Evolving Fuzzy Systems (EFS)

Fundamentals, Reliability, Interpretability, Useability and Applications

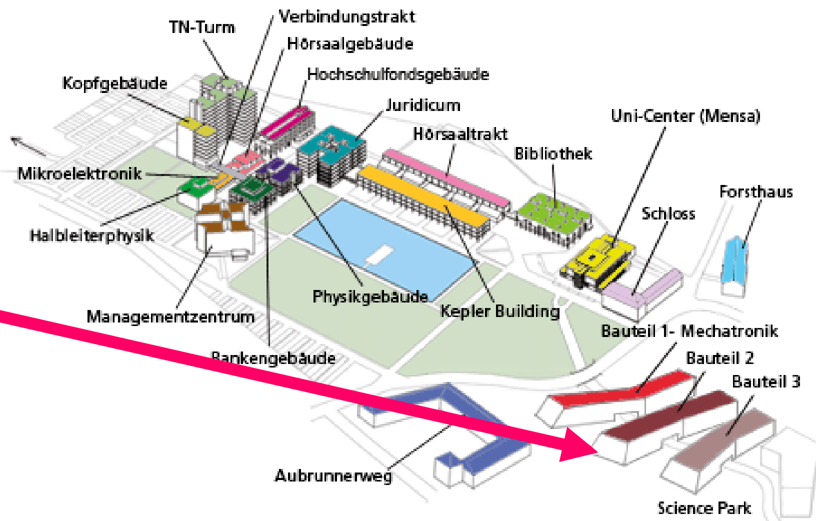
Dr. Edwin Lughofer
Department of Knowledge-Based Mathematical Systems
Johannes Kepler University, Linz, Austria
http://www.flll.jku.at/staff/edwin

Our Department – Location in Linz AND Hagenberg (Upper Austria)

Softwarepark Hagenberg

Fuzzy Logic Labor

Main Campus at the University of Linz



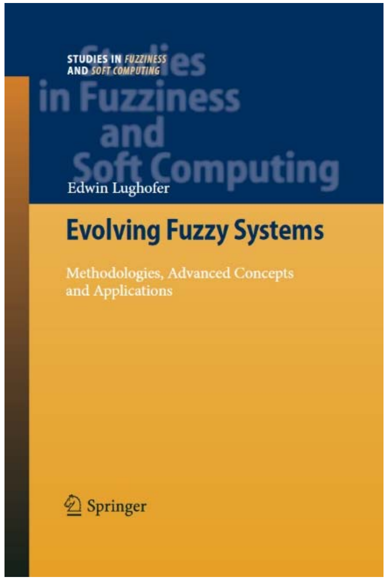
Surveys about EFS

Edwin Lughofer

Evolving Fuzzy Systems

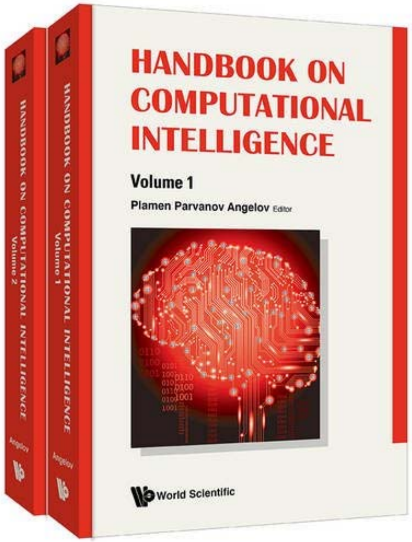
Methodologies
Advanced Concepts
Applications

Springer, Heidelberg, 2011
(approx. 150 figures, 250 images, 460 pages)



+ book chapter (60 pages) in

Handbook on Computational Intelligence



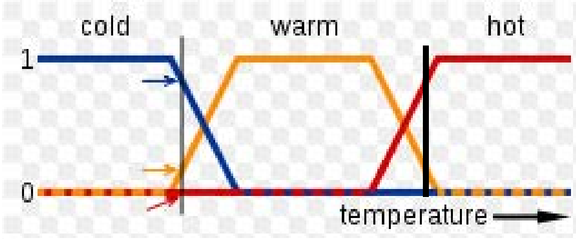
Editor: Plamen Angelov

Title: EFS – Fundamentals, Reliability, Interpretability, Useability and Apps

Publisher: World Scientific, January 2016

Properties of Fuzzy Systems

• **Rule-Based Models** which build upon the concept of **Fuzzy Logic** (going back to Zadeh: truth values are fuzzified in [0,1] by Fuzzy Sets)



20 25 28 => to 0.2 degree" Warm" to 0.8 degree „Hot"

• **Formation of Linguistic Rules** by Conjunction (t-norm) of Fuzzy Sets over Input Vars, e.g. (example from premise price modeling):

IF Area is LOW **AND** Age is OLD Then Price is LOW

•**Tradeoff** between

» **Universal Approximation:** Models with arbitrary degree of non-linearity by piecewise local approx.

» **Interpretability** (Readable Linguistic Terms and Rules)

• **Data-Driven Case => Gray-Box Models (ranging from light to dark grey)**

• Similarities Between Certain Architectures with Certain Types of **Neural Networks** (e.g. TS with Gaussian sets ⇔ RBF networks)

Research Directions in Fuzzy Sys Design

FUZZY.LOGIC.LAB.LINZ

- **Old-School (70ties, 80ties):** Knowledge-Based Design (White Box)
- **New-School (90ties, 00s):** Data-Driven Learning with Machine Learning/Optimization Techniques (genfis2+3, LOLIMOT (tree-like structures), FMCLUST, ANFIS, Genetic FS …) --- **batch design!**

- **Emerging Topic 1 (in infants): Hybrid Design**
  - » Input: Expert-based Design + Measurement Data
  - » Output: Refined Fuzzy Systems meeting Interpretability Constraints
  - » Variants:
    - – Movement of Expert-based Partitioning
    - – Optimization of Fuzzy Set Combinations (=> New Rules)
    - – Model Transfer (Mamdani=>Takagi-Sugeno, consequent re-learning)

- **Emerging Topic 2: Adaptive Evolving from Streams**

The Idea of Evolving Fuzzy Systems
(emerged approx. 2004/05)

Evolving $\neq$ Evolutionary (=>Genetic FS)

Learning Fuzzy Systems in **(Single-Pass) Incremental** and **Evolving Manner** from Streaming (Block/Sample-wise Loaded) Data

Characterization of a Data Stream (Gama, 2010):

- The data samples or data blocks are continuously arriving on-line over time
- The data samples are arriving in a specific order, over which the system has no control.
- Data streams are usually not bounded in a size
- Once a data sample/block is processed, it is usually discarded immediately, afterwards

Concepts in Evolving Fuzzy Systems

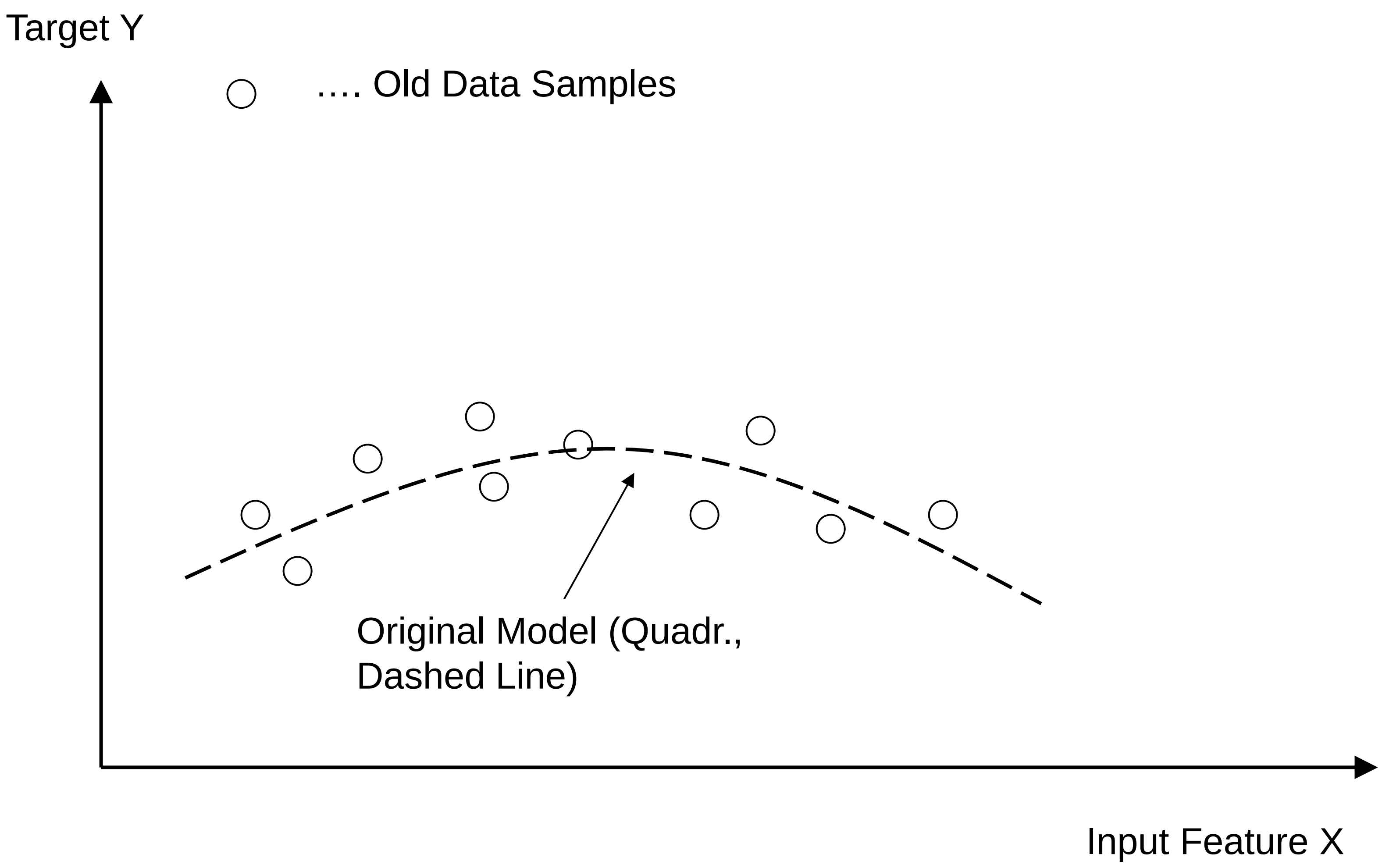
- **Incrementality** accounts for a step-wise (sample or block) processing of data and model building, omitting time-intensive re-training (**on-line capability**)

- **Adaptivity** accounts for **(recursively)** adapting parameters with newly loaded samples

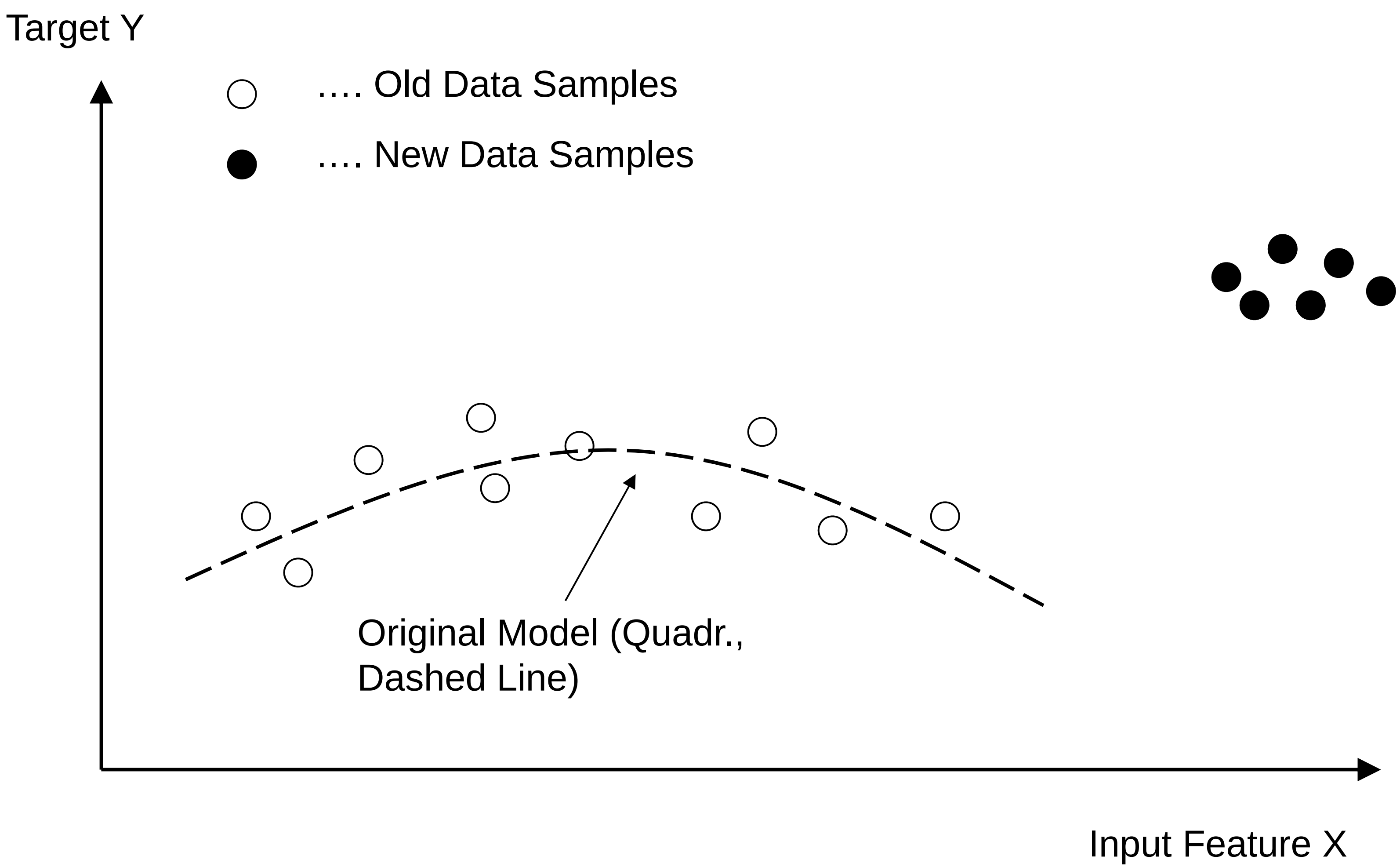
- **Evolving** means that **structural components** (rules, neurons) are **added on demand** due to new system states, operating conditions etc.

- **Single-Pass Capability** = Sample is loaded, sent into the incremental learning engine and discarded immediately, afterwards (**achieving low computation time and virtual memory demand**)

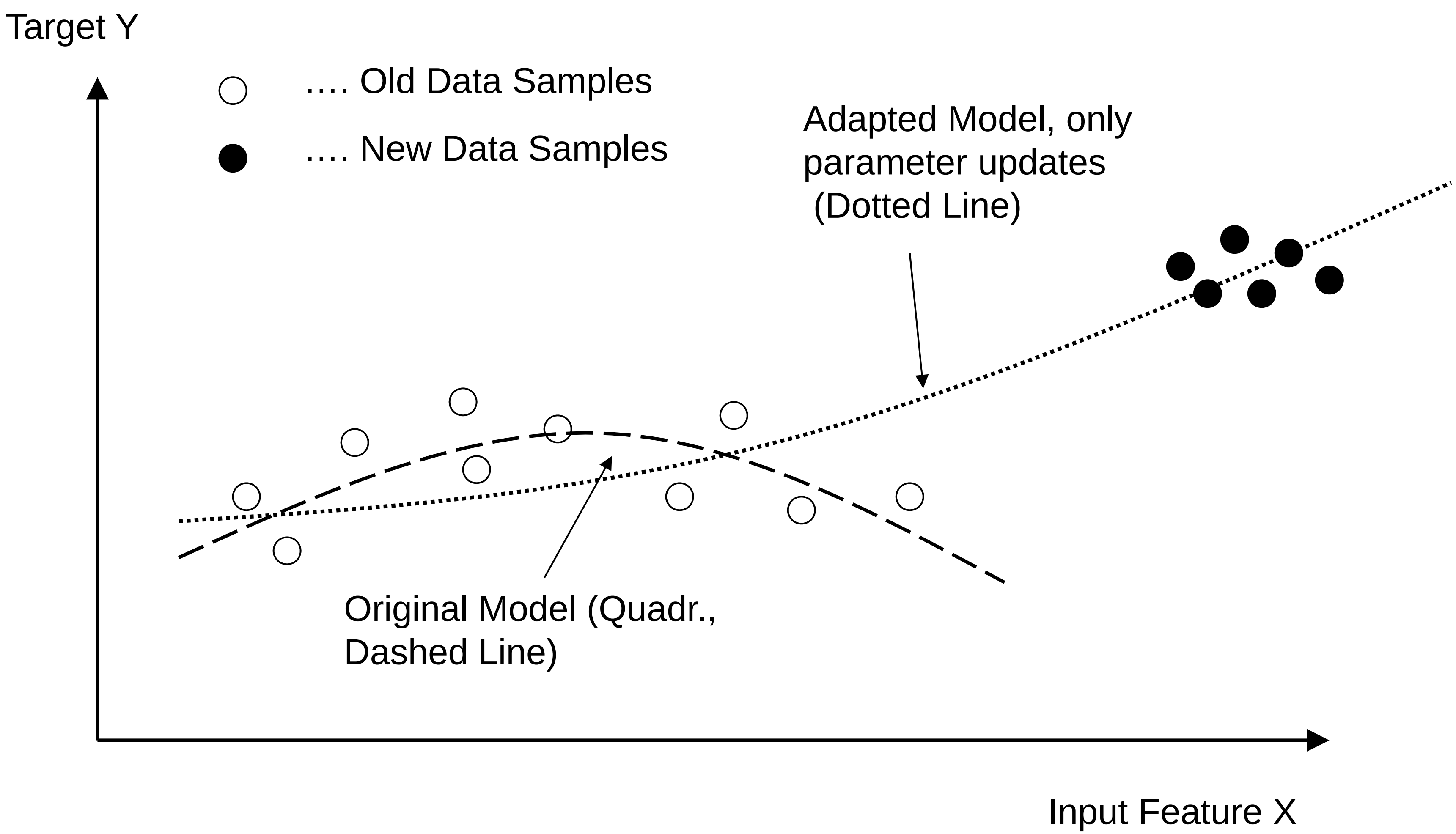
Structural Evolution vs. Parameter Adaptation



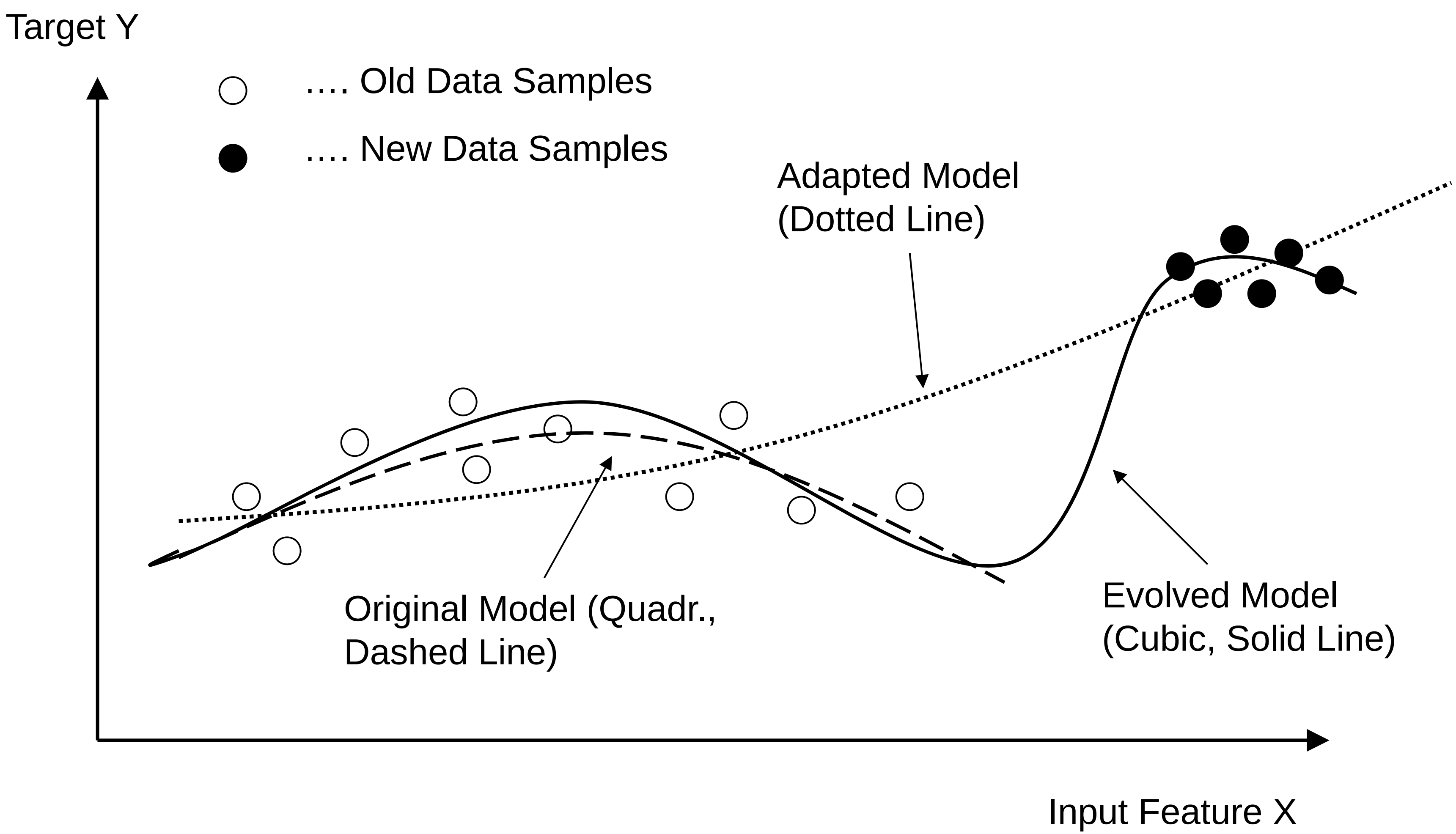
Structural Evolution vs. Parameter Adaptation



