy Instance Based Model (F-IBM)
4. Prediction of Best Units for Asset Selection: Case Study
5. Future Work

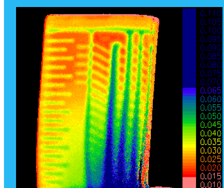
1. PHM at GE Global Research

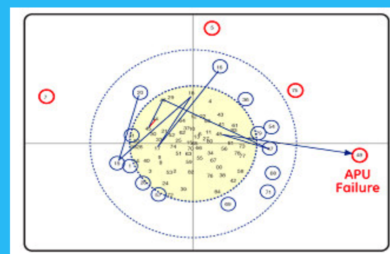
- PHM Technical Synergies
- PHM Elements
- PHM Installed Basis
- PHM Technology

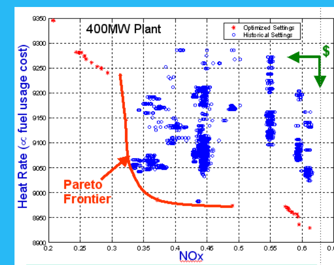
imagination at work

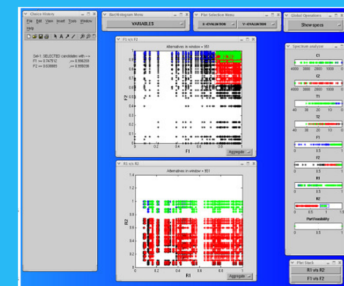
PHM Technical Synergy

Prognostics and Health Management (PHM)







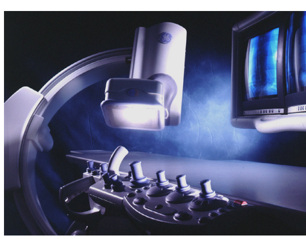


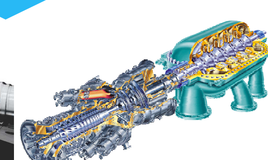
On-board Sensors & Off board Inspections

Anomaly Detection, Diagnostics & Prognostics Alg.

On-board / Off board Optimization Alg.

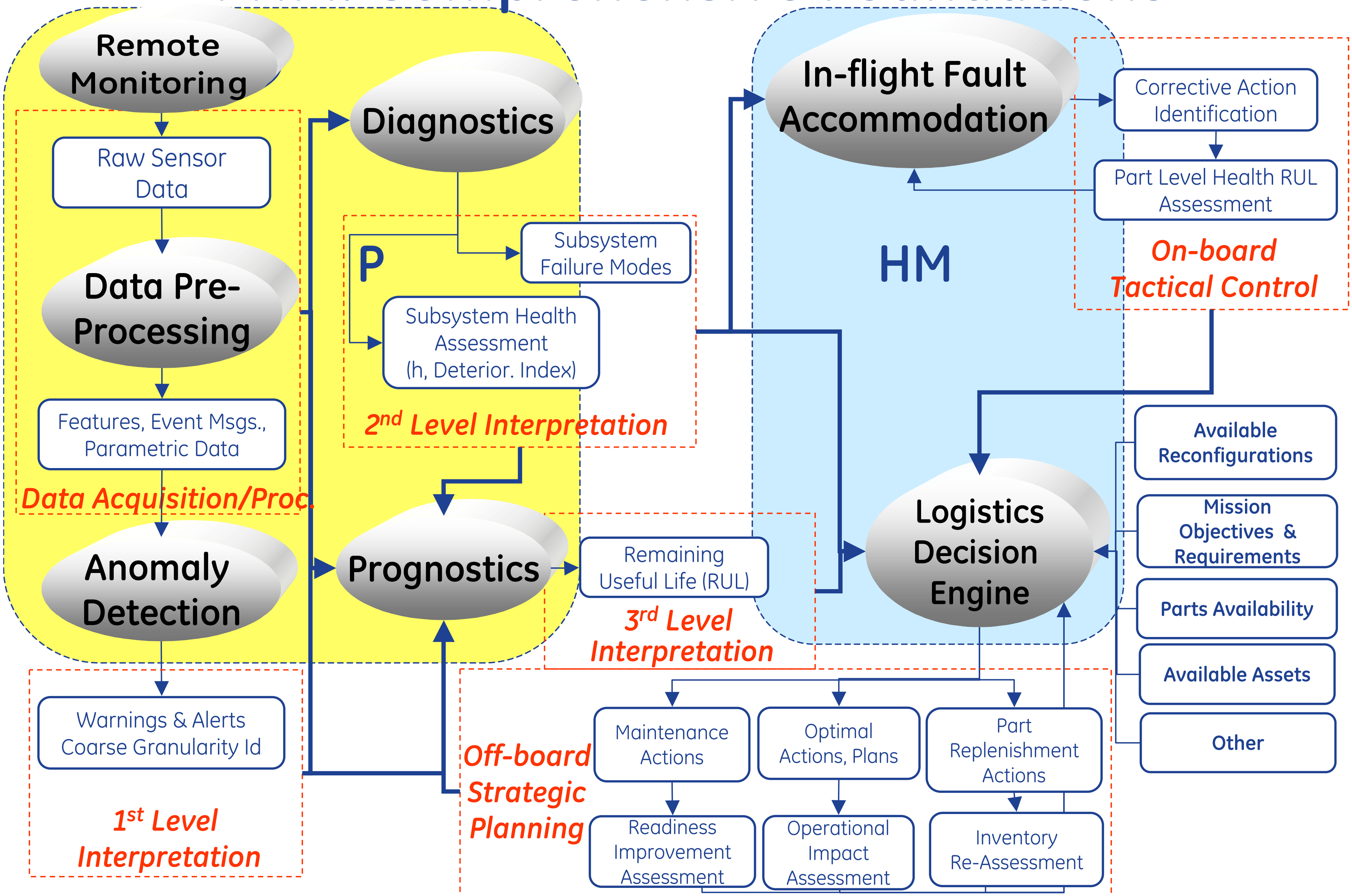
Visualization & Multi-Criteria Decision-Making Systems





GE Healthcare **GE Aviation** **GE Energy** **GE Oil & Gas** **GE Rail** **LM**

PHM: Comprehensive Foundations

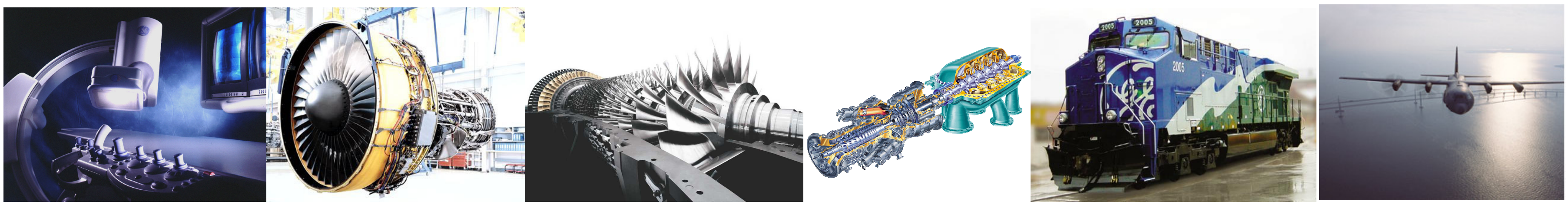


PHM Requires a Holistic Approach

PHM: From RM&D to Prognostics & Optimization

GE RM&D/PHM Installed Base

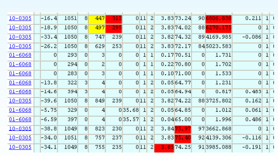
- ~20,000 steam/gas/hydro/wind turbines
- ~17,500 aircraft engines
- ~9,000 rail locomotives
- ~60,000 medical imaging machines



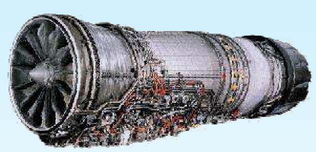
GE Healthcare GE Aviation GE Energy GE Oil & Gas GE Rail Lockheed Martin

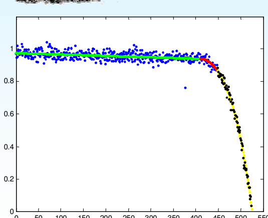
PHM Elements

Diagnostics

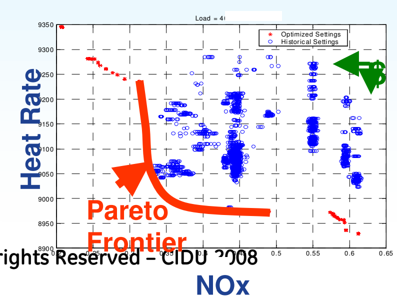


Prognostics

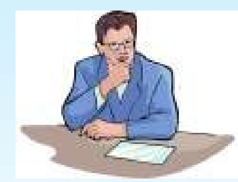




Optimization



Maintenance decisionmaking

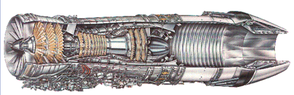
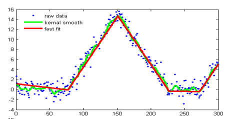
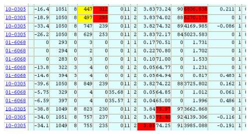
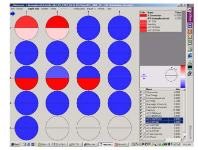
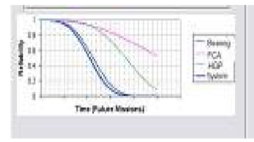
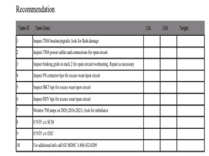


Mission Objectives & Requirements

Parts Availability

imagination at work

PHM Technologies

Data Acquisition	State Detection	Diagnostics	Prognostics	Advisory Generation	Optimization	Feedback & Learning
						

Functions

Data Acquisition	State Detection	Diagnostics	Prognostics	Advisory Generation	Optimization	Feedback & Learning
On-platform sensors Data recorders **Off-platform** sensors Inspection / NDE Data repository Data de-noising Data smoothing Data compression Data transmission IT infrastructure	**Features** Generation/Selection Manual/Automated **Anomaly Detection** Normal / Abnormal structures (trajectory, clusters, states, patterns) **Anomaly Resolution / Id.** Resolving sensors faults, op. transients, model faults, system faults	**State Interpretation** Manual/Automated Classification **Fault mode Id.** **Health Assessment**	**RUL Prediction** Manual/Automated RUL forecast **Health Assessment Projection** **Lifing** Fault propagation	**On-board Fault Accommodation** **Off-board DSS**	**Maintenance Optimization** Workscope (build level) Shop scheduling **Logistics Optimization** Supply Chain Mgmt **Operational Optimization** Asset/mission allocation	**Performance metrics dashboard** **Ground truth acquisition** **Structured feedback**

Key Technologies

Data Acquisition	State Detection	Diagnostics	Prognostics	Advisory Generation	Optimization	Feedback & Learning
Electronics Signal Processing Communication Information Theory Statistics Nondestructive Inspection	Physics-based model Machine Learning (unsupervised) Statistics (Clustering, RF) Pattern Recognition Information Theory Soft Computing (NN, Fuzzy, Probabilistic, Covariance-based) Signal Processing	Physics-based model Machine Learning (Sup.) Statistics (CART, RF, MARS) Pattern Recognition Soft Computing (NN, Fuzzy Classifiers Prob. Classifiers) AI (case-based, rule-based)	Physics-based Lifing (fault prop.) Physics-based Model (Cycledeck) Machine Learning (Lazy learning) Statistics (Kernel-based regressions) Pattern Rec. (ARMA) Soft Computing (NN regressions, Fuzzy (TS), EA tuned models) AI (case-based) NDE Inspection	Adaptive control EA-designed control Human Interface Usability Multi Criteria DM (MCDM)	Multi-Objective Optimization (MOO) Evolutionary Algorithms Operations Research Multi Criteria DM (MCDM)	Automated Performance Monitoring Supervisory learning

GR Labs / GE Business *[a partial list]*

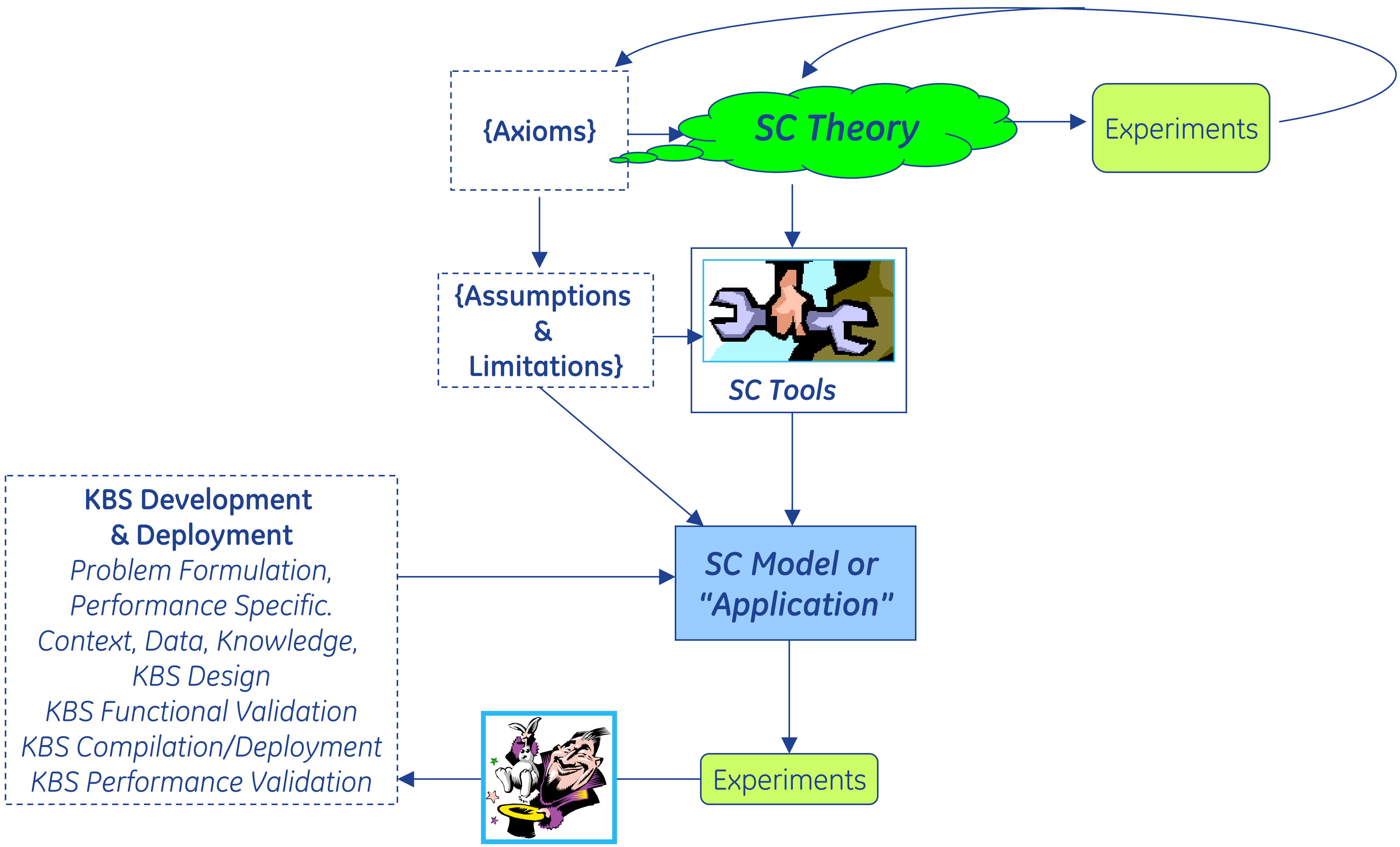
Data Acquisition	State Detection	Diagnostics	Prognostics	Advisory Generation	Optimization	Feedback & Learning
NDT Lab Integrated Device Lab Communications Lab Service Oriented Arch Lab GE Aviation Systems	iAI Lab Automation & Controls Lab GE Aviation Systems	iAI Lab Automation & Controls Lab GE Aviation Systems	Lifing Technologies Lab iAI Lab Automation & Controls Lab NDT Lab GE Aviation Systems	iAI Lab Automation & Controls Lab Risk and Value Mgmt. Lab	iAI Lab Risk and Value Mgmt Lab.	iAI Lab

