intellitech
INTELLIGENT TECHNOLOGIES

xtractis®

Augmented Fuzzy Cognitivist AI Robots for Robust Predictive Knowledge Discovery

Is it still possible to challenge them?

WIEF Forum

May 15, 2017 v1.1

Prof. Zyed ZALILA
President-CEO

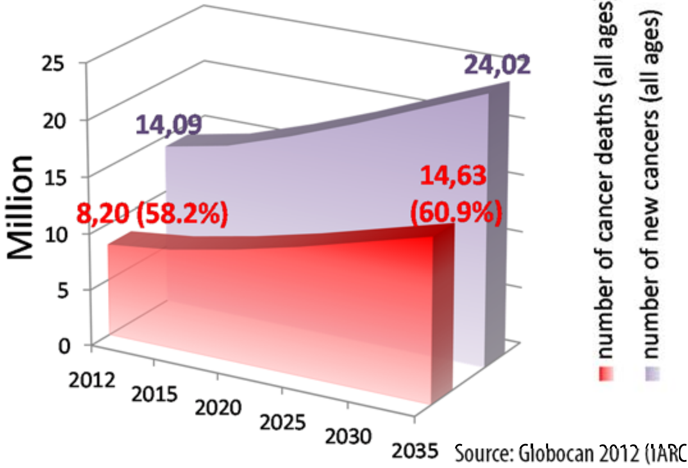
xtractis® application

most complex issue: Predictive Medicine

Medicine of 21st century

- the **Right Diagnosis**
- the **Right Drug** in the **Right Dose** to the **Right Patient**
- Epigenetics, Oncogenetics
- Cancer: 2nd / 3rd fatal disease worldwide

Projected number of Cancer incidence and mortality worldwide 2012-2035



Source: Globocan 2012 (IARC)

Hugely complex process

- ≈25K genes (often **weak** expression intensities)
- ≈1K metabolic & environmental variables (food, stress, location, happiness, work, well-being,…)
- ∝ interactions (genes ↔ genes, genes ↔ environment,…)

Unattainable for a human brain

- Limitation of human comprehension (1-3 to 7-9 variables simultaneously)

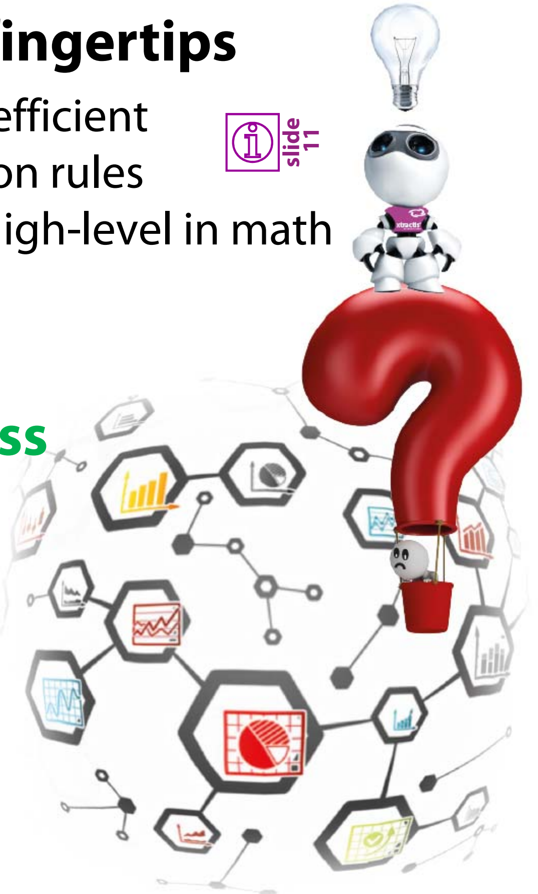
xtractis® Robots

an Exobrain

Robust Predictive Analytics at everyone's fingertips

- New Scientific Approach: automatic design of most efficient learning strategies for automatic induction of decision rules
  ➨ Find out w/o *a priori*, w/o decomposition, w/o a high-level in math
- Co-operate with your Exobrain

ⓘ slide 11



Discover the real complexity of your process

- Make explicit the tacit implicit knowledge
- Complex ≠ Complicated ([Descartes 1637] was wrong!)
- Complexity cannot be reduced to few dimensions

ⓘ slide 13

ⓘ slide 19

Hire your AI Robots!

- No best Human experts could beat best AI Robots
- AI Robot can reason 24/7/365 and solve very complex problems **faster and better than a Human**
  ➨ no tiredness, no holidays, no strike, at a low hourly cost

AI Robots

what ethical/unethical social impacts?

1/2

on Employment

- Less specialist jobs: drivers, doctors, actuaries, traders…
- Always manual expert jobs: building workers, surgeons…
- More AI scientists jobs (Masters, PhD) ➡ to invent new generation of Robots

on Sciences (Virtual Scientist)

- New discoveries ➡ next Nobel Prize, an AI Robot for sure!
- PhD (experimental sciences) in few months instead of 3-4 years

on Finance/Insurance (Virtual Banker/Insurer)

- Bank advisor available 24/7/365 ➡ mandatory for the digital natives!
- Credit/Insurance approval in few seconds, at the lowest price (digital service)

on Transport (Virtual Driver)

- More safety and comfort ➡ “Rolling Lounge”
- Fewer deaths and injuries **BUT** potentially: vehicle hacking (terrorist attacks)

AI Robots

what ethical/unethical social impacts?

2/2

on Marketing/Industry (Virtual Marketer/Engineer)

- New optimal products, fitting customer preferences

on Health (Virtual Doctor)

- Early diagnosis of pathologies (cancer, diabetes, Alzheimer…)
  BUT potentially: refusal of insurance/credit/hiring… and eugenics risk!!
- Predictive Personalized Medicine (drug discovery)
  BUT potentially: design of chemical weapons

on Defense (Virtual Soldier)

- Armed drones to save soldiers' life
  BUT potentially: killing authorization without remorse (≈ video game)

on Justice/Security/Cybersecurity (Virtual Judge/Policeman)

- Guarantee of an impartial decision
- Early detection of fraudulency and malicious behaviors
  BUT potentially: hacking of the defender's strategy (reverse engineering)

AI Robots

how to reduce the concerns?

Teaching Ethics

- To AI students/specialists (≈ Hippocratic/judicial oath ≈ Do no harm!)

Certifying AI Robots

- To ensure their efficiency and ethics/honesty

  ➡ ID cards and **diplomas** for AI Robots given by Human Regulator

  (DeepMind AlphaGo won the "Divine" 9th Dan in Go Game - 03/2016)

  ➡ Universities for AI Robots? Do they still need Humans to learn?

  (x**tractis**® Robot builds autonomously its efficient learning strategies for each new predictive problem it has to solve)

- Trust and use AI Robots with the most renowned diplomas

- **Paradox**: to certify complex strategies, Human must rely on an AI Robot Regulator **BUT** who will certify the AI Regulator?