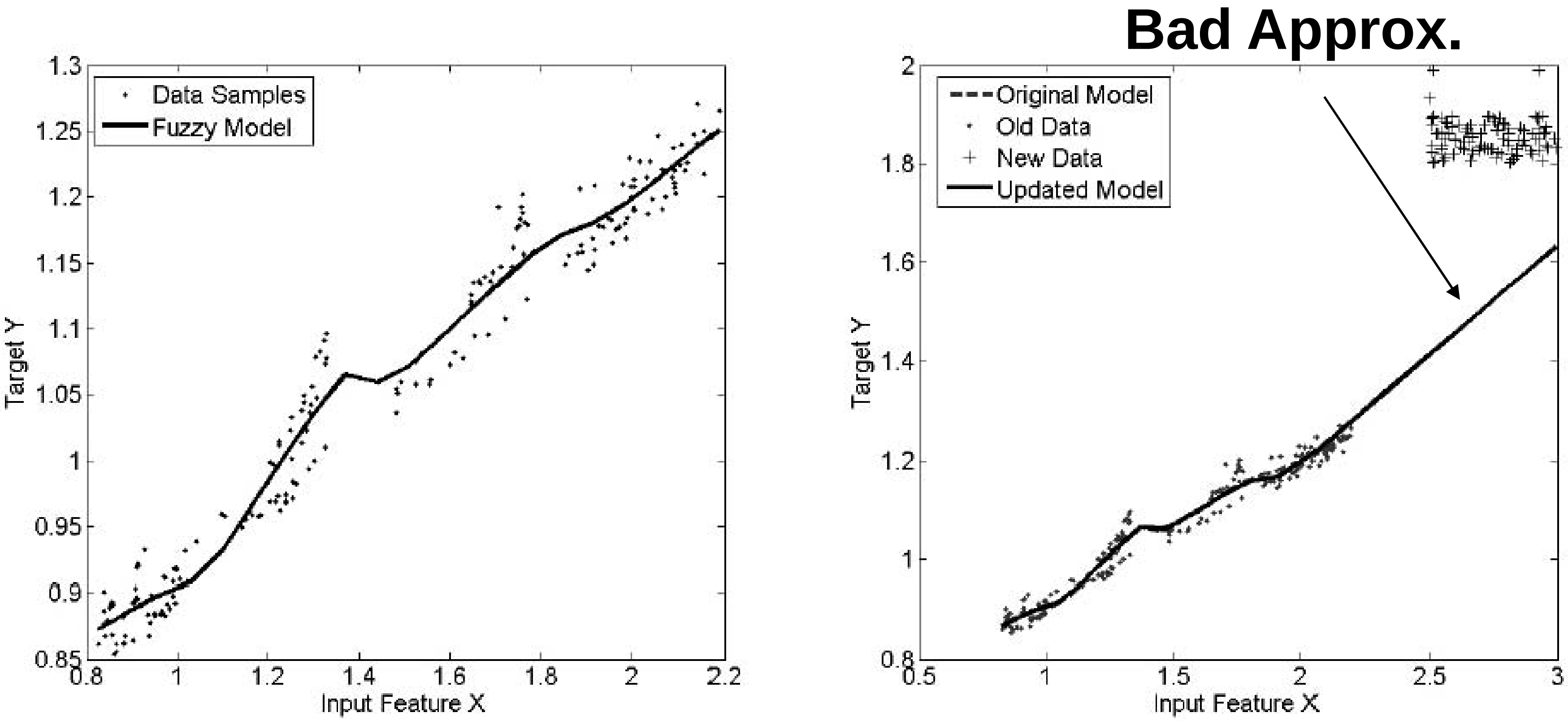
Structural Evolution vs. Parameter Adaptation



Structural Evolution vs. Parameter Adaptation



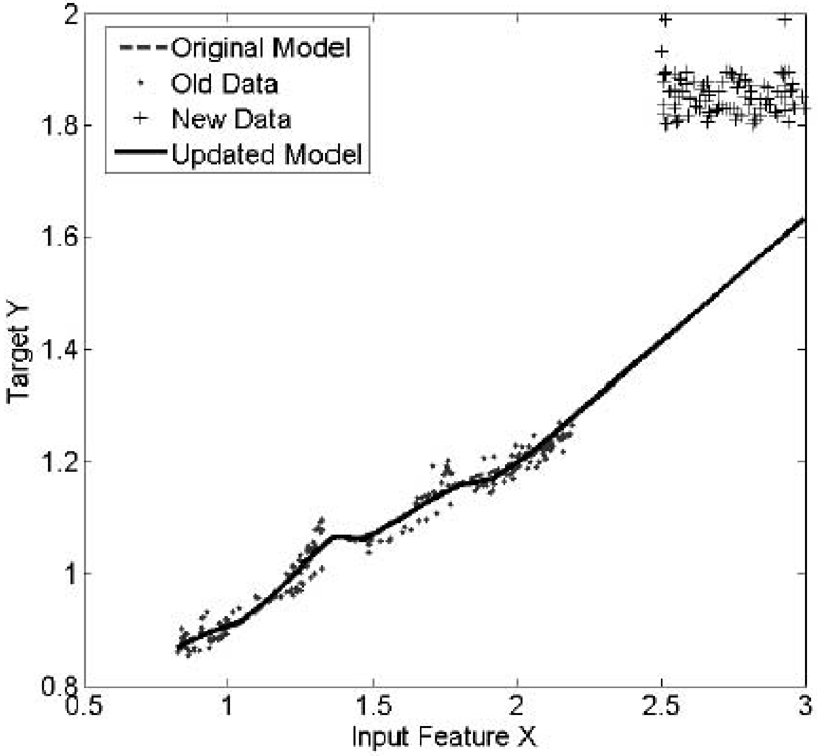
Failure Case (Parameter Adaptation)



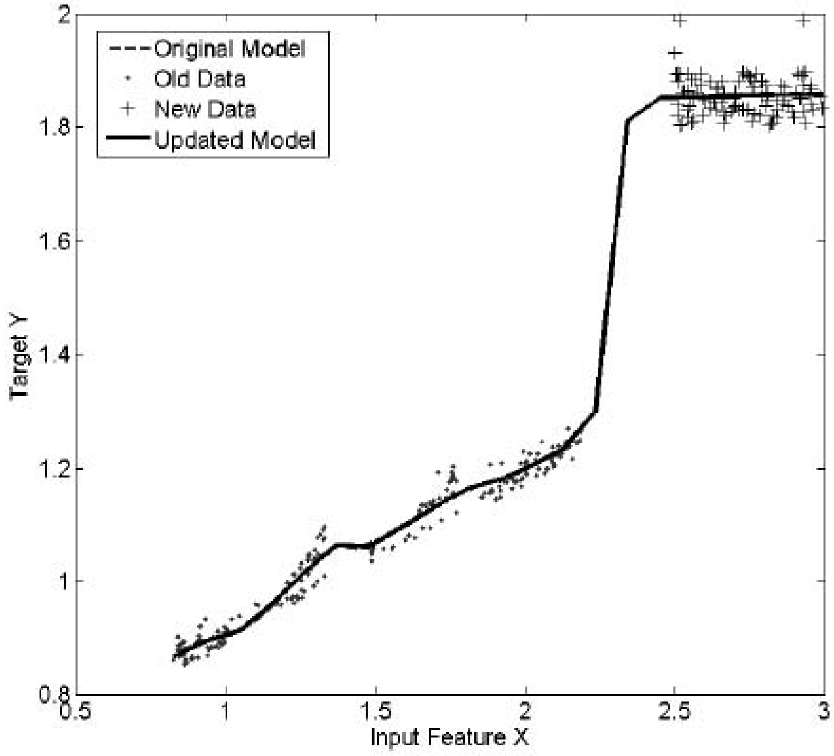
(a) (b)

Significant Range Extension

Success of Update with EFS (evolving rule base)



(a)

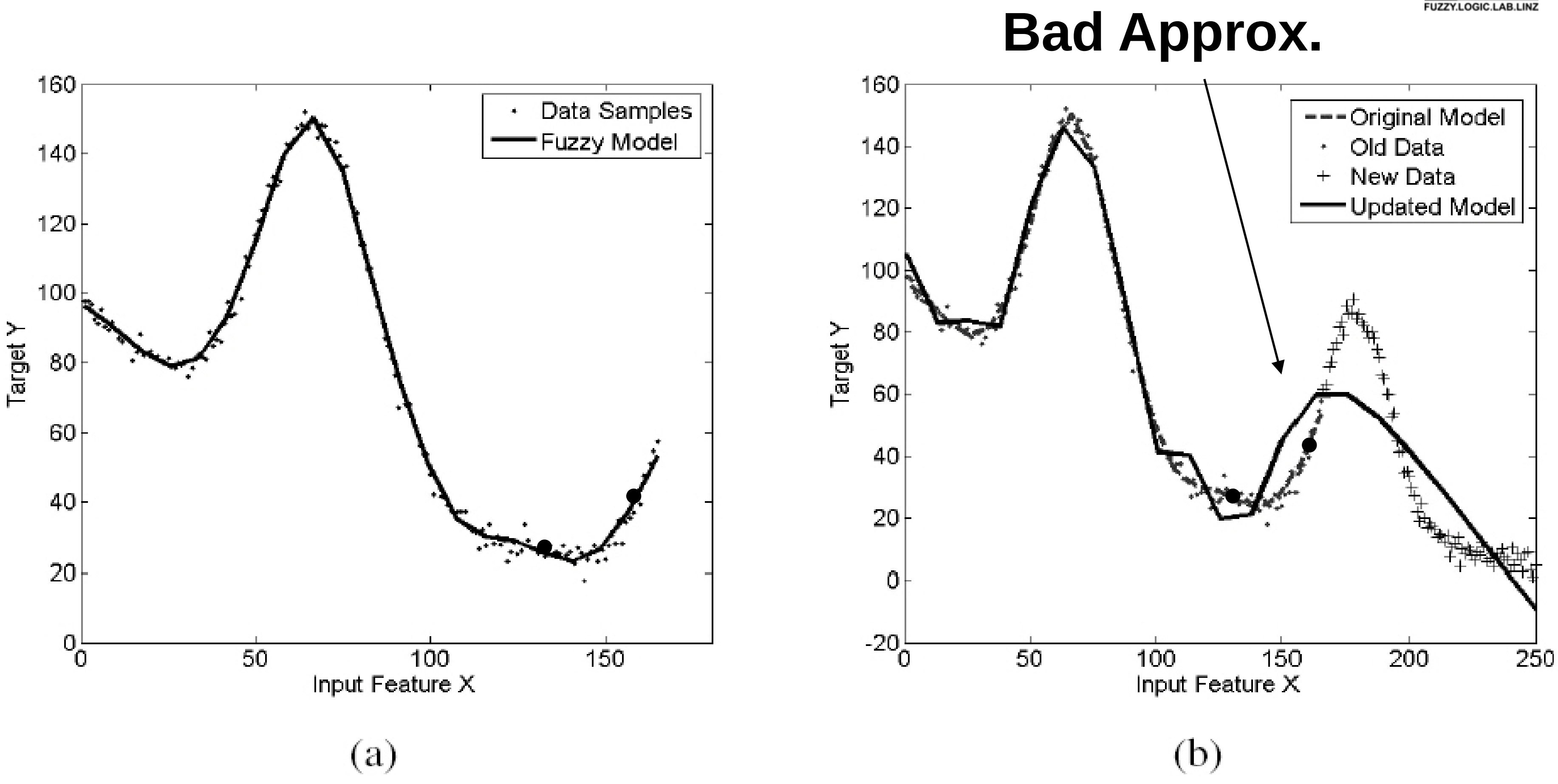


(b)

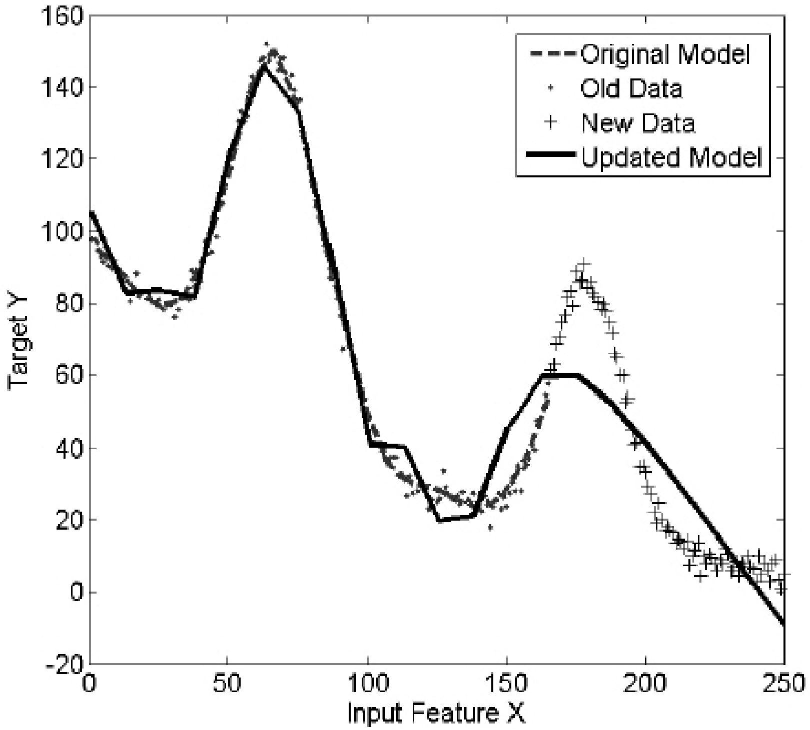
Update Params only

Update Structure + Params

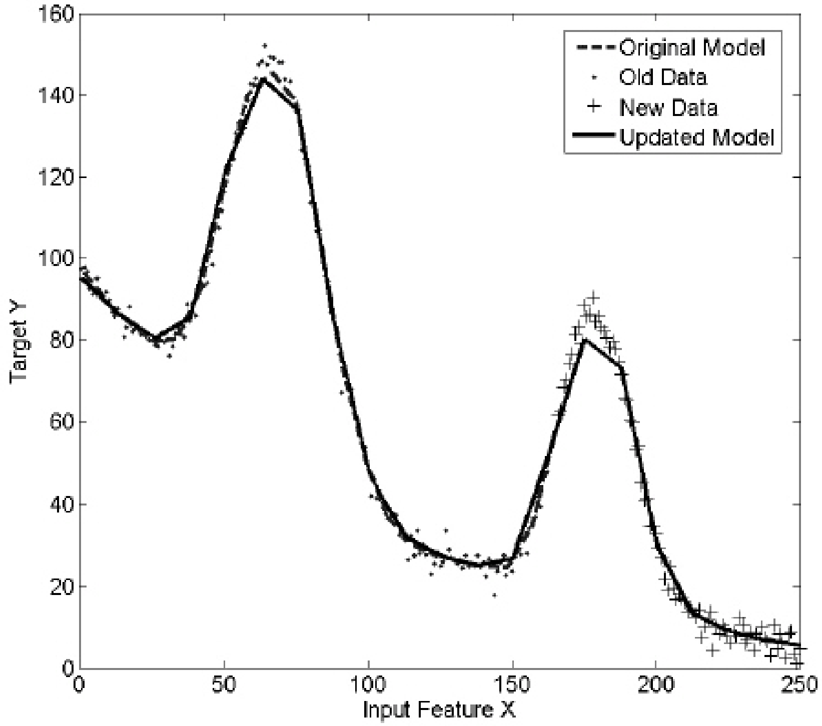
Failure Case (Parameter Adaptation)



Success of Update with EFS (evolving rule base)



(a)



(b)

Update Params only

Update Structure + Params

Some (Industrial) Requirements

- **Fast Online Identification** of Models/Classifiers (from scratch), usage in **On-line Production/Decision/Plaus.**
- **Updating and extending models to dynamic changes, new operating conditions, environmental influences etc.** => improvement of generalization capability of models

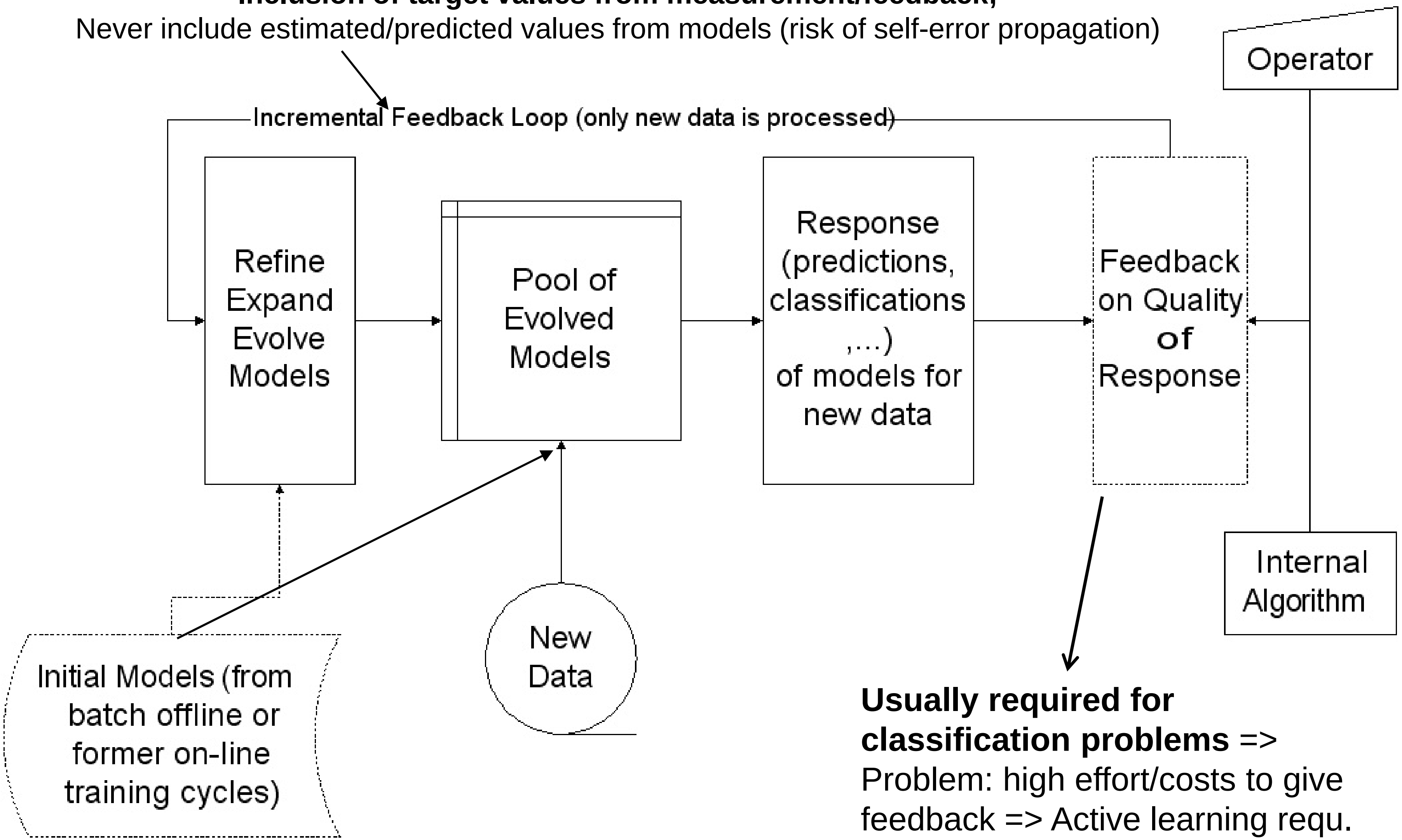
- **Hybrid Modelling**: Refining Knowledge-based Models (Rule-Base Systems) with Data

- **Huge Data Bases [Big Data]:** data which cannot be loaded into virtual memory at once => has to be processed block-wise

- **Enhanced Human-Machine Interaction Scenarios** (Operators give feedback and provide their expertise)

Evolving Modelling Framework

Inclusion of target values from measurement/feedback;
Never include estimated/predicted values from models (risk of self-error propagation)



Usually required for classification problems =>
Problem: high effort/costs to give feedback => Active learning requ.