2. Issues in Model Lifecycle

- Handcrafting Models
- Lifecycle: Build, Use, Update & Maintain

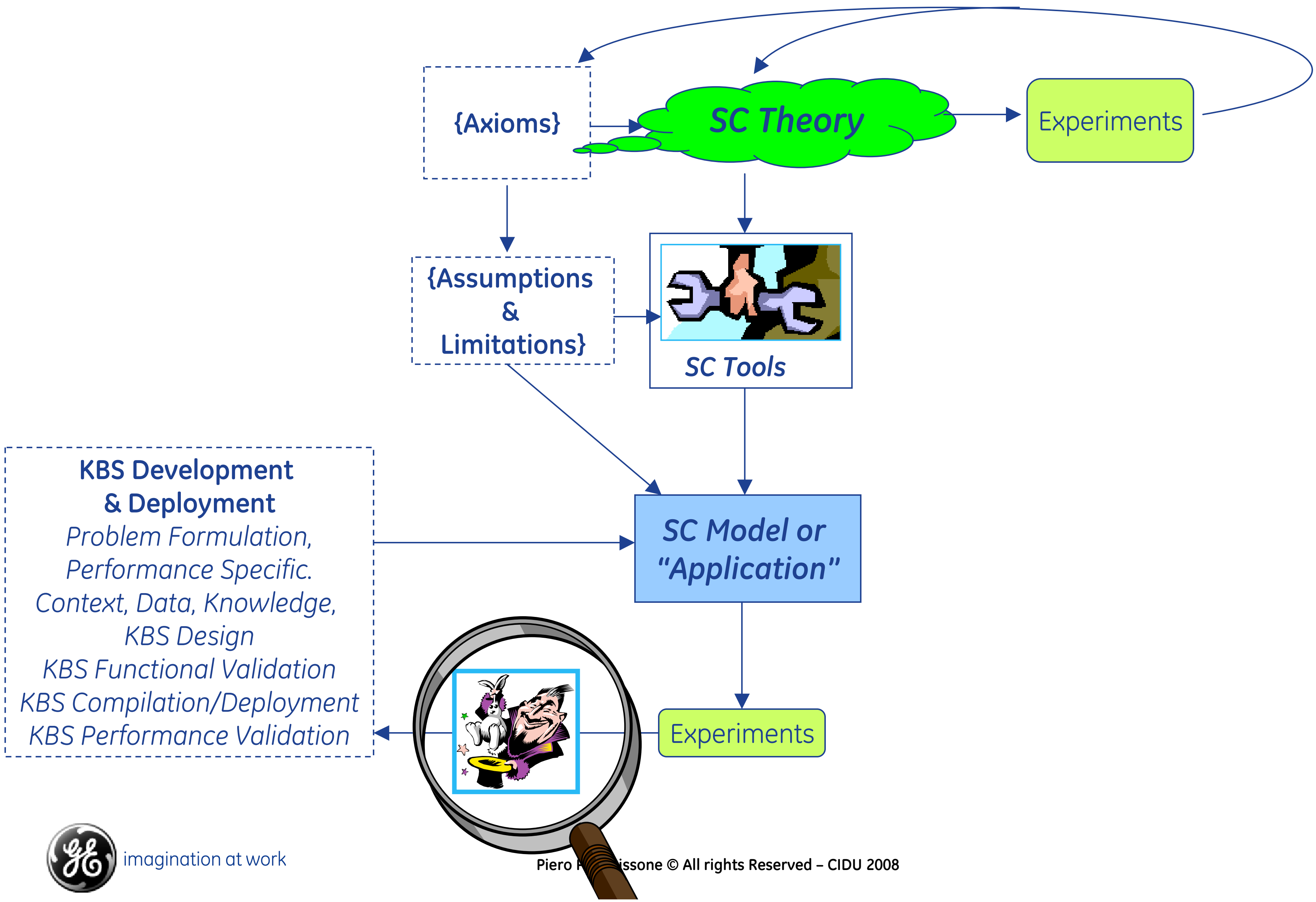
imagination at work

Model Generation Process: Pictorial

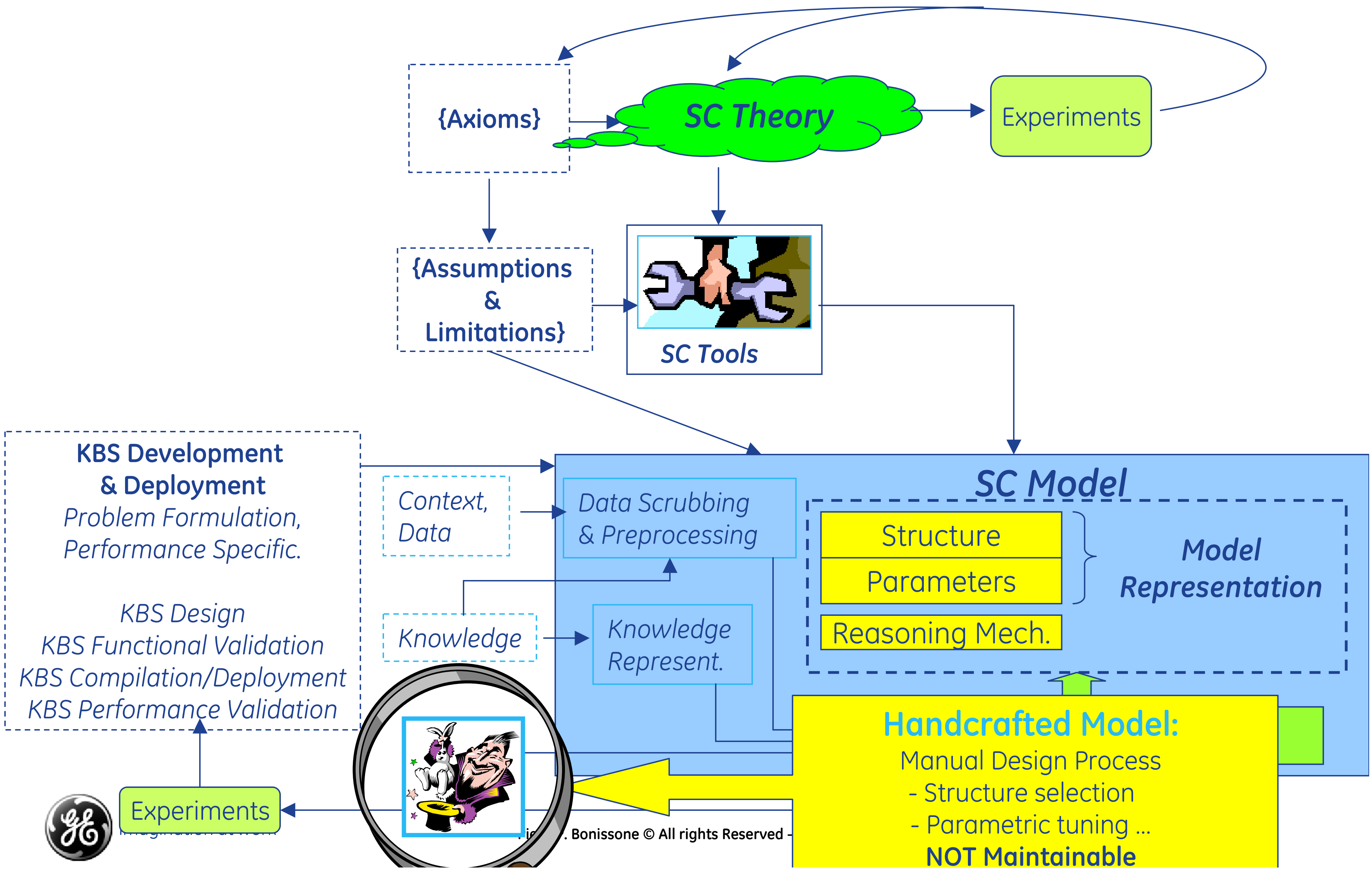


imagination at work

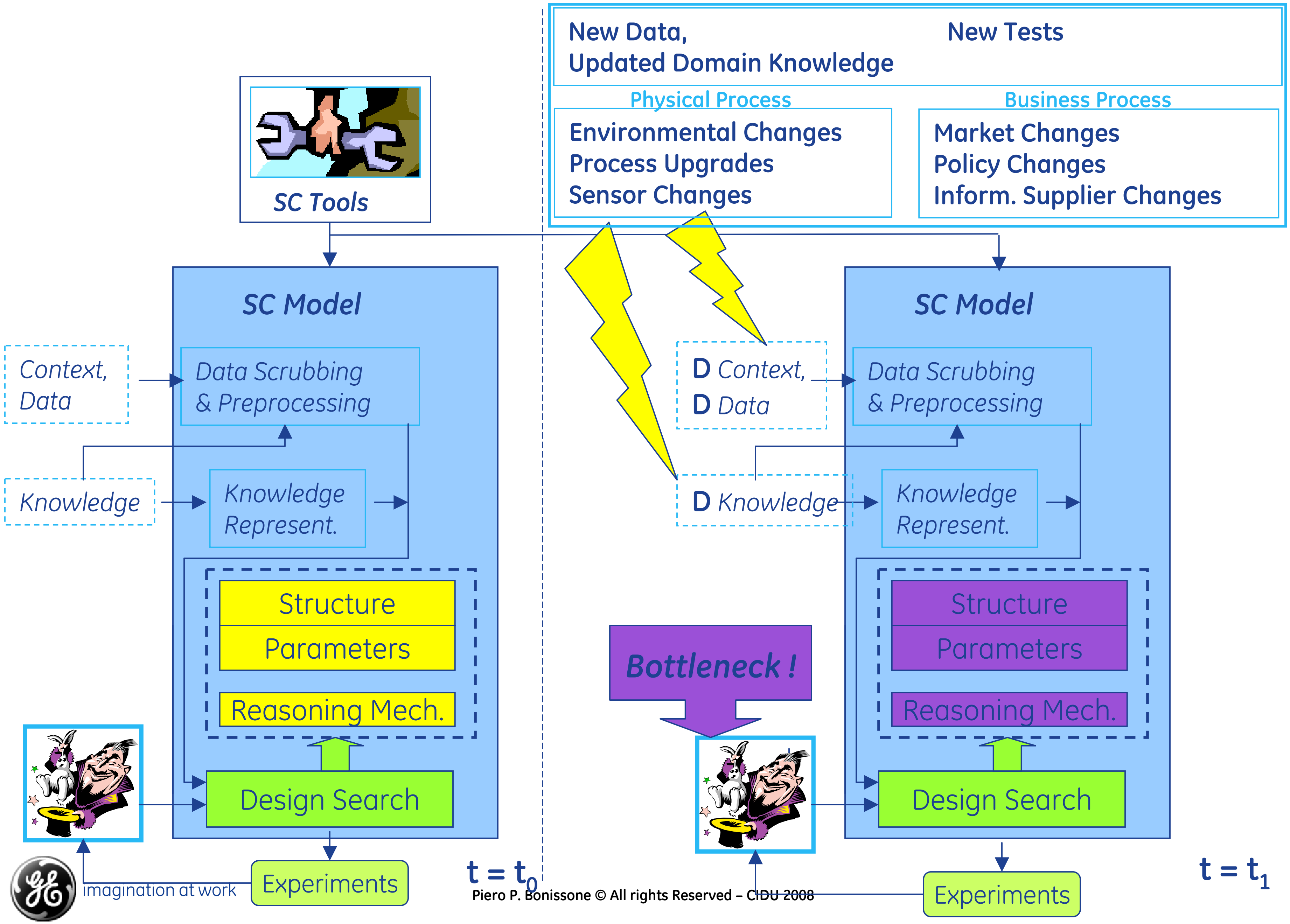
What is Wrong with this Picture?



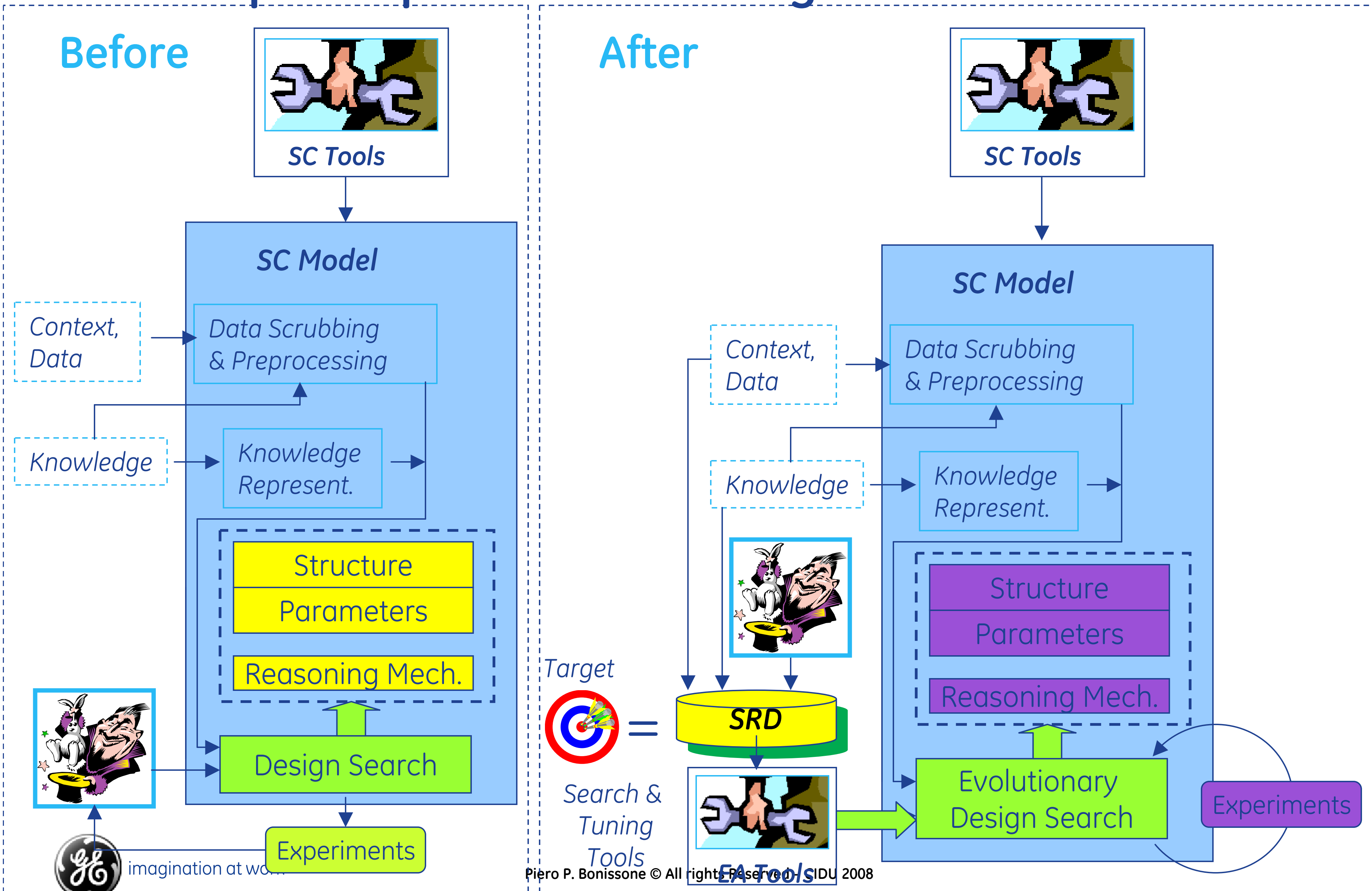
Let's Analyze the Problem



What Will Happen Over Time?

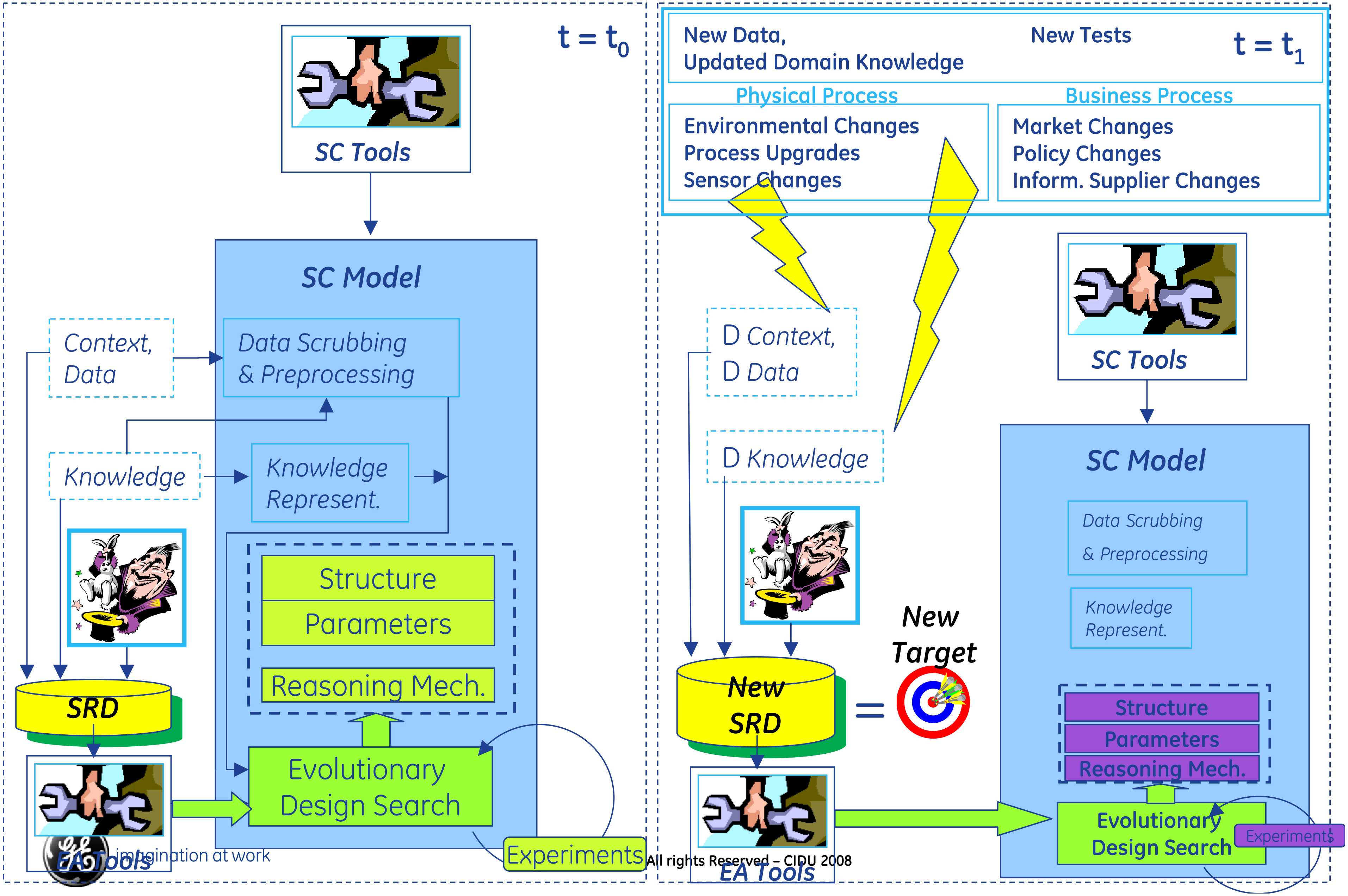


Addressing the Lifecycle of a KBS
Example: Supervised Learning



imagination at work

Addressing the Lifecycle of a KBS (cont.)
Example: Supervised Learning



3. Computational Intelligence and Model Generation

- Representation, Reasoning, and Search
- Functional Approximation (ANFIS) vs Fuzzy Instance Based Model (F-IBM)

imagination at work

Representation, Reasoning & Design Search

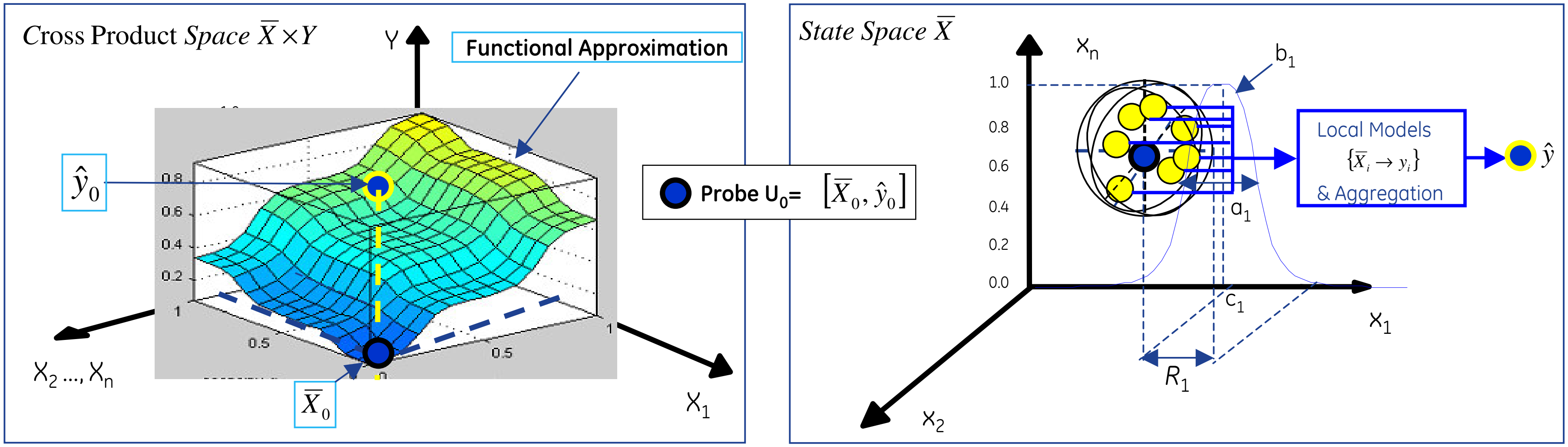
REPRESENTATION

Modeling Techniques	***CLASSICAL:*** *Linear Differential Equations*	***CI:*** *Bayesian Belief Networks*	***CI:*** *Neural Networks*	***CI:*** *Fuzzy Systems TSK/ ANFIS*	***CI:*** *Fuzzy Instance Based Reasoning*
Model Structure	**Order**	**Topology**	**Topology**	**Rule Set**	**Attribute Space**
Model Parameters	**Coefficients**	**Prior Prob.** **Conditional Prob.**	**Biases** **Weights**	**Term sets** **Scaling Factors** **Coefficients**	**Attribute weights** **Similarity parameters**
Reasoning Mechanism	**Solve equations** **- Closed form** **- Approximation**	**Propagation**	**Propagation**	**Node evaluation & Propagation**	**Local Model evaluation & Output combination**
Design Search Method	**First principle** **Energy balance methods** **(Bond Graphs)**	**Manual** **EM** **EA** **…**	**Manual** **EA** **Backpropagation** **Conjugate gradient,** **…**	**Manual** **EA** **Backpropagation** **…**	**Manual** **EA** **…**

ANFIS

Fuzzy Instant Based Model

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

Statistics

(Continuous) CBR

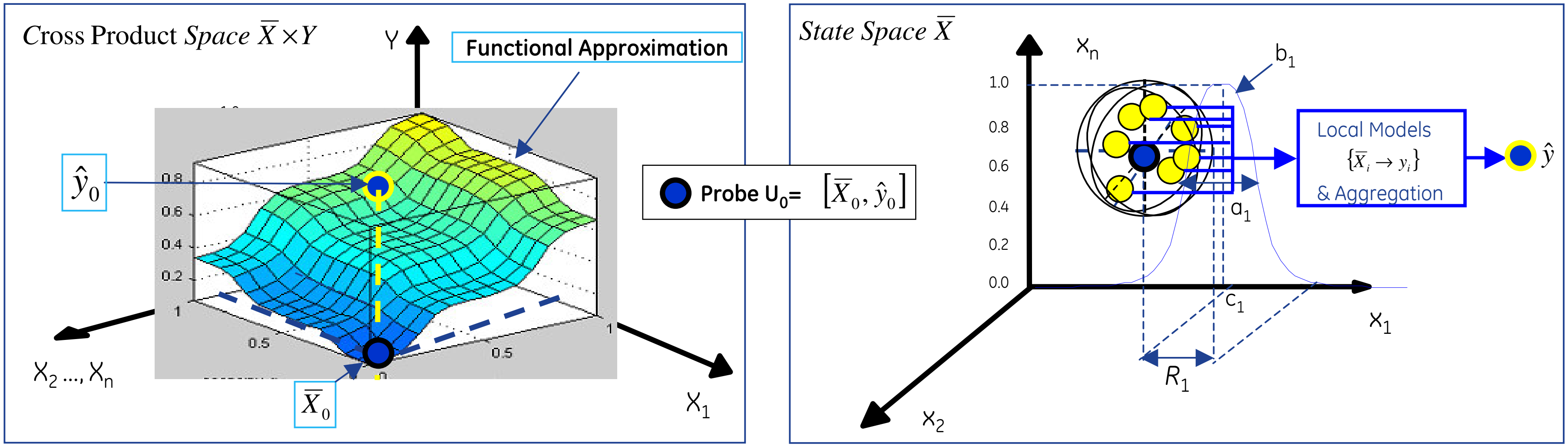
Lazy Learning

Kernel-Based Models

Fuzzy Instance Based Models (F-IBM)

imagination at work

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*

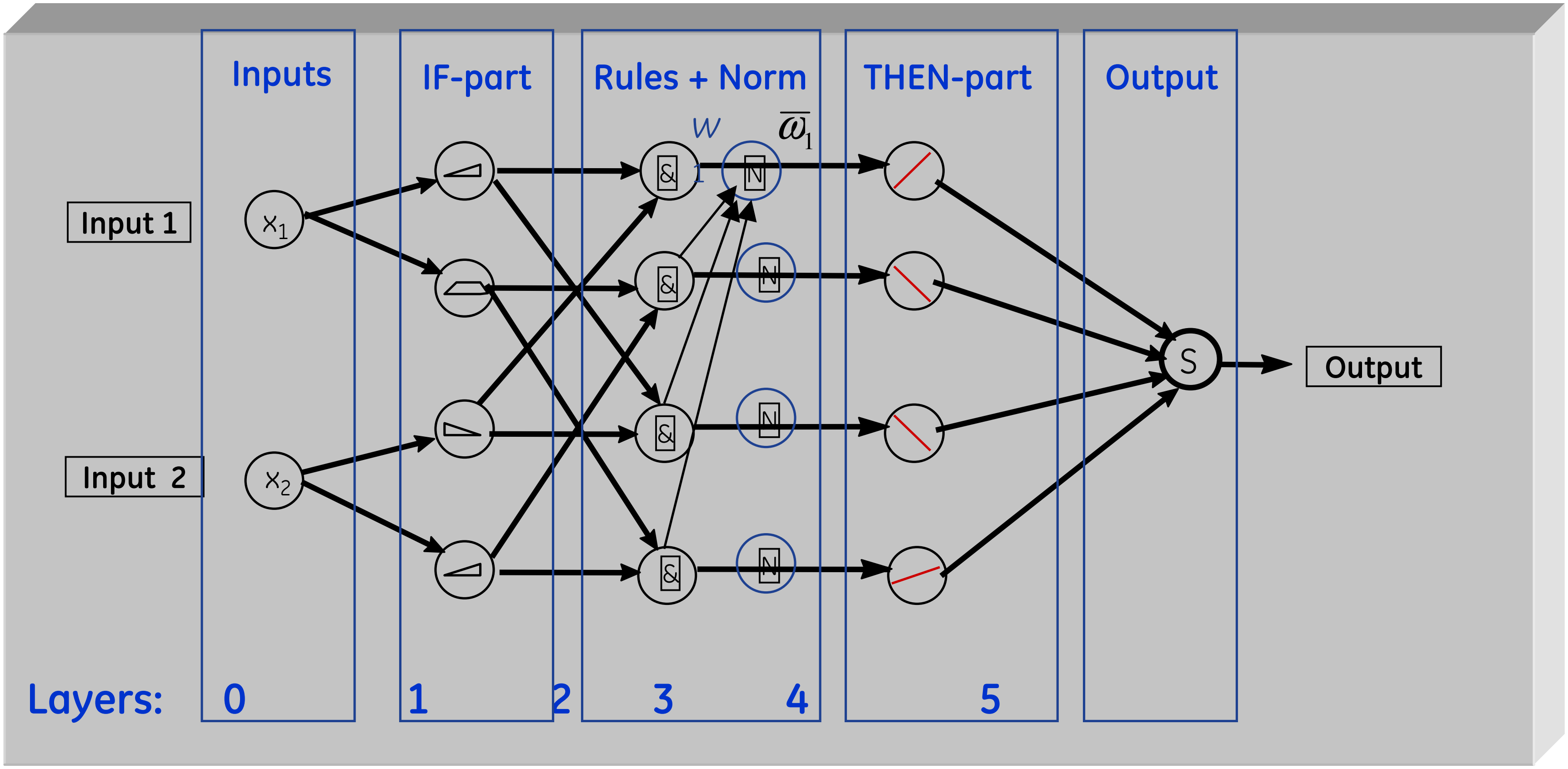
Statistics

Kernel-Based Models

Fuzzy Instance Based Models (F-IBM)