Legal Recognition of AI Robots

- **Rights & Duties**, Tax on the value created by AI Robots
- **Paradox**: **IP of discoveries** made by an AI Robot: granted to the Robot, to the User of the Robot or to the inventor of the Robot?

intellitech
INTELLIGENT TECHNOLOGIES

xtractis®

Augmented Fuzzy Cognitivist AI Robots
for Robust Predictive Knowledge Discovery

Is it still possible to challenge them?

www.xtractis.ai

xtractis@intellitech.fr

Warning

Zalila, Z. & al (2012-2017) *xtractis® Augmented Fuzzy Cognitivist AI Robots for Robust Predictive Knowledge Discovery. Is it still possible to challenge them?*, intellitech [intelligent technologies], May 2017, Compiegne, France, 22p.

Fuzzy model

classification – Prostate Cancer

1/2

UPD
1705

Prostate Cancer Diagnosis

- 102 cases
- 12,600 gene expression levels
- Variable to predict: 52 (51%) Tumor (1) / 50 (49%) Normal (0) diagnosis
- Independent testing set from a different experiment: 25 tumor and 9 normal samples
- Database source: D. Singh & *al.*, Department of Adult Oncology, Brigham and Women's Hospital, Harvard Medical School, [http://www-genome.wi.mit.edu/mpr/prostate]

xtractis® model

- type 5 Combined model (CB5) – 500 fuzzy models
  - ➡ 70 variables (2 to 11 variables per model)
  - ➡ 1,000 rules (2 rules per model)

Performances

CB5
(majority voting)

Decision	**Classification error**	**Matthews Correlation**	**Min. Sensitivity Specificity**	**Unavailable**
Training (100s x750g x 70%)	0.00%	1.000	100.00%	0 (0.00%)
Validation (100s x750g x 15%)	**1.96%**	0.961	**98.00%**	0 (0.00%)
Testing (100s x750g x 15%)	3.92%	0.922	96.00%	0 (0.00%)
Ext. Validation	**2.94%**	0.930	**96.00%**	0 (0.00%)

CB5
(majority voting) Actual class

Predicted class

Decision	0	1
0	50 (100.00%)	0 (0.00%)
1	0 (0.00%)	52 (100.00%)
Unavailable	0 (0.00%)	0 (0.00%)

Training Confusion matrix

CB5
(majority voting) Actual class

Predicted class

Decision	0	1
0	49 (**98.00%**)	1 (1.92%)
1	1 (2.00%)	51 (**98.08%**)
Unavailable	0 (0.00%)	0 (0.00%)

Validation Confusion matrix

CB5
(majority voting) Actual class

Predicted class

Decision	0	1
0	9 (**100.00%**)	1 (4.00%)
1	0 (0.00%)	24 (**96.00%**)
Unavailable	0 (0.00%)	0 (0.00%)

External Testing Confusion matrix

Results from **xtractis**® Generate 9.2.19175

Fuzzy model *vs.* KSVM

classification – Prostate Cancer 2/2

UPD
1705

KSVM model performance

Decision	**Classification error**	**Matthews Correlation**	**Minimum Sensitivity Specificity**
Accuracy	0.00%	1.000	100.00%
1,000 x MC 15%	**8.23%**	0.836	89.67%
External Testing	**26.47%**	**0.566**	**64.00%**

Comparison

KSVM

Training

Actual class

Predicted class: **Decision**	0	1
0	50 (100.00%)	0 (0.00%)
1	0 (0.00%)	52 (100.00%)

error = 0.00%

External Testing

Actual class

Predicted class: **Decision**	0	1
0	9 (100.00%)	9 (**36.00%**)
1	0 (0.00%)	16 (**64.00%**)

error = **26.47%**

xtractis®

CB5
(majority voting)

Training

Actual class

Predicted class: **Decision**	0	1
0	50 (100.00%)	0 (0.00%)
1	0 (0.00%)	52 (100.00%)
Unavailable	0 (0.00%)	0 (0.00%)

error = 0.00%

CB5
(majority voting)

External Testing

Actual class

Predicted class: **Decision**	0	1
0	9 (**100.00%**)	1 (4.00%)
1	0 (0.00%)	24 (**96.00%**)
Unavailable	0 (0.00%)	0 (0.00%)

error = **2.94%**

➡ **xtractis® beats KSVM by 800%**

(based on increase of External Testing Error)

Results from R 3.2.0
packages: mlr 1.1.18, imputeR 1.0.0, kernlab
Tuning and Robustness Assessment modules
by **intelli**tech

Results from **xtractis**® Generate 9.2.19175

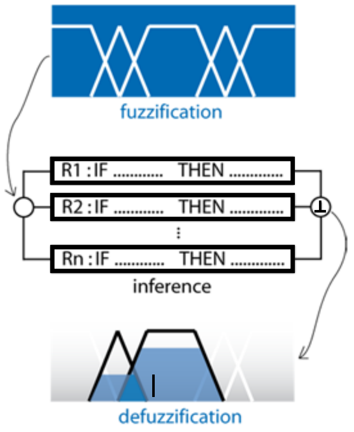
xtractis
BY INTELLITECH

xtractis® Robot

a smart knowledge discoverer

intellitech
INTELLIGENT TECHNOLOGIES

Robust Predictive Modeling (KDD, DDM)



- Automatic discovery of hidden laws ruling the Real World from big multidimensional, heterogeneous, structured datasets
- Proprietary AI algorithms: Fuzzy Theory + Machine Learning
- Holistic approach handling **weak signals** & non-linearity

ⓘ slide 12 ⓘ slide 19

Advantages

- Interpretability (“If…Then” decision rules) ➡ human validation, certification ⓘ slide 13
- Better **predictive capacity** (robustness) *vs.* Open Source algorithms ⓘ slide 16
  (Random Forest, CART Decision Tree, Boosted Trees, SVM, Neural Networks/Deep Learning, Polyn. Regression, Logistic Regression, PLS)
- Fully **automated** ➡ neither coding, nor framework ⓘ slide 21