Definition of Fuzzy Rule Types in EFS (Regression)

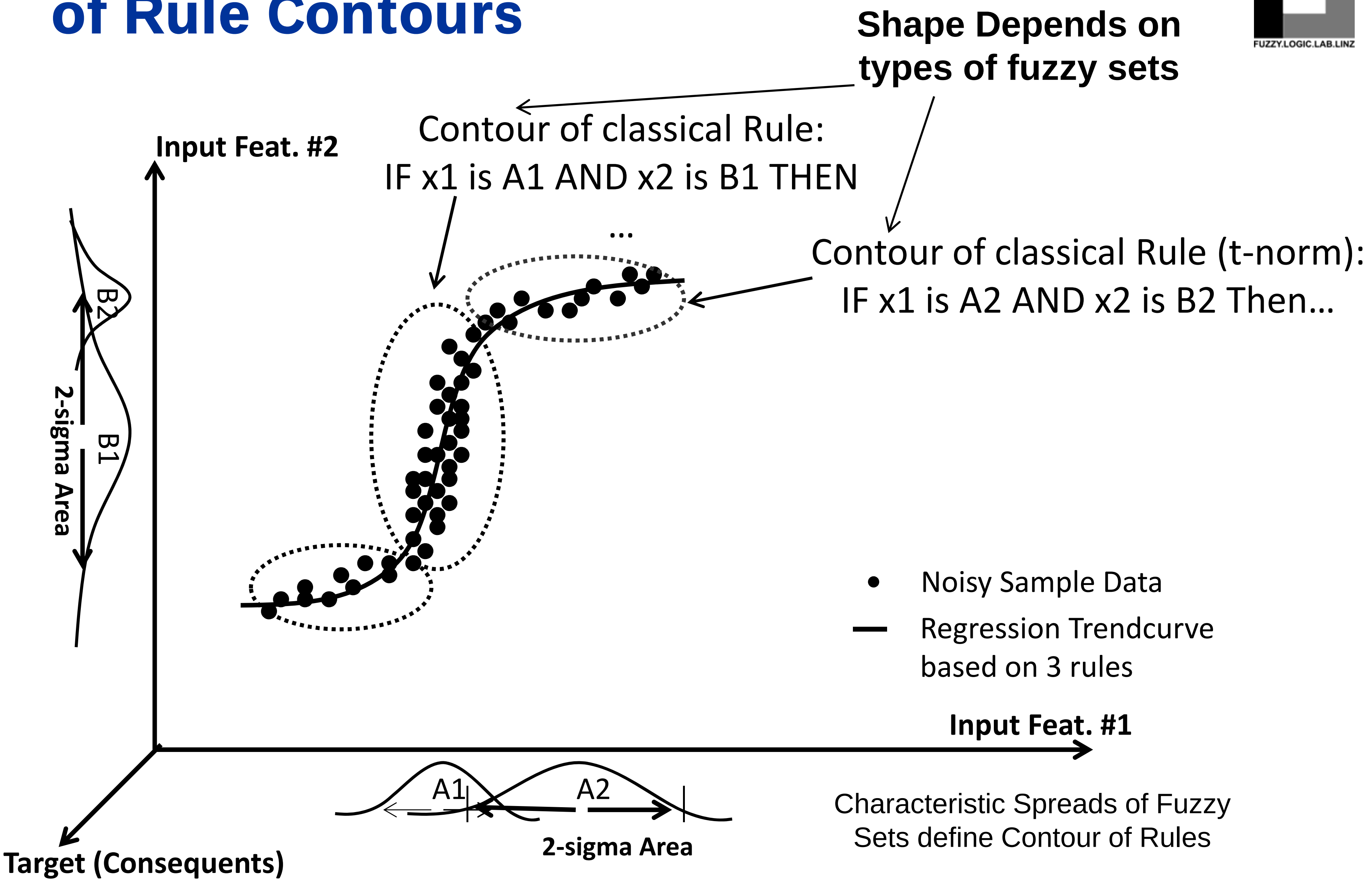
Arbitrary many ANDs

$$\text{Rule}_i: \quad \text{IF } (x_1 \text{ IS } \mu_{i1}) \text{ AND...AND } (x_p \text{ IS } \mu_{ip}) \text{ THEN } l_i \text{ IS } \Phi_i$$

- **Mamdani** (Mamdani, Assilian, FSS, 1977)
  - » Φ_i : fuzzy set
- **Sugeno**
  - » Φ_i : singleton numerical (real) value
- **Takagi-Sugeno** (Takagi and Sugeno, IEEE SMC, 1985)
  - » Φ_i : linear function (hyperplane)
- **Takagi-Sugeno-Kang** (Sugeno and Kang, FSS, 1988)
  - » Φ_i : polynomial functions, Gamma, Kernels (local SVM)
- **Generalized Takagi-Sugeno – non-axis parallel contours!** New developm. in Lughofer, Cernuda et al., Evolving Sys, 2015
  - » Antecedent part is a multidimensional kernel, e.g. $\exp(-(X - C_i)^T \Sigma_i^{-1} (X - C_i))$

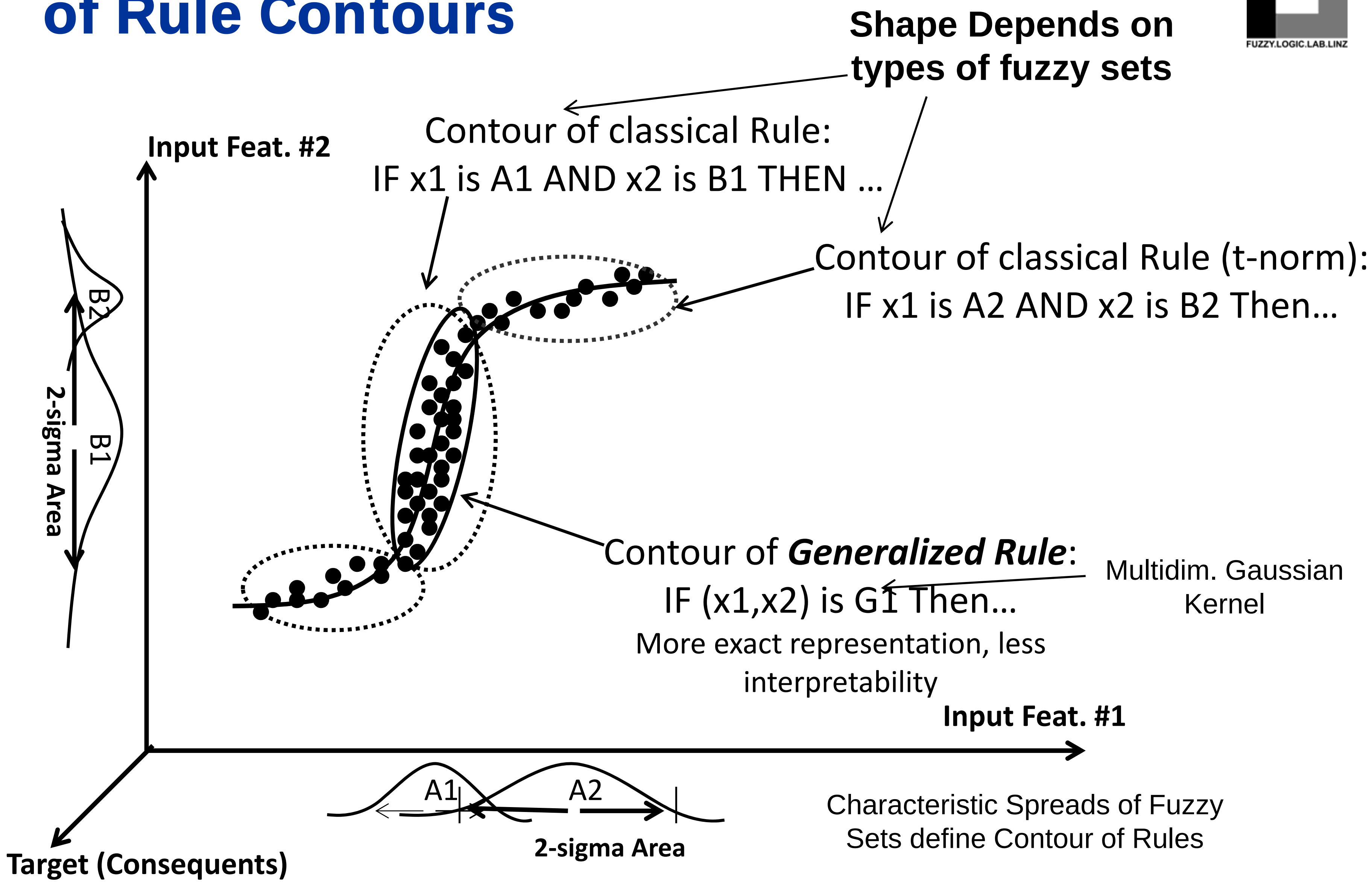
Covariance Matrix

Basic Geometric Interpretation of Rule Contours

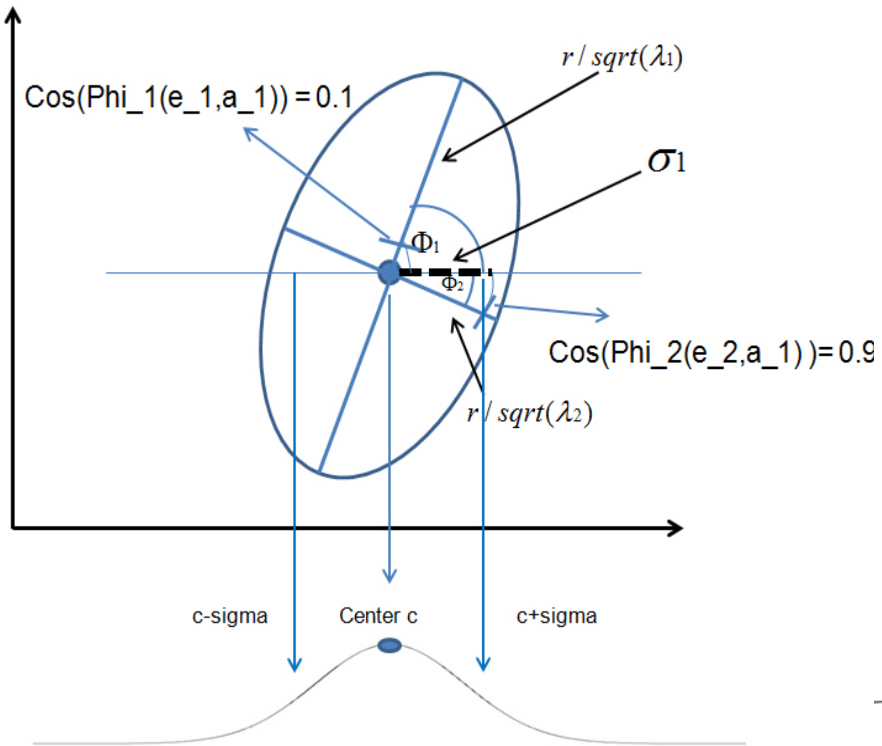


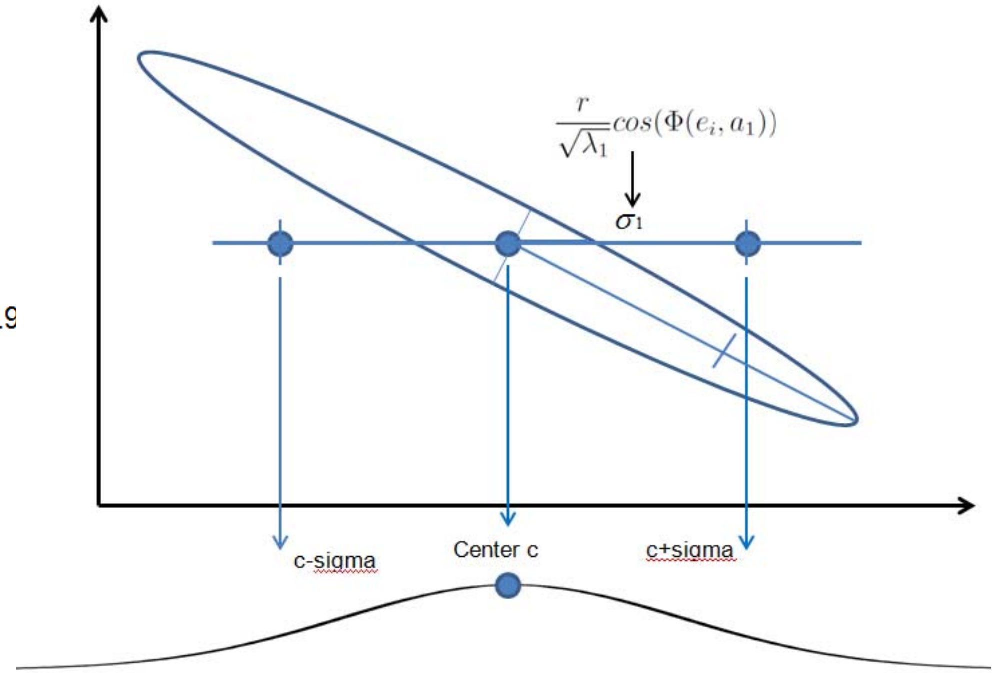
Characteristic Spreads of Fuzzy Sets define Contour of Rules

Basic Geometric Interpretation of Rule Contours



Extended Projection Concept to Assure Interpretability (Lughofer, Cernuda, Pratama, ES, 2015)





Eigenvalue Eigenvector

Case: Large Span Rules

$$\sigma_i = \max_{k=1,..,u} \left(\frac{r}{\sqrt{\lambda_k}} \cos(\phi(e_k, a_i))\right)$$

$$a_i = (0,0,...,1,...0)$$

Evolving Fuzzy Classifiers

- ## Single Model Architecture
  (classical: Kruse, Nauck, Ishibuchi), two options
  - » Φ_i : Singleton consequent labels
  - » Φ_i : Confidence Vector for each Class
    (able to represent class overlaps in rules)

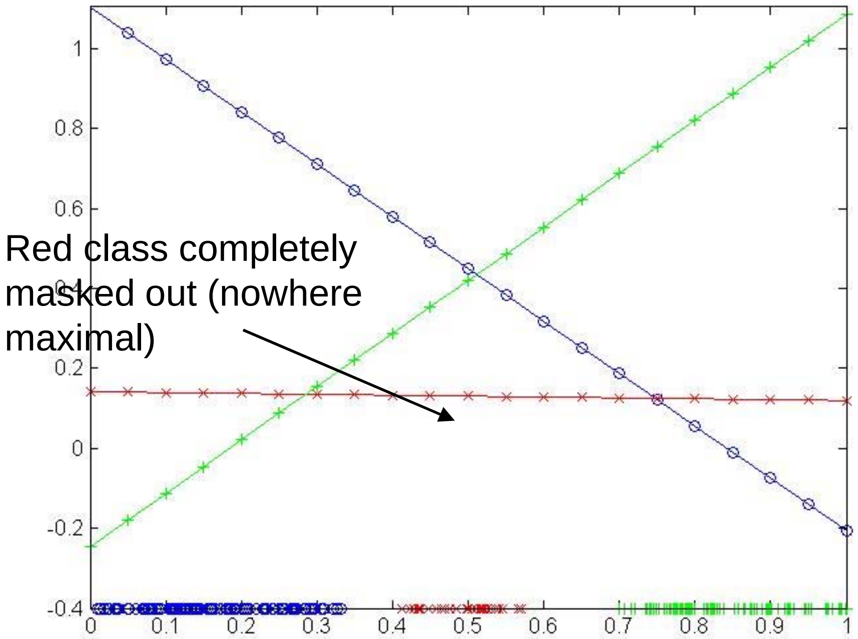
- ## Multi-Model Architecture for Class Decomposition:
  - » **Regression-based on Indicator = one-versus-rest** (Lughofer, Angelov, Zhou, FSS, 2008) resolving masking effect of linear version
  - » **All-Pairs** (Lughofer and Buchtala, IEEE TFS, vol. 21 (4), 2013) - binary classifier between each class pair!
    - – **less complex decision boundaries,**
    - – **new upcoming classes can be integrated quicker**
      (lower class imbalance effect)
    - – **enhanced output interpretation based on preference relation matrix**

Preference of
Class #1 over all othe

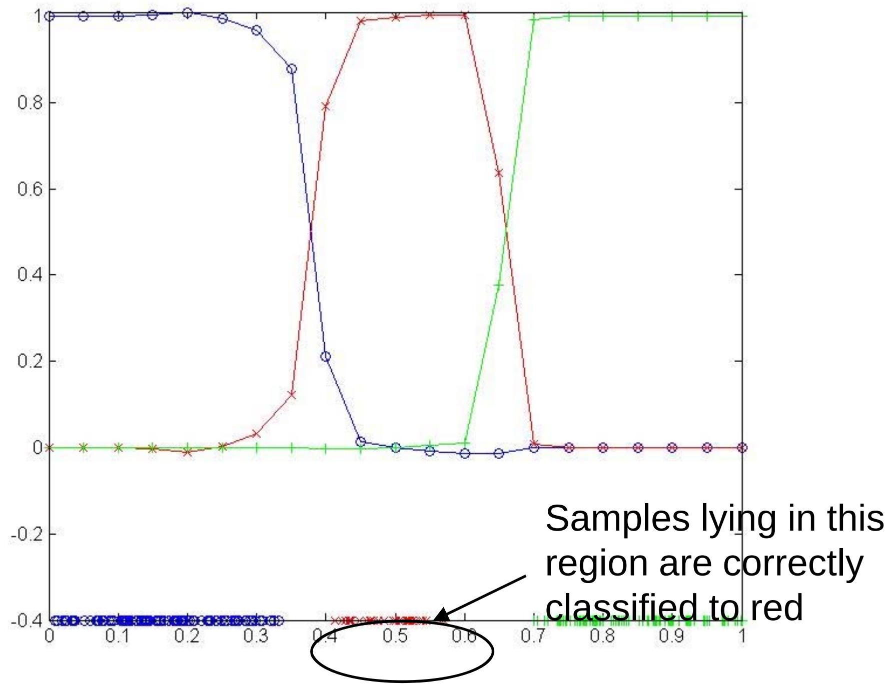
$$R = \begin{bmatrix} \boxed{0 \quad conf_{1,2} \quad conf_{1,3} \quad \ldots \quad conf_{1,K}} \\ conf_{2,1} & 0 & conf_{2,3} & \ldots & conf_{2,K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ conf_{K,1} & conf_{K,2} & conf_{K,3} & \ldots & 0 \end{bmatrix}$$

Preference degree of Class #2 over #1

Resolving Masking Effect in Indicator-Based Multi-Class Classification



Linear Regr. By Indicator matrix, 3 classes



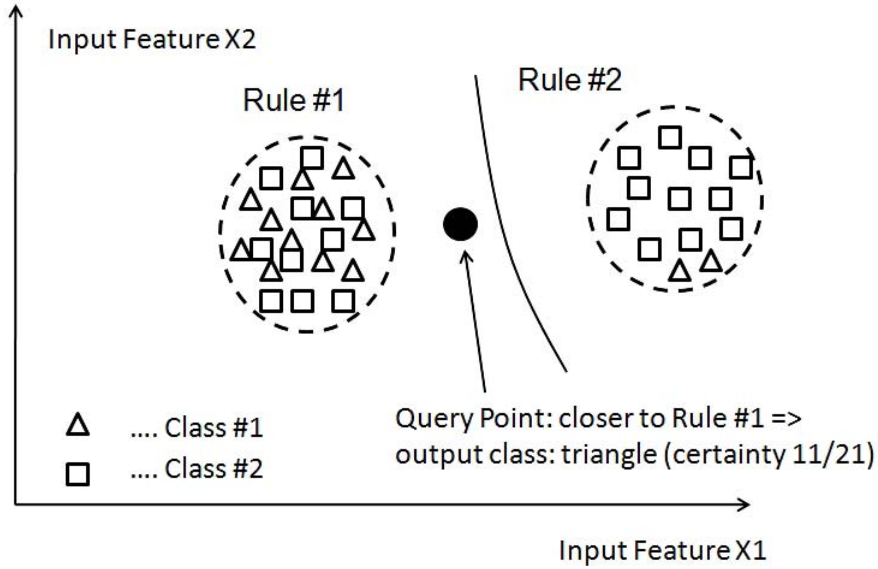
Fuzzy regression by indicator matrix, 3 classes

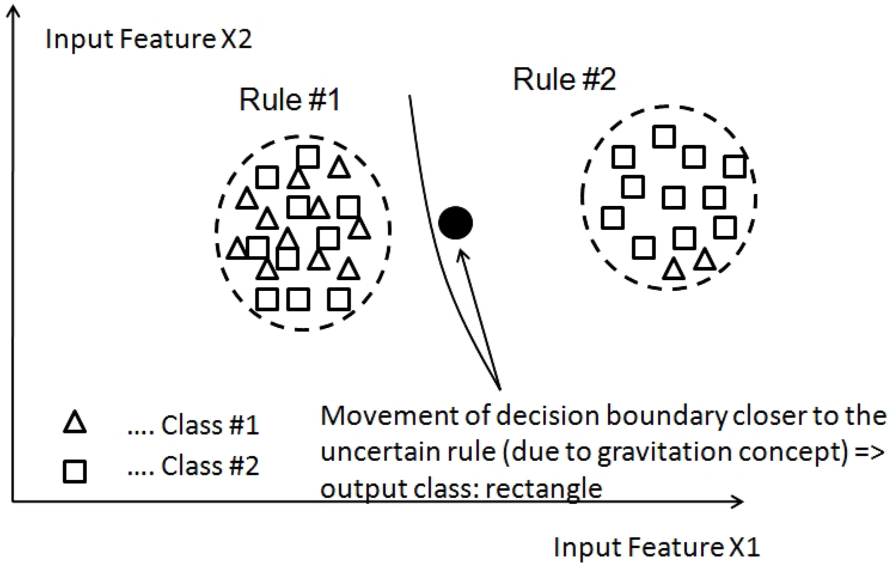
from all K models:

$$conf = \frac{\max_{m=1,\ldots,K} \hat{g}_m(\vec{x})}{\sum_{m=1}^{K} \hat{g}_m(\vec{x})}$$

where $\hat{g}_m(\vec{x}) = \hat{f}_m(\vec{x}) + |\min(0, \min_{m=1,\ldots,K} \hat{f}_m(\vec{x}))|$.

Geometric Interpretation of Classification in EFC





Winner-Takes-it-all

New Weighted Average Voting
(including class confidences)
Lughofer, Evolving Systems, 2012
Lughofer, Weigl et al. ASOC, 2015

$$L = k^* \text{ with } k^* = \text{argmax}_{1 \le k \le K} \left(\frac{\mu_1(\mathbf{x}h *_{1,k} + \mu_2(\mathbf{x})h*_{2,k}}{\mu_1(\mathbf{x}) + \mu_2(\mathbf{x})} \right)$$

$$conf_L = \frac{\mu_1(\mathbf{x})h *_{1,L} + \mu_2(\mathbf{x})h*_{2,L}}{\mu_1(\mathbf{x}) + \mu_2(\mathbf{x})}$$

Membership of nearest rule

Membership Second nearest rule with **different majority class** label L_2 than nearest rule

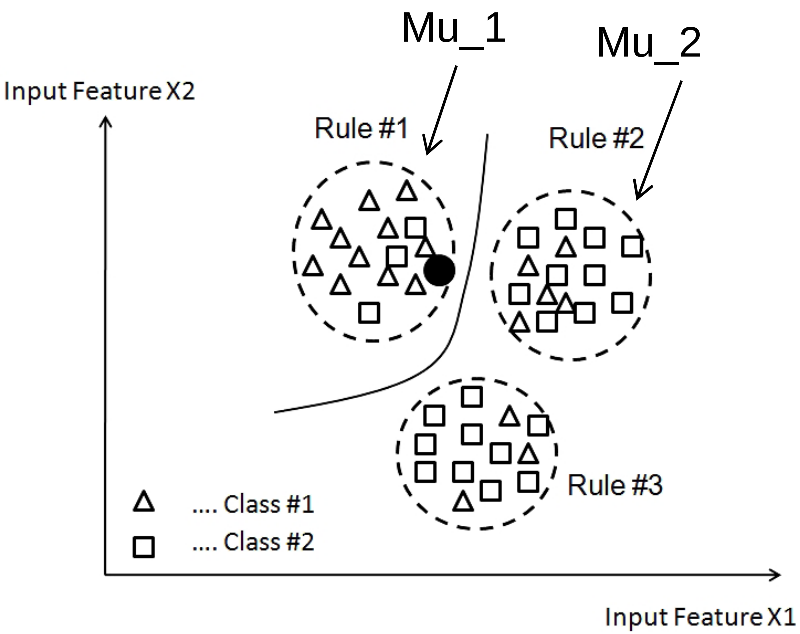
$$h*_{1,L} = \frac{h_{1,L}}{h_{1,L} + h_{1,L_2}} \qquad h*_{2,L} = \frac{h_{2,L}}{h_{2,L} + h_{2,L_2}}$$

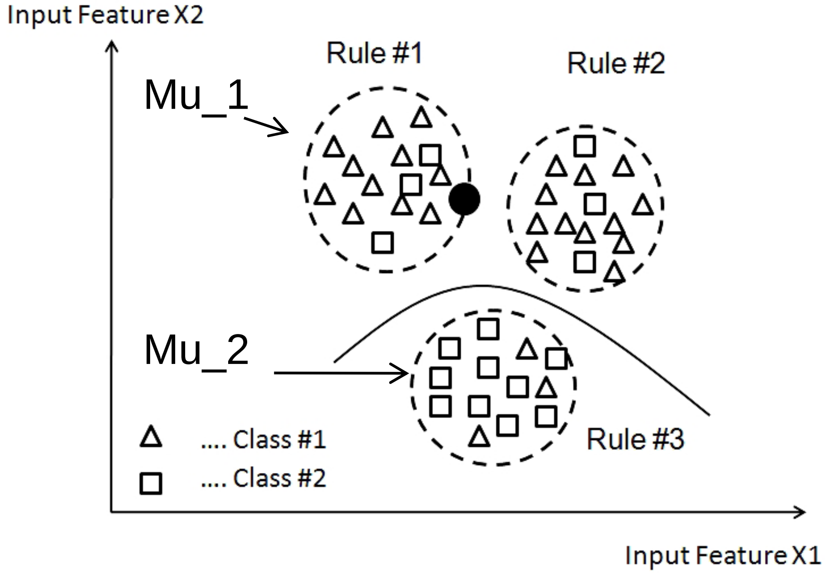
Behavior Decision Boundary

$$conf_L = \frac{\mu_1(\mathbf{x})h*_{1,L} + \mu_2(\mathbf{x})h*_{2,L}}{\mu_1(\mathbf{x}) + \mu_2(\mathbf{x})}$$

$$h*_{1,L} = \frac{h_{1,L}}{h_{1,L} + h_{1,L_2}} \quad h*_{2,L} = \frac{h_{2,L}}{h_{2,L} + h_{2,L_2}}$$