imagination at work

ANFIS Network

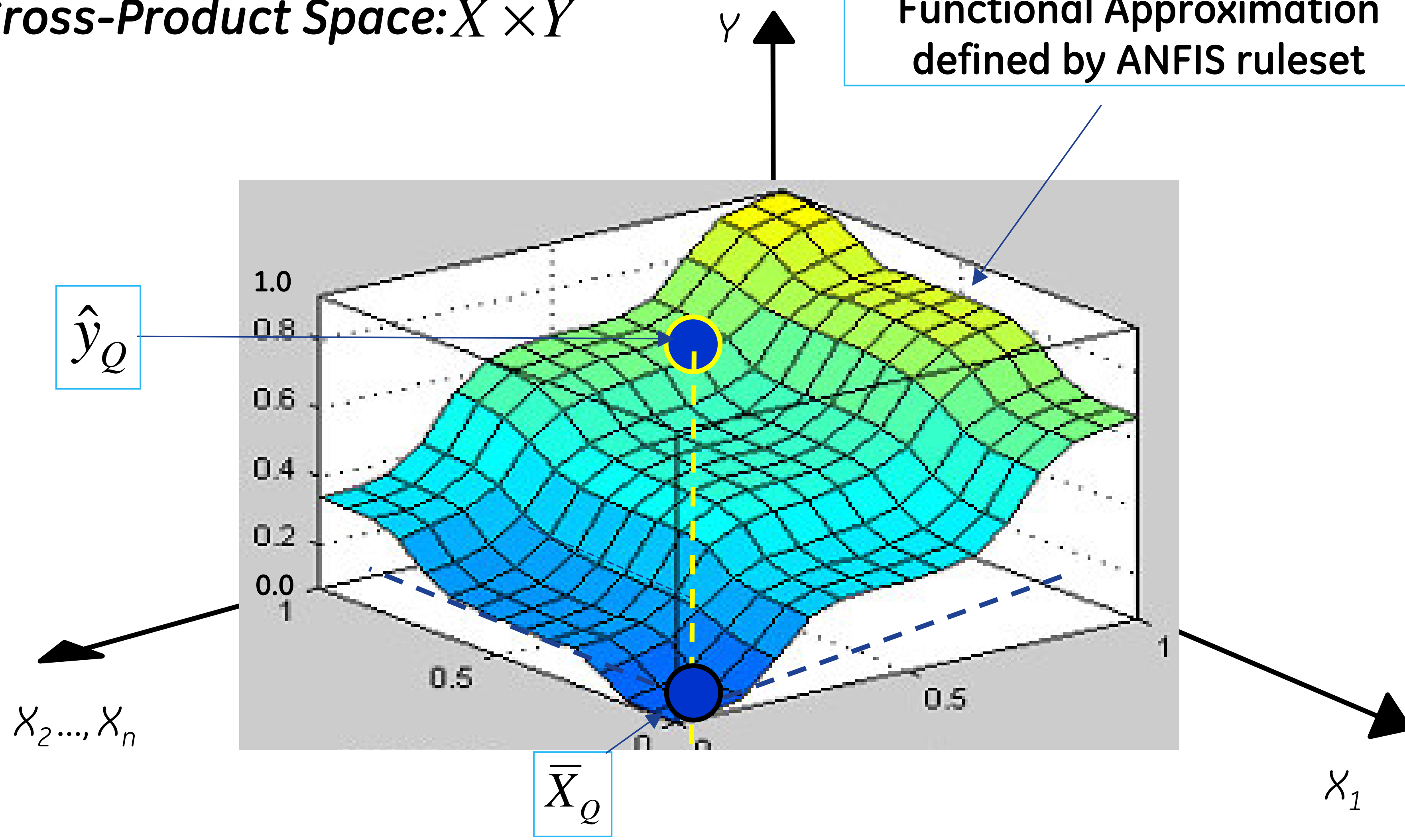


L_0: *Inputs layer nodes:* State variables
L_1: *Values layer nodes:* State variable Termsets, computing the membership value of inputs
L_2: *Rules layer nodes:* Fuzzy Rules, using product to compute rule matching factor w_i
L_3: *Normalization layer nodes:* Each w_i is normalized (to add up to one) generating $\overline{\omega}_i$
L_4: *Function layer nodes*: Linear regressions f_i are evaluated, generating rule outputs y_i that are weighted by normalized rule matching factors $\overline{\omega}_i$
L_5: *Output layer node:* Sum of weighted outputs – (completing weighted average of outputs y_i)

imagination at work

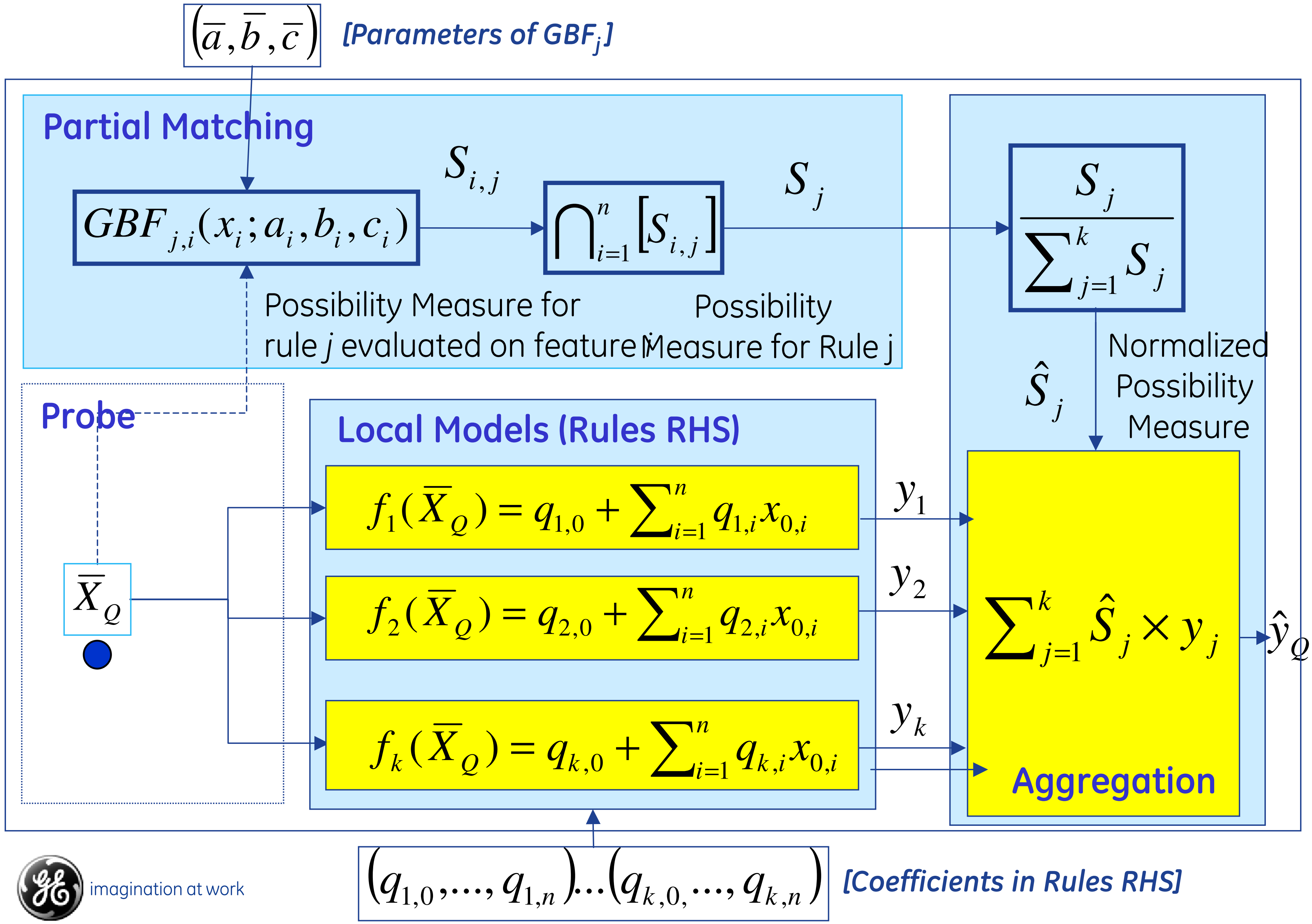
ANFIS Reasoning: Geometrical Interpretation

Cross-Product Space: $\overline{X} \times Y$

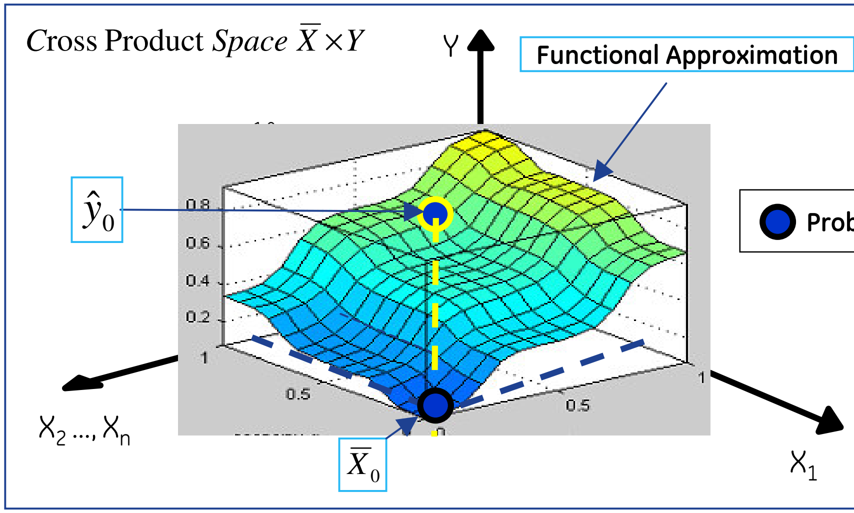


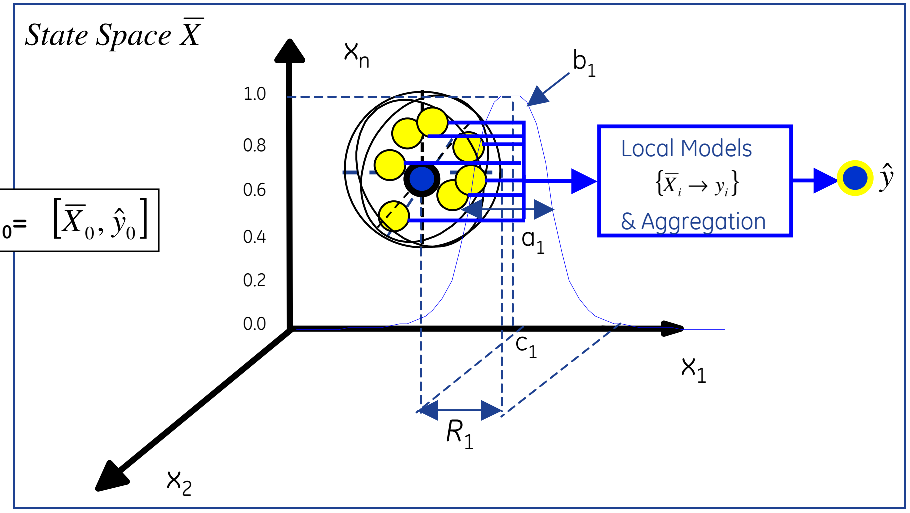
Probe Q = $\left[\overline{X}_Q, \hat{y}_Q\right]$

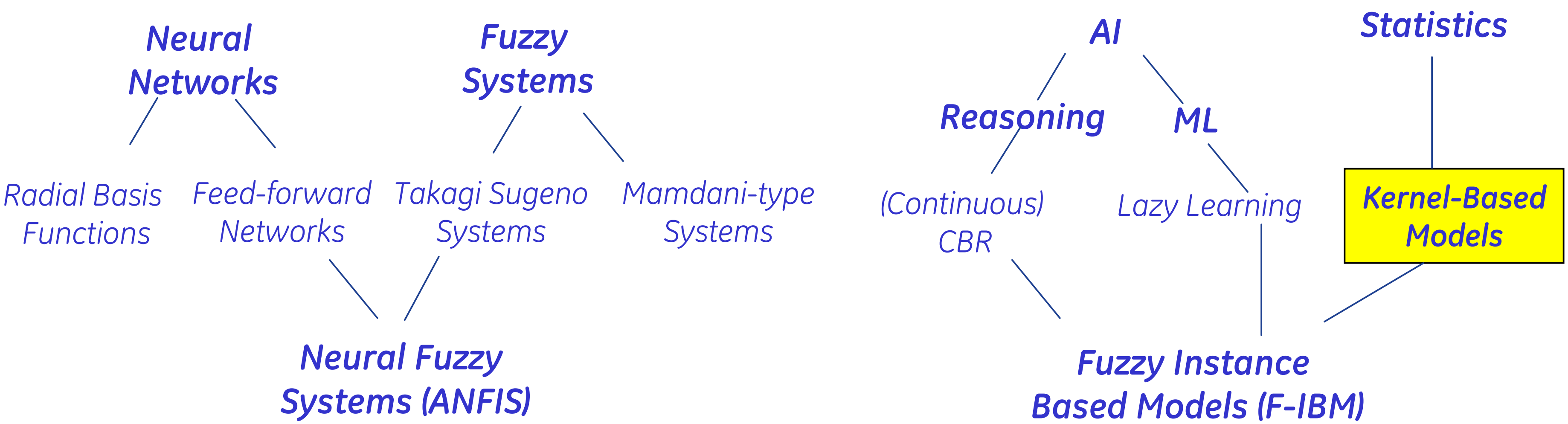
ANFIS Reasoning: Algebraic Interpretation



Functional Approximation vs. Instance-Based







To FIBM

imagination at work

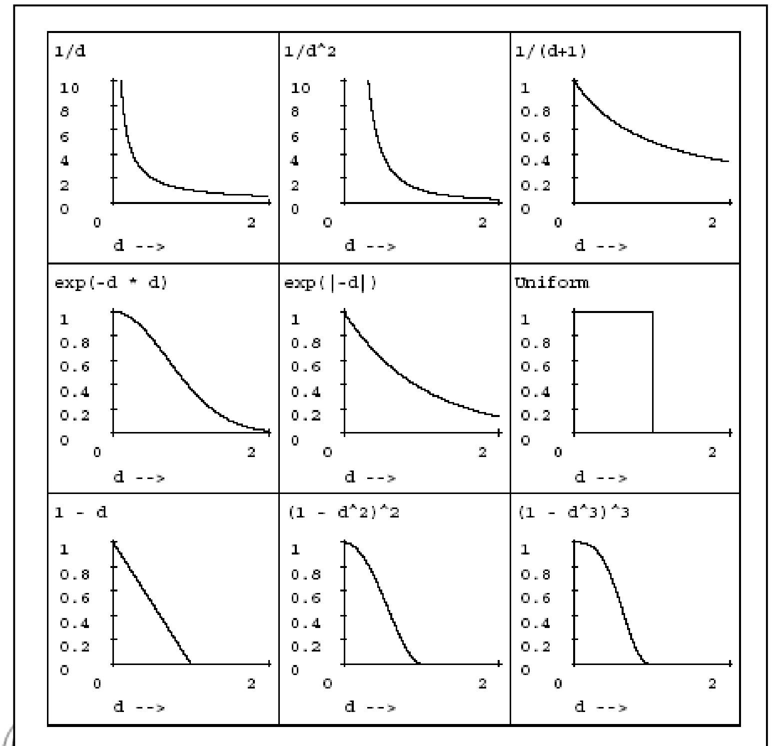
Kernel-based Models

Classical Kernel-based Models

We have:

- a collection of points $\{u_j\}$: $u_j = [\bar{X}_j; y_j] = [x_{1,j}, x_{2,j}, \ldots, x_{n,j}; y_j]$ where $j = 1, \ldots, k$
- a Query Q defined in the $\bar{X}$ space $Q = [\bar{X}_Q; \hat{y}_Q] = [x_{1,Q}, x_{2,Q}, \ldots, x_{n,q}; ?]$ where $|\bar{X}_Q| = |\bar{X}_j| = n$
- a Kernel function of the distance between each data point and the query, and a smoothing parameter h $K\left(\frac{d(\bar{X}_j, \bar{X}_Q)}{h}\right)$

Examples of Kernel Functions *K*



Examples of Distances *d*
(from Atkeson, Moore, Schaal, 1996)

- **Unweighted Euclidean distance:**

$$d_E(\mathbf{x}, \mathbf{q}) = \sqrt{\sum_j (\mathbf{x}_j - \mathbf{q}_j)^2} = \sqrt{(\mathbf{x} - \mathbf{q})^T(\mathbf{x} - \mathbf{q})} \quad (24)$$

- **Diagonally weighted Euclidean distance:**

$$d_m(\mathbf{x}, \mathbf{q}) = \sqrt{\sum_j (m_j(\mathbf{x}_j - \mathbf{q}_j))^2} = \sqrt{(\mathbf{x} - \mathbf{q})^T\mathbf{M}^T\mathbf{M}(\mathbf{x} - \mathbf{q})} = d_E(\mathbf{Mx}, \mathbf{Mq}) \quad (25)$$

where m_j is the feature scaling factor for the jth dimension and $\mathbf{M}$ is a diagonal matrix with $\mathbf{M}_{jj} = m_j$.

- **Fully weighted Euclidean distance:**

$$d_\mathbf{M}(\mathbf{x}, \mathbf{q}) = \sqrt{(\mathbf{x} - \mathbf{q})^T\mathbf{M}^T\mathbf{M}(\mathbf{x} - \mathbf{q})} = d_E(\mathbf{Mx}, \mathbf{Mq}) \quad (26)$$

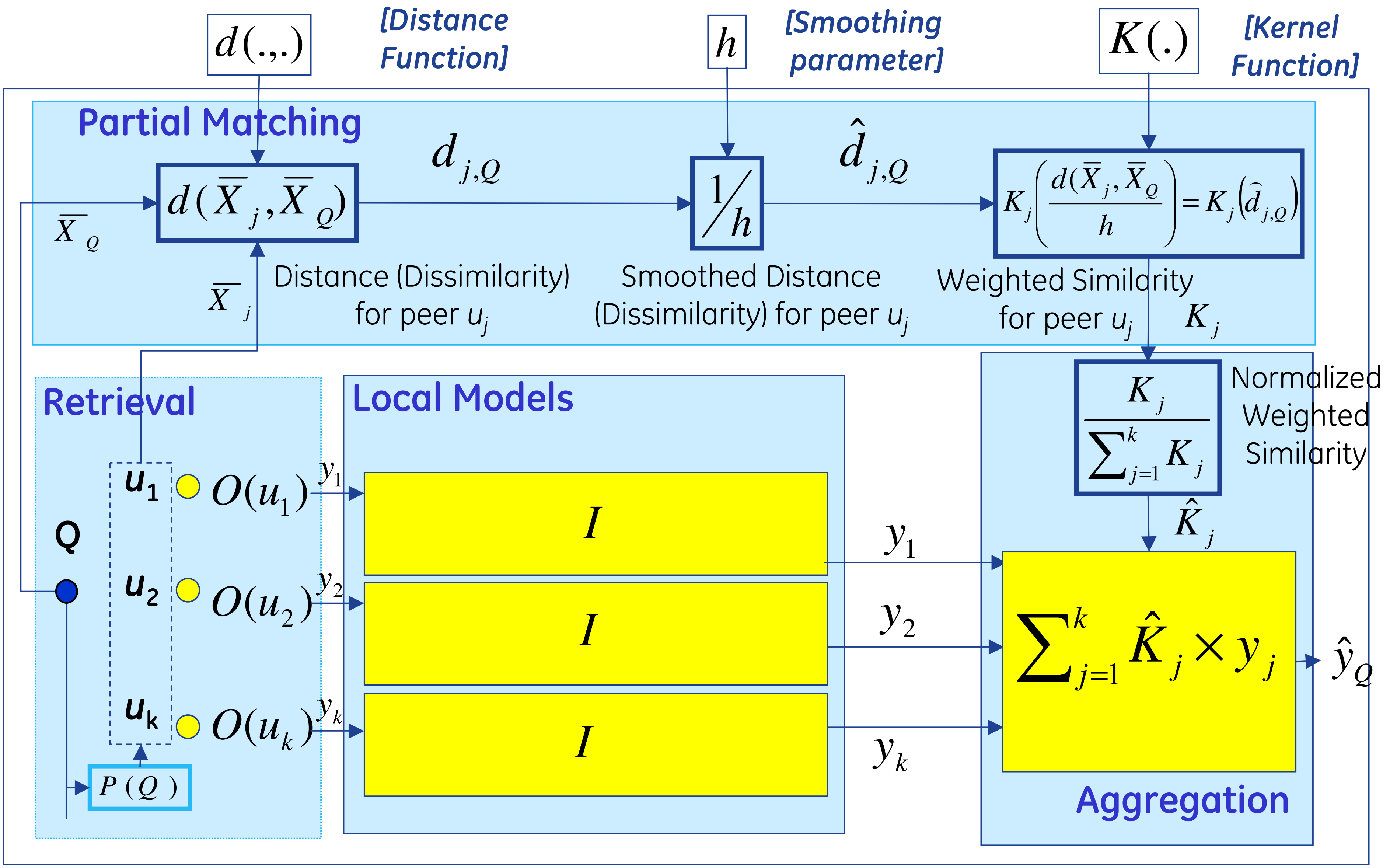
where $\mathbf{M}$ is no longer diagonal but can be arbitrary. This is also known as the Mahalanobis distance (Tou and Gonzalez, 1974; Weisberg, 1985).

- **Unweighted L_p norm (Minkowski metric):**

$$d_p(\mathbf{x}, \mathbf{q}) = \left(\sum_i |\mathbf{x}_i - \mathbf{q}_i|^p\right)^{\frac{1}{p}} \quad (27)$$

- **Diagonally weighted and fully weighted L_p norm:** The weighted L_p norm is $d_p(\mathbf{Mx}, \mathbf{Mq})$.

Kernel-based Models: Nadaraya-Watson Estim.