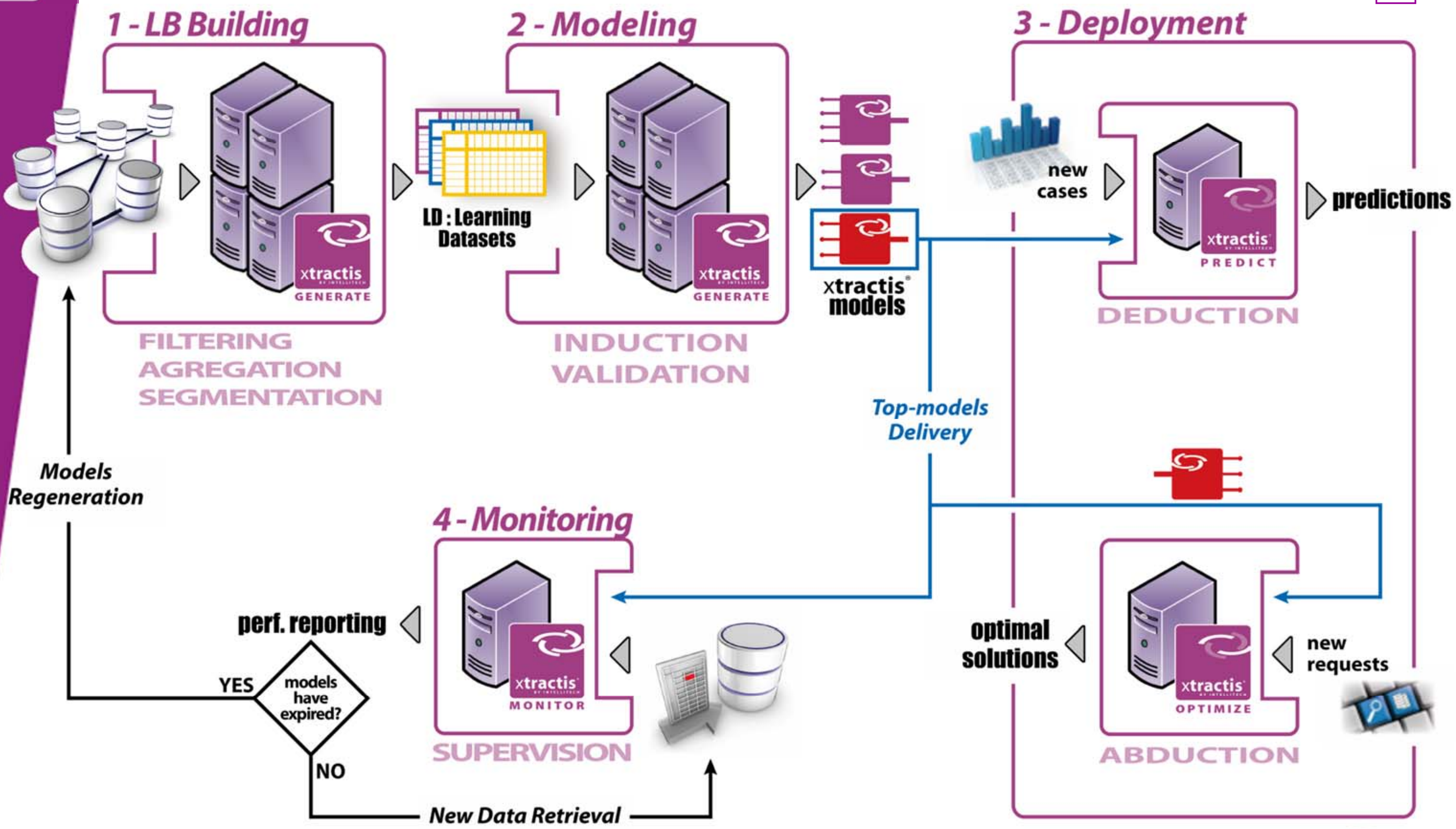
Clients & Awards

- Groupe Mars / Mars Petcare / Mars Food, Groupe Bel, L’Oréal, Essilor International, Georgia-Pacific, SCA, Technip / Flexi France, Groupe Engie / CPCU, Technip / Cybernetix, Groupe BPCE / Crédit Foncier de France, Crédit Logement, Groupe Pierre Fabre, Groupe Decathlon, Renault-Nissan, PSA Groupe, Groupe Thales / Thales Alenia Space, Groupe Arkema / Bostik, Groupe Areva, CNES, ESA, Union Européenne …
- Deloitte Technology Fast 500 – EMEA 2011 & 2012 Winner

xtractis®

4 software Robots for an "in-house" complete solution

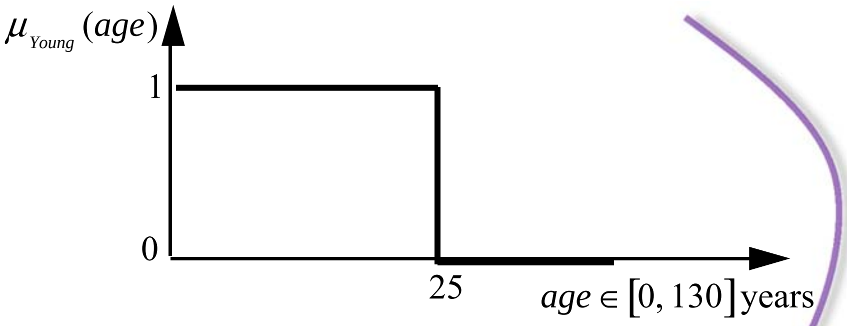
slide 20
slide 13
slide 3



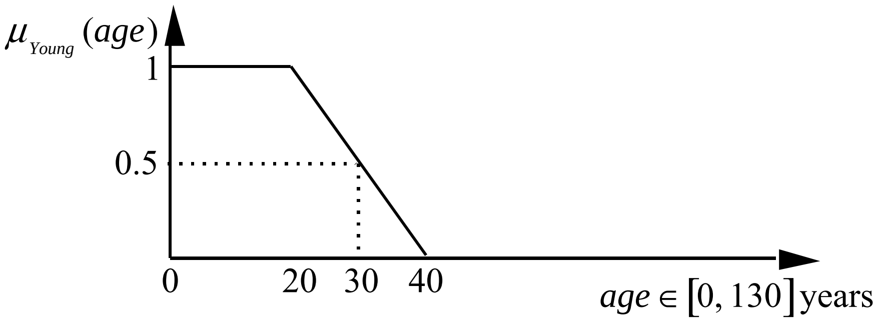
Fuzzy Mathematics

suitable for real world data

Fuzzy set

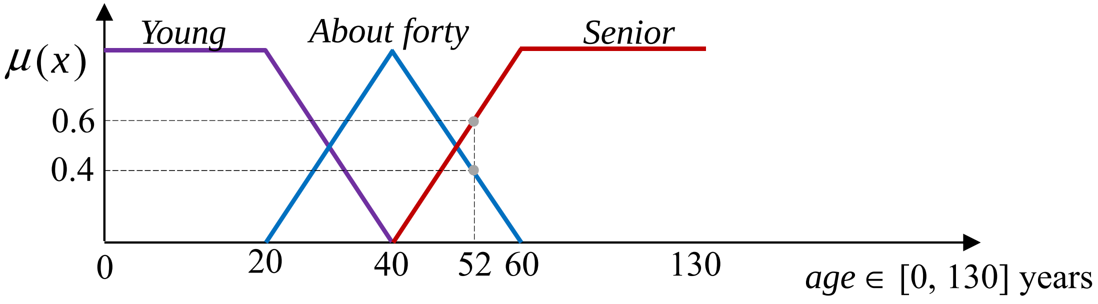


Crisp set: **Young** traveler
(for airlines)



Fuzzy set: **Young** traveler
(for Customs & Border Police)

$$X = [0,130]\ years;\ \forall x \in X, \quad \mu_{Young}(x) = \begin{cases} 1, & if\ x \in [0,20] \\ \frac{40-x}{20}, & if\ x \in [20,40] \\ 0, & otherwise \end{cases}$$