FUZZY.LOGIC.LAB.LINZ





Example: two nearest rules **having different majority classes** => decision boundary inbetween, query point with high conflict

Example: two nearest rules **having same majority class** => decision boundary between these two and a third rule with different majority class
=> query point with low clonflict

Evolving Fuzzy Systems Approaches in Literature (2004-2015)

Common Baseline: RFWLS + Variants

Method	Architecture	Ant. Learning	Cons. Learning	Rule Pr.	Forget.	Dim. Red.	# of P.	Note
AHLTNM[119]	TS fuzzy system	Incremental Split-and-Merge (own)	RFWLS	Yes	Yes, consequ. only	No	1-2	
DENFIS[112]	neuro-fuzzy system	Evolving Clustering (ECM)	RFWLS with distance weights	No	No	No	1-2	one of the first approaches
EFC-AP[90]	all-pairs classifiers	Evolving Clustering (eVQ)	RFWLS or Consequent labeling	Yes	No	No	1-2	Very high accuracy
EFP[116]	Mamdani + TS fuzzy system	Evolving Clustering (own) + RLM	RLM + RLS (global)	Yes	Yes, consequ. only	No	1-2	first attempt of recursive learning of non-linear params
eFPT[70]	Tree-based structure	Dynamic substitution with neighbor tree	None	Yes	No	Yes	3	good interpretability, exploits architecture from[130]
eFT[71]	Tree-based structure	Replacement of leaves with subtrees	RFWLS double-weight.	No	No	Yes	4	first approach of evolving fuzzy decision tree
eFuMo[113]	dynamic TS fuzzy system (lags)	Evolving Clustering (Gustafson-Kessel)	RFWLS double-weight.	Yes	Yes, consequ + ante.	No	3-5	capable of splitting of rules
eMG[48]	generalized TS fuzzy system	Participatory Learning	RFWLS double-weight.	Yes	No	No	4	first approach using generalized rules
ENFM[117]	generalized TS fuzzy system	Recursive Clustering (Gath-Geva)	RFWLS	Yes	Yes, consequ. only	No	3	
eClass[85]	single model, one-versus-rest classifiers and MIMO	Evolving clustering (eClustering)	RFWLS + or Consequent labeling	No	Yes, consequ. only	No	1-2	first one-vs-rest in EFC
ePL[124]	TS fuzzy system	Participatory Learning	RLS (global)	Yes	No	No	5	first usage of participatory learning[123]
eT2FIS[54]	Type-2 Mamdani	Native activation levels, parameter updates	same as antecedent learning	Yes	No	No	3	first evolving Type-2 Mamdani system
eTS(+)[114,115]	TS fuzzy system	Evolving clustering (eClustering)	RFWLS double-weight.	Yes	Yes, later in[131]	Yes	1-2	one of the pioneering methods (2004)
eTS-LS-SVM[55]	extended TS fuzzy system	Recursive clustering, suitability	Recursive Gauss-Newton	Yes	No	No	6 (2+4)	first approach for evolving extended TS fuzzy system

Huge Diversity

Evolving Fuzzy Systems Approaches in Literature (2004-2015)

Common Baseline: RFWLS + Variants

Method	Architecture	Ant. Learning	Cons. Learning	Rule Pr.	Forget.	Dim. Red.	# of P.	Note
FLEXFIS(++)[101,132]	TS fuzzy system	Evolving clustering (eVQ)	RFWLS	Yes	Yes, consequ. + ante.	No	1-2	first approach forgetting in ante, enhanced robustness
FLEXFIS-Class[85,133]	single model, one-vs-rest classifiers	Evolving clustering (eVQ)	RFWLS	No	Yes, consequ. only	Yes (smooth)	1-2	first one-vs-rest in EFC
FPA[134]	single model classifiers	incremental constraint-based optimization	weight vector update	No	No	No	2	
Gen-FLEXFIS[54]	generalized TS fuzzy system	Evolving clustering (extended eVQ)	RFWLS	Yes	Yes, consequent only	Yes (smooth)	1-2	first joint concept on-line dim. red. for gen. rules + pruning
GENEFIS[51]	generalized TS fuzzy system	Gen. adaptive resonance theory + stat. influence	FWGRLS	Yes	No	Yes (crisp)	3-4	gen. rules + on-line dim. red.
HAW-NFS[135]	neuro-fuzzy system	inc opt. (nonlinear mod. of Kaczmarz[136])	RLS (global)	No	No	No	3	first wavelet-based neuro-fuzzy
LOLIMOT inc.[126]	Tree-based structure	Node replacement (recursive split)	Recursive non-linear least squares	No	No	No	1-2	first approach for an incremental LOLIMOT[33]
PANFIS[50]	generalized TS fuzzy system	Extended Self-Organizing Map, statistical influence	FWGRLS	Yes	No	No	2	first proj. concept of gen. rules, stability proofs
eGK[85]	TS fuzzy system	Recursive clustering (GK)	RLS (global), RFWLS (local)	No	Yes, consequ. only	No	3-8	no rule evolution
(E)SAFIS[120,137]	TS fuzzy system (MIMO)	statistical influence, distance criterion	RLS (global), inc. stability with Lyapunov	Yes	No	No	3	one of the pioneering methods (2005-2006)
SAFIS-Class	single model using TS fuzzy system	statistical influence, distance criterion	RLS (global)	Yes	No	No	6	
SEIT2FNN[63]	Type-2 TS fuzzy system	coverage criterion, incremental steepest descent	RLS (global)	No	Yes, consequ. only	No	3	first approach for an evolving Type-2 fuzzy system
SOFMLS[26]	Mamdani	Evolving clustering (nearest neighborhood)	modified RLS with increased stability	Yes	No	No	4	first approach of an evolving Mamdani fuzzy system
SOFNN (imp)[103,118]	neuro-fuzzy system	coverage and system error criteria, rule enlargement	modified RLS (global) with weight parameter	Yes	No	No	2	one of the pioneering methods (2005)

Evolving Fuzzy Systems — Fundamentals, Reliability, Interpretability, Useability, Applications29

Huge Diversity

Algorithmic Key Steps in EFS (Common Denominator)

Algorithm 1.1. Key Steps in an Evolving Fuzzy Systems Learning Engine

(1) Load new data sample $\vec{x}$
(2) Pre-process data sample (e.g. normalization)
(3) **If** rule-base is empty, initialize the first rule with its center to the data sample $\vec{c} = \vec{x}$ and its spread (range of influence) to some small value; goto Step (1).
(4) **Else**, perform the following steps (5-10):
(5) Check if rule evolution criteria are fulfilled
  (a) If yes, evolve a new rule (Section 3.4) and perform body of Step (3) (without if-condition).
  (b) If no, proceed with next step.
(6) Update antecedents parts of (some or all) rules (Sections 3.4 and 3.2)
(7) Update consequent parameters (of some or all) rules (Sections 3.3 and 3.1)
(8) Check if the rule pruning/merging criteria are fulfilled
  (a) If yes, prune or merge rules (Section 3.4); goto Step (1).
  (b) If no, proceed with next step.
(9) Optional: Perform corrections of parameters towards more optimality.
(10) Goto Step (1).

Algorithmic Key Steps in EFS (Common Denominator)

Algorithm 1.1. Key Steps in an Evolving Fuzzy Systems Learning Engine

(1) Load new data sample $\vec{x}$
(2) Pre-process data sample (e.g. normalization)
(3) **If** rule-base is empty, initialize the first rule with its center to the data sample $\vec{c} = \vec{x}$ and its spread (range of influence) to some small value; goto Step (1).
(4) **Else**, perform the following steps (5-10):
(5) Check if rule evolution criteria are fulfilled — **Very approach dependent**
  (a) If yes, evolve a new rule (Section 3.4) and perform body of Step (3) (without if-condition).
  (b) If no, proceed with next step.
(6) Update antecedents parts of (some or all) rules (Sections 3.4 and 3.2) — **Very approach dependent**
(7) Update consequent parameters (of some or all) rules (Sections 3.3 and 3.1) — **Common Baseline: RFWLS**
(8) Check if the rule pruning/merging criteria are fulfilled
  (a) If yes, prune or merge rules (Section 3.4); goto Step (1).
  (b) If no, proceed with next step.

Very approach dependent: Generalization Concept in Lughofer et al., EUSFLAT, ES, 2011

(9) Optional: Perform corrections of parameters towards more optimality.
(10) Goto Step (1).

Consequent Learning (More or less Standardized)

- **Updating (Linear) Consequent Parameters** in TS(K)-fuzzy Systems by **recursive fuzzily weighted least squares**
  - » => induces a **local learning** effect
  - » **Converges in one iteration step** (=> global optimality) due to parabola of objective function!
  - » => Fast and robust convergence to hypothetical batch case
  - » **Recent extensions introduce:**
    1.) weight decay regularization (GFWLS)
    2.) adaptation regulation factor (based on the current approximation error), 3.) forgetting of older samples

Incremental Learning of Consequent Parameters (in TS fuzzy systems)

Local learning by a **recursive fuzzily weighted least squares** (consequ. func. for each rule separately)

$$e_i(k) = y(k) - \hat{y}_i(k)$$

Objective Function $$J_i = \sum_{k=1}^{N} \Psi_i(\vec{x}(k)) e_i^2(k) \longrightarrow \min_{w_i}$$

$$\hat{\mathbf{w}}_\mathbf{i}(k+1) = \hat{\mathbf{w}}_\mathbf{i}(k) + \gamma(k)(y(k+1) - \mathbf{r}^T(k+1)\hat{\mathbf{w}}_\mathbf{i}(k))$$

$$\gamma(k) = P_i(k+1)\mathbf{r}(k+1) = \frac{P_i(k)\mathbf{r}(k+1)}{\frac{1}{\Psi_i(\mathbf{x}(k+1))} + \mathbf{r}^T(k+1)P_i(k)\mathbf{r}(k+1)}$$

$$P_i(k+1) = (I - \gamma(k)\mathbf{r}^T(k+1))P_i(k)$$

start values:
alpha*I / 0-vec
from nearest rule (Cernuda, Lughofer 2012) or
weighted average from all rules
(Angelov, Filev 2004)

Membership degree to Rule #i

Weighted Inverse Hessian Matrix:
X^TQ_iX with Q_i weighting matrix

Local Verus Global Learning (All Params in All Rules at Once)

- **More Flexibility during learning**: Rules can be added, pruned or merged by not ‚disturbing' the recursive parameter update of the others
(Global learning requires a spec. Handling of the inverse Hessian in such cases as size changes)

- Comp. Complexity: **O(C(p+1)^2) versus O((C(p+1))^2)**
- **More stability** as dealing with smaller Hessian matrices
- **Local feature weighting capability** (per local region)
- **Better Interpretability** (piecewise linear approx. along real trend of curve, Yen, Wang and Gillespie, IEEE TFS, 1998)

Antecedent Learning - Principles

- **Incremental Partitioning of the Feature Space:**
  - » Concept 1: **Rule Evolution** (expansion of the knowledge base to explore new regions) – distance and density-based criteria
  - » Concept 2: **Rule Pruning** (contraction of the knowledge base to remove redundancies, obsolete rules etc.) => more compact, less time-intensive for on-line – distance and geometric criteria

- **Incremental Adaptation of Non-Linear Parameters:**
  1.) **analytical (incremental optimization, e.g. RLM),** - concept by Wang, Vrbanek, IEEE TFS, 2008!
  2.) **heuristics-based** (movement of rules, fuzzy sets), often done with inc. Clustering (many possibilities: eVQ, incremental fuzzy c-means, recursive Gustafsson.Kessel, recursive subtractive clustering, potentials as denisity estimators, evolving participatory learning, …)

Incremental Learning of Non-Linear Antecedent Parameters (for LS Problem)

Recursive LM (used in EFP by Wang and Vrbanek, 2008)

$$\Phi_{nonlin}(k+1) = \Phi_{nonlin}(k) + P(k)Jac^T(k)e(k) \tag{2.87}$$

with

$$P(k) = \frac{1}{\lambda_k}(P(k-1) - P(k-1)US^{-1}U^T P(k-1) \tag{2.88}$$

λ_m a forgetting factor (set to 1 in case of no forgetting) and

$$S = \lambda_k V + U^T P(k-1)U \tag{2.89}$$

The matrix S is a 2×2 matrix, therefore easily and quickly calculated in each update step; this is because

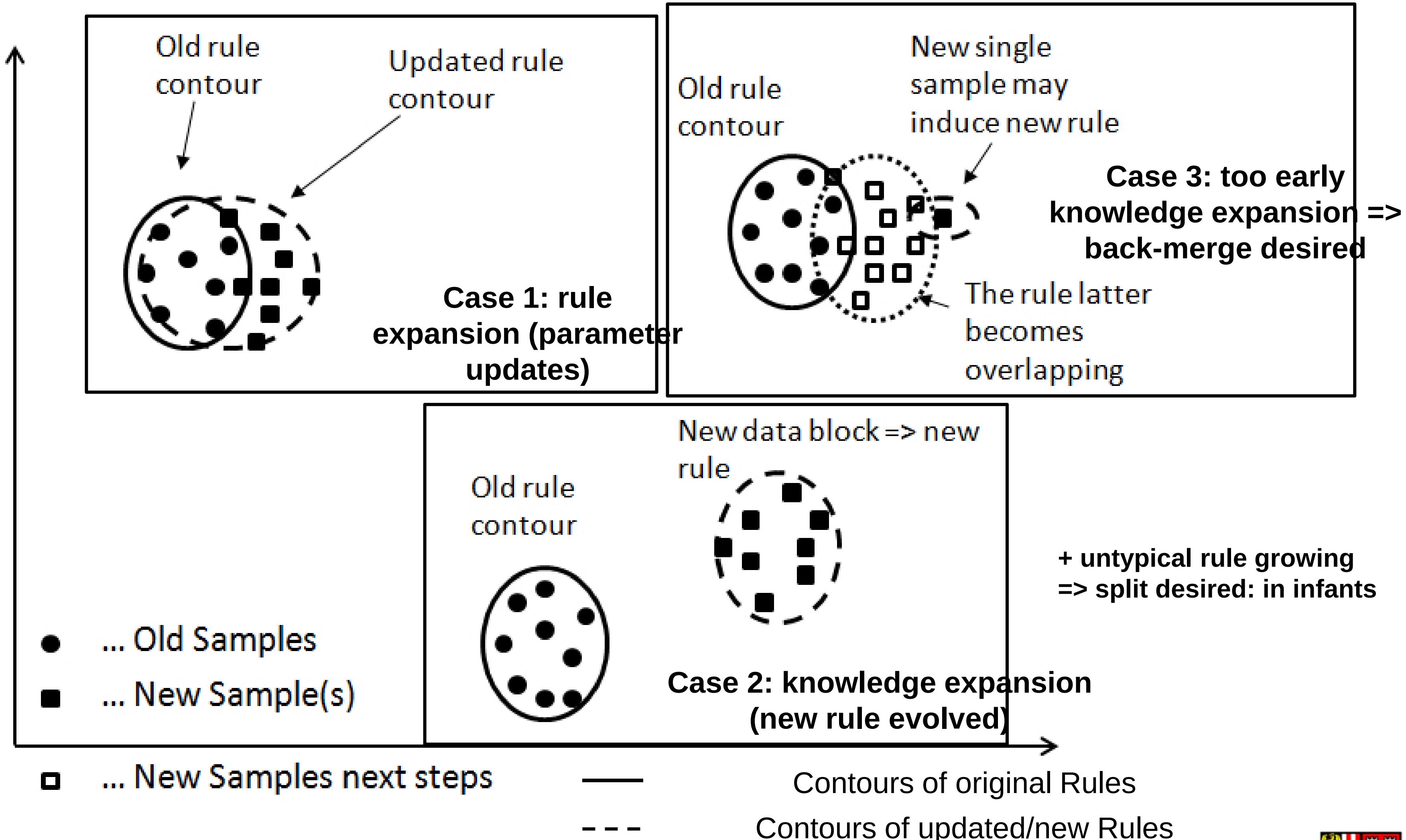
$$U^T = \begin{bmatrix} Jac^T(k) \\ 0\ldots010\ldots0 \end{bmatrix}$$

and

Update of only one param in one cycle

$$V^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & D\alpha_k \end{bmatrix}$$

Principal Cases of Data Stream Influence on Rules (over time)



Concrete Example: Antecedent Learning in FLEXFIS (Lughofer, IEEE TFS, 2008)

- is performed in the Clustering Space **=> each cluster is associated with one rule (local region)**
- **Based on evolving version of Vector Quantization (**eVQ, Lughofer, PR, 2008)
- **Three central steps (**Start: first center = first sample**):**
  - Checking whether a data sample x fits into cluster partition
    - Main Pos (Euclid): distance to nearest cluster higher than vigilance
    - Arb. Pos (Mahal): Statistical Tolerance Region
  - If yes, Movement of winning = nearest center towards sample

    $\mathbf{c}_{win}^{(new)} = \mathbf{c}_{win}^{(old)} + \eta(\mathbf{x} - \mathbf{c}_{win}^{(old)})$, Update ranges of influence of win. Cluster

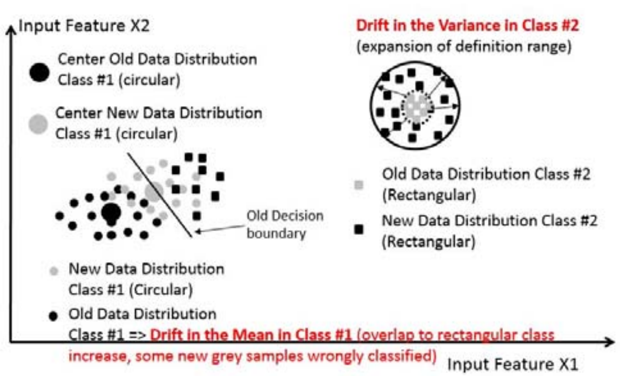
Ellipsoids in Main Position (Euclid) : $(k_i+1)\sigma_{ij}^2 = k_i\sigma_{ij}^2 + (k_i+1)\Delta c_{ij}^2 + (c_{ij} - x_{kj})^2 \quad \forall j = 1, ..., p+1$

Ellipsoids in Arbitrary Position (Mahal) : $\Sigma^{-1}(k+1) = \frac{\Sigma^{-1}(k)}{1-\alpha} - \frac{\alpha}{1-\alpha}\frac{(\Sigma^{-1}(k)(\mathbf{x}-\mathbf{c}))(\Sigma^{-1}(k)(\mathbf{x}-\mathbf{c}))^T}{1+\alpha((\mathbf{x}-\mathbf{c})^T\Sigma^{-1}(k)(\mathbf{x}-\mathbf{c}))}$

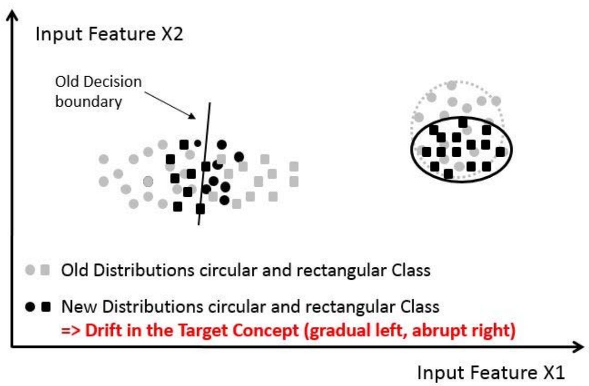
  - If no, evolve new Rule $\quad c_{C+1} = x \quad \Sigma_{C+1}^{-1} = \frac{\sum_{i=1}^{C}\Sigma_i^{-1}}{C} \quad \Sigma_1^{-1} = diag(\frac{frac}{\mathbf{range}^2})$

Concepts for Increased Stability

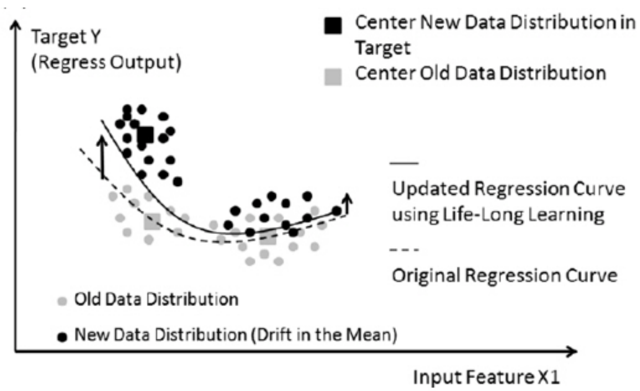
- **Automatic Drift handling (detection + reaction)**
  - » Older learned relations become obsolete



Drift in Input Space (Mean, Variance)



Drift in Target (classification)



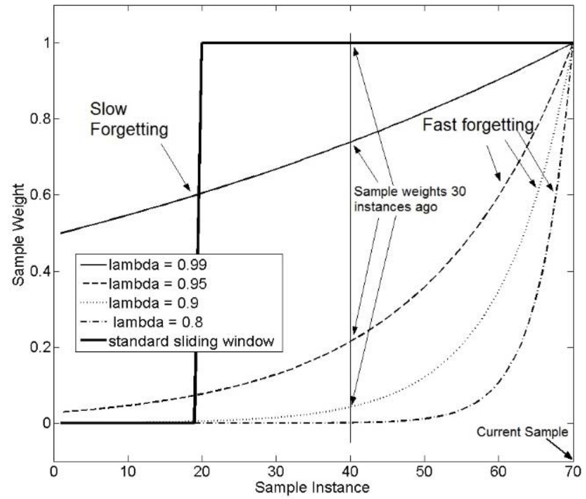
Drift in Target (regression)