Kernel-based Models: Weighted Regression

Kernel-based Regressions

Data Weighting Process:

$$w_j = \sqrt{K\left(\frac{d(\overline{X}_j, \overline{X}_Q)}{h}\right)}$$

- Use the Kernel function (evaluated at each point u_j) as a weight
- Create a diagonal *[k x k]* matrix ***W*** with diagonal elements $W_{jj} = w_j$ and zeros elsewhere
- Create a *[k x n]* matrix ***X*** with the original state data. The i^{th} row of ***X*** contains the *n* coordinates of the i^{th} point, i.e., $\overline{X}_i$
- Create a *[k x 1]* vector ***Y*** with the original output data. The i^{th} row of ***Y*** contains the value of y_i
- Weigh matrix ***X*** and vector ***Y*** using the weights ***W:***

Z = WX and ***V=WY*** ***Z is*** *[k x n] and* ***V*** *is [k x 1]*

- We want to solve the equation:

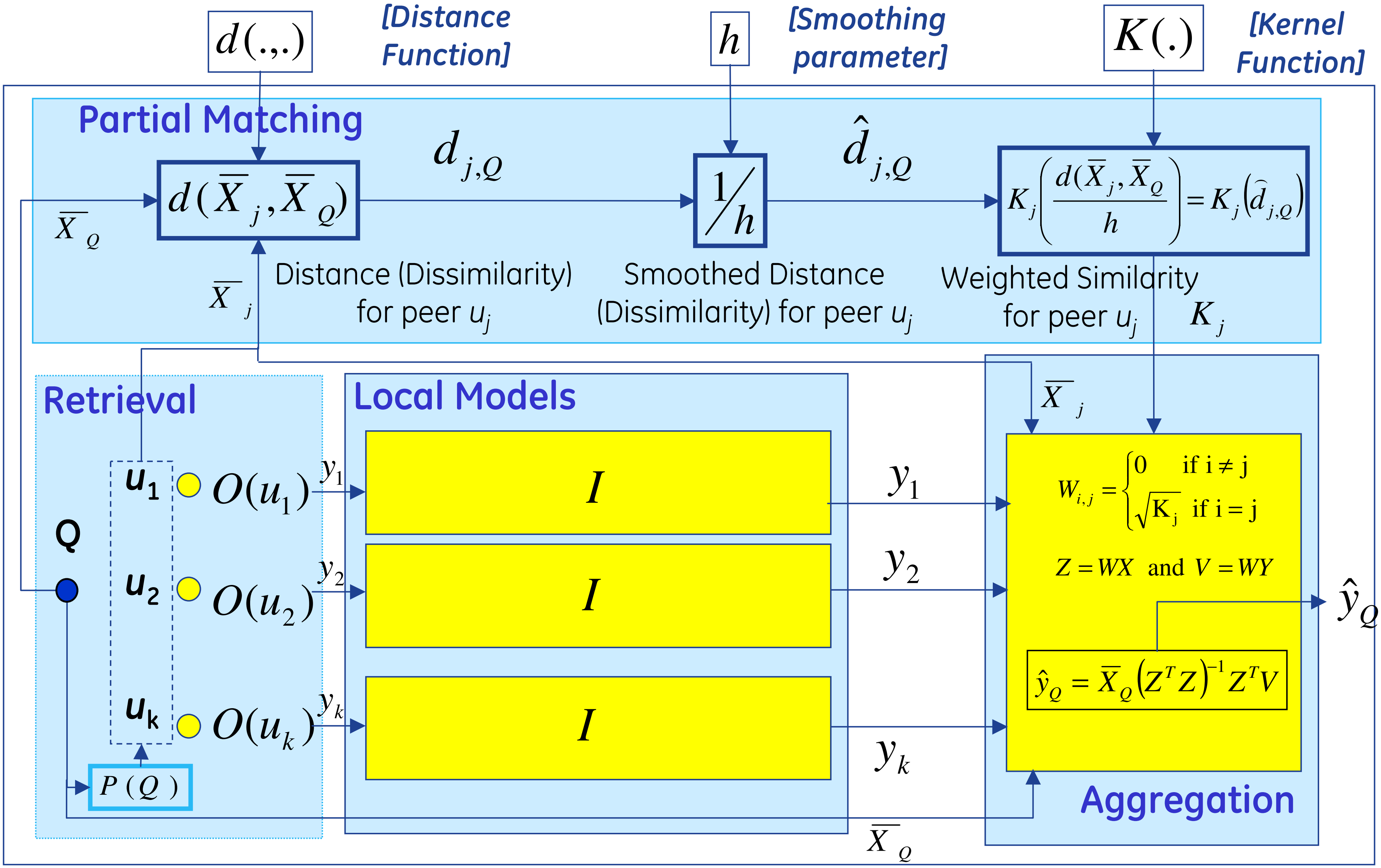
$$\boxed{\hat{y}(Q) = \overline{X}_Q\left(Z^T Z\right)^{-1} Z^T V}$$

which is predicated on having a non-singular *[n x n]* matrix: $\left(Z^T Z\right)^{-1}$

- When the matrix $\left(Z^T Z\right)^{-1}$ is singular we can resolve the issue using ridge regressions

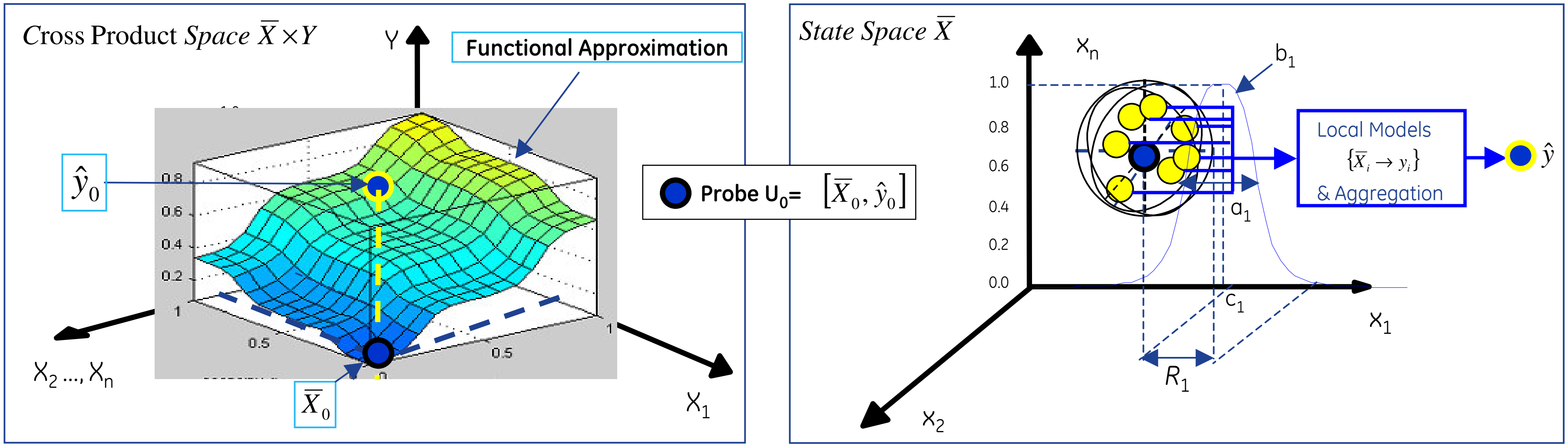
imagination at work

Kernel-based Models: Weighted Regression



imagination at work

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*

Statistics

(Continuous) CBR — *Lazy Learning* — *Kernel-Based Models*

Fuzzy Instance Based Models (F-IBM)

4. Fuzzy Instance Based Model (F-IBM)

- Retrieval
- Similarity Evaluation
- Creation of Local Models
- Outputs Aggregation
- Evolutionary Search for Designing a F-IBM

imagination at work

Fuzzy Instance Based Models (F-IBM's)

Comparison with Case Based Reasoning

- IBM's rely on a collection of previously experienced data that can be store in their raw representation
- Unlike Case-based Models (CBM's), IBM's ***do not need*** to have data refined, abstracted and ***organized as cases***
- Like CBM's, IBM's are based on analogical reasoning, as they rely upon finding previous instances of ***similar*** objects (or points) and use them to create an ensemble of local models

Fuzzy Instance Based Models (F-IBM's)

Comparison with Case Based Reasoning

- IBM's rely on a collection of previously experienced data that can be store in their raw representation
- Unlike Case-based Models (CBM's), IBM's ***do not need*** to have data refined, abstracted and ***organized as cases***
- Like CBM's, IBM's are based on analogical reasoning, as they rely upon finding previous instances of ***similar*** objects (or points) and use them to create an ensemble of local models

Similarity Measure

- The definition of **similarity** plays a **critical** role in IBM's performance
- Similarity will be a **dynamic** concept and will change over time. Therefore, it is important to apply **learning** methodologies **to define it and adapt it.**
- Furthermore, the concept of similarity is not crisply defined, creating the need to allow for some degree of **vagueness** in its evaluation

Fuzzy Instance Based Models (F-IBM's)

Comparison with Case Based Reasoning

- IBM's rely on a collection of previously experienced data that can be store in their raw representation
- Unlike Case-based Models (CBM's), IBM's ***do not need*** to have data refined, abstracted and ***organized as cases***
- Like CBM's, IBM's are based on analogical reasoning, as they rely upon finding previous instances of ***similar*** objects (or points) and use them to create an ensemble of local models

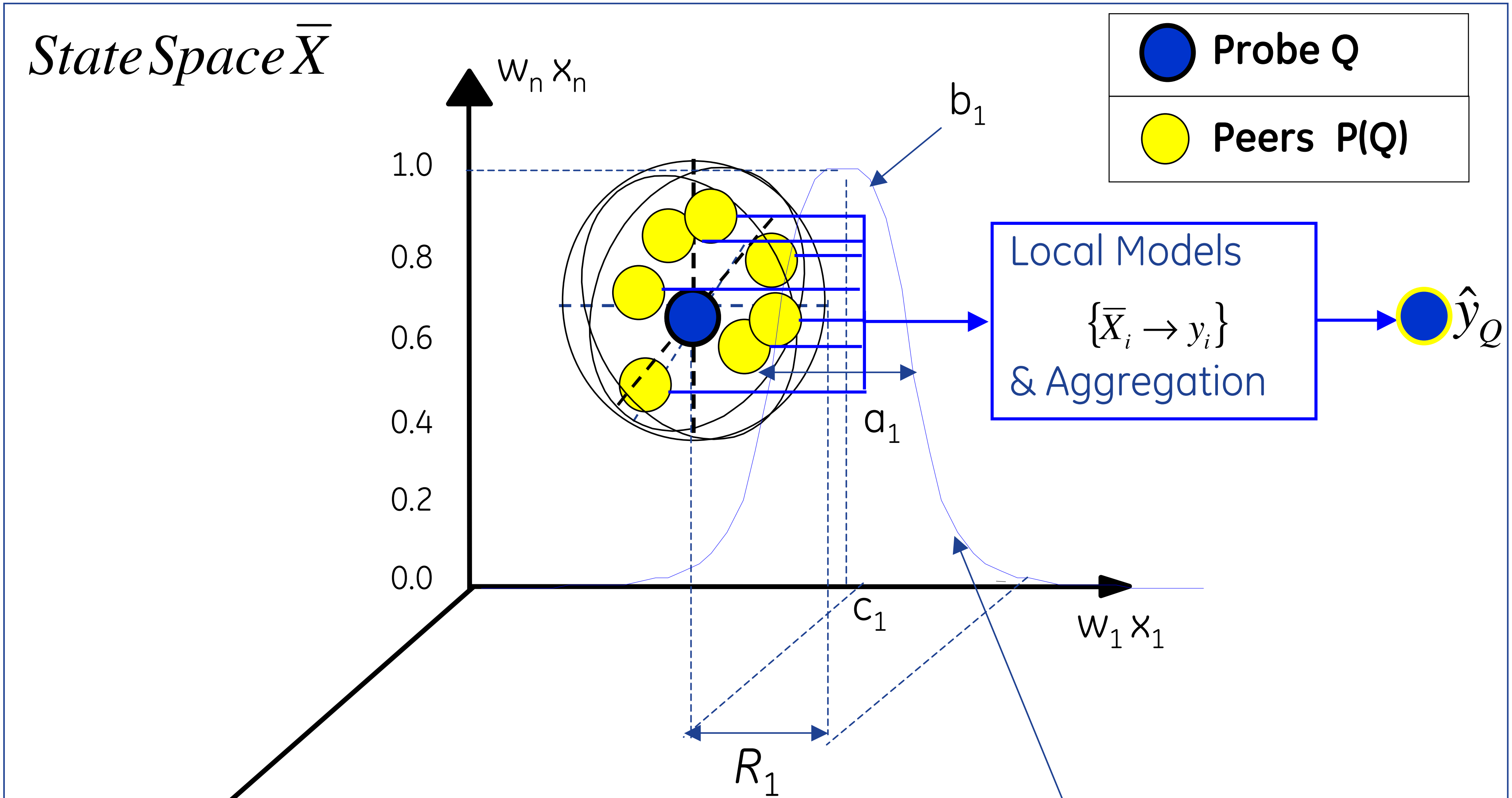
Similarity Measure

- The definition of **similarity** plays a **critical** role in IBM's performance
- Similarity will be a **dynamic** concept and will change over time. Therefore, it is important to apply **learning** methodologies **to define it and adapt it.**
- Furthermore, the concept of similarity is not crisply defined, creating the need to allow for some degree of **vagueness** in its evaluation

Solution:

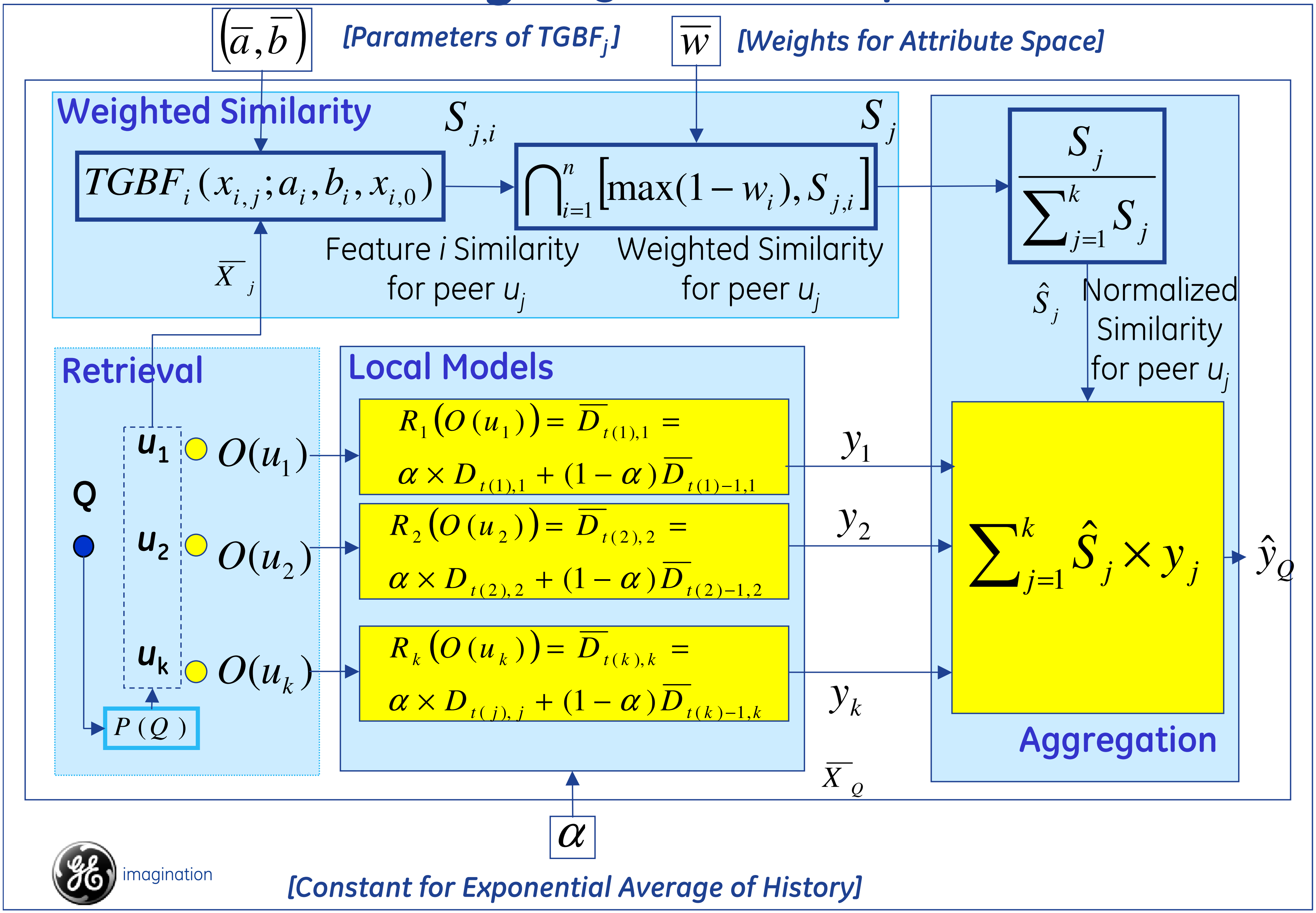
- By using a ***wrapper*** approach, we ***evolve*** the ***design of the similarity*** function and the ***design of the attribute space*** in which the similarity is to be evaluated.

F-IBM Reasoning: Geometrical Interpretation



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

F-IBM Reasoning: Algebraic Interpretation



Fuzzy Instance Based Models (F-IBM's)

We address this design issues by evolving the design of a similarity function in conjunction with the design of the attribute space in which similarity is evaluated. Specifically we us the following four steps:

1) Retrieval of similar instances from the Data Base

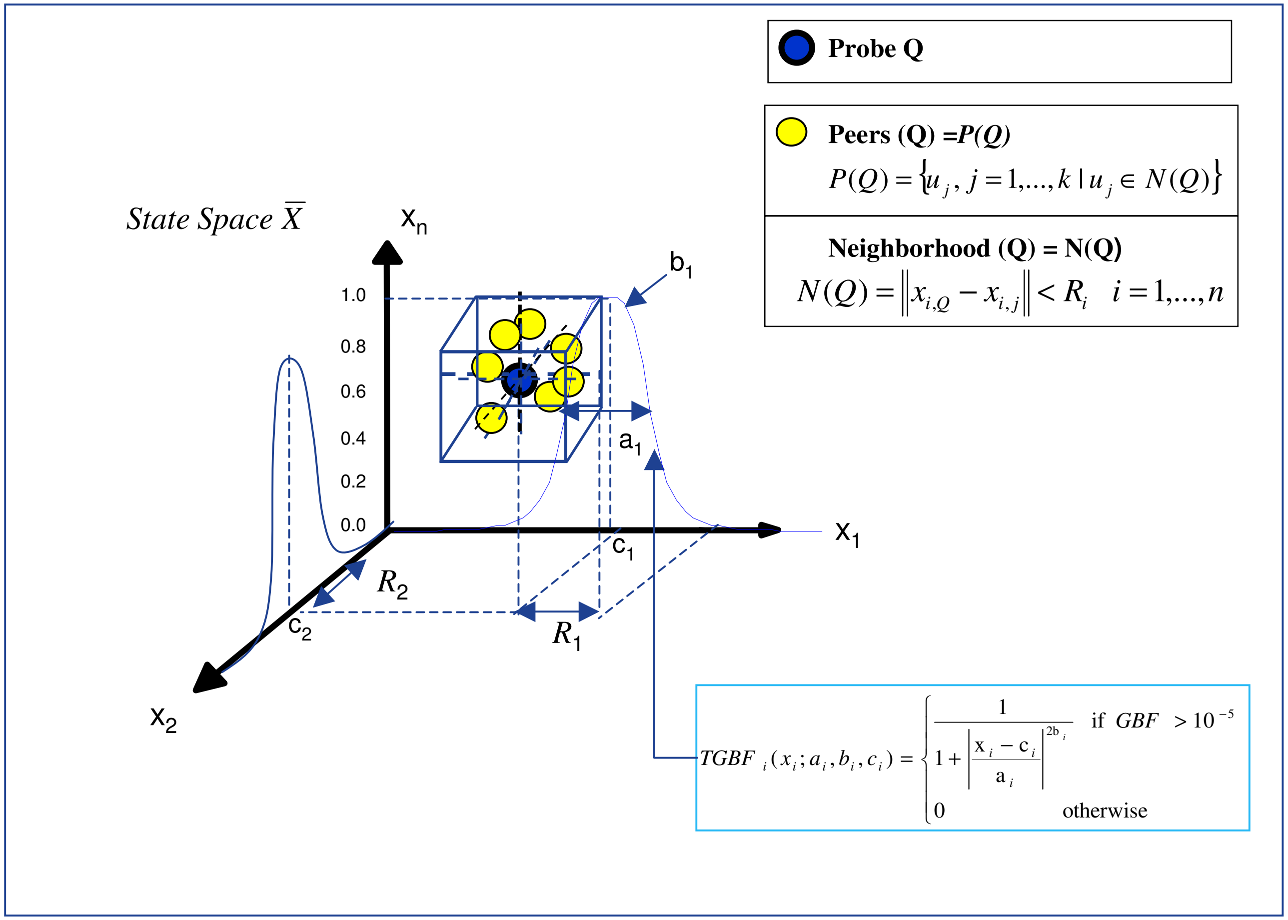
2) Evaluation of similarity *measure* between the probe and the retrieved instances

3) Creation of local models using the most similar instances

4) Outputs Aggregation (weighted by their similarities)

Let us explain the four steps

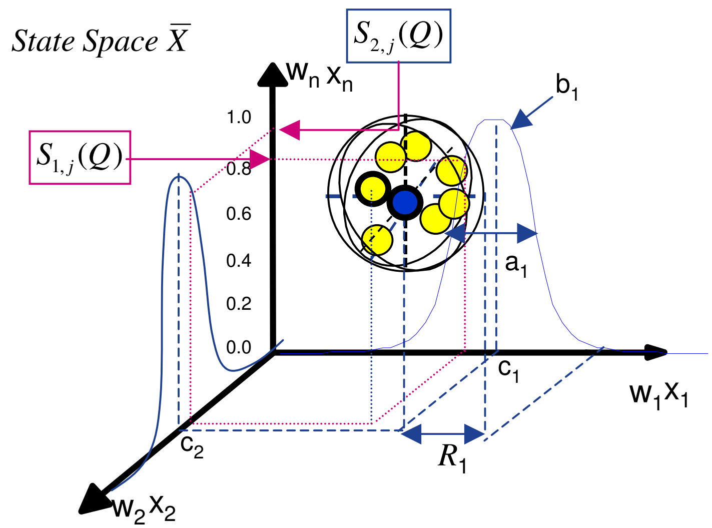
(1) Retrieval



Representation	Feature Space	History	
Probe $Q = [\overline{X}_Q; O(Q)]$	$\overline{X}_Q = [x_{1,Q},\dots,x_{n,Q}]$	**History of Probe**	$O(Q) = [D_{1,Q},\dots,D_{t(Q),Q}]$
Peer $u_j = [\overline{X}_j; O(u_j)]$	$\overline{X}_j = [x_{1,j},\dots,x_{n,j}]$	**History of Peer**	$O(u_j) = [D_{1,j}, D_{2,j},\dots,D_{t(j),j}]$

imagination at work

(2) Similarity Evaluation



Probe Q

Peers (Q) =*P(Q)*

$P(Q) = \{u_j, j = 1,\ldots,k \mid u_j \in N(Q)\}$

Neighborhood (Q) = N(Q)

$N(Q) = \left\| x_{i,Q} - x_{i,j} \right\| < R_i \quad i = 1,\ldots,n$

Peer u_j $u_j \in P(Q)$

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

S_j* = Similarity between Probe Q and Peer *u_j

$$S_j = Min_{i=1}^{n}\{S_{i,j}\} = Min_{i=1}^{n} TGBF(x_{i,j};a_i,b_i,x_{i,Q})$$

S_j* = Weighted Similarity between Probe Q and Peer *u_j

$$S_j = Min_{i=1}^{n}\{Max[(1-w_i), S_{i,j}]\} = Min_{i=1}^{n}\{Max[(1-w_i), TGBF_i(x_i;a_i,b_i,x_{i,Q})]\}$$

(3) Creation of Local Models

Local Models

Rather than using a pre-constructed model, as in ANFIS or other functional approximators, we use local models, as in memory-based approaches, kernel-based regressions, and lazy-learning.