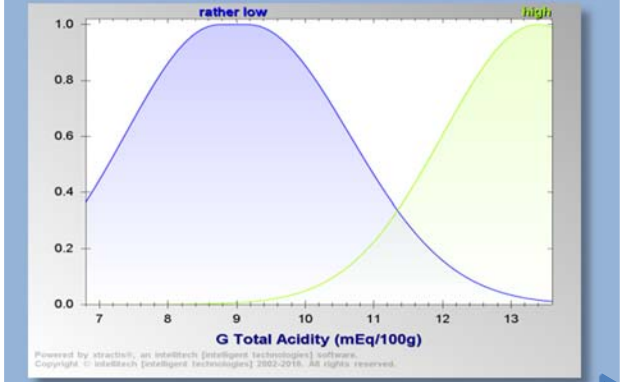
3-class **Fuzzy partition**
(for Customs & Border Police)

Fuzzy model

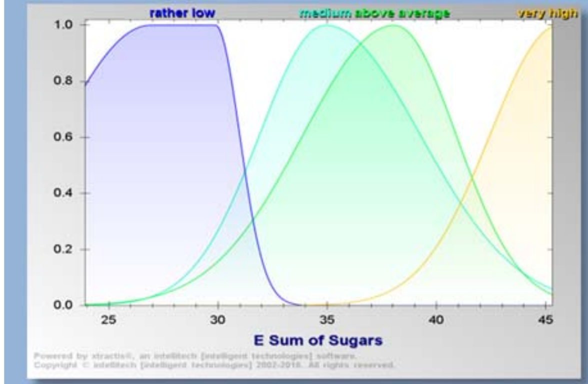
regression – interpretable model

Sweet perception of a fresh tomato: 2 variables, 4 rules (complexity 33.0)

Classes



Input1: *Total Acidity*



Input 2: *Sum of Sugars*

Rules

Rule ❶

If *Total Acidity* is *rather low*
And *Sum of Sugars* is *rather low*
Then *Sweet* equals *3.39*

Rule ❷

If *Total Acidity* is *rather low*
And *Sum of Sugars* is *medium*
Then *Sweet* equals *7.19*

Rule ❸

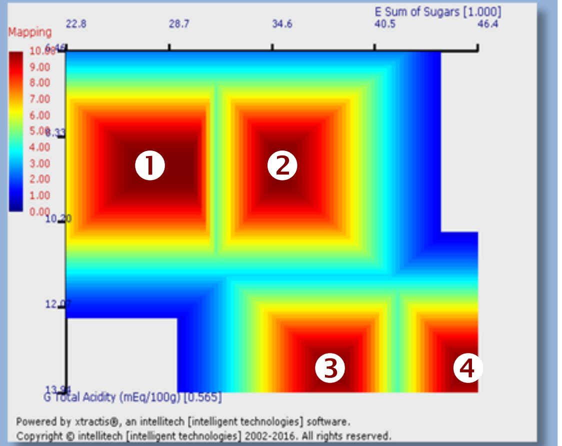
If *Total Acidity* is *high*
And *Sum of Sugars* is *above average*
Then *Sweet* equals *3.30*

Rule ❹

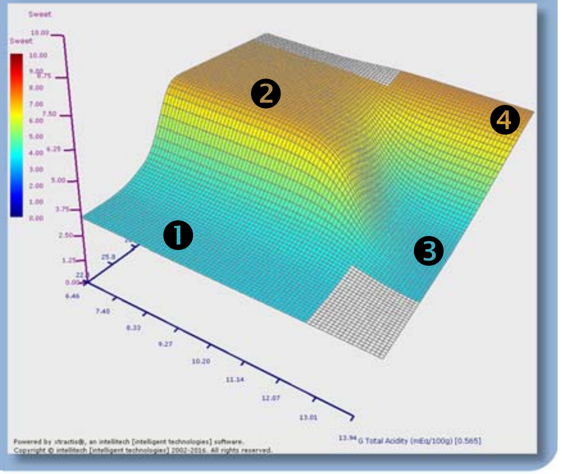
If *Total Acidity* is *high*
And *Sum of Sugars* is *very high*
Then *Sweet* equals *7.49*

Inference

Mapping



Decision surface



Data: INRA (Institut National de la Recherche Agronomique) et CTIFL (Centre Technique Interprofessionnel des Fruits et Légumes) – 7th Sensometrics Conf., July 2004, Davis, CA, USA

Results from **xtractis**® Generate 9.1.16325

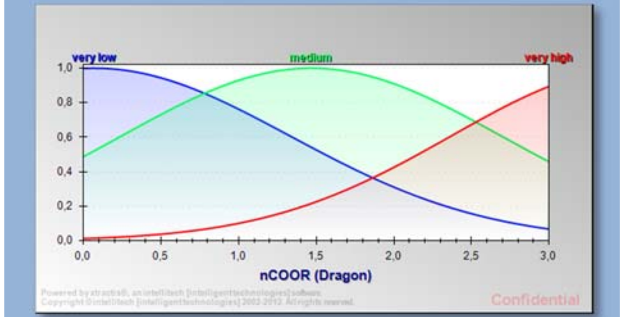
Fuzzy model

regression – complex model

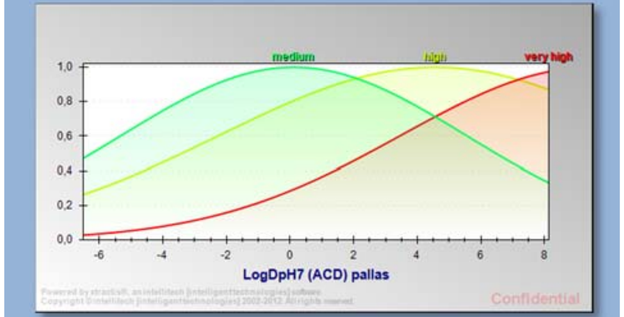
slide 3
slide 10

Pesticide toxicity on trout: 11 variables, 4 rules (complexity 99.3)

Partitions



Input 1: *nCOOR*



Input 2: *LogDpH7*

Rules

Rule ❶

If *nCOOR* is *very low*
And *LogDpH7* is *medium*
And …
then *Toxicity_Trout* equals *-1.78*

Rule ❷

If *nCOOR* is *medium*
And *LogDpH7* is *very high*
And …
then *Toxicity_Trout* equals *5.96*

Rule ❸

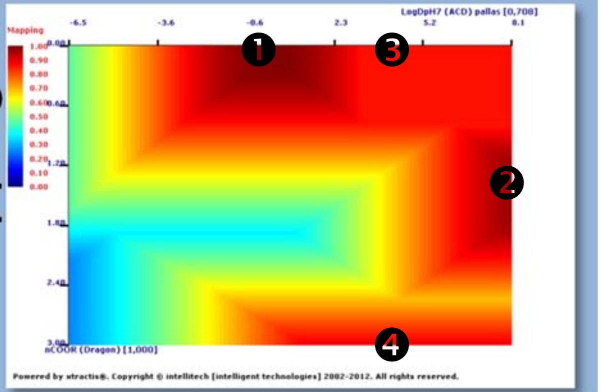
If *nCOOR* is *very low*
And *LogDpH7* is *high*
And …
then *Toxicity_Trout* equals *6.28*