- » **=> More flexibility in model updates required => Forgetting**
- » **Global and Local Drift Handling (different intensity! per Rule):**
  Shaker, Lughofer, Evolving Systems, 2014

$$\lambda_{i,t} = \lambda_{i,t-1} - direction(drift\_indicator) C_t A_{i,t}$$

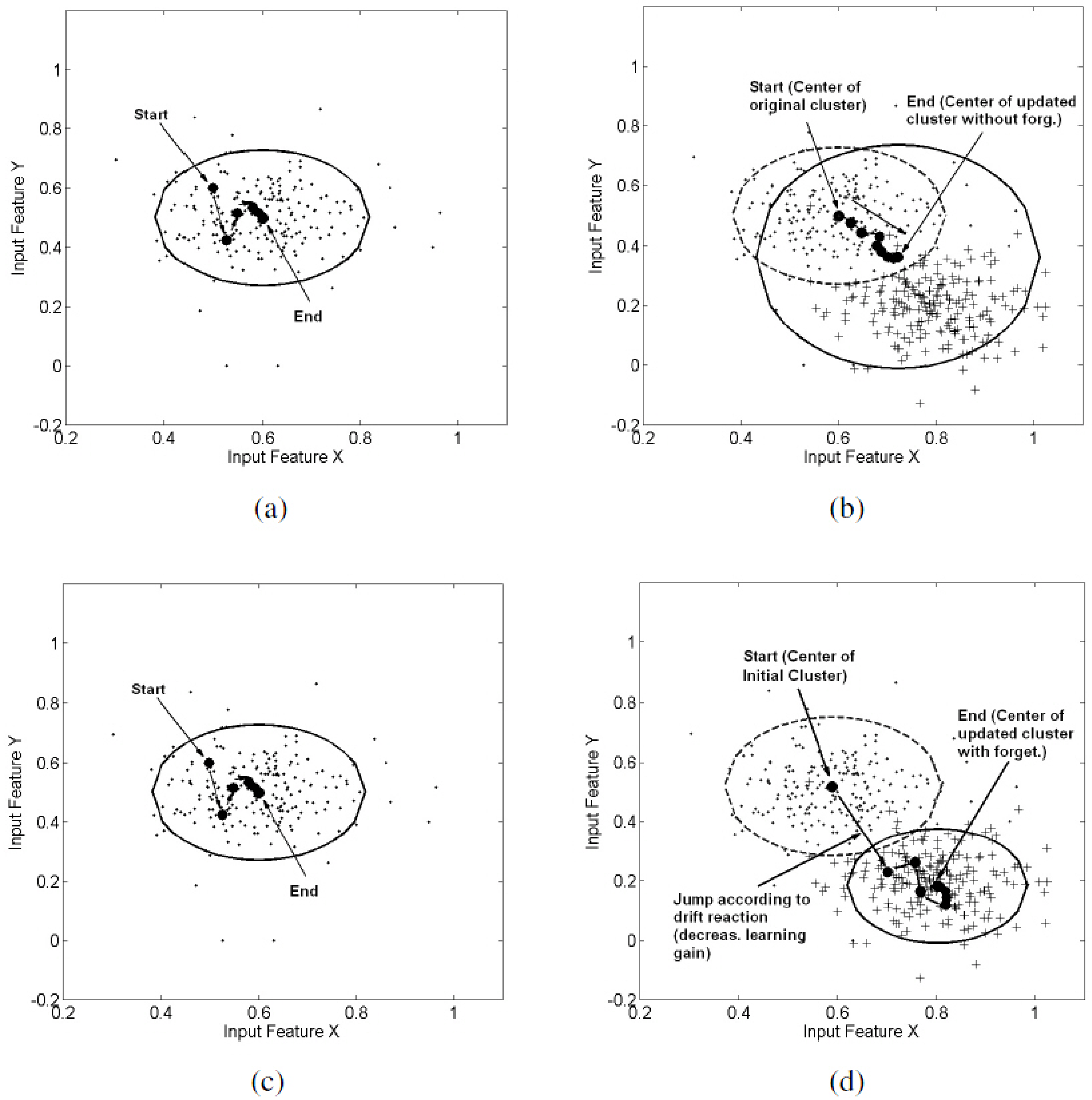
$$C_t = \frac{drift\_accum_{t-1} - drift\_accum_t}{\rho * rmse_t}$$

$$A_{i,t} = \frac{Err_i}{\sum_{k=1}^{C} Err_k}$$



Exponential forget

Drift Reaction - Example

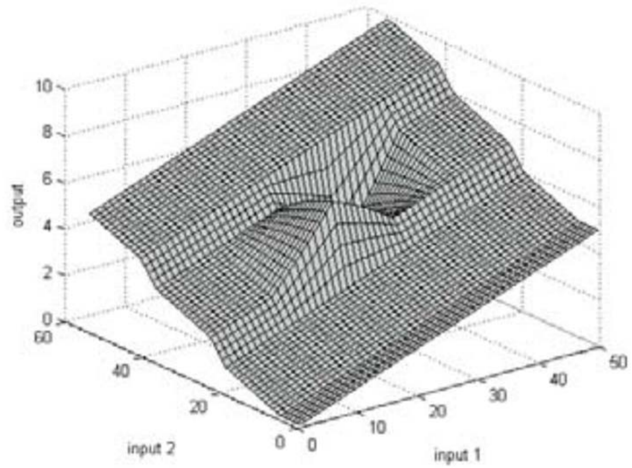


Conventional Update => cluster joins both distribution

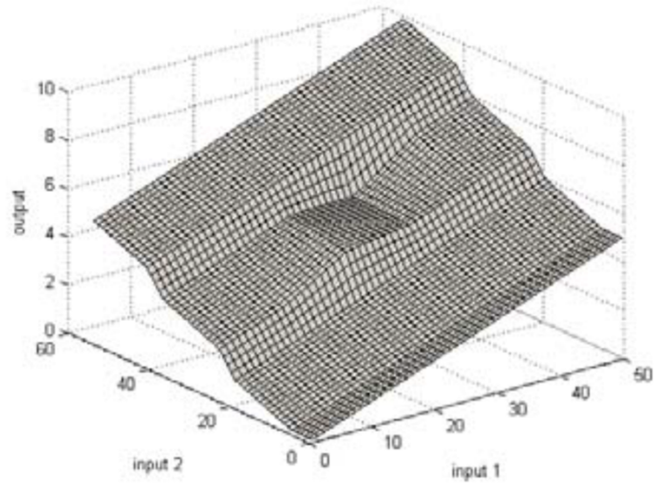
Update with forgetting => Cluster covers new distribution (old samples completely forgotten

Concepts for Increased Stability

- **Dynamic Curse of Dimensionality Reduction**
  - » Next slide
- **Incremental Smoothing (Regularization)**
  - » Rosemann, Brockmann, INS, 2012



Non-Smoothed Surface



Smoothed Surface after Several Incremental Steps

- **Convergence Analysis / Ensurance (FLEXFIS, PANFIS)**
  - » Bounded Sub-optimality in the least squares sense (correction terms integration), Proofed bounds on the system error

Idea of Dynamic Smooth Dim. Reduction

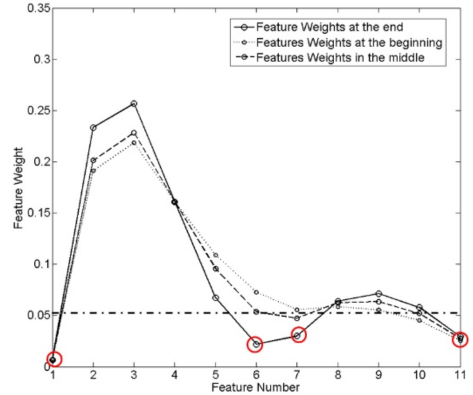
- **Introduction of the Concept of Feature Weighting**

  (EFS Approach Independent, Lughofer, Pratama, ES, 2015)

  » Criterion: expected statistical contribution $Cont_i = \lim_{N\to\infty} |w_{i0}| \frac{\sum_{k=1}^{N} \Psi_i(\mathbf{x}_k)/N}{\sum_{j=1}^{C} \sum_{k=1}^{N} \Psi_j(\mathbf{x}_k)/N}$

  » Important Features => High Weights (in [0,1])

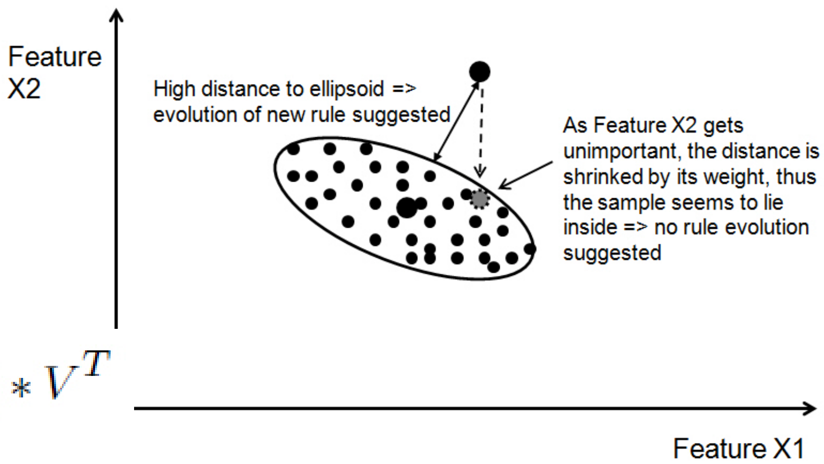
  » Update Feature Weights over Time =>
  smooth change in the importance of features –
  No crisp selection/deletion
  (as features may become important later again!)



- **Integration of Feature Weights in Rule Evolution Criterion**
  (Approach Dependent)

  » In Gen-Smart-EFS: weights into Mahalnobis distance

  » => re-scaled version of inverse covariance matrix:



$$\Sigma_i^{-1}(rescaled) = V * diag(\boldsymbol{\lambda}*) * D * diag(\boldsymbol{\lambda}*) * V^T$$

$$diag(\lambda*) = diag(\Lambda_*) = diag(V^T * diag(\lambda) * V)$$

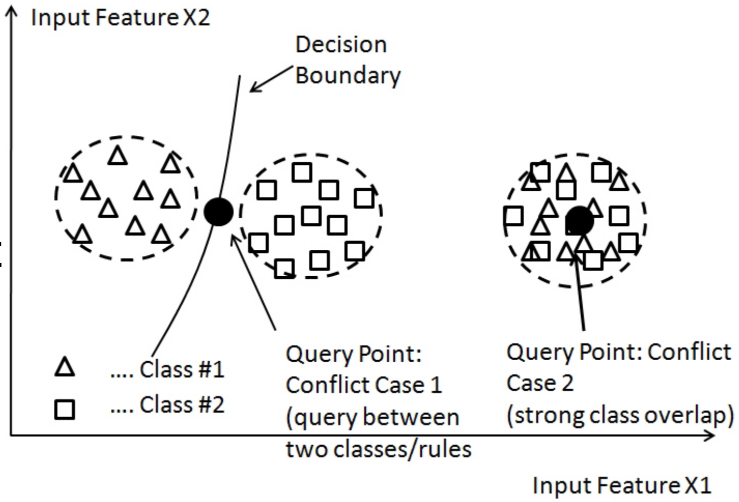
Original feature weights

Concepts for Increased Useability

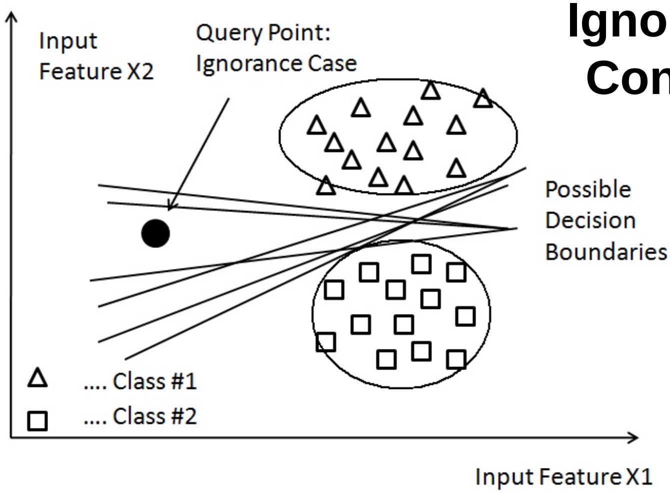
- **Interpretability:** understanding the model components and behavior
  - » **Position Paper:**
    Edwin Lughofer, Online Assurance of Interpretability Criteria in Evolving Fuzzy Systems - Achievements, New Concepts and Open Issues, *Information Sciences*, vol. 251, pp. 22-46, 2013
  - » **Criteria examined:** Distinguishability, Simplicity, Consistency, Coverage, Input/Output Behavior, Feature Weights, Rule Lengths, Rule Weights, Local Property, Interpretation of Consequents (red = essential for transparency)

- **Reliability:** interpretation of model predictions (useable for AL)

Conflict Concept



Ignorance Concept



+ preference relation interpret. in case of AP

Merge Criteria for Assuring Distinguishability and Compactness

FUZZY.LOGIC.LAB.LINZ

Final Criterion: AND Connection of 1, 2 and 3:

1

$$d(c_A, c_B) \leq \frac{\sum_{j=1}^{p} |c_{A;j} - c_{B;j}|(fac * \sigma_{A;j} + fac * \sigma_{B;j})}{\sum_{j=1}^{p} |c_{A;j} - c_{B;j}|} + \epsilon$$

Touching (eps=0)
Slightly Overlapping (eps<0 big)
Close (eps>0 small)

Euclidean distance between centers

2

$$V_{merged} \leq p(V_A + V_B)$$

Dimensionality

Homogenuity Criterion (restricting the volume of the merged ellipsoid)

3 **For Regression:** Similarity of Consequents high (dihedral angle)

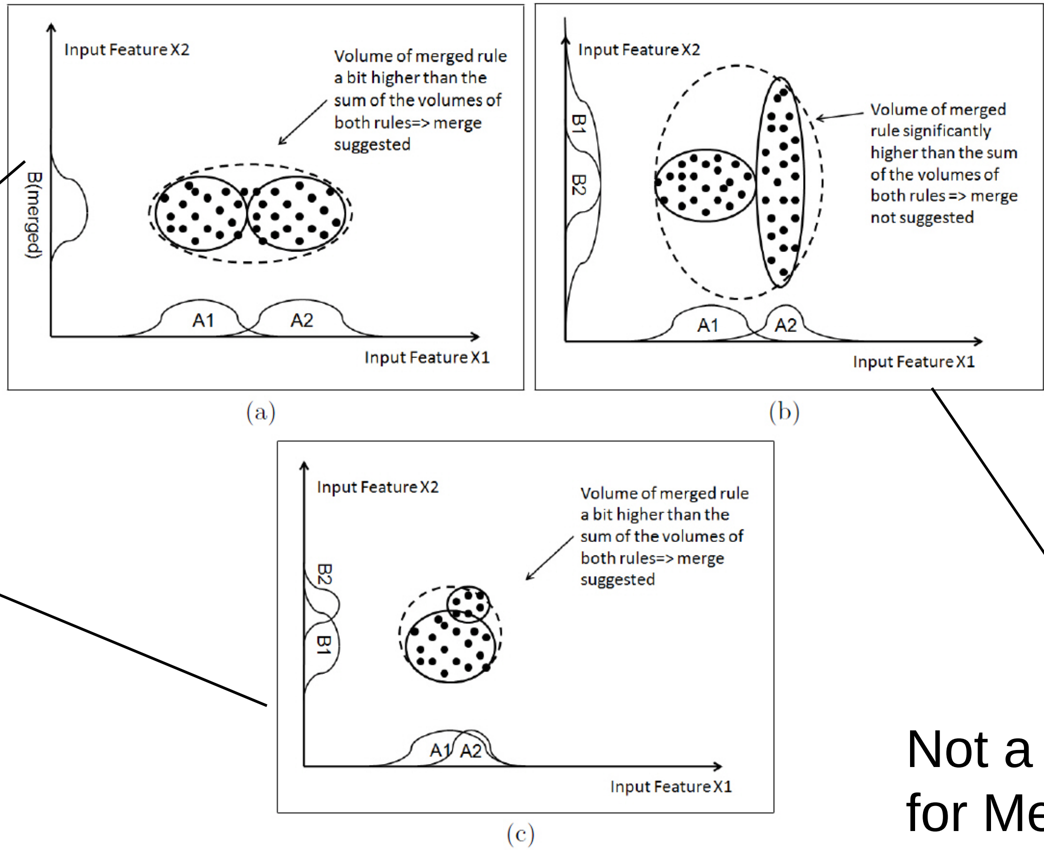
$$S_{cons}(\vec{w}_A, \vec{w}_B) \geq thresh$$

3 **For Classification:** majority class in both rules are the same

Homogenuity Criterion: Impact

» Basic Idea: **Merging of rules which are touching,** slightly overlapping, very close to each other and **fulfilling homogenuity criterion**

Candidate for Merging

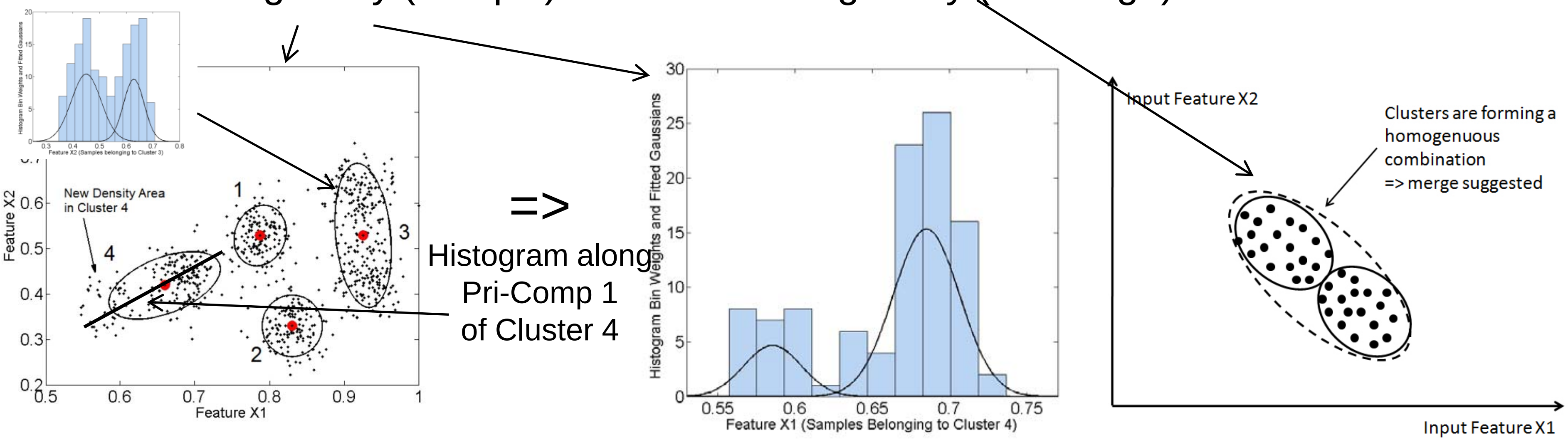


(a) (b) (c)

Not a Candidate for Merging

Concepts for Increased Useability

- **Single-Pass Active Learning**
  (Lughofer, PR / ES, 2012 for classification; Cernuda, Lughofer et al. 2014 for regression)
  - » **Reducing Annotation / Measurement Effort**
  - » Based on Reliability Concepts and Beyond (certainty-based sampling)
- **(Towards) Plug-and-Play Functionality: Compensating „unlucky“ initial settings of learning parameter(s)**
  - » **Possibility 1:** adaptive learning parameters (e.g. steering rule evolution vs. Rule update) according to the current data stream characteristics
  - » **Possibility 2:** dynamic split-and-merge of clusters (rules), resolving intra-heterogenuity (=> split) and inter-homogenuity (=> merge)



Applications of EFS
and
Some Results
(from own Projects)

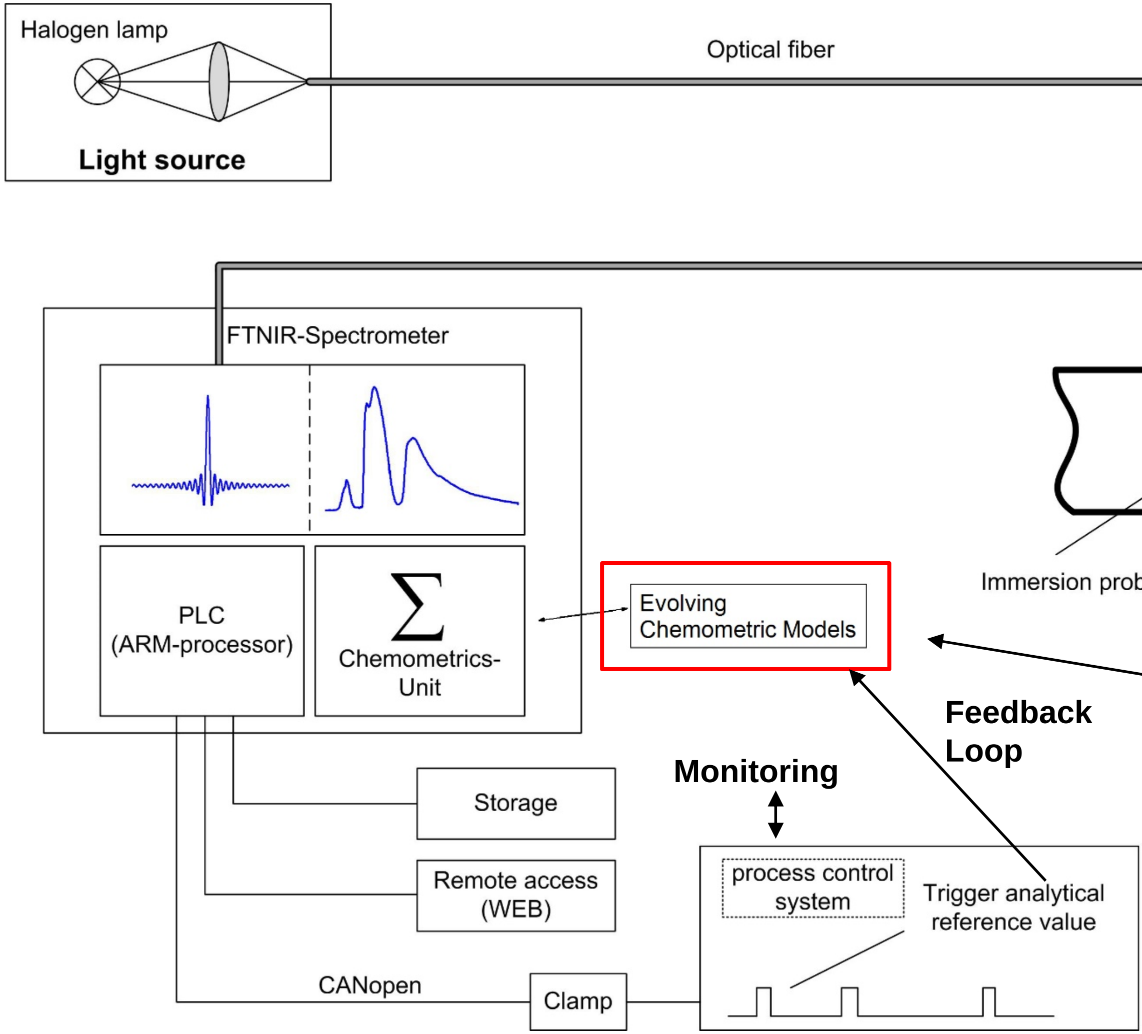
Applications of EFS – OverView

Application Type/Class (alphabetically)	EFS approaches (+ refs)	Comment
Active learning / human-machine interaction	FLEXFIS-Class,[191,192] EFC-AP,[86] FLEXFIS-PLS[148]	reducing the annotation effort and measurement costs in industrial processes
Adaptive on-line control	evolving PID and MRC controllers in,[193] eFuMo,[115] rGK,[194] self-evolving NFC,[109] adaptive controller in[139]	design of fuzzy controllers which can be updated and evolved on-the-fly
Bioinformatics	EFuNN[195]	specific applications such as ribosome binding site (RBS) identification, gene profiling
Chemometric Modeling and Process Control	FLEXFIS++;[38,104] the approach in[37]	The application of EFS onto processes in chemical industry (high-dim. NIR spectra)
EEG signals classification and processing	eTS,[196] epSNNr[197]	time-series modeling with the inclusion of time delays
Evolving Smart Sensors (eSensors)	eTS+[168] (gas industry),[169,198] (chemical process industry), FLEXFIS[199] and PANFIS[53] (NOx emissions)	evolving predictive and forecasting models in order to substitute cost-intensive hardware sensors
Forecasting and prediction (general)	AHLTNM[121] (daily temp.), eT2FIS[67] (traffic flow), eFPT[72] (Statlog from UCI), eFT[73] and eMG[51] (short-term electricity load), FLEXFIS+[172] and GENEFIS[54] (house prices), LOLIMOT inc.[128] (maximum cylinder pressure), rGK[194] (sales prediction) and others	various successful implementations of EFS