In this example, each retrieved object u_j has an associated time-series:

$$O(u_j) = [D_{1,j}, D_{2,j}, \ldots, D_{t(j),j}]$$

One of the simplest local models that we can create is the ***exponential average***, that will "discount" the oldest data, using a forgetting factor a.

The peer u_j of the probe will produce an output y_j representing the prediction for the next point in the time-series:

$$\begin{aligned} y_j &= D_{t(j)+1,j} \\ &= \overline{D}_{t(j),j} = \alpha \times D_{t(j),j} + (1-\alpha) \times \overline{D}_{t(j)-1,j} \quad [\text{where } \overline{D}_{1,j} = D_{1,j}] \\ &= (1-\alpha)^{k(j)-1} D_{1,j} + \sum_{i=2}^{t(j)} (1-\alpha)^{t(j)-i} \times \alpha \times D_{i,j} \end{aligned}$$

imagination at work

(4) Outputs Aggregation

Outputs Aggregation *(weighted by their similarities)*

We need to combine the individual predictions $\{D_{t(j)+1,j}\}$ $(j=1,\ldots,k)$ obtained from the peers $u_j(Q)$ to generate the next prediction $D_{Next,Q}$ for the probe Q

To this end, we compute the weighted average of the peers' individual predictions using their normalized similarity to the probe as a weight:

$$\hat{y}_Q = D_{Next,Q} = \frac{\sum_{j=1}^{k} S_j \times D_{t(j)+1,j}}{\sum_{j=1}^{k} S_j} = \frac{\sum_{j=1}^{k} S_j \times y_j}{\sum_{j=1}^{k} S_j}$$

If we define the normalized weights as: $\hat{S}_j = \frac{S_j}{\sum_{j=1}^{k} S_j}$

then the above expression can be rewritten as: $\hat{y}_Q = \sum_{j=1}^{k} \hat{S}_j \times y_j = \langle \hat{S}, Y \rangle$

imagination at work

(4) Outputs Aggregation (cont.)

Outputs Aggregation *(weighted by their similarities)*

$$\hat{y}_Q = D_{Next,Q} = \frac{\sum_{j=1}^{k} S_j \times y_j}{\sum_{j=1}^{k} S_j}$$

Note that this expression (a convex sum of local models outputs using the similarities as weights) is similar to the structure of the ***Nadaraya-Watson*** *[1964] estimator for non-parametric regressions using locally weighted averages – where the weights are the values of a kernel function K:*

$$\hat{y}_Q = \frac{\sum_{j=1}^{k} K(x - x_j) \times y_j}{\sum_{j=1}^{k} K(x - x_j)}$$

From this analogy, we can see a structural similarity between the Similarity measures S_j used by the Instance-based method and the Kernel functions K evaluated on the distance between the probe and each point, i.e.:

$$S_j \approx K(x - x_j)$$

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 $[w_1, w_2, \ldots, w_n]$.

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.

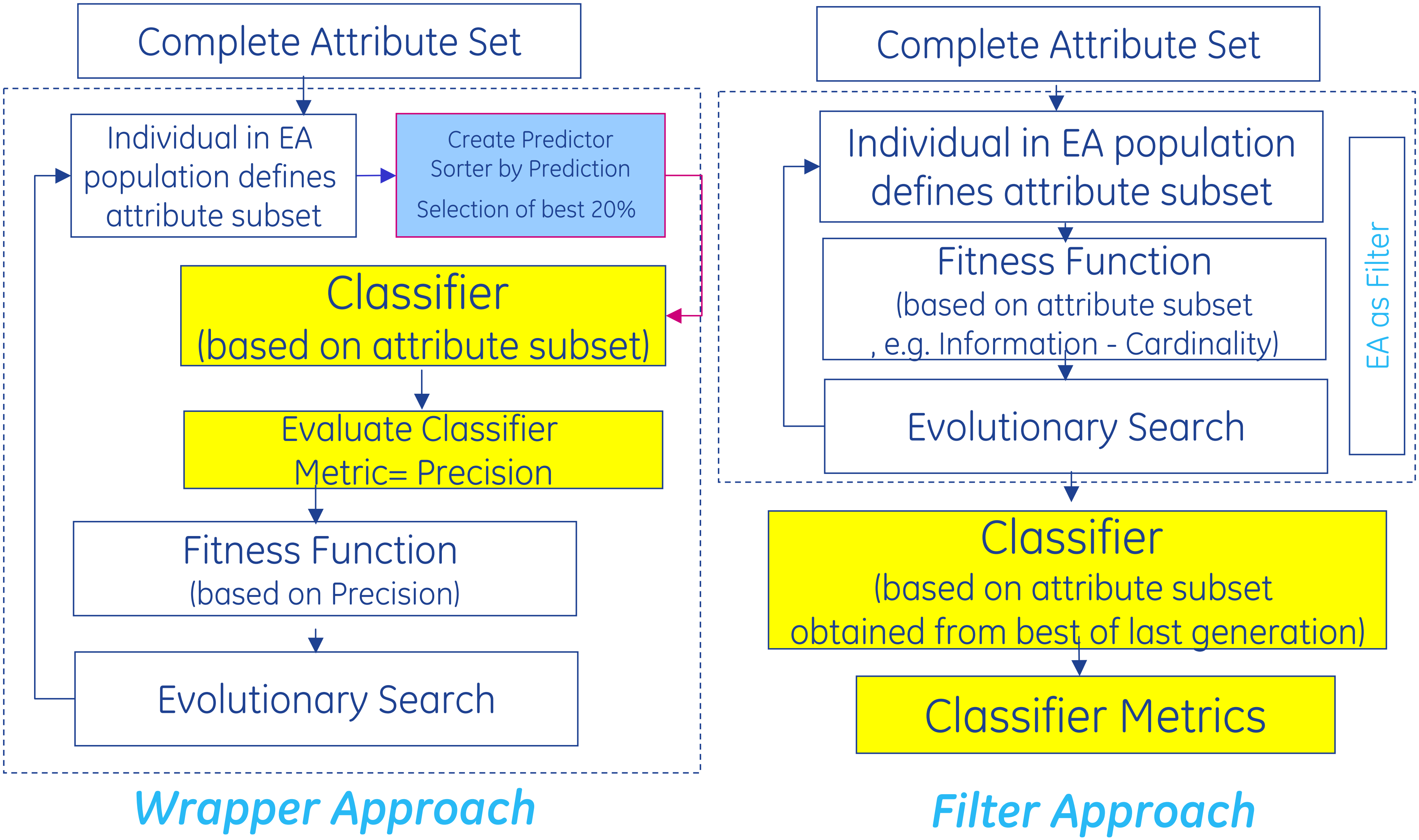
If $w_i \in [0,1]$ we perform *attribute weighting*, i.e., we define a fuzzy subset from the universe of potential attributes

Chromosome representation:

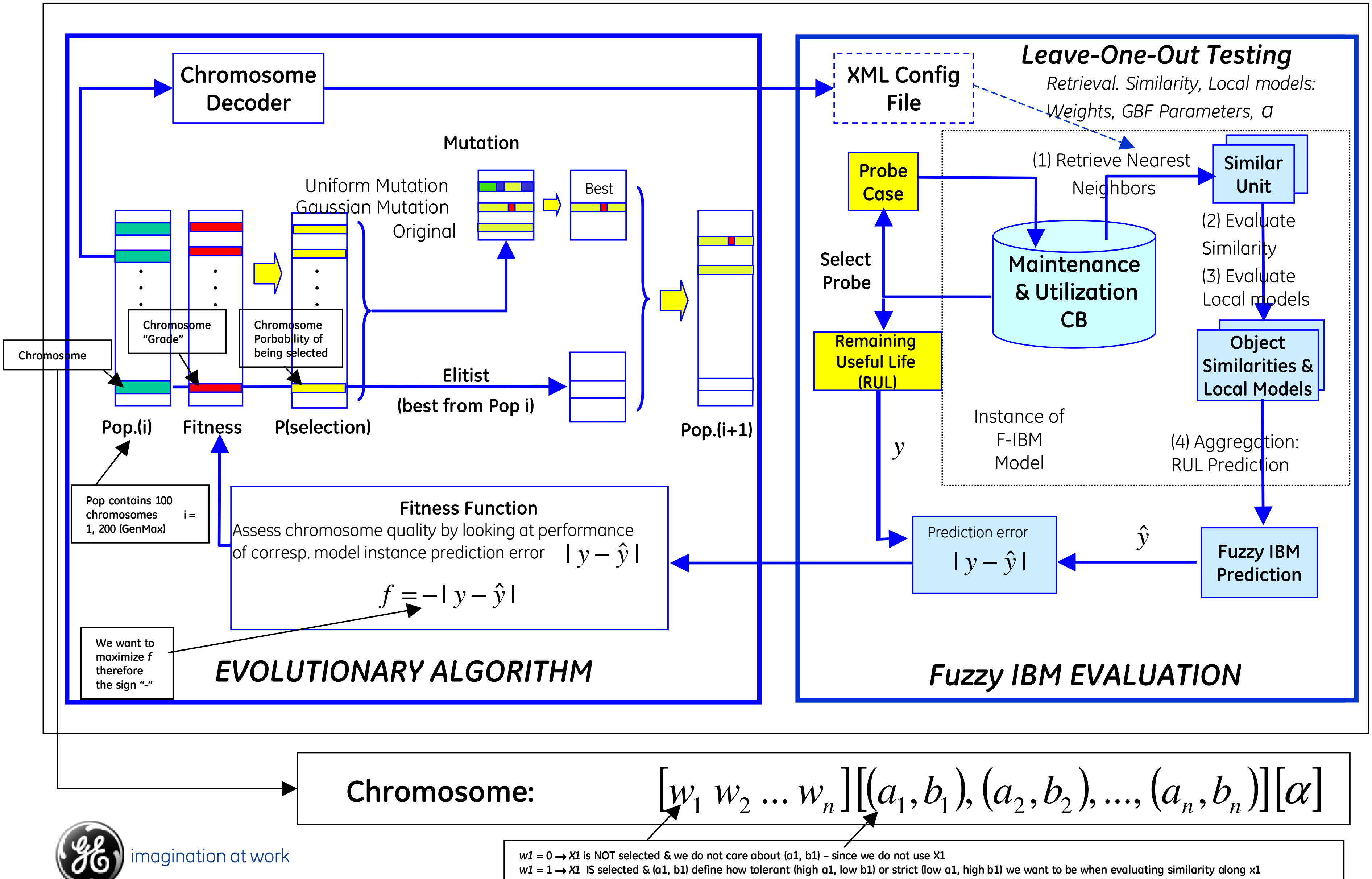
$$[w_1\ w_2\ \ldots\ w_n][(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)][\alpha]$$

w1 = 0 → *X1* is NOT selected & we do not care about (a1, b1) – since we do not use X1
w1 = 1 → *X1* IS selected & (a1, b1) define how tolerant (high a1, low b1) or strict (low a1, high b1) we want to be when evaluating similarity along x1

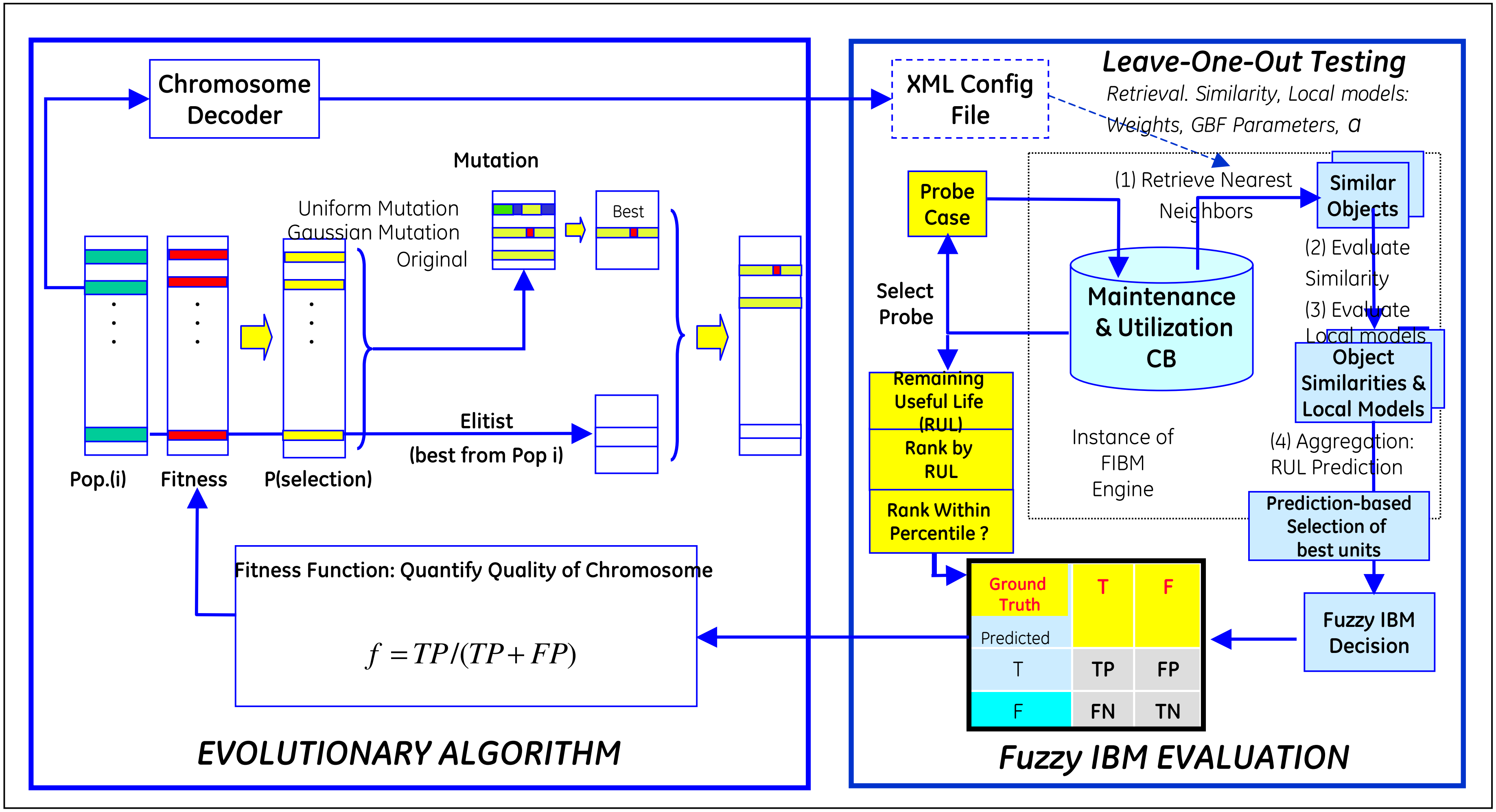
Wrapper and Filter Approaches



Evolutionary Search for Designing A Fuzzy IBM (Predictor) using a Wrapper Approach



Evolutionary Search for Designing A Fuzzy IBM (Classifier) using a Wrapper Approach

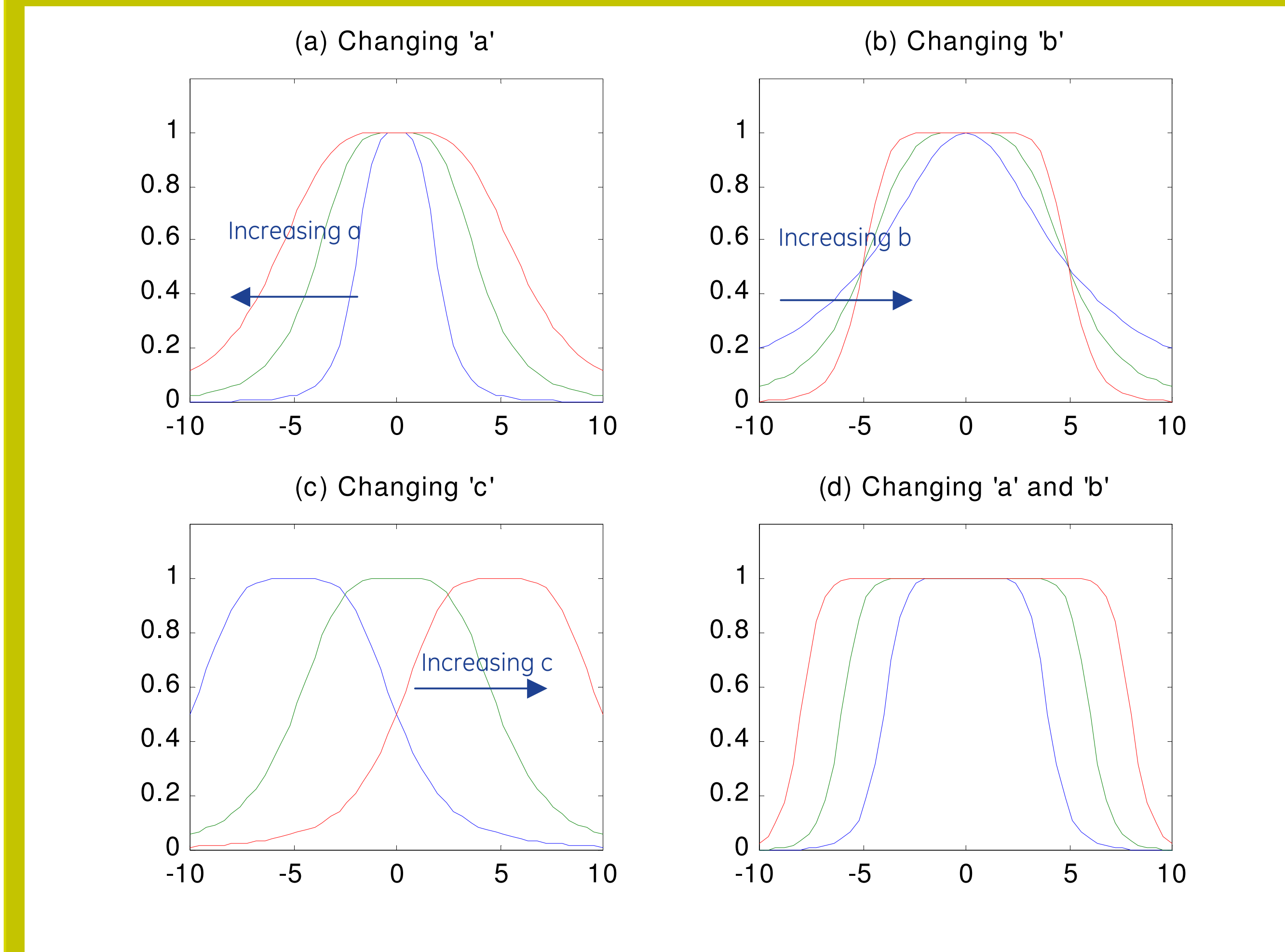


Chromosome: $[w_1\ w_2\ \ldots\ w_n][(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)][\alpha]$

To Example

Effect of Changing Parameters {a,b,c}

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b}}$$



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

5. Case Study: Prediction of Best Units for Asset Selection

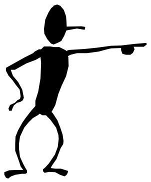
- Asset Selection for Mission Reliability: A locomotive example
- Data Collection, Baselines, and Experiments
- Peer Learning Methodology
- Results & Conclusions

imagination at work

Asset Selection Outline

- Problem Description
- Data Collection and Experiments Set-Up
- Proposed Peer-Learning Methodology
- Results
- Conclusions

Asset Selection Outline



- **Problem Description**
  - *Mission Reliability*
  - *The Selection Problem*
- Data Collection and Experiments Set-Up
- Proposed Peer-Learning Methodology
- Results
- Conclusions

The Premise

New learning approaches can address early phases of new platform deployment with limited operational and sensor data.

-Learning using **existing utilization and maintenance history** improves our ability to select the best units for a mission, leading to better utilization of existing assets.

-**Learning from peers** is a robust approach to rapid learning from limited data.

-**Evolutionary Learning** provides a comprehensive framework for automating the design and maintenance of a classifier used for selection of the best assets.