Rule ❹

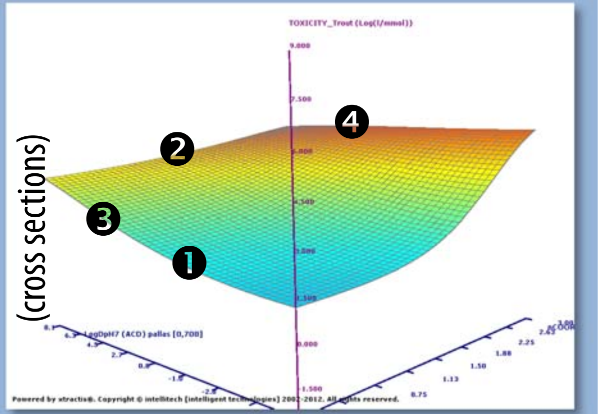
If *nCOOR* is *very high*
And *LogDpH7* is *high*
And …
then *Toxicity_Trout* equals *7.74*

Inference

Mapping



Decision surface (cross sections)



Data: US EPA, German BBA, EU-project SEEM

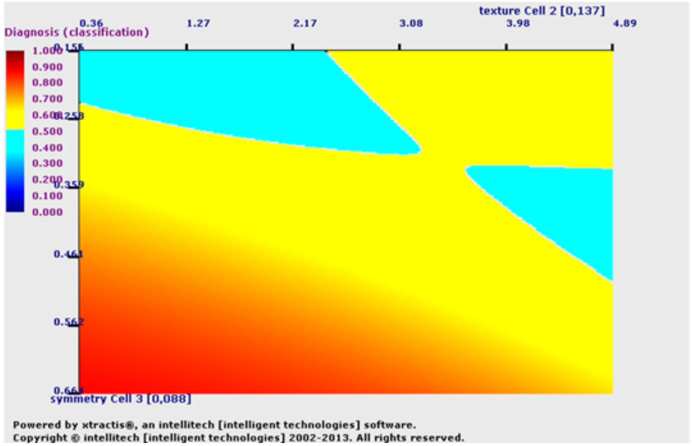
Fuzzy model

classification

slide 3
slide 19

Breast cancer diagnosis

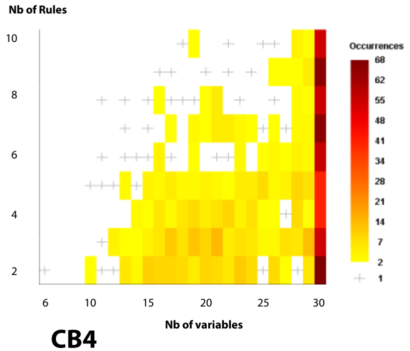
- 569 patient images
- 30 potential predictors
- Variable to predict: 357 (62.7%) Malignant diagnosis (1)
  212 (37.3%) Benign diagnosis (0)
- Data: Dr. William H. Wolberg, W. Nick Street, Olvi L. Mangasarian – University of Wisconsin [UCI Machine Learning Repository]



Decision surface
(cross section on variables *texture cell 2* & *symmetry cell 3*)

xtractis® model

- type 4 Combined model (CB4) – 1,000 models
  - ➡ 30 variables
  - ➡ 5,159 rules



CB4
(Absolute majority)

Decision	**Classification error**	**Matthews Correlation**	**Min. Sensitivity Specificity**	**Refused**
Training (100s x 1,000g x70%)	1.41%	0.970	98.11%	0.00%
Validation (100s x 1,000g x15%)	**1.23%**	0.974	98.10%	0.47%
Testing (100s x 1,000g x15%)	**1.93%**	0.959	97.17%	0.00%

Performances

CB4
(Absolute majority)

	Decision	0	1
Predicted class	0	98.88%	1.89%
	1	1.12%	98.11%
	Refused	0.00%	0.00%

Training Confusion matrix

CB4
(Absolute majority)

	Decision	0	1
Predicted class	0	**99.16%**	1.90%
	1	0.84%	**98.10%**
	Refused	0.00%	0.47%

Validation Confusion matrix

CB4
(Absolute majority)

	Decision	0	1
Predicted class	0	**98.60%**	2.83%
	1	1.40%	**97.17%**
	Refused	0.00%	0.09%

Testing Confusion matrix

Results by **xtractis**® Generate 9.1.16139

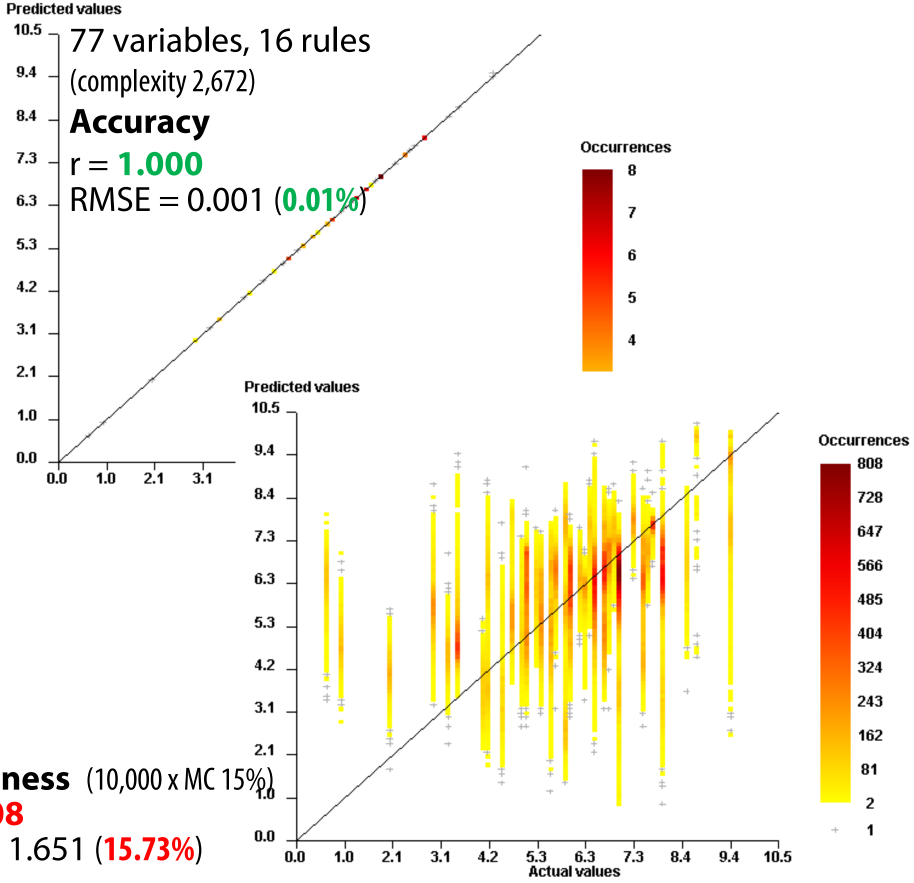
Specificities of xtractis®

robustness through cross-validation

slide 18

Evaluation of the generalization capacity of predictive models

- Accurate model is **not necessarily** robust
- Robust model: difficult to obtain but **mandatory**