Applications of EFS - Overview

Financial domains	eT2FIS,[67] evolving granular systems,[200] ePL,[201] PANFIS,[53] SOFNN[202]	time-series modeling with the inclusion of time delays
Identification of dynamic benchmark problems	DENFIS,[114] eT2FIS,[67] eTS+,[117] FLEXFIS,[103] SAFIS,[123] SEIT2FNN,[66] SOFNN[202]	Mackey-Glass, Box-Jenkins, etc.
On-line fault detection and condition monitoring	eMG for classification,[127] FLEXFIS++,[35,203] rGK[194]	EFS applied as SysID models for extracting residuals
On-line monitoring	eTS+[168] (gas industry)	supervision of system behaviors
Robotics	eTS+[204]	in the area of self-localization
Time-series modeling	DENFIS,[205] ENFM[119] and eTS-LS-SVM[58] (sun spot)	local modeling of multiple time-series versus instance-based learning
User behavior identification	eClass and eTS,[206,207] eTS+,[208] FPA[136]	analysis of the user's behaviors in multi-agent systems, on computers, indoor environments etc.
Video processing	eTS, eTS+[209,210]	including real-time object id., obstacles tracking and novelty detection
Visual quality control	EFC-AP,[92] FLEXFIS-Class[211],[80] pClass[89]	image classification tasks based on feature vectors

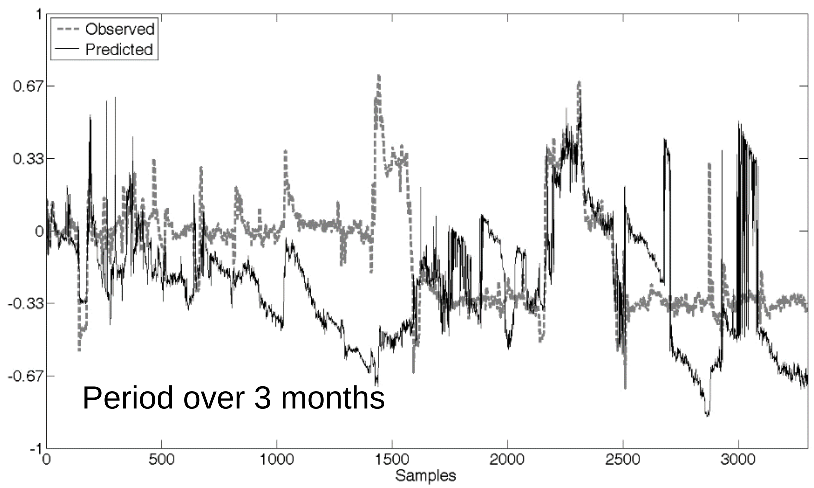
Evolvable Chemometric Models in Chemical Production Systems



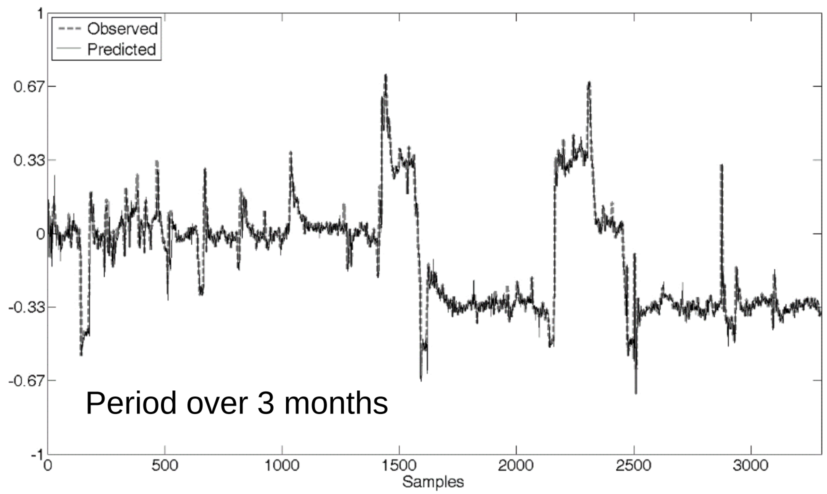
eChemo paradigm
(recent development
2012-2015)

Some Results (eChemo paradigm)
(Cernuda, Lughofer et al, ACA 2012 - EFS + PLS for Chemo models)

- Dynamic Adaptation of Chemometric Models in **Viscose Production**



Prediction On-line Data with Static PLS Models (State of the Art)

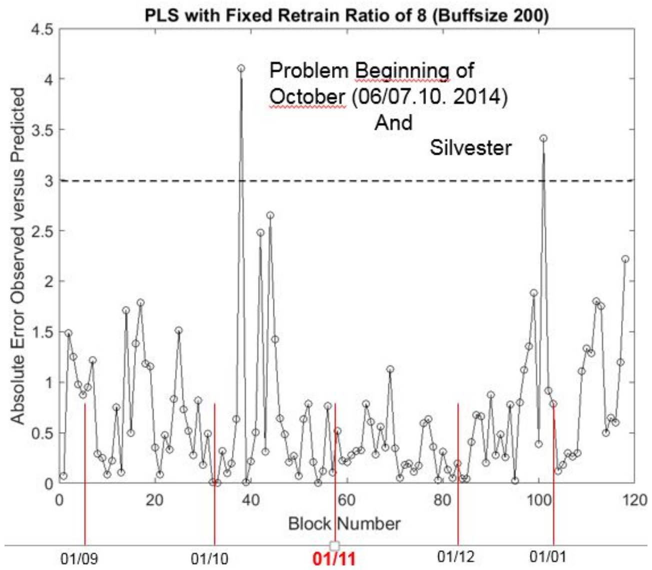


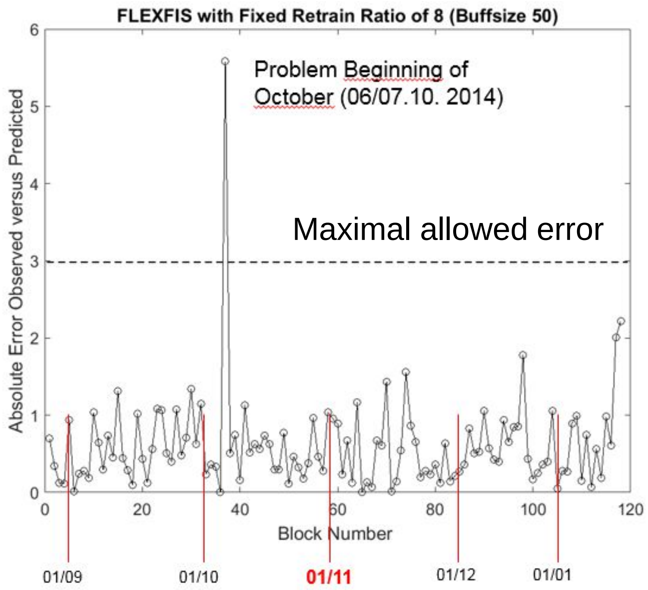
Prediction On-line Data with Dynamic Evolving Fuzzy Models (eChemo paradigm)

- Coupling with Active Learning => **Reduction of Measurements + Model Updates by approx. 98%** => Significant Cost Reduction (Cernuda, Lughofer et. al, CILS, 2014)

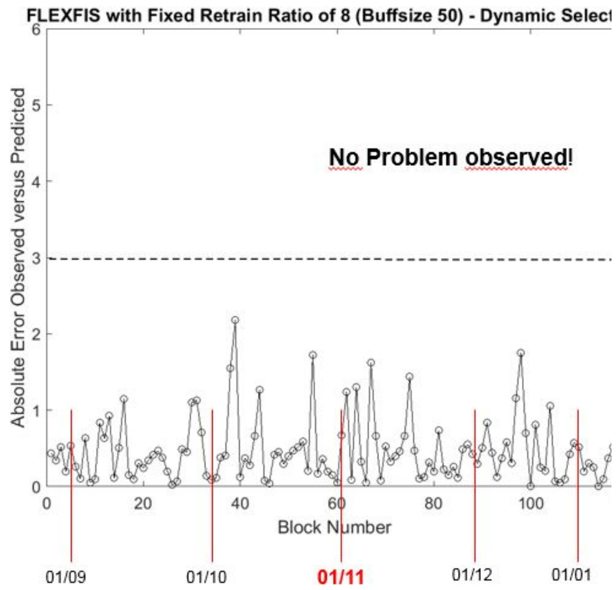
Active Learning Results in On-line Melamin Resine Production
(Sliding Window Based Approach)

No Active Sample Selection, blind equidistant adaptation





Active Sample Selection

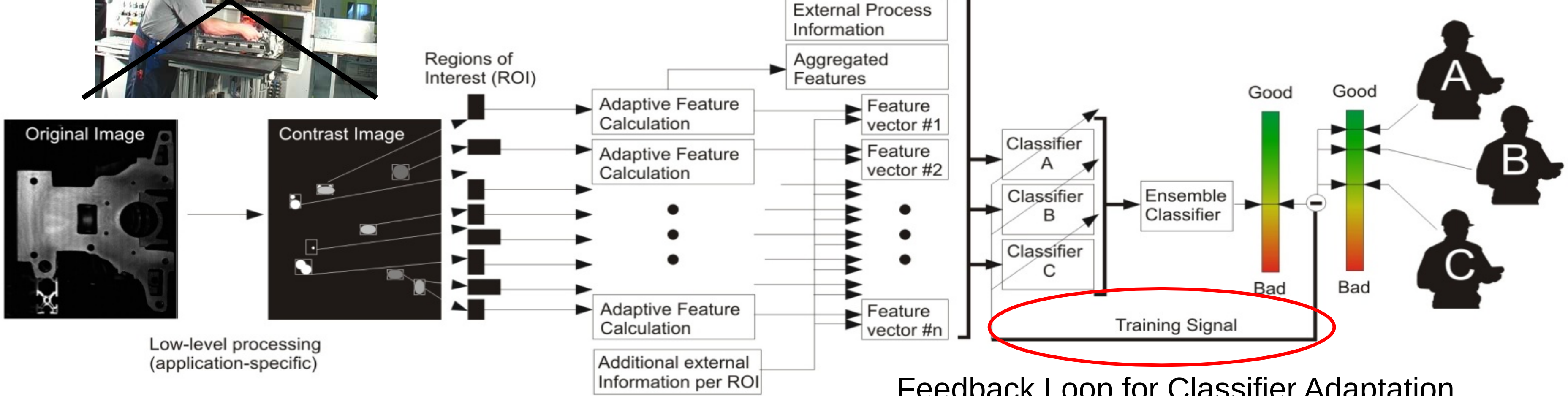


Savings: ~13% (~87% Samples used for Selection)

Automatic On-line Surface Inspection with Adaptivity of Image Classifiers (binary: good/bad reward)

Eitzinger, Lughofer et. al, MVA, 2010
Lughofer, Smith et al., IEEE SMC-A, 2009

Getting Rid of Manual Checks!



Feedback Loop for Classifier Adaptation
(based on native good/bad reward)

- **Achievements:**
  - » Common Interfaces between Images and Classifiers
  - » **Extracting Arbitrary Regions of Interests with Clustering Methods**
  - » **Machine-learning Classifiers with Accuracy > 99%** (on three real-world surface inspection scenarios)
  - » **Resolving Contradicting Input** from Several Operators (Ensembles)
  - » **On-Line Adaptation and Evolution** of Image Classifiers (with EFC)

On-line Surface Inspection - Results

(Lughofer, Eitzinger, Guardiola, On-line Quality Control, in: Learning in Non-Stationary Environments, Springer, New York, 2012)

FUZZY.LOGIC.LAB.LINZ

	CD imprint (#2)	*Eggs*	*Rotor*	*Bearing*
Static Image Classifiers (trained in off-line mode)				
eVQ-Class variant A	75.69	91.55	66.67	63.75
eVQ-Class variant B	88.82	90.11	66.67	64.65
EFC SM	78.82	95.20	66.67	60.73
EFC MM	73.53	95.89	54.67	55.59
k-NN	79.61	91.51	53.33	58.30
CART	78.82	91.78	52.00	65.26
Evolved Image Classifiers (updated in on-line mode)				
eVQ-Class variant A	89.22 (+13.5)	91.12(-0.4)	86.67 (+20)	67.67 (+3.9)
eVQ-Class variant B	90.39 (+1.6)	93.33 (+3.2)	86.67 (+20)	67.98 (+3.3)
EFC SM	78.82 (+0.0)	96.21 (+1.0)	64.00 (-2.6)	63.14 (+2.4)
EFC MM	87.65 (+14.1)	97.19 (+1.3)	78.67 (+24)	65.56 (+10.0)
k-NN (re-trained)	90.98 (+11.4)	96.06 (+4.6)	74.67 (+21.3)	59.52 (+1.2)
CART (re-trained)	90.59 (+11.8)	97.02 (+5.2)	52.00 (+0.0)	69.18 (+3.9)
+max %	14.1	5.2	21.3	10.0

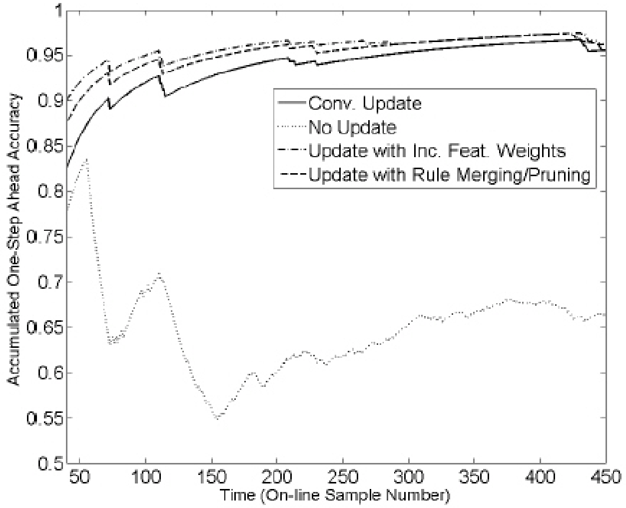
Very slow – retraining on each sample, just for benchmark purposes

Multi-class problem

Data Set	*EFC SM*	*EFC MM*	*EFC AP*	*eVQ-Class A*	*eVQ-Class B*	*CART*	*k*-NN
CD imprint	62.0 ± 2.5	73.1 ± 1.1	$\mathbf{82.6 \pm 1.5}$	64.1 ± 2.4	74.9 ± 1.6	75.81 ± 2.1	73.85 ± 2.9

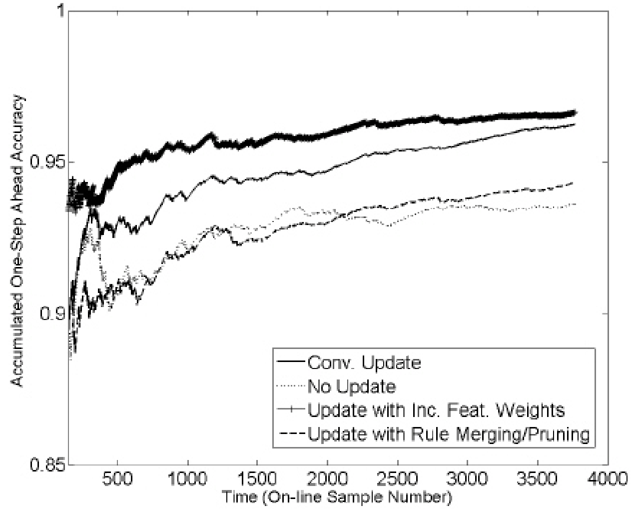
Accumulated Accuracies

CD-Imprint with merging and feat weights



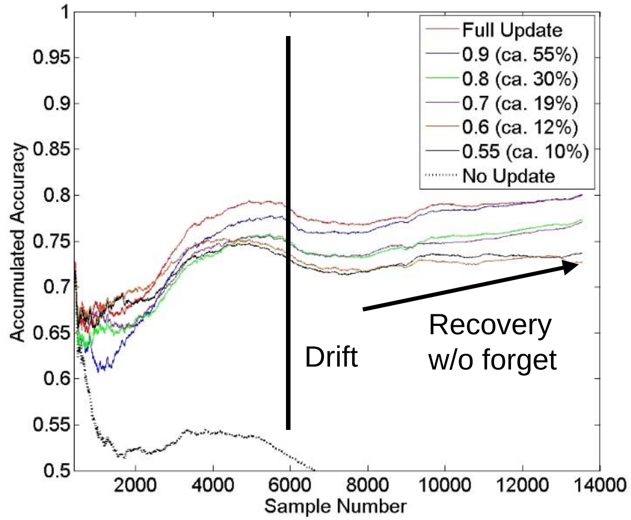
(a)

Eggs with merging and feat weights



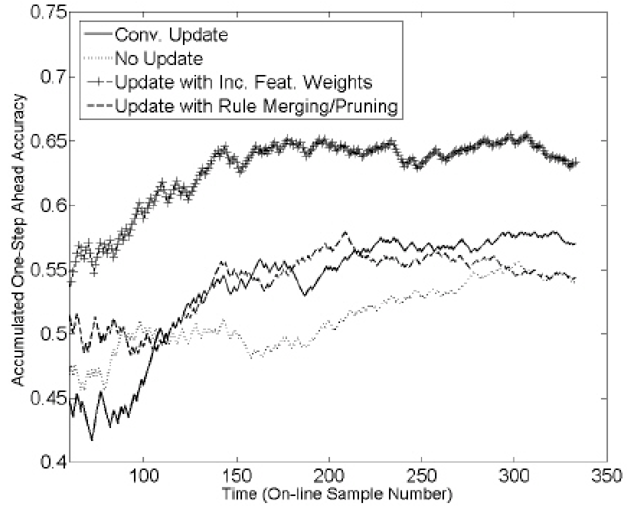
(b)

Micro Fluid Chips with Active Learning



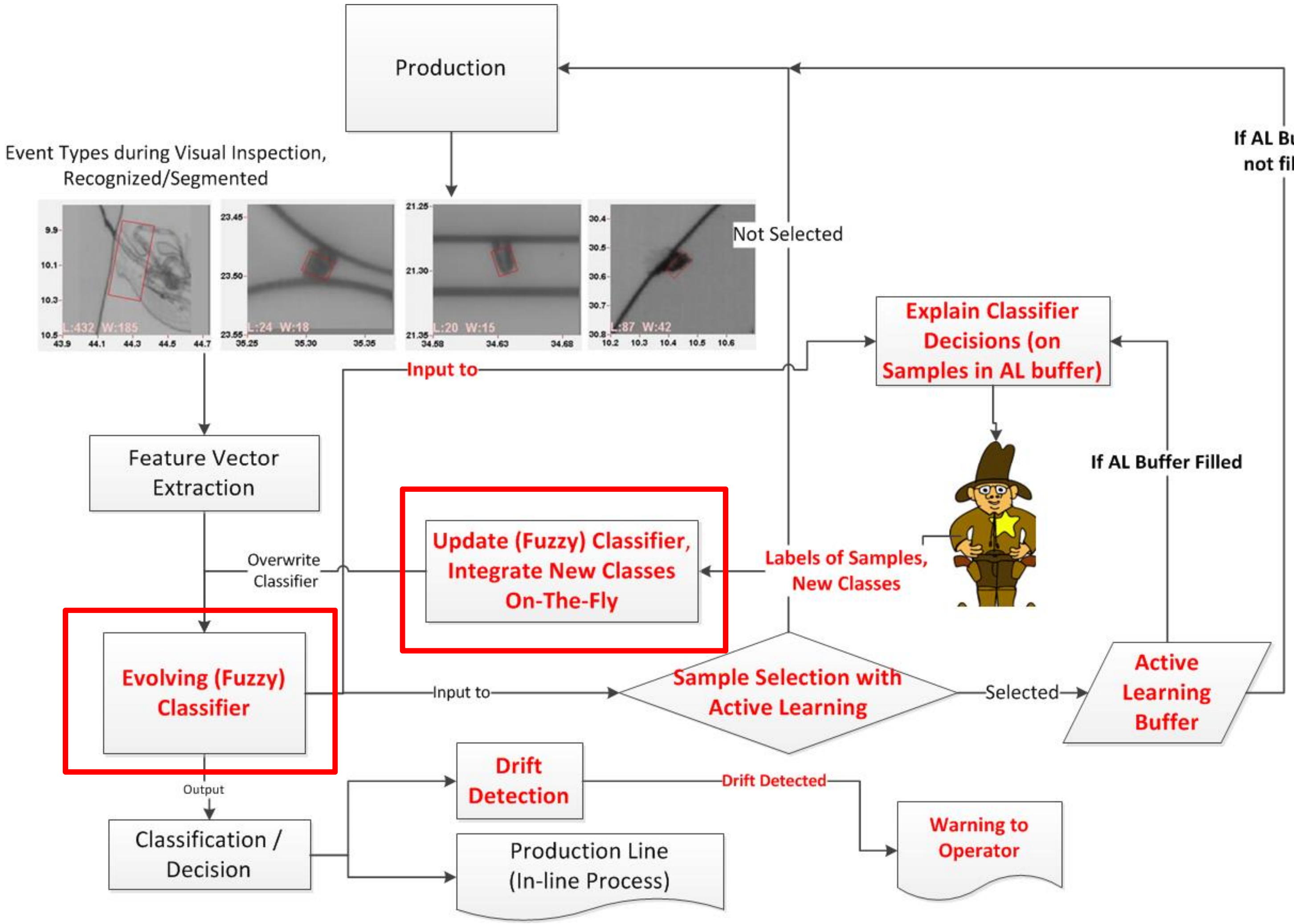
(c)