Asset Selection Based on Predicted Life

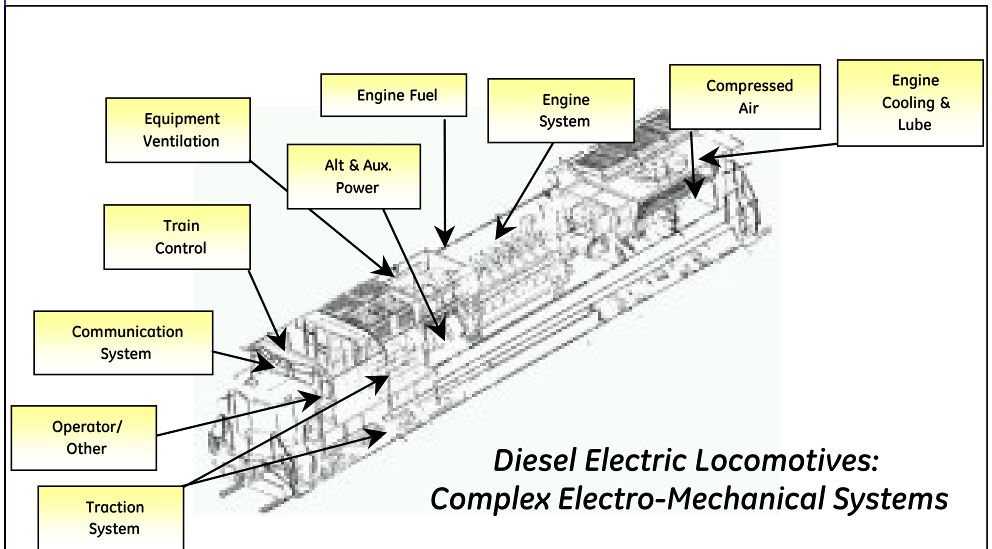
Months

Weeks

Days

Half Day

Collective Mind for Equipment Prognostics and Pulse Reliability

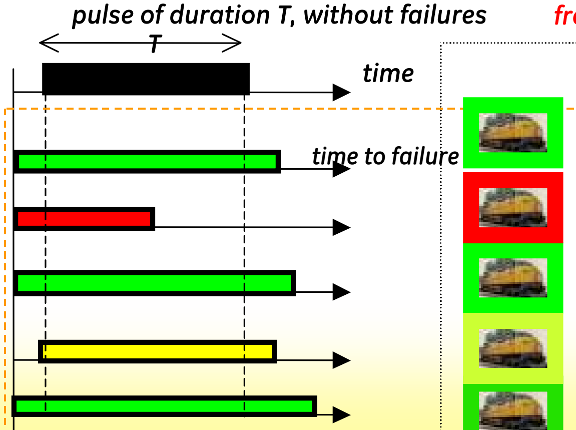


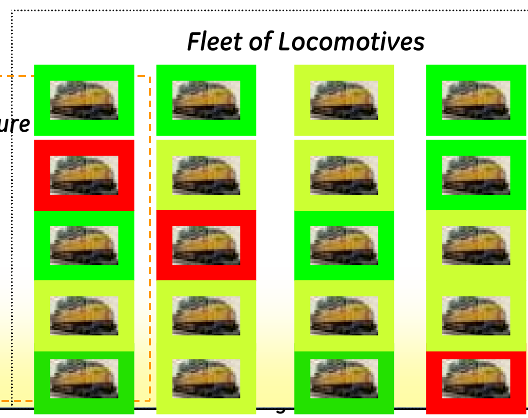
Diesel Electric Locomotives: Complex Electro-Mechanical Systems

- ***Locomotive Fleet***
  - 7,000 units (60% with remote monitoring and satellite uplink) to provide time-stamped failure codes and state info.
- ***Fleet, Units, and Collective***
  - Fleet is a collection of points in a sparsely populated, high dimensional feature space
  - Each point (unit in the fleet) has a record in a DB, describing model, configuration, operational and maintenance information
  - Each point has a collective of peers
  - *Collective of peers* are units whose individual experience can be composed to better predict the unit's performance
  - To adapt to changes in units or peers, we will re-assess the collective over time

Pulse Reliability Problem *: Select best N units from fleet to perform during an operational pulse of duration T, without failures*

Approach *: Predict individual units' time-to-failure based on unit track record and peers from collective*

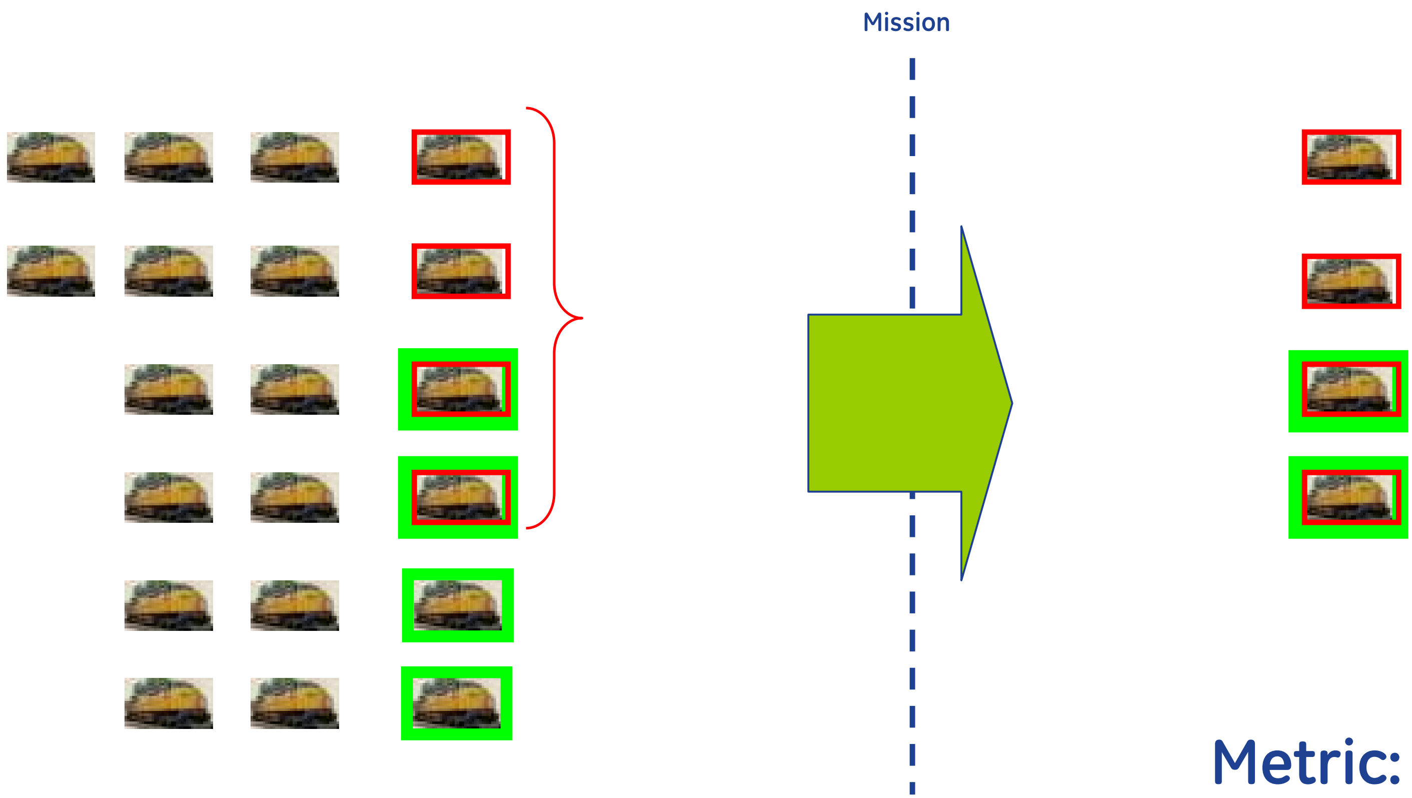




Sensor Data
Utilization Data
Maintenance Logs
Repair Data

Fuzzy + EA

Metric of Success: Selection Precision



20 units

Pick best 20% = 4 units

Metric:

Precision: $\dfrac{\#\text{actual best}}{\#\text{selected}}$

= 50% (in this case)

imagination at work

Asset Selection Outline

- Problem Description
- **Data Collection and Experiments Set-Up**
  - *Data Sources*
  - *Experiments: Metrics and Baselines*
- Proposed Peer-Learning Methodology
- Results
- Conclusions

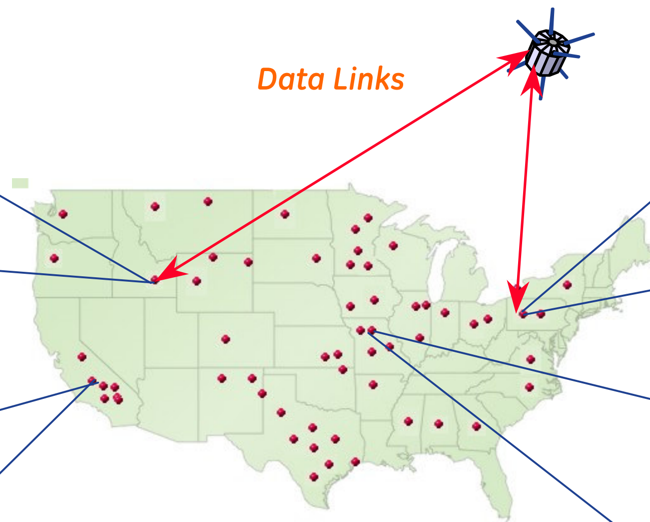
Data Compilation and Experiments

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)*

Locomotives with EOA Service

Fault codes
(3 uploads/day)



GE Rail Locomotives Services

Recommendations
(4-8 Rx/yr)

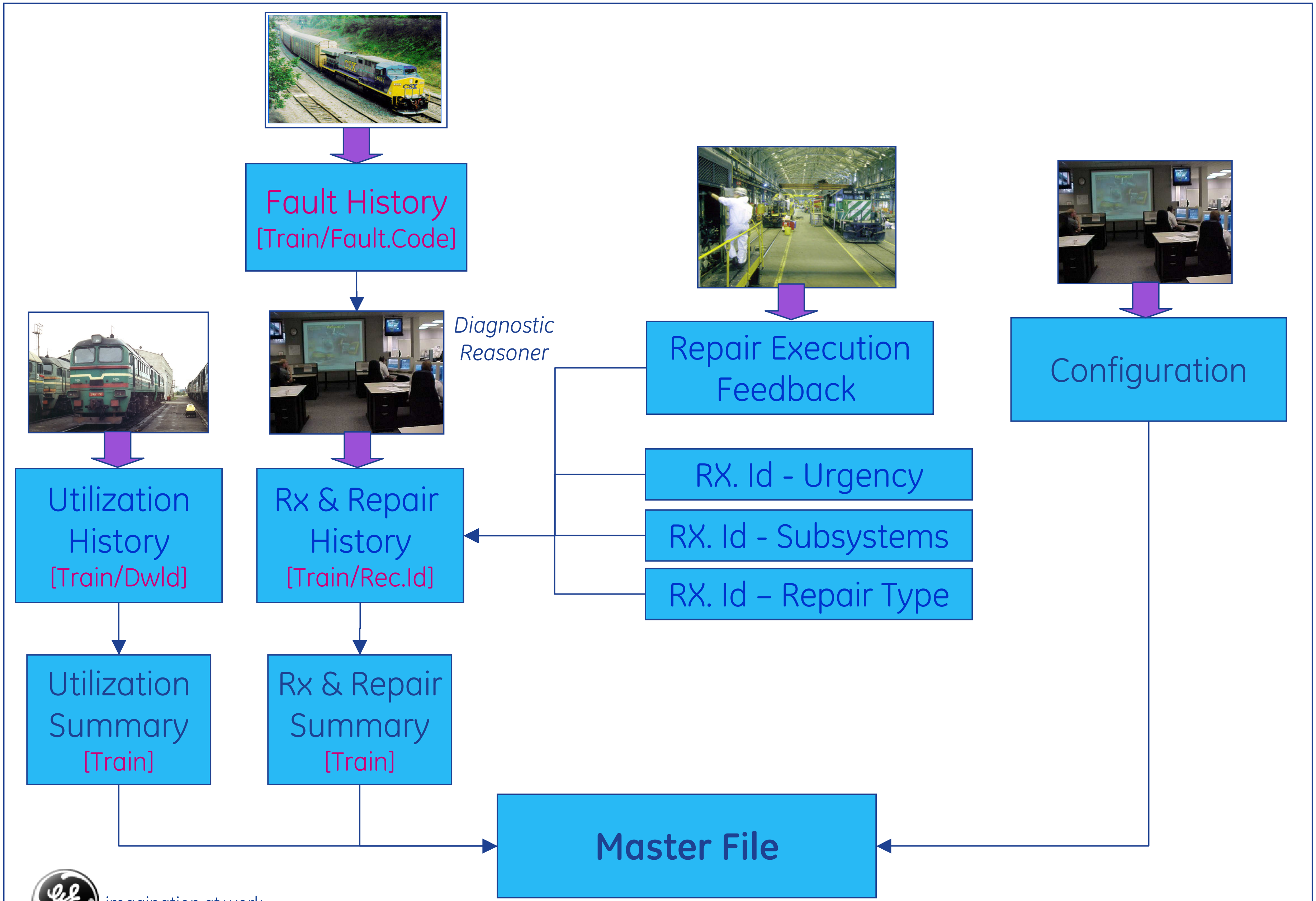
Railroad Yards

Utilization Information
(1 download/~ 30 days)

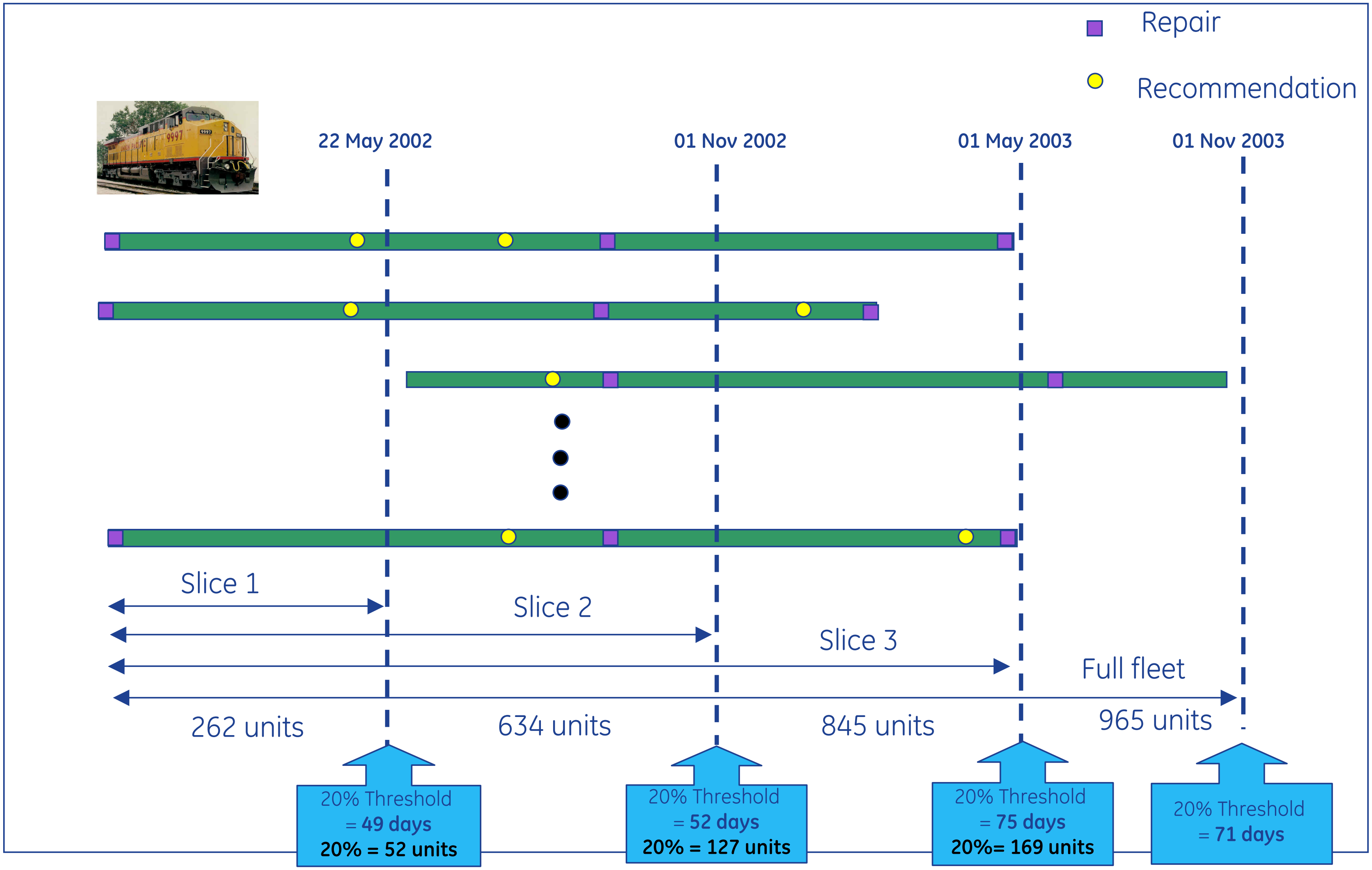
GE Rail/Railroads Repair Shops

Repair Execution
(4-8 Repairs/yr)

Data Collection and Compilation: Tables Relationships



Data Slices

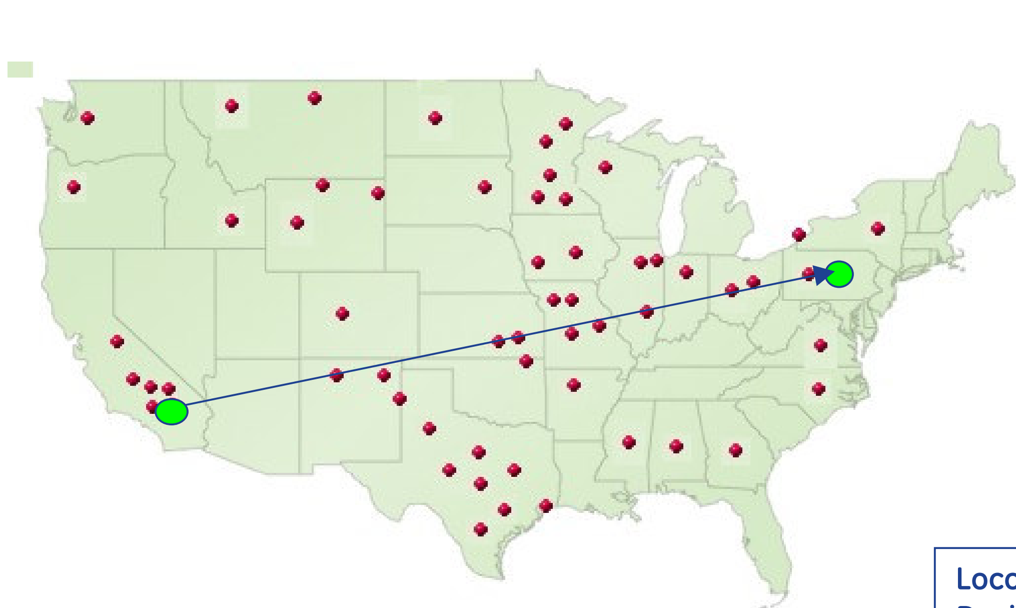


imagination at work

Asset Selection Outline

- Problem Description
- Data Collection and Experiments Set-Up
- **Proposed Peer-Learning Methodology**
  - *Definition of Peers*
  - *Fuzzy Instance-based Classifier (FIBC)*
  - *Evolution of FIBC Design*
- Results
- Conclusions

Current Location of All Available Locomotives



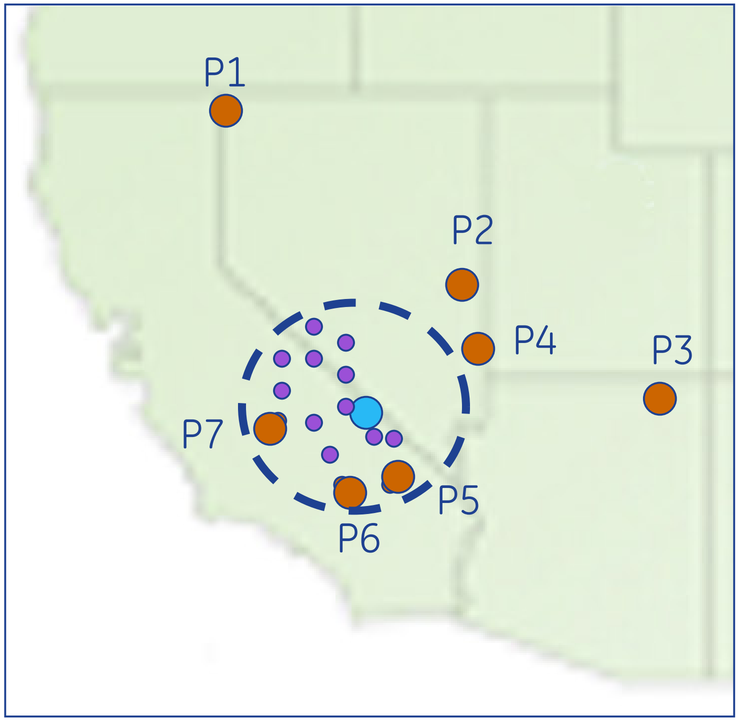
Mission Requirements

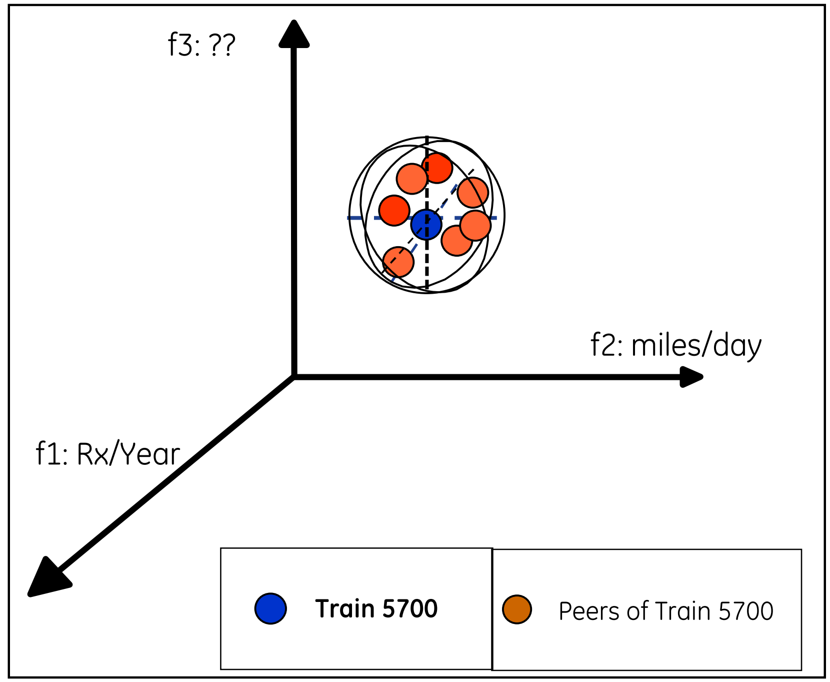
# Locomotives needed:	12
Duration:	9 days
Start date:	+ 48 hrs
Distance:	2,435 miles
Average Miles/day	304 m/day Max
Grade/Elevation:	2%
Climate, Tunnel operation:	desert, hot

Loco Number:	*5700*
Design and Configuration	
Type:	*AC4400*
Electrical System:	*Bosch*
...	
Utilization Information	
Age:	*2.9 years*
Mileage:	*247,567 mi.*
Average miles/day:	*299*
Maintenance Information	
Time elapsed since last repair:	*10 days*
Median time between repairs	*60 days*
Median time from repair to next recommendation (Rx)	*52 days*

imagination at work

Identifying Loco 5700's Peers





= Peers of Loco 5700

Experiments focused on defining the relevant attributes (f1, f2, …, fn) for defining peers

The cluster of peers defines the *collective* for this unit

imagination at work

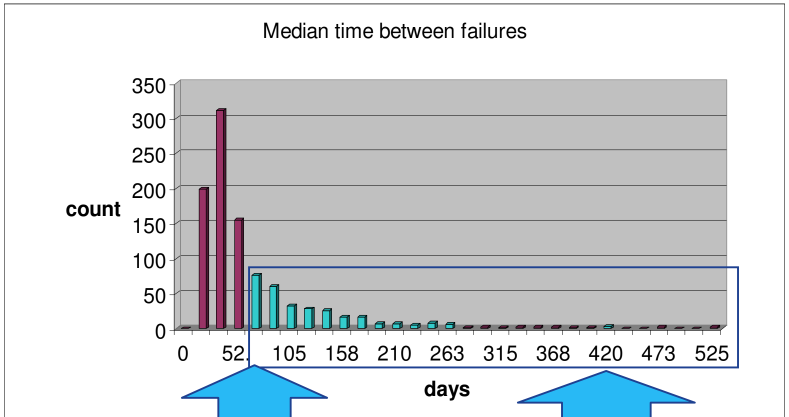
Asset Selection Outline

- Problem Description
- Data Collection and Experiments Set-Up
- Proposed Peer-Learning Methodology
- **Results**
  - *Experiment 1: Top 20% Current Performers*
    - *Robustness*
  - *Experiment 2: Top 52 units Current Performers*
  - *Experiment 3: Top 20% Future Performers*
  - *Experiment 3 bis: Top 20% Future Performers*
    - *Static vs Dynamic Models*
- Conclusions

Baseline – Single-Heuristic based Selection

- Heuristic selection of N1 units (based on full fleet)

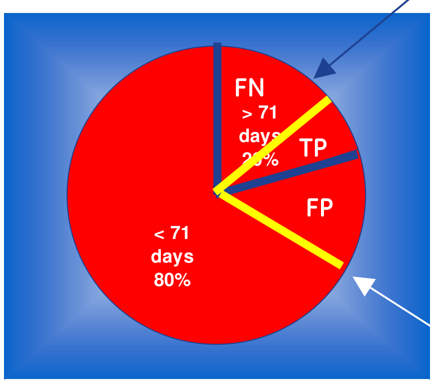
Objective: Create heuristic knowledge driven baseline.