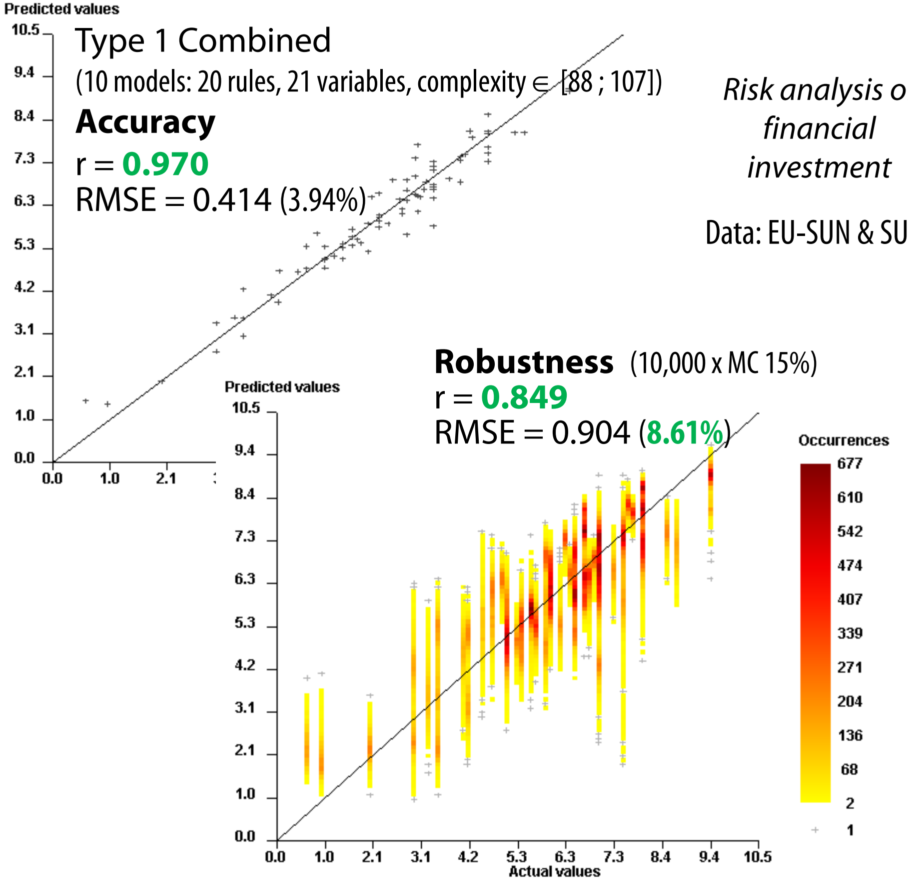


77 variables, 16 rules
(complexity 2,672)

Accuracy
r = **1.000**
RMSE = 0.001 (**0.01%**)

Robustness (10,000 x MC 15%)
r = **0.408**
RMSE = 1.651 (**15.73%**)

Trivial result:
Accurate model but **non-robust**



Type 1 Combined
(10 models: 20 rules, 21 variables, complexity ∈ [88 ; 107])

Accuracy
r = **0.970**
RMSE = 0.414 (3.94%)

Robustness (10,000 x MC 15%)
r = **0.849**
RMSE = 0.904 (**8.61%**)

Risk analysis of financial investment

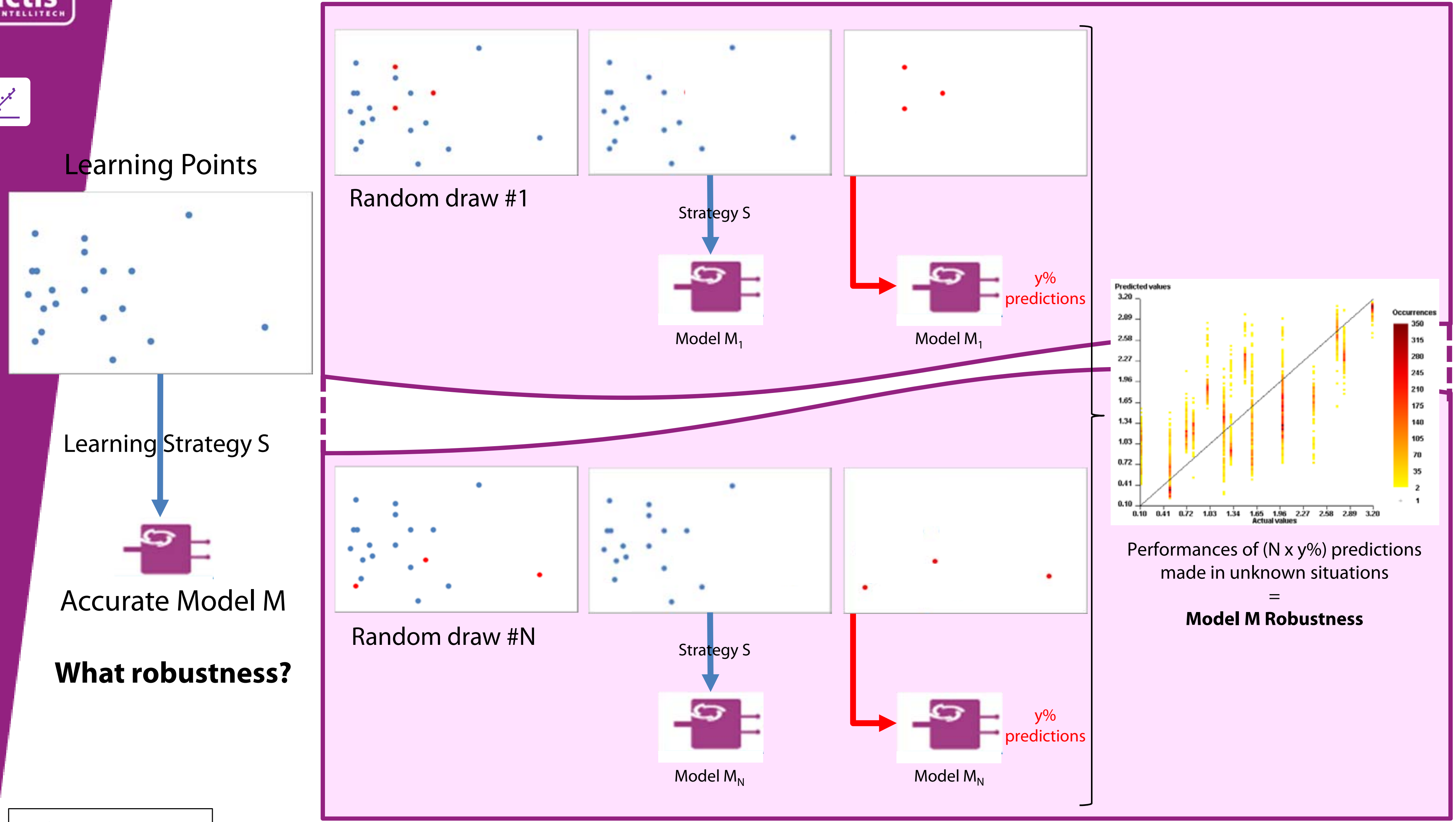
Data: EU-SUN & SUP

Use of robustness estimators:
Less accurate model but **robust**

Results by **xtractis**® Generate 9.1.16139

Robustness

Training/Validation cross validation (Monte-Carlo)



Learning Points

Learning Strategy S

Accurate Model M

What robustness?