Rotor with merging and feat weights



(d)

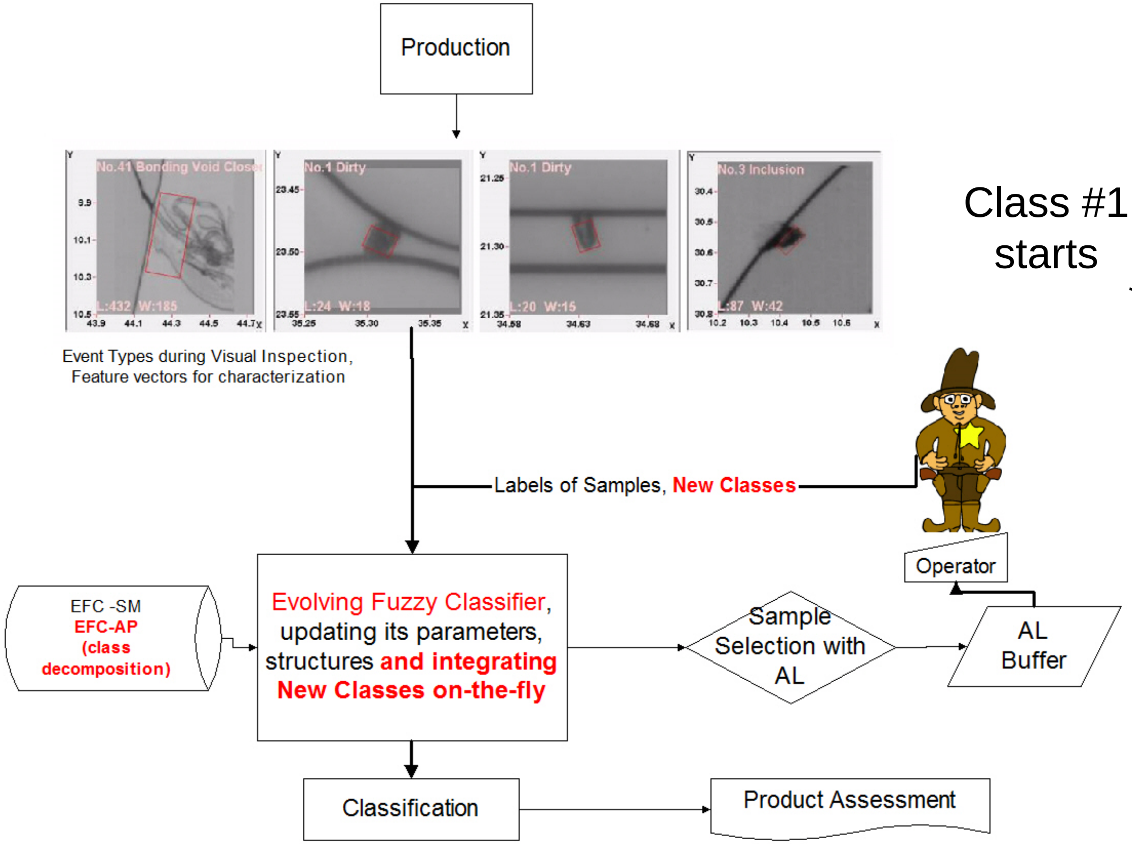
On-line Human-Machine Interaction @ VISystems with Evolving Fuzzy Classifiers, AL and Drift Detection embedded (Weigl, Lughofer et al., MVA, 2015)



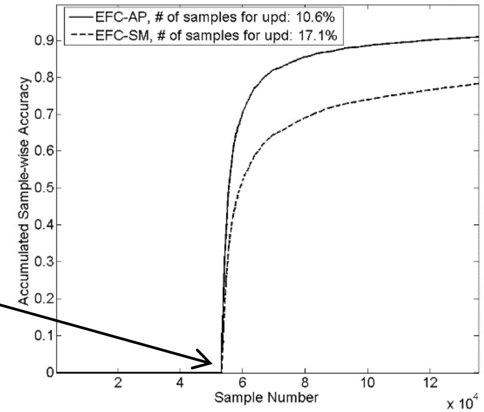
„Adding New Event Types/Classes On-the-Fly“ (On-line Visual Inspection)

Lughofer, Weigl et al., ASOC, 2015

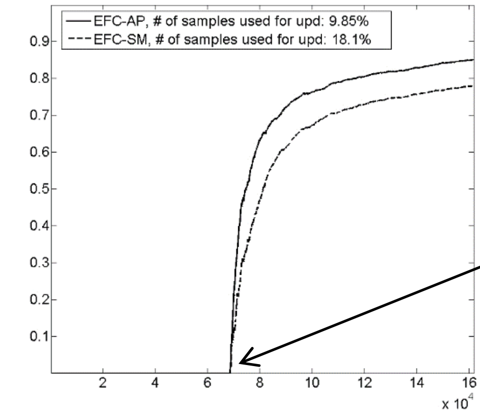
- **Operator labels from time to time new samples** within active learning cycles => **new classes** may be defined
- => **Extension in EFCs**, **class decomposition in EFC-AP favorable** (K new binary classifiers introduced)



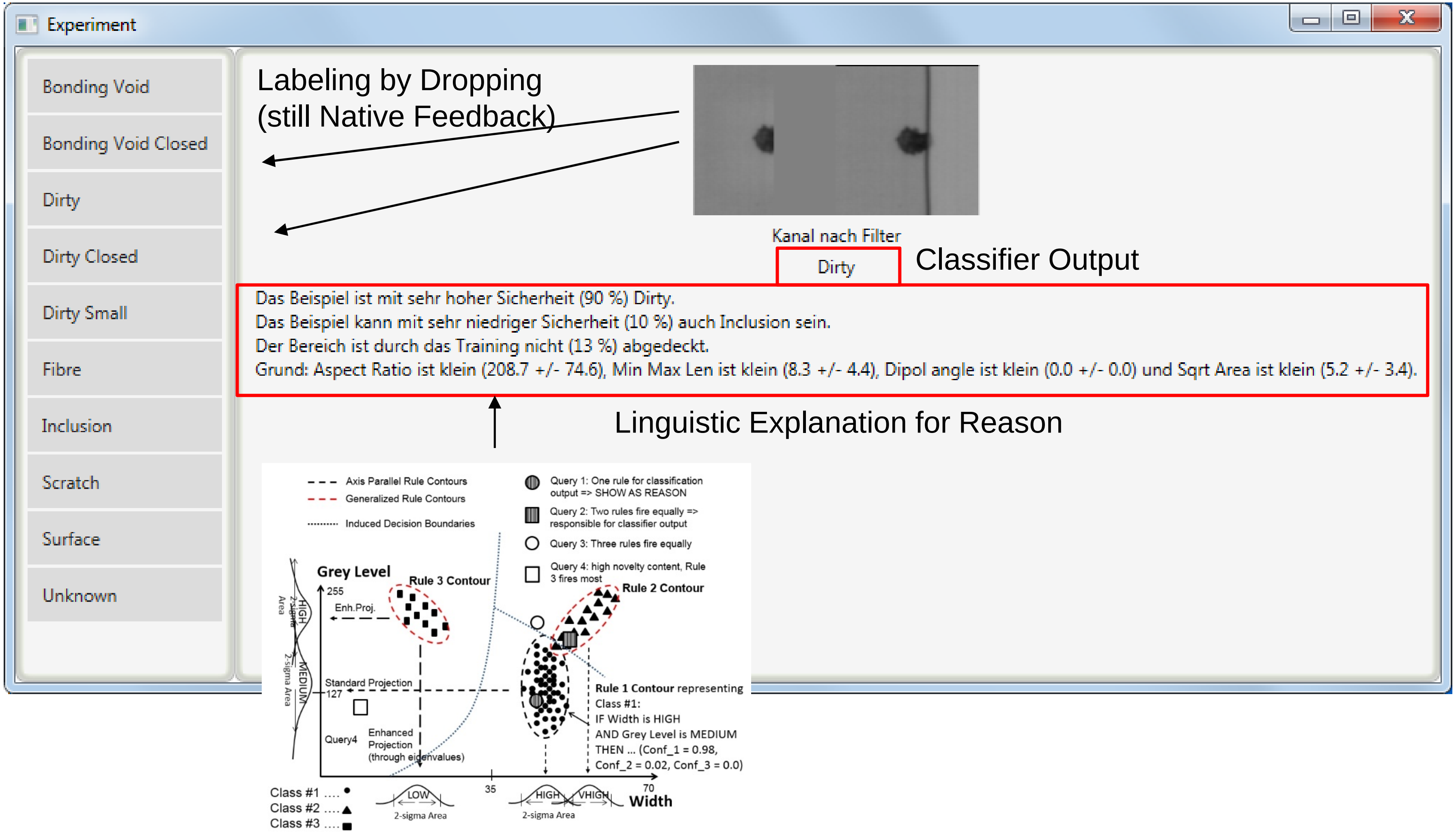
Class #1 starts



Class #7 starts



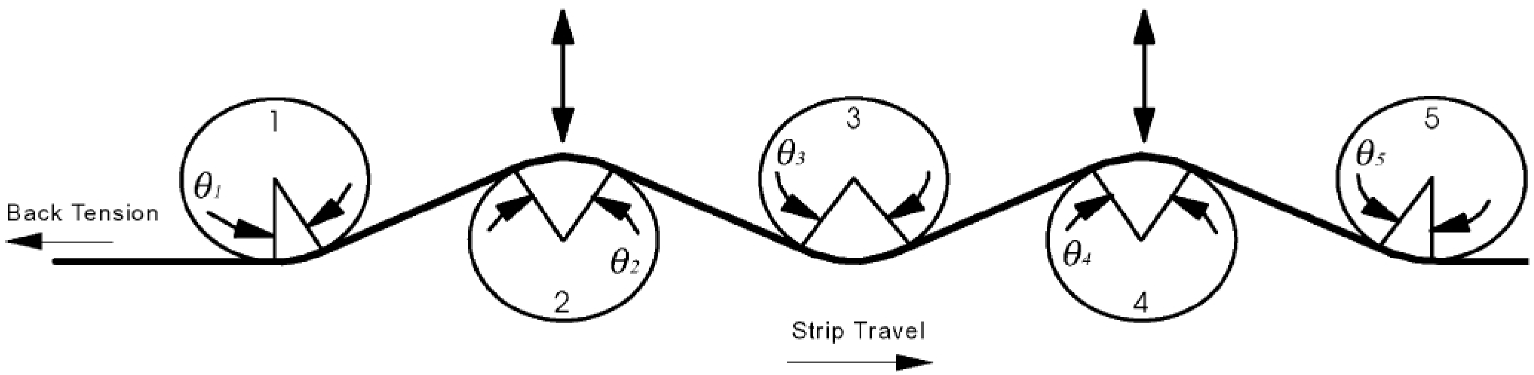
Explaining Decisions Linguistically
(Enhanced Model Output Interpretation)
Lughofer, Richter et al., ESWA, submitted, 2016



Condition Monitoring in Rolling Mills

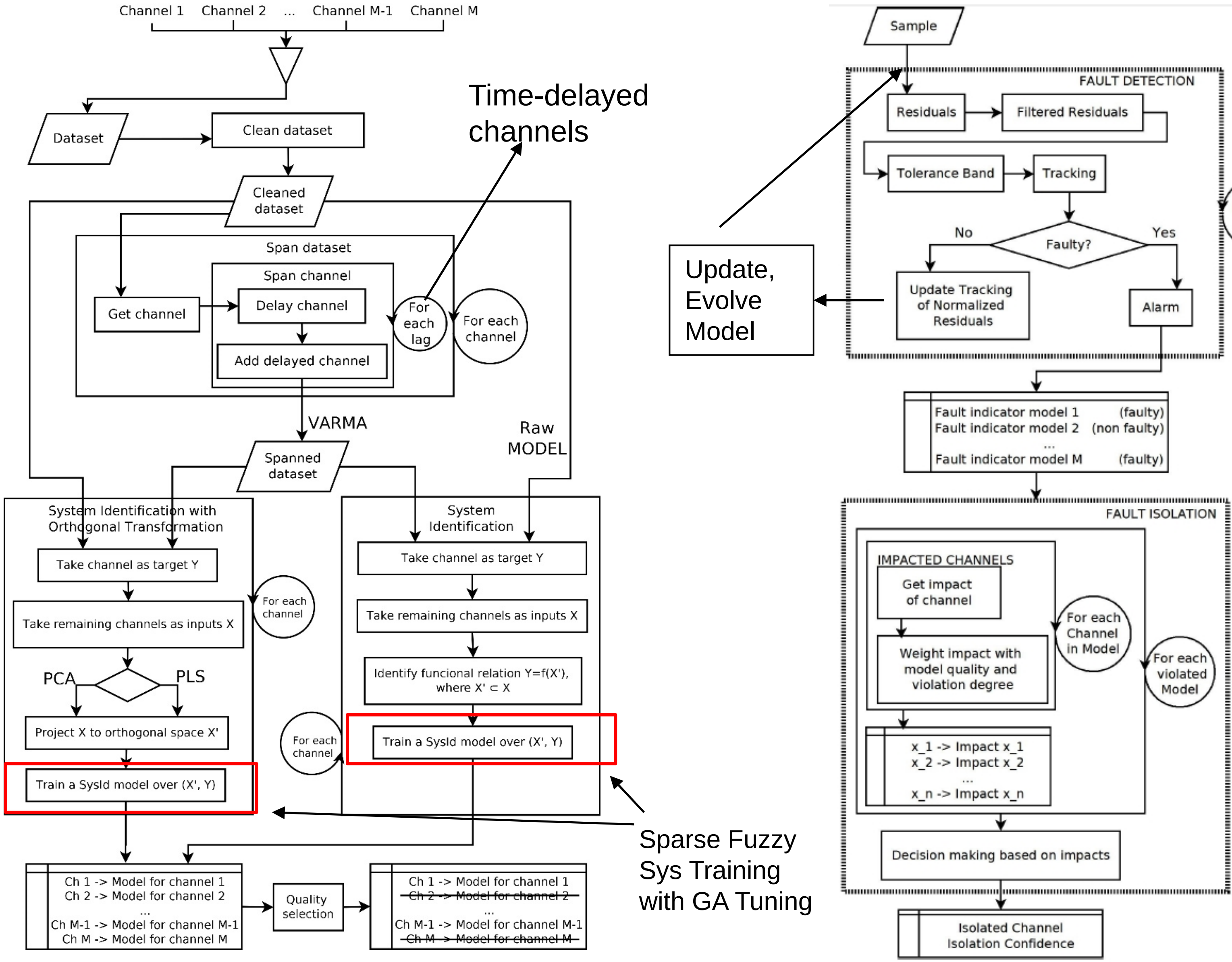
Serdio, Lughofer et al, INS, 2014/15 (FD + FI)
Serdio, Lughofer et al, INF, 2014 (dynamic models with lags)