Target: time between failures > 71 days

Random selection in target area: **~20%**

Ground Truth



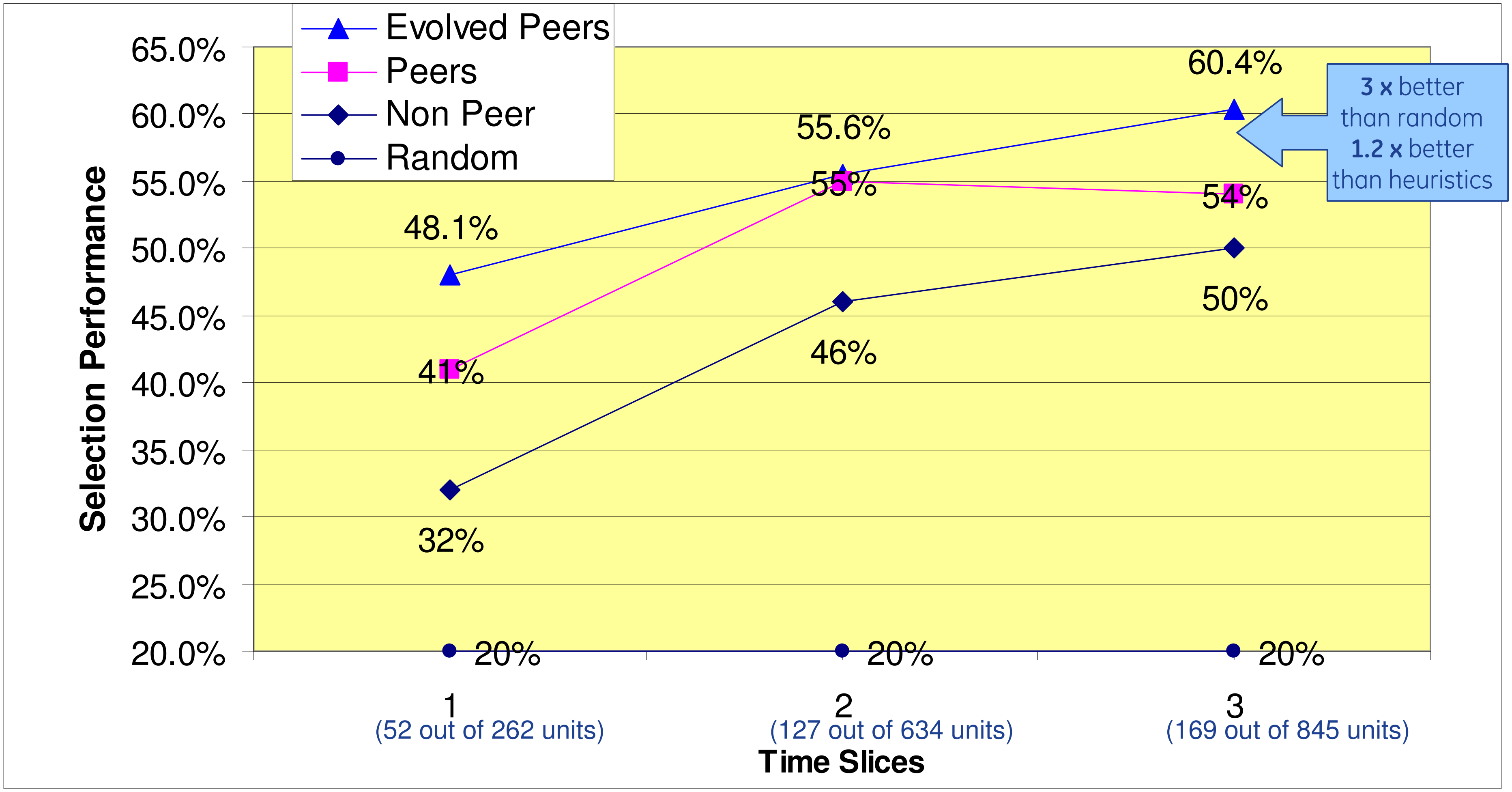
Ground Truth → / Predicted ↓	T	F
T	TP	FP
F	FN	TN

Predicted

Single Heuristic	% of Correctly Classified Units = TP/20%	% Data Correct
Lowest Mileage	**17%**	**81%**
Newest Units	**18%**	**100%**
Random	20%	N/A
Highest Energy (MWHRS) generated	**24%**	**79%**
Highest Miles/ Hours Moving	**26%**	**97%**
Highest Percentage Hours Moving	**29%**	**58%**
Lowest Percentage of: Subsystem 10 Failures	**38%**	**95%**
Lowest Ratio: Recommendations / Age [Rx/yr]	**49%**	**100%**

imagination at work

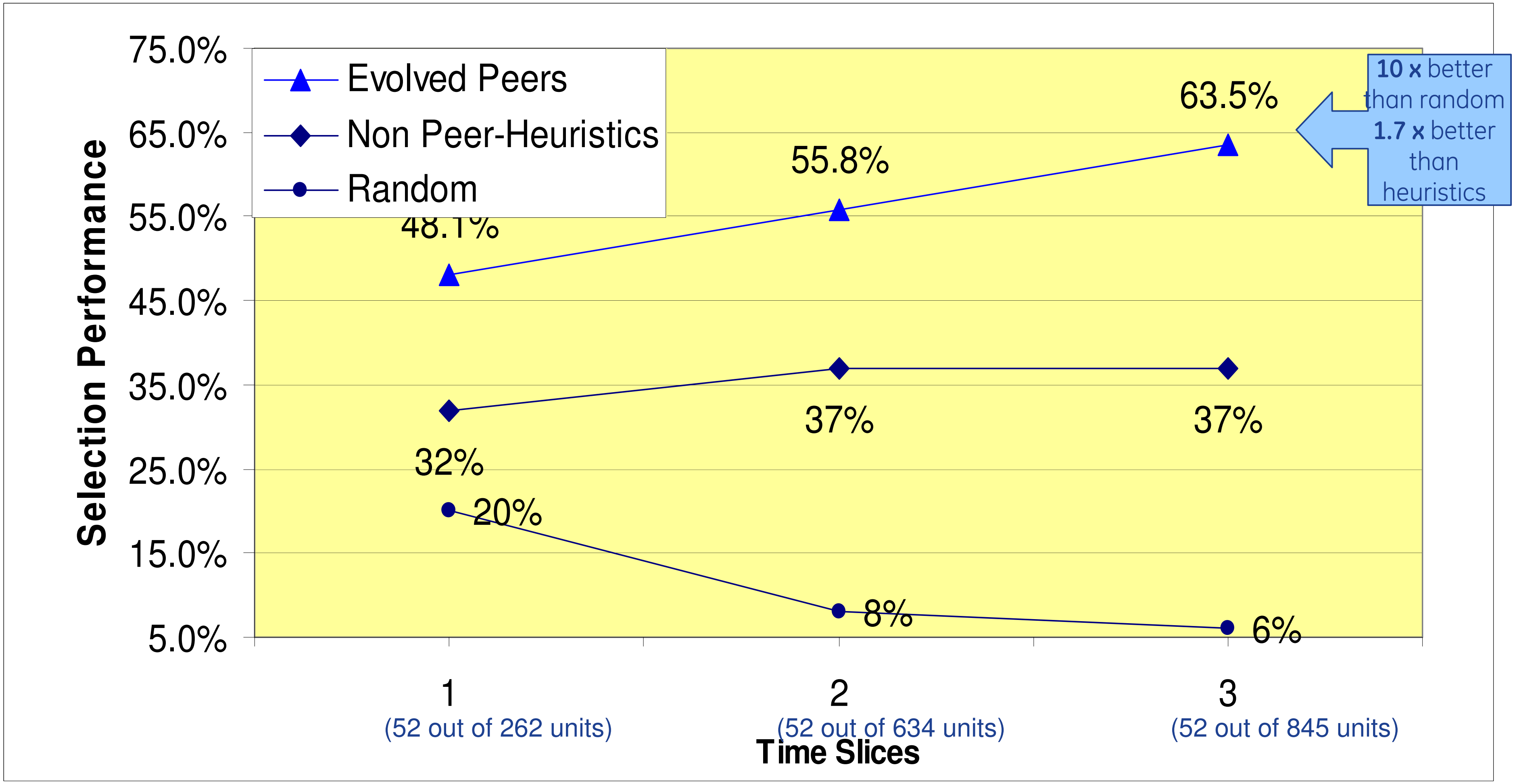
Exp 1: Peers Evolution over Time:
Estimating Best Current performers (20%)



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

imagination at work

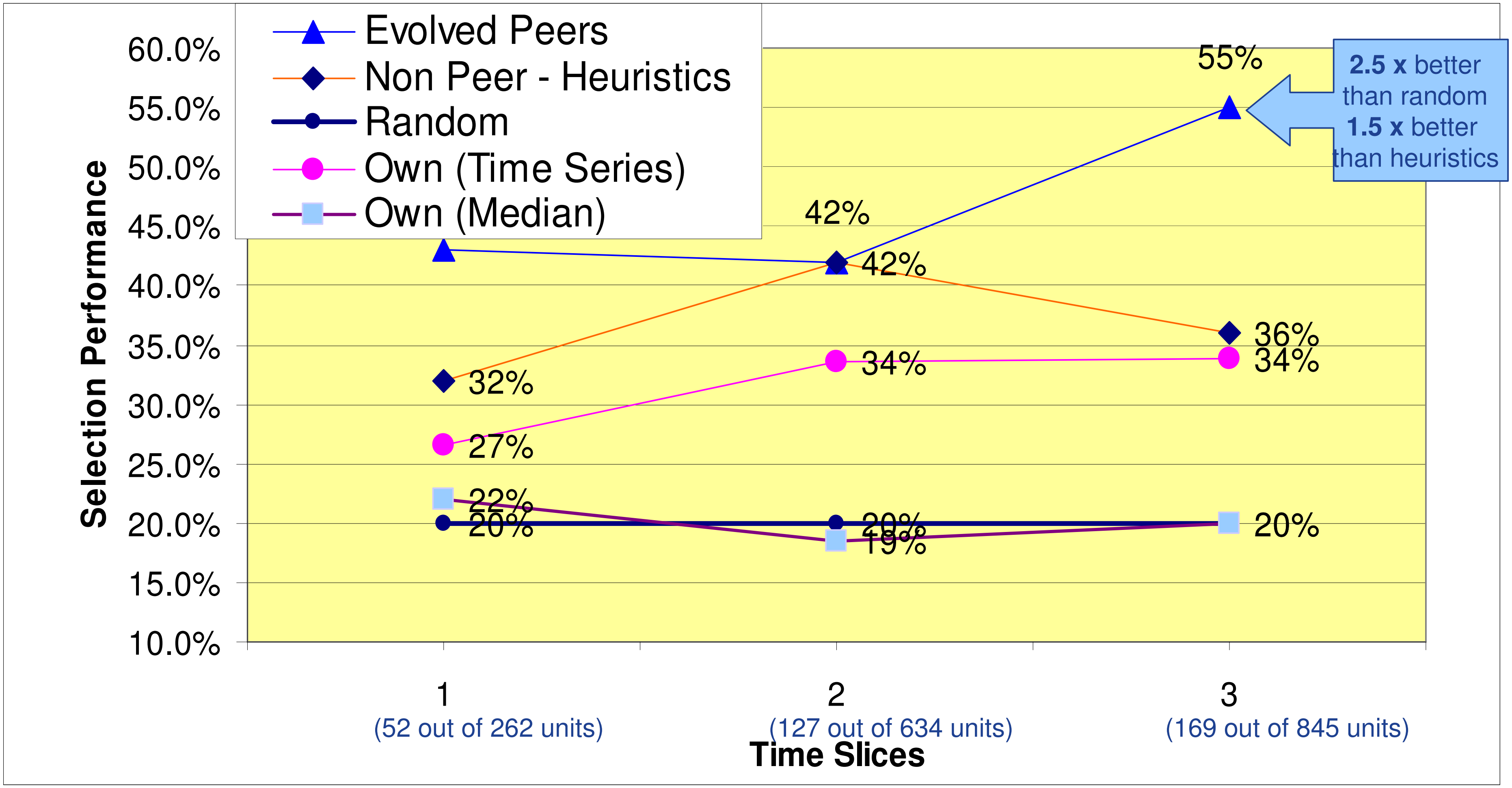
Exp 2: Peers Evolution over Time:
Estimating best Current performers (52 units)



(Targeting Top 52 units for each Time Slice)

imagination at work

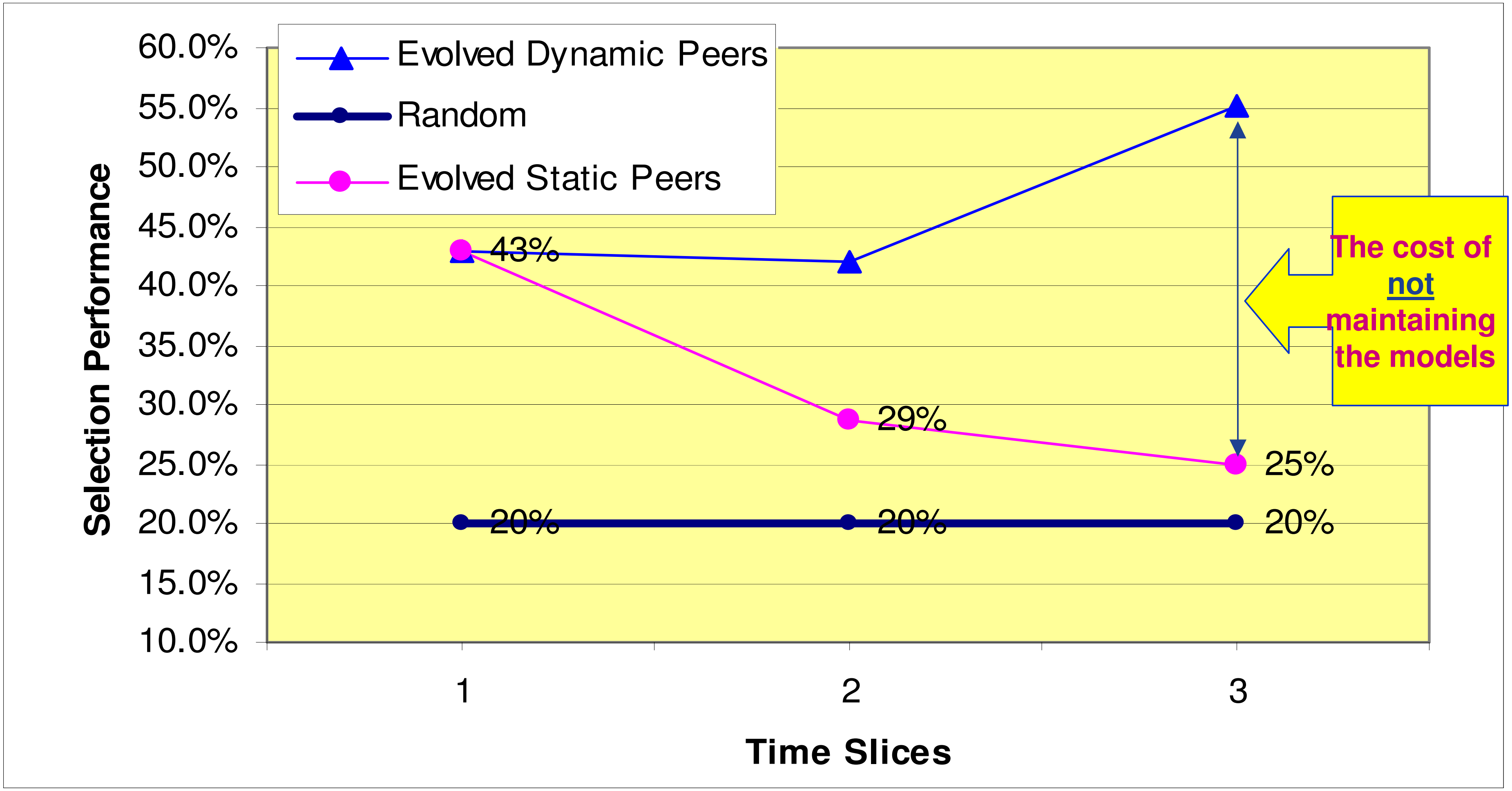
Exp 3: Peers Evolution over Time: Estimating best future performers



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

imagination at work

Experiment 4: Comparing Evolved, Dynamic (updated) Peers with Evolved Static (Non-updated) Peers: Estimating Best Future Performers



The Benefit of Automated Peer Redesign/Update

Asset Selection Outline

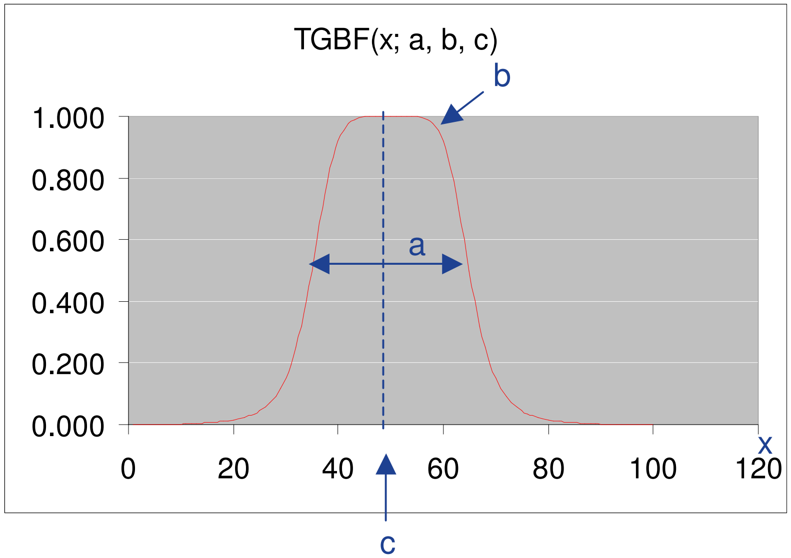
- Problem Description
- Data Collection and Experiments Set-Up
- Proposed Peer-Learning Methodology
- Results
- **Conclusions**
  - *Summary*

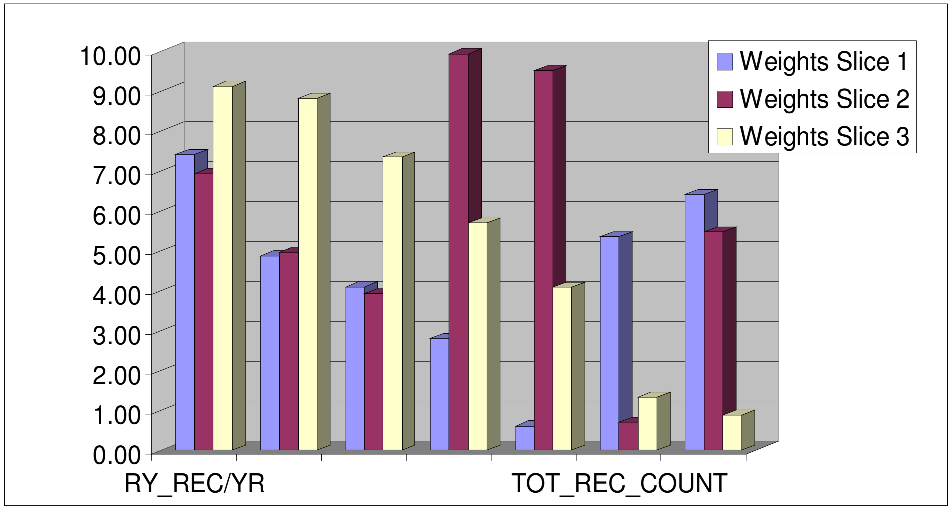
To Future Work.

imagination at work

Attribute Weighing using GA [Slice 3] (for prediction, ranking, and selection of best units)

Feature	Weight	a	b
RY_REC/YR	9.11	5.55	3.73
RY_REC/100K_MILES	8.81	5.35	2.57
RY_REC/100K_ENGINE_HRS	7.35	35.94	2.67
REC_COUNT_RY	5.69	8.49	3.37
RY_REC/100K_ENG_HRS_MOVE	4.08	10.06	2.75
TOT_REC_COUNT	1.33	18.70	3.20
RED_REC_COUNT	0.87	0.99	3.16





$$TGBF = \begin{cases} \dfrac{1}{1+\left|\dfrac{x-c}{a}\right|^{2b}} & \text{if } GBF > 10^{-5} \\ 0 & \text{otherwise} \end{cases}$$