N cycles of y%

(100-y)% training pts

y% validation pts
(unknown situations)

Random draw #1

Strategy S

Model M_1

Model M_1

y%
predictions

Random draw #N

Strategy S

Model M_N

Model M_N

y%
predictions

Performances of (N x y%) predictions
made in unknown situations
=
Model M Robustness

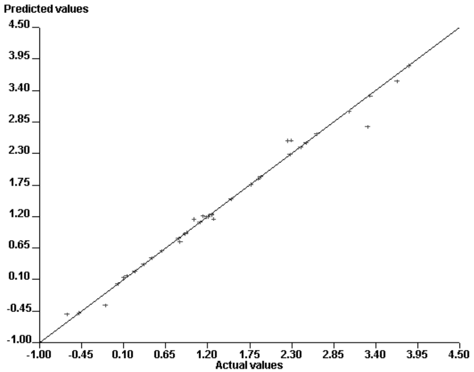
→ Learning

→ Predictions

Specificities of xtractis®

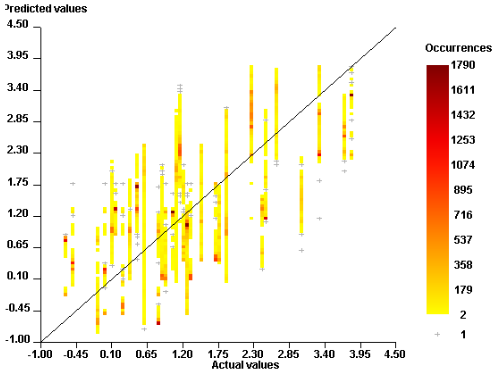
noise detection

slide 10
slide 16



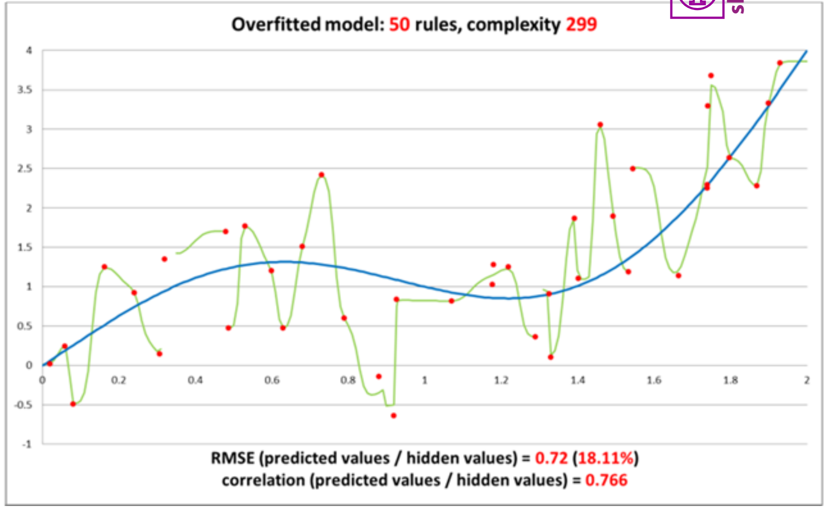
Accuracy

Accuracy: r = **0.995**, RMSE = 0.12 (**3.00%**)



Robustness 10,000 x MC 25%

Robustness: r = **0.643**, RMSE = 0.93 (**23.25%**)



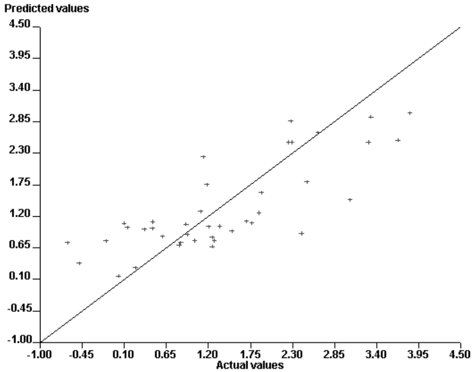
Overfitted model: 50 rules, complexity 299

RMSE (predicted values / hidden values) = 0.72 (18.11%)
correlation (predicted values / hidden values) = 0.766

Gaussian noise, σ=20% x I([0;4])

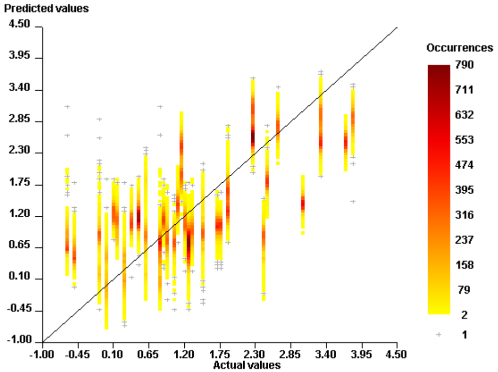
RMSE (learning points / hidden values) = 0.74 (18.39%)
Correlation (learning points / hidden values) = 0.753

● 40 learning points
— hidden law: y = Sin (πx) + x²
— xtractis®



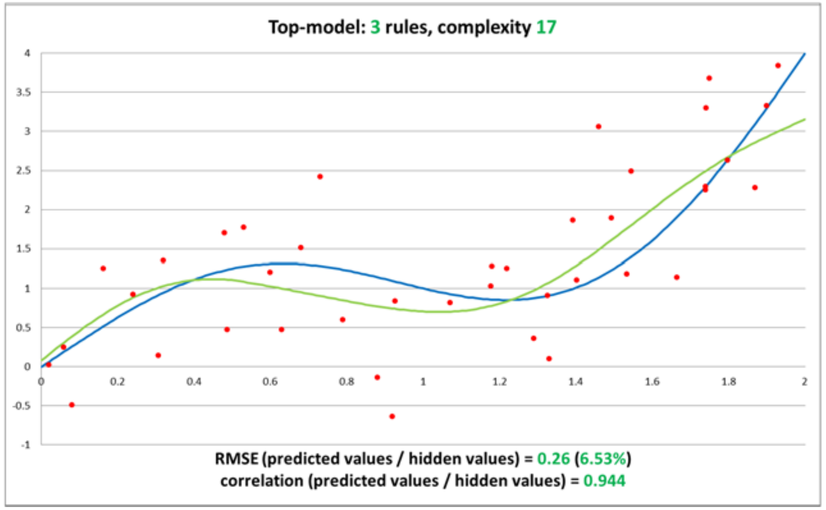
Accuracy

Accuracy: r = 0.792, RMSE = 0.69 (17.25%)



Robustness 10,000 x MC 25%

Robustness: r = 0.706, RMSE = 0.79 (19.75%)



Top-model: 3 rules, complexity 17

RMSE (predicted values / hidden values) = 0.26 (6.53%)
correlation (predicted values / hidden values) = 0.944

Modeling with noisy data
(**without** and **with** robustness analysis)

Results by **xtractis**® Generate 9.1.16419

Weak signals

importance of predictors with weak individual influence

slide 3
slide 10
slide 15

UPD
1501

Breast Cancer Diagnosis

(30 potential predictors)

Top-model: 21 predictors, 3 rules

Actual class

Predicted class

Decision	0	1
0	98.98%	**4.78%**
1	**1.02%**	95.22
Non-mapped	0.41%	2.55%

Robustness 1,000 x MC 15%

Rank	Var. ID	Label	dividual influen	Missing value
1	**11**	**radius Cell 2**	1	0,00%
2	**22**	**texture Cell 3**	0,436	0,00%
3	**8**	**concave points Cell 1**	0,274	0,00%
4	**29**	**symmetry Cell 3**	0,147	0,00%
5	**23**	**perimeter Cell 3**	0,12	0,00%
6	**28**	**concave points Cell 3**	0,105	0,00%
7	**21**	**radius Cell 3**	0,103	0,00%
8	**15**	**smoothness Cell 2**	0,078	0,00%
9	**2**	**texture Cell 1**	0,066	0,00%
10	**16**	**compactness Cell 2**	0,063	0,00%
11	**1**	**radius Cell 1**	0,052	0,00%
12	**25**	**smoothness Cell 3**	0,049	0,00%
13	**12**	**texture Cell 2**	0,046	0,00%
14	**18**	**concave points Cell 2**	0,045	0,00%
15	**10**	**fractal dimension Cell 1**	0,044	0,00%
16	**24**	**area Cell 3**	0,039	0,00%
17	**3**	**perimeter Cell 1**	0,039	0,00%
18	**27**	**concavity Cell 3**	0,033	0,00%
19	**4**	**area Cell 1**	0,025	0,00%
20	**7**	**concavity Cell 1**	0,022	0,00%
21	**30**	**fractal dimension Cell 3**	0,021	0,00%

Predictors with weak individual influence

Best learning strategy applied on the 2 predictors presenting the strongest individual influences (Cartesian approach)

2 predictors, 8 rules

Individual influence

Rank	Var. ID	Label	Individual influence	Missing value
1	**11**	**radius Cell 2**	1,000	0,0 %
2	**22**	**texture Cell 3**	0,570	0,0 %

Actual class

Predicted class

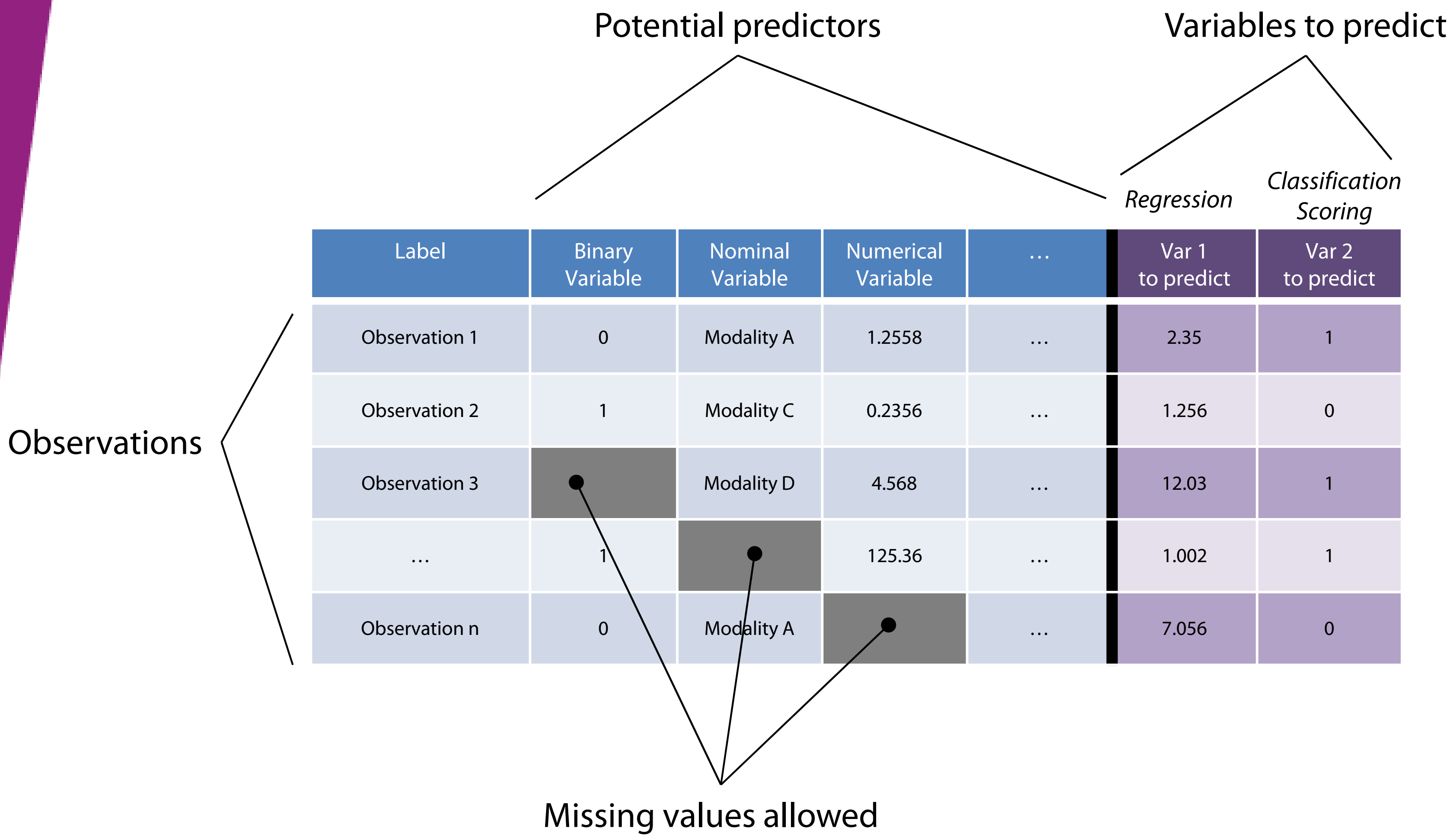
Decision	0	1
0	82.04%	**22.46%**
1	**17.96%**	77.54%
Non-mapped	0.88%	1.04%

Robustness 1,000 x MC 15%

Results by **xtractis**® Generate 8.0.10349

Data for xtractis®

structured, quantitative/qualitative



Potential predictors

Variables to predict

Regression