FUZZY.LOGIC.LAB.LINZ



Condition Monitoring in Rolling Mills

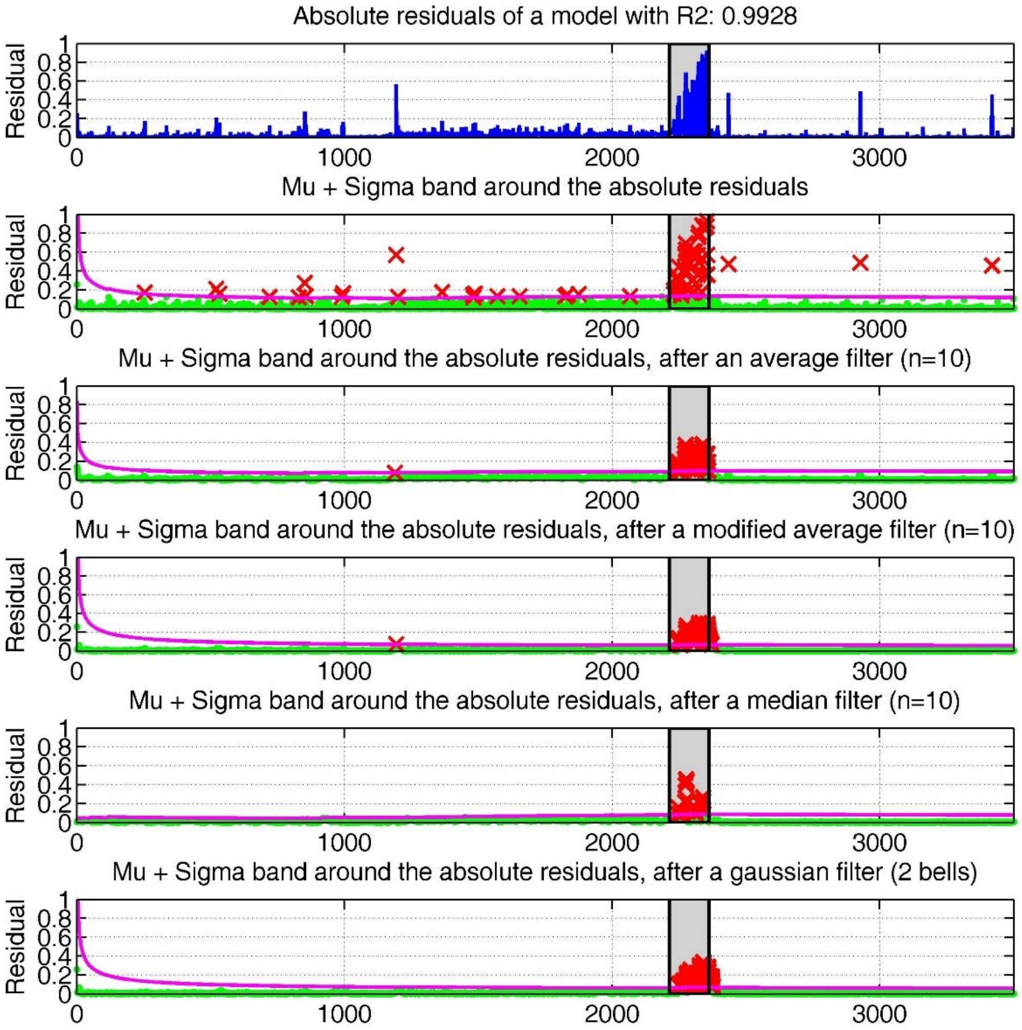
IDENTIFICATION



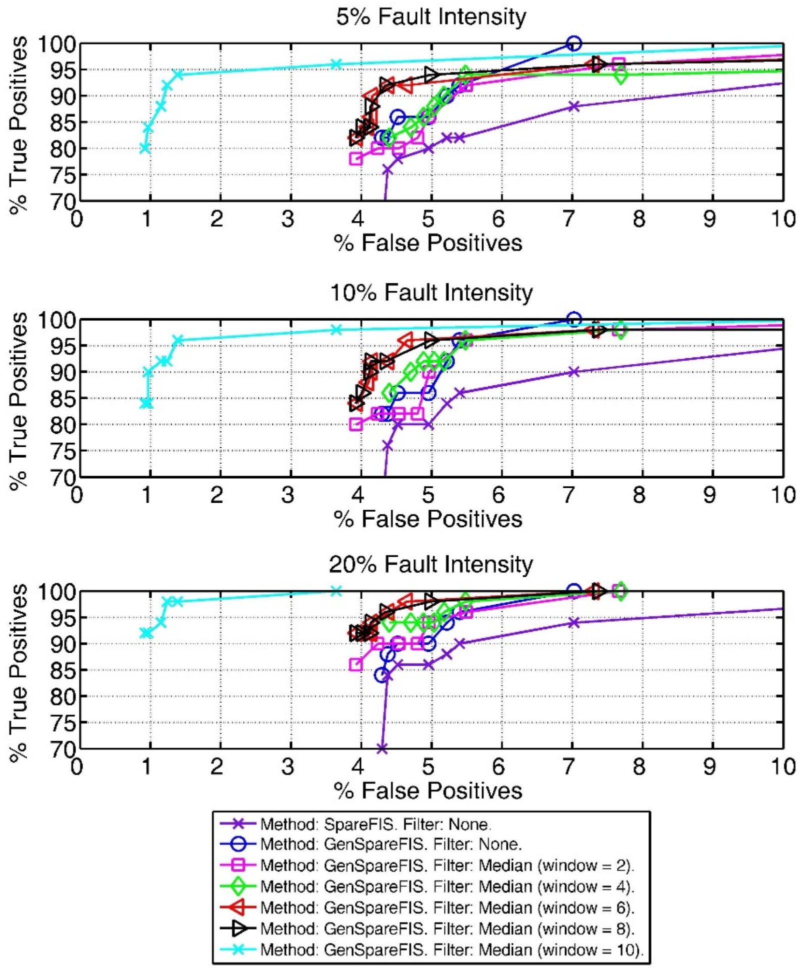
MONITORING + FDI

Results from Condition Monitoring

Effect of Adaptive Filters

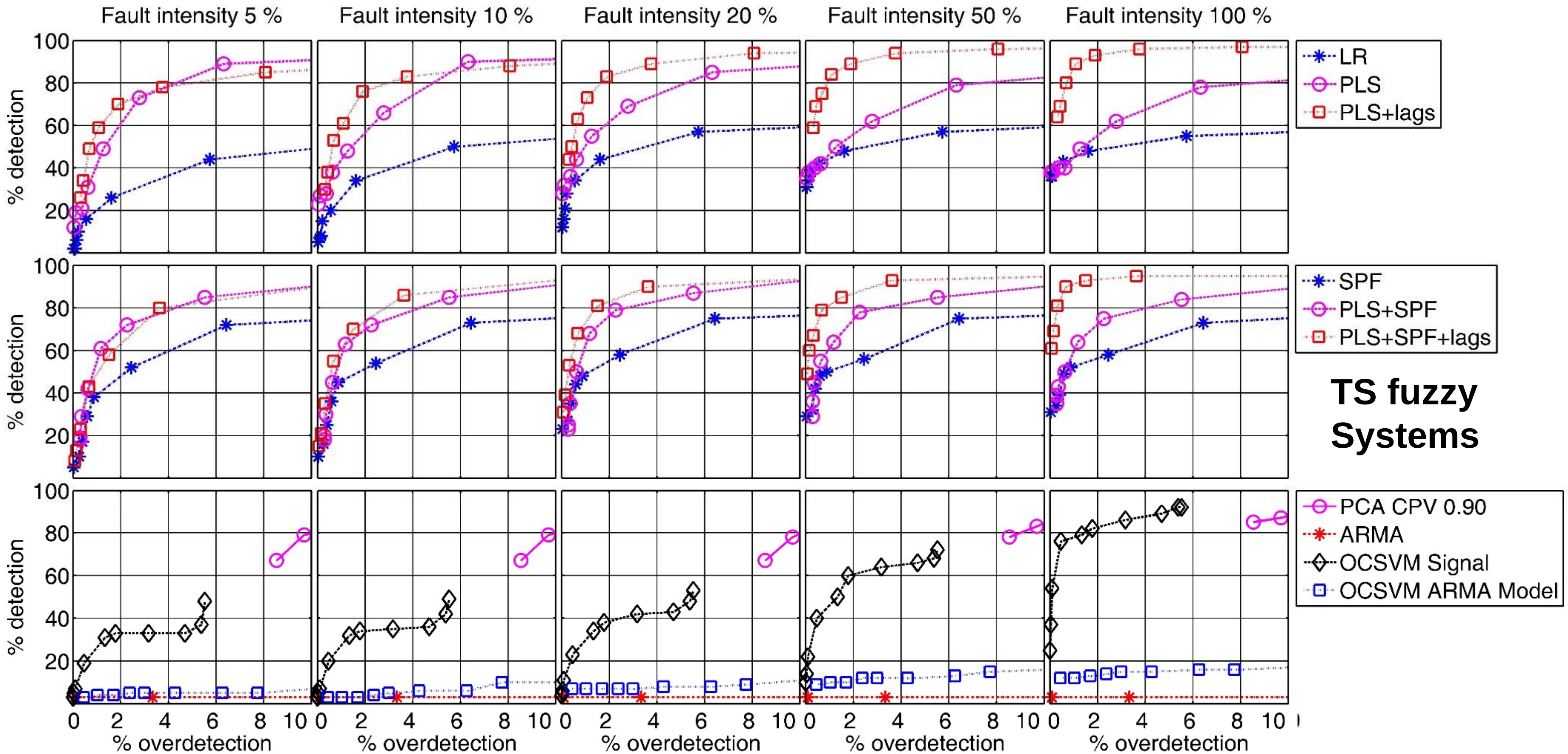


Effect of Genetic Tuning



Results from Condition Monitoring

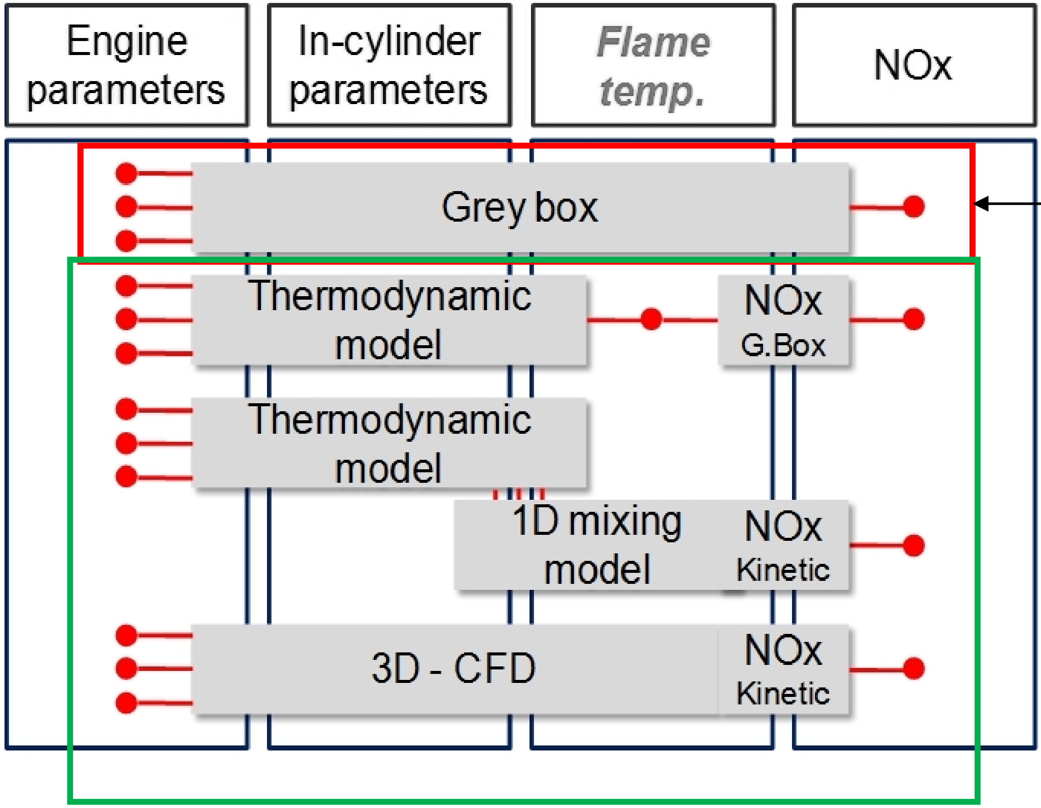
Effect of Lags



TS fuzzy Systems

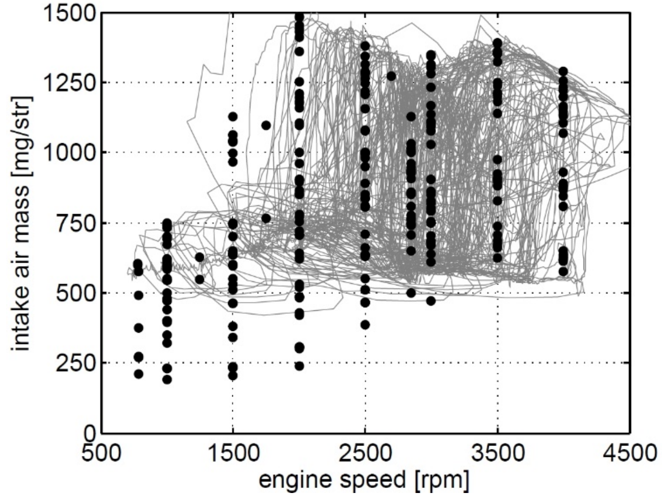
NOx Emission Modeling (eSensor)

Lughofer, Guardiola et al., ASOC, 2011



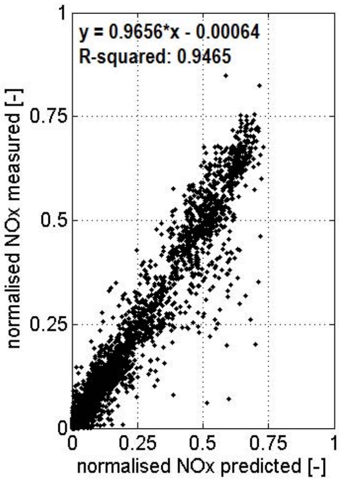
Evolving Fuzzy Systems Modeling with Compact Rulebases

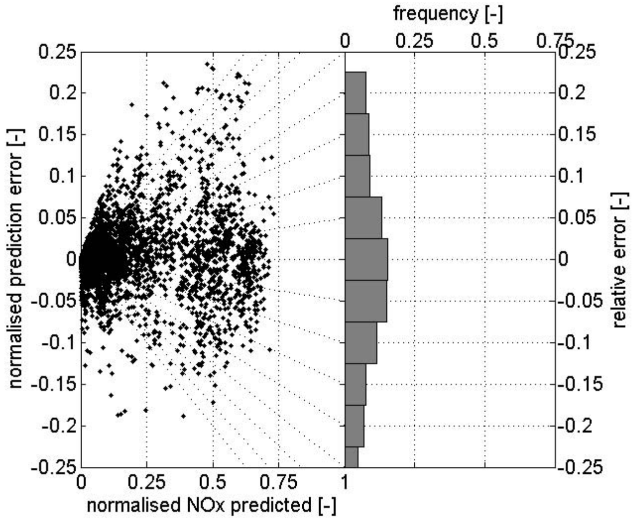
Analytical Models requiring long development and setup phases for new types of engines!

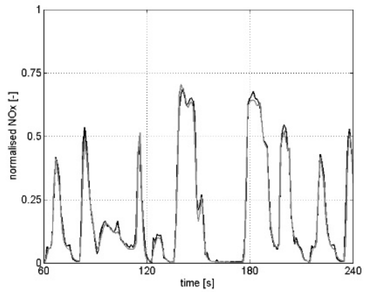


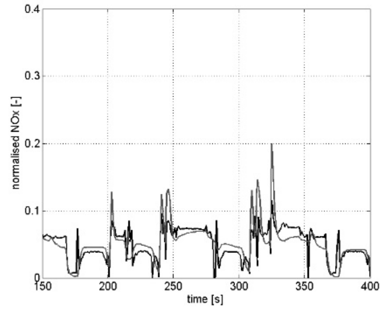
Static Points of Engine Map (off-line test case)
Dynamic Driving Trajectory (real on-line case)

Results eSensor for NOx
(separate test set)





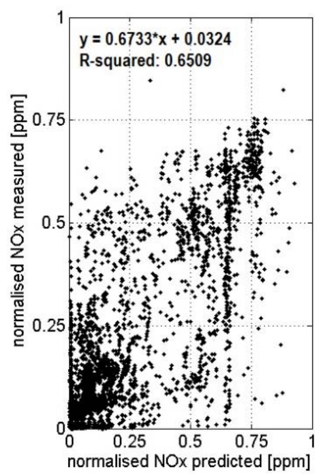


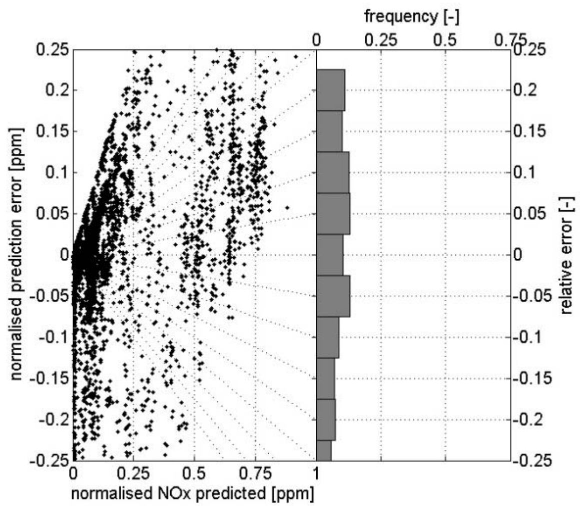


Error Analysis
Evolving Fuzzy Systems

Method	MAE Static	MAE Dyn.	MAE Mixed / Static	MAE Mixed / Dyn.
Fuzzy	**1.32**	**2.04**	**1.61**	**2.46**
Physical	1.57	2.23	NA	NA
Ridge Regr.	2.91	3.04	5.76	2.76
SVR	3.61	3.44	4.94	4.61
ANFIS	2.37	3.26	4.04	4.74
NN	1.49	2.65	7.06	3.49

Outperformance
of analytical model

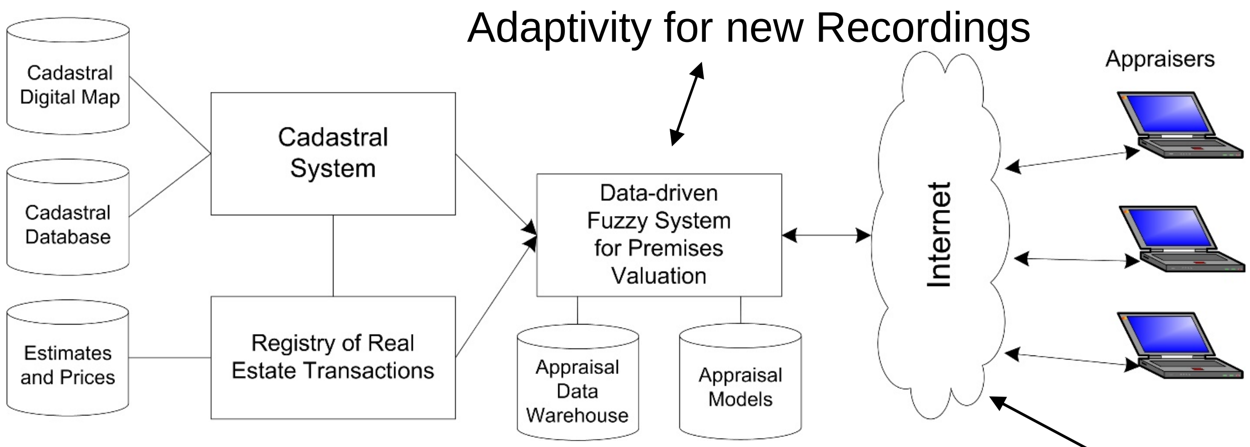




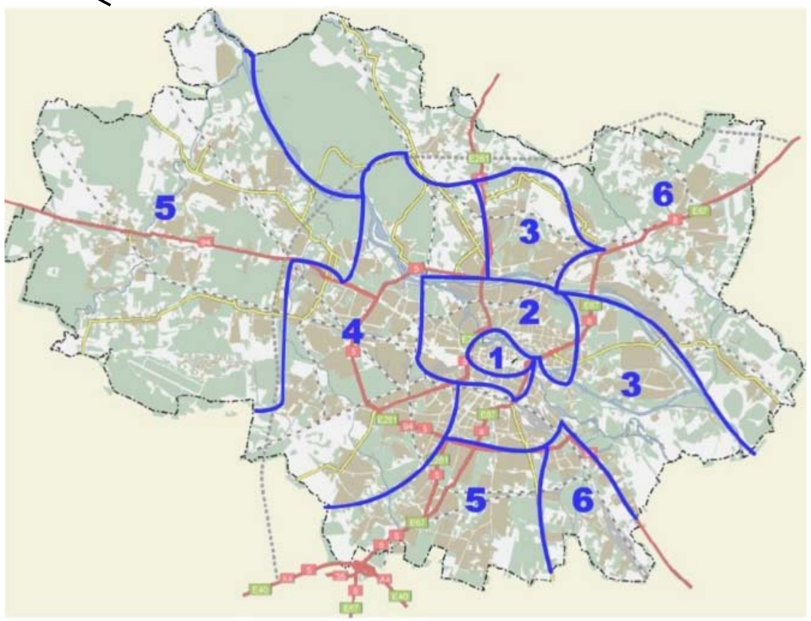
Error Analysis
Static Fuzzy Systems

Adaptive Dynamic House Pricing

(Lughofer, Trawinski, Trawinski, Information Sciences, 2011)



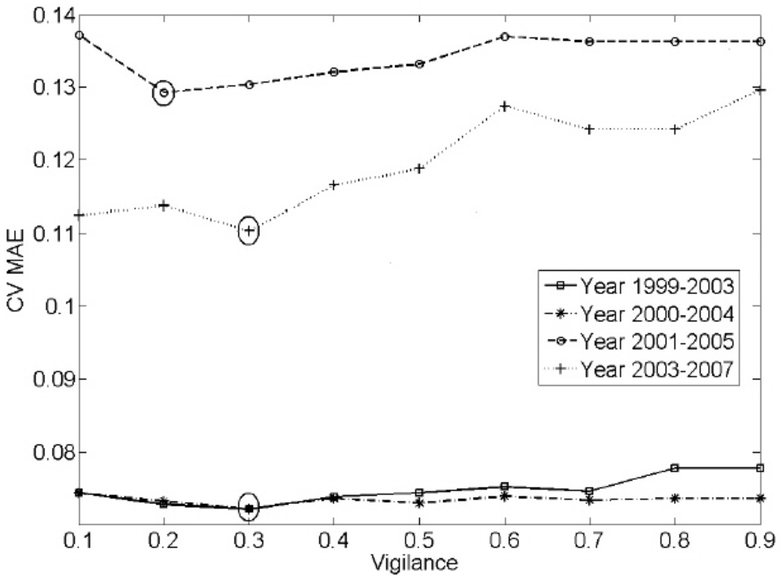
Quality Zones



State-of-the-Art before:
hard coded Expert System,
No (flexible) data-driven model at all

Results from Prediction of House Prices

Tr. / Test	*FLEXFIS*	*FLEXFIS*+pr.	Expert	# Rules	# R. pr.
1998-2002 / 2003	**0.049**	**0.049**	0.078	9	9
1999-2003 / 2004	**0.072**	0.074	0.106	12	10
2000-2004 / 2005	0.072	**0.068**	0.121	12	9
2001-2005 / 2006	0.130	**0.125**	0.134	10	8
2002-2006 / 2007	**0.089**	**0.089**	0.138	14	10
2003-2007 / 2008	**0.110**	0.115	0.145	13	7



Default vigilance of 0.3 optimal in 3 out of 4 cases!

Tr. / Test	evolved TS	equiv. Mamdani
1998-2002 / 2003	0.049	0.059
1999-2003 / 2004	0.072	0.113
2000-2004 / 2005	0.072	0.093
2001-2005 / 2006	0.130	0.158
2002-2006 / 2007	0.089	0.126
2003-2007 / 2008	0.110	0.118

Model Insights House Price Prediction

Nice Part.
With on-line
merging of local
redundancies
(Lughofer
et al, ES,
2011)

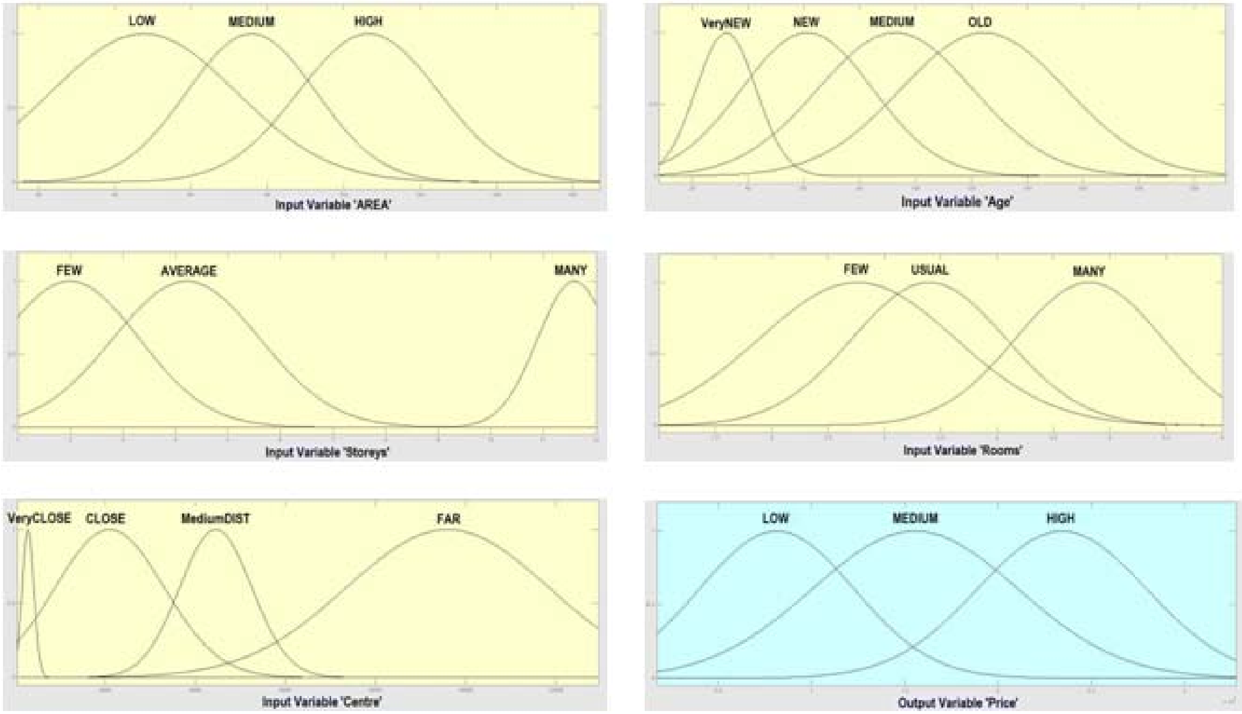


Figure 1. Fuzzy partitions for the five input variables (a) to (e): Area, Age, Storey, Rooms, Centre and the output variable Price (f)

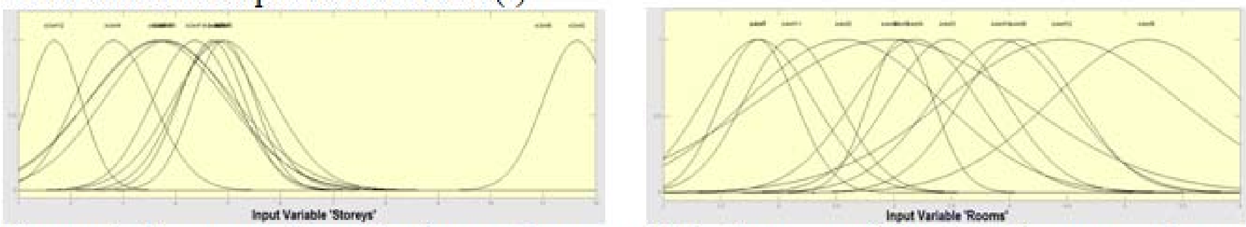


Figure 2. Fuzzy partitions for the two input variables Storeys and Rooms when not using any merging/pruning option in FLEXFIS – compare with those in Figure 1.

Rules House Price Prediction

Rule 1: If Area is LOW and Age is NEW and Storeys is AV and Rooms is FEW and Centre is CLOSE
Then Price is LOW

Rule 2: If Area is LOW and Age is VeryNEW and Storeys is MANY and Rooms is FEW and Centre is CLOSE
Then Price is LOW

Rule 3: If Area is LOW and Age is NEW and Storeys is AV and Rooms is USUAL and Centre is CLOSE
Then Price is LOW

Rule 4: If Area is LOW and Age is OLD and Storeys is AV and Rooms is FEW and Centre is CLOSE
Then Price is LOW

Rule 5: If Area is LOW and Age is NEW and Storeys is FEW and Rooms is FEW and Centre is FAR
Then Price is LOW

Rule 6: If Area is LOW and Age is MEDIUM and Storeys is MANY and Rooms is FEW and Centre is VeryCLOSE
Then Price is MEDIUM

Rule 7: If Area is MEDIUM and Age is NEW and Storeys is FEW and Rooms is USUAL and Centre is MEDIUMDist
Then Price is MEDIUM

Rule 8: If Area is HIGH and Age is MEDIUM and Storeys is AV and Rooms is MANY and Centre is CLOSE
Then Price is HIGH

Rule 9: If Area is MEDIUM and Age is MEDIUM and Storeys is AV and Rooms is USUAL and Centre is CLOSE
Then Price is HIGH

Thank you for your attention!