Conclusions

- **Validating the Premise**
  - Combining limited utilization and maintenance history **improves our ability to select the best units for a mission**.
  - Evolved Peers provided the **best overall accuracy** (60.35% = over 3 times better than random selection) for past performance. When the selection was limited to a small fixed number of units, Evolved Peers provided an accuracy of 63.5% (**over 10 times better than random selection**) for past performance.
  - Evolved Peers provided the **best overall accuracy** (55% = **2.7 times better than random selection** and 1.5 x better than best heuristics) for future performance
  - Construction of local fuzzy models does not require computing a distance in the n-dimensional feature space

- **Operational Impact:**
  - **Robustness** to information loss exhibited by peer-based approach will enable mission reliability for minimally instrumented platforms operating with limited bandwidth

6. Future Work

- Improving the Feature Space: GP for Attribute Construction
- Improving Fitness Function: Accuracy, Confidence, Info. Theory
- Improving Aggregation: Adapting Kernel Based Regressions

imagination at work

Future Work

- Extend evolutionary search
  - from attribute selection and weighting to attribute construction
- Improve fitness function
  - to address classifier accuracy and confidence, and representation parsimony
- Improve aggregation models

To References

imagination at work

Future Work

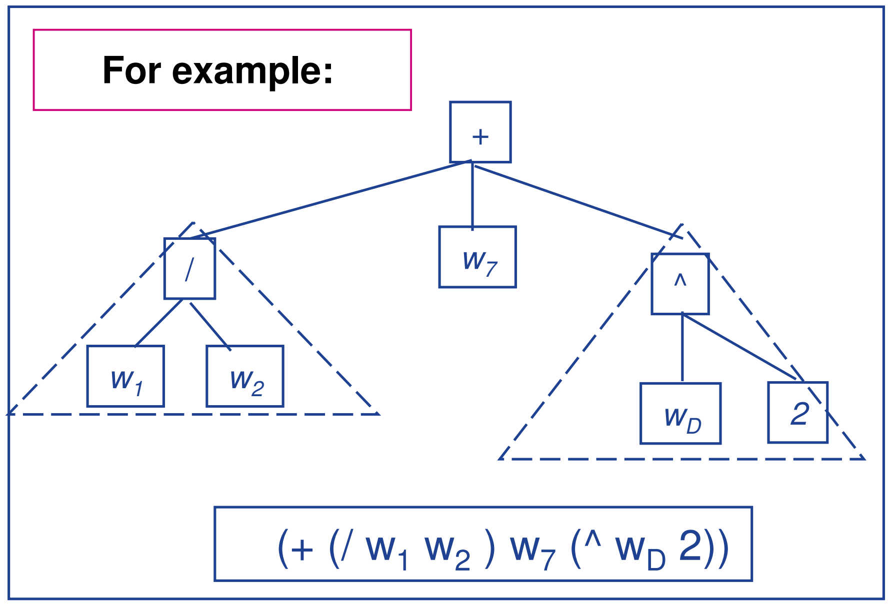
- **Extend evolutionary search from attribute selection and weighting to *attribute construction*.**

  Use **Genetic Programming** to automate attribute construction and evolve attribute space with functional compositions of primitive attributes

GA: $[(w_1, w_2, \ldots, w_D)][(a_1, b_1), \ldots, (a_D, b_D)][\alpha]$

GP: $[f_1 * g_1 * \ldots * h_1(w_1, w_2, \ldots, w_D), \ldots, f_D * g_D * \ldots * h_D(w_1, w_2, \ldots, w_D)][(a_1, b_1), \ldots, (a_D, b_D)][\alpha]$

A sentence derived from a grammar that defines a syntactically correct functional composition of attributes to replace a primitive attribute



$$\left(\frac{w_1}{w_2} + w_7 + w_D^2 \right)$$

imagination at work

Future Work: Improving the Fitness Function

- Extend evolutionary search.
- **Improve *fitness function* to address classifier accuracy and confidence, and representation parsimony**

classifier accuracy, confidence and representation parsimony

$$f1 = \alpha_1 \left[\frac{\text{TotPred} - \text{CorrPred}}{\text{TotPred}} \right] + \alpha_2 \left[\frac{\sum_{i=1}^{TotPred} \frac{\text{nmin}_i}{K_i}}{\text{TotPred}} \right] + \alpha_3 \left| \frac{\left(\sum_{j=1}^{D} w_j \right) - 1}{(D-1)} \right| =$$

$$= \alpha_1 [Classifier\_Inaccuracy] + \alpha_2 [Average(\text{Retrieval\_Inaccuracy})] + \alpha_3 \left| \frac{d-1}{D-1} \right|$$

Predicted → / Ground Truth ↓	T	F
T	**TP**	**FP**
F	**FN**	**TN**

Assumptions:

Two-class problem or

Multiple class problem with indistinguishable errors

Where:

Classifier_Inaccuracy = (FN+FP)/(TP+FP+FN+TN)

TotPred = *(TP+FP+FN+TN)*

$nmin_i$ *= number of minority decisions in i^{th} retrieval*

K_i *= number of points retrieved for case I (Cardinality of i^{th} retrieval)*

Retrieval_Inaccuracy (i)= ($nmin_i$)/ K_i = Cardinality of Non-Mode/TotalCardinality for i^{th} retrieval

Typical values: $\alpha_1 = 2; \quad \alpha_2 = 0.2; \quad \alpha_3 = 0.1$

Future Work

- Extend evolutionary.
- Improve *fitness function*
- **Improve aggregation models**

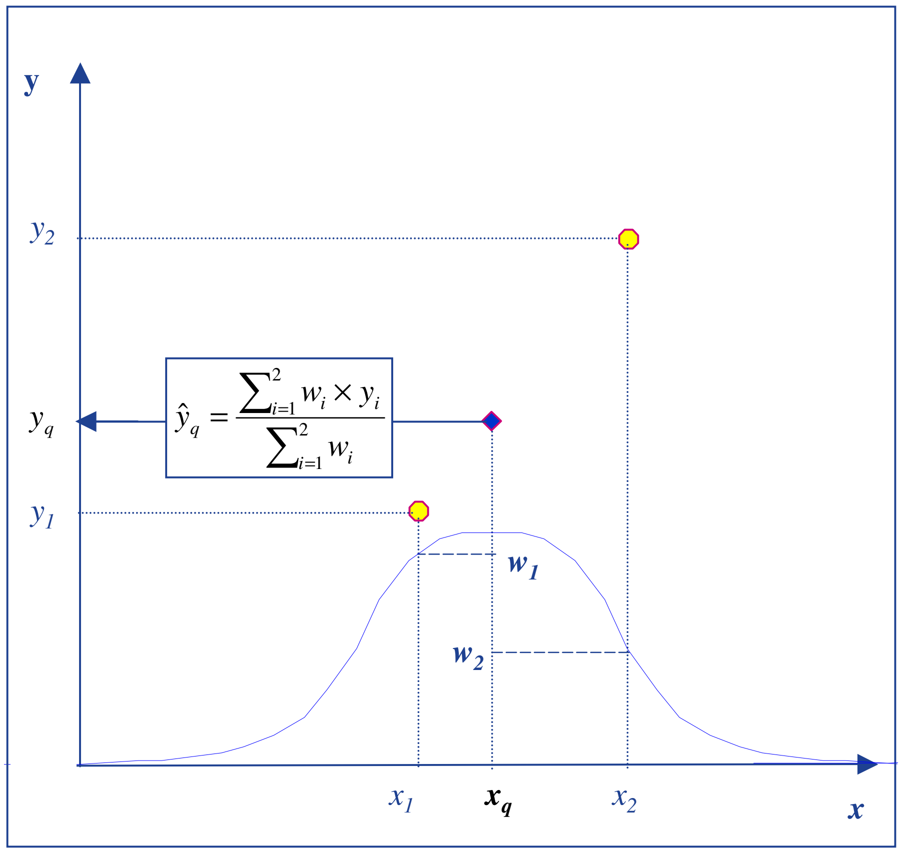
We could use the similarity measures as ***kernel functions*** for each dimension *i* and create ***kernel-based regressions***

We would use local search methods to obtain the parameters of the kernel-based regression, within each trial of the EA's.

Future Work: Improving Aggregation

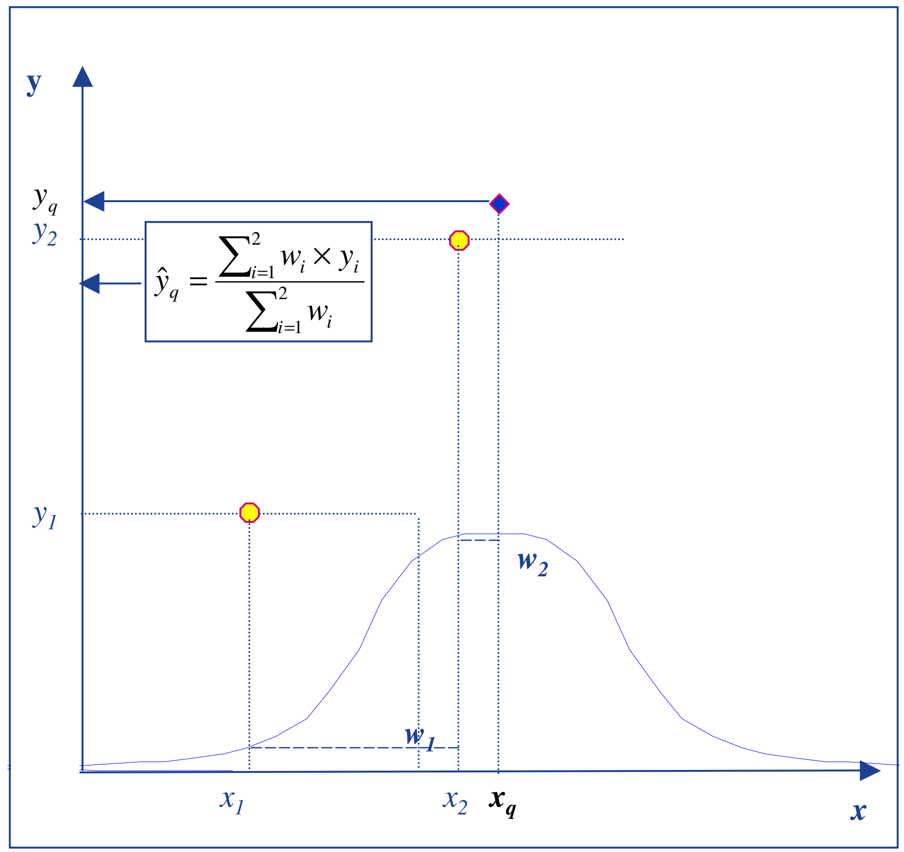
0th Order Approximation (weighted average)

- It works well when **interpolating**



0th Order Approximation (weighted average)

- It is lousy when **extrapolating**
- Bounded by the **minimum** & **maximum** of its arguments

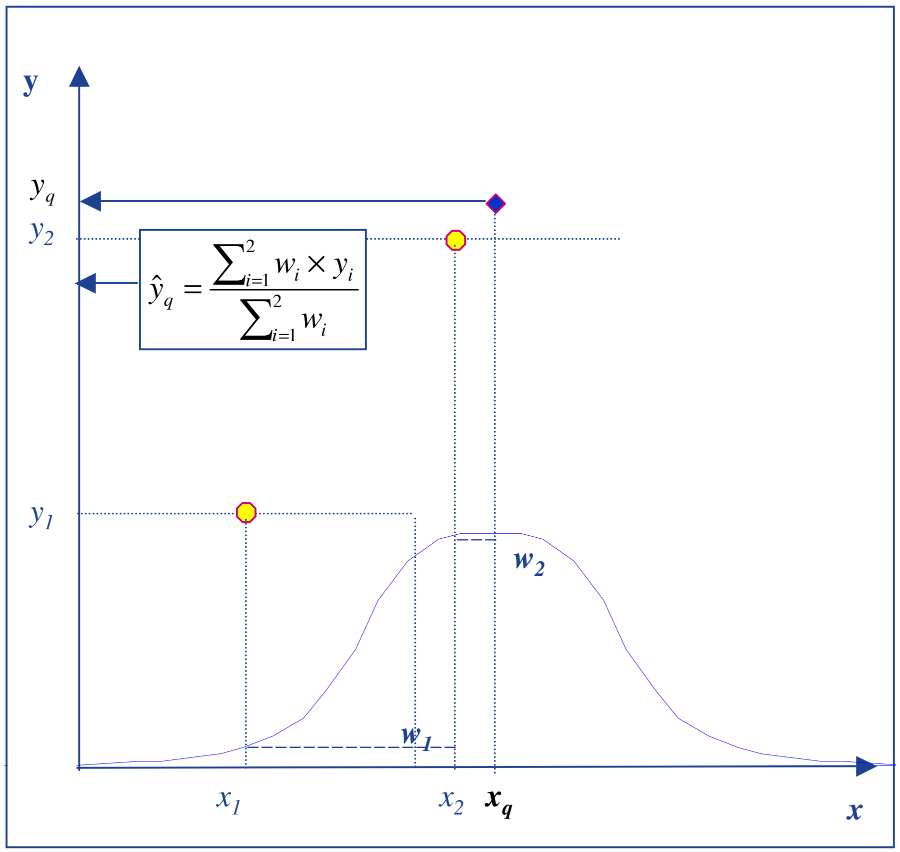


imagination at work

Future Work: Improving Aggregation

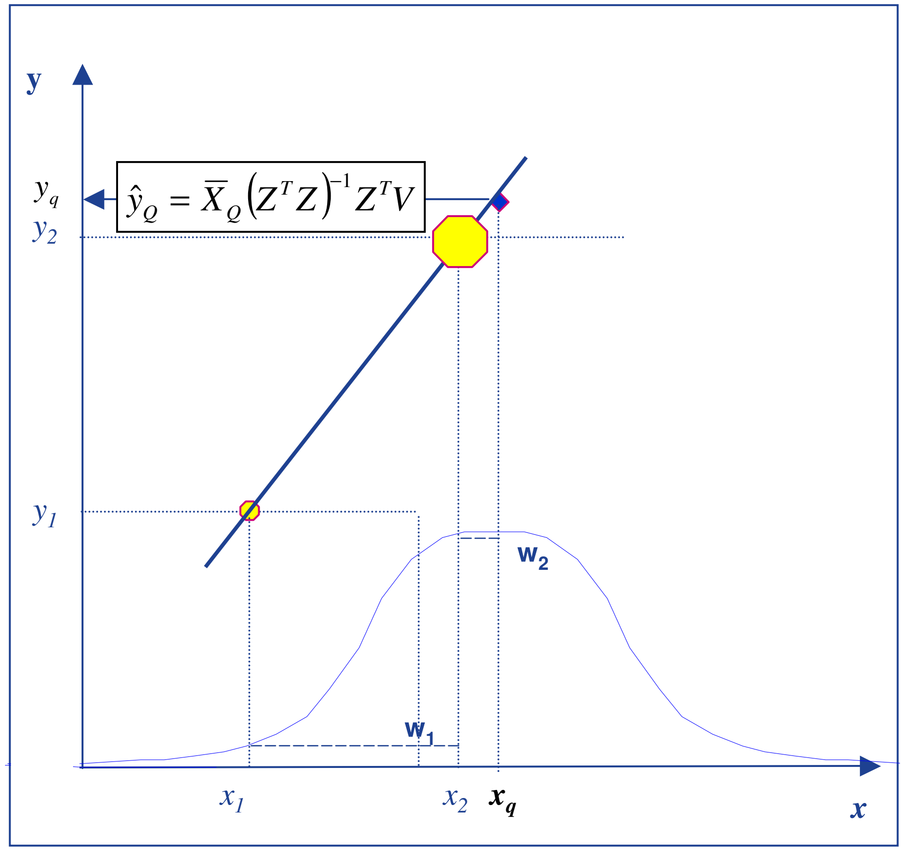
0th Order Approximation (weighted average)

- It is lousy when **extrapolating**
- Bounded by the **minimum** & **maximum** of its arguments



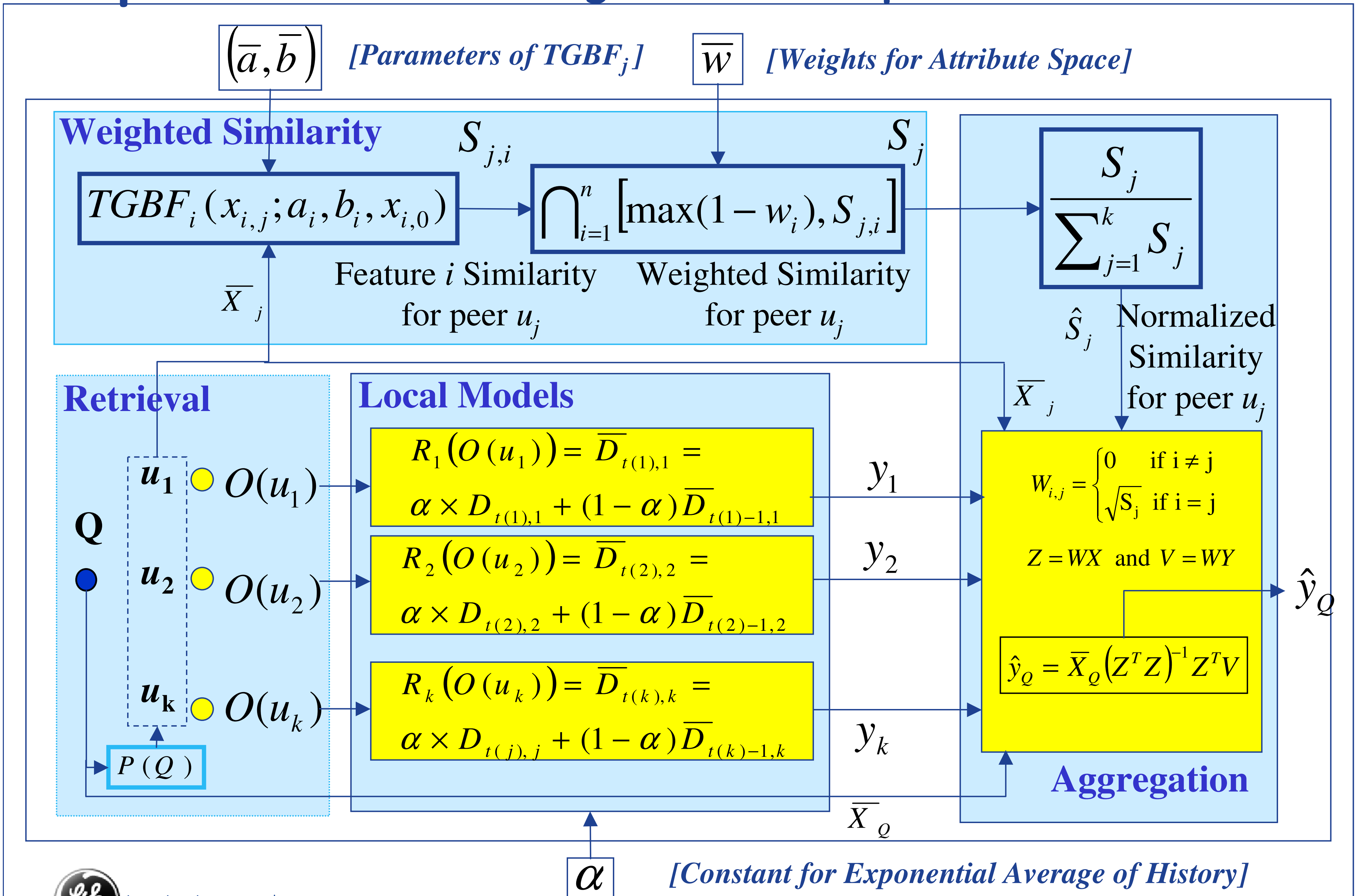
1st Order Approximation (Weighted Linear Regression)

- Weight each data point by its similarity degree with the probe Q



imagination at work

Improved F-IBM: Algebraic Interpretation



imagination at work

Further Resources

From my Web site: www.rpi.edu/~bonisp

Go to: http://www.rpi.edu/~bonisp/publications-new.html

Relevant papers that you can download:

"Hybrid Soft Computing Systems: Industrial and Commercial Applications", P. P. Bonissone, Y-T Chen, K. Goebel and P. S. Khedkar, *Proceedings of the IEEE*, pp 1641--1667, vol. 87, no. 9, Sept. 1999. http://www.rpi.edu/~bonisp/NASA-course/hybridSC99.pdf

"When will it break? A Hybrid Soft Computing Model to Predict Time-to-break Margins in Paper Machines", P. Bonissone and K. Goebel, *Proc. SPIE 2002* , pp. 53--64, Aug. 2002, Seattle, WA http://www.rpi.edu/~bonisp/NASA-course/webbreakage.pdf

"Evolutionary Optimization of Fuzzy Decision Systems for Automated Insurance Underwriting", P. Bonissone, R. Subbu, and K. Aggour, *Proc. FUZZ-IEEE 2002*, pp. 1003 - 1008, May 2002, Honolulu, HI. http://www.rpi.edu/~bonisp/NASA-course/wcci2002.pdf

"Automating the Quality Assurance of an On-line Knowledge-Based Classifier By Fusing Multiple Off-line Classifiers", P. Bonissone, *Inform. Proc. & Management of Uncertainty (IPMU),* Perugia, Italy, July 2004 http://www.rpi.edu/~bonisp/NASA-course/IPMUv8.pdf

"Development and Maintenance of Fuzzy Models in Financial Applications", P. Bonissone, *Proc. SMPS 2004,* , Oviedo, Spain, September 2004. http://www.rpi.edu/~bonisp/NASA-course/Oviedo2004.pdf

"Six Sigma Quality Applied Throughout the Lifecycle of and Automated Decision System", A. Patterson, P. Bonissone, andM. Pavese, *Journal of Quality and Reliability International*, 21:275-292, 2005 http://www.rpi.edu/~bonisp/NASA-course/SixSigma.pdf

"An Evolutionary Process for Designing and Maintaining a Fuzzy Instance-based Model (FIM)", P. Bonissone, A. Varma, K. Aggour, *1st Workshop Genetic Fuzzy Systems (GFS 2005)*, Granada, Spain, March 2005. http://www.rpi.edu/~bonisp/NASA-course/piero-gfsmod.pdf

Predicting the Best Units within a Fleet: Prognostic Capabilities Enabled by Peer Learning, Fuzzy Similarity, and Evolutionary Design Process", P. Bonissone, A. Varma,. *FUZZ-IEEE 2005*, pp 312-318, Reno NV, May 22-25, 2005. http://www.rpi.edu/~bonisp/NASA-course/fuzz05anilfinalv27.pdf

Questions ?

Thank You!

imagination at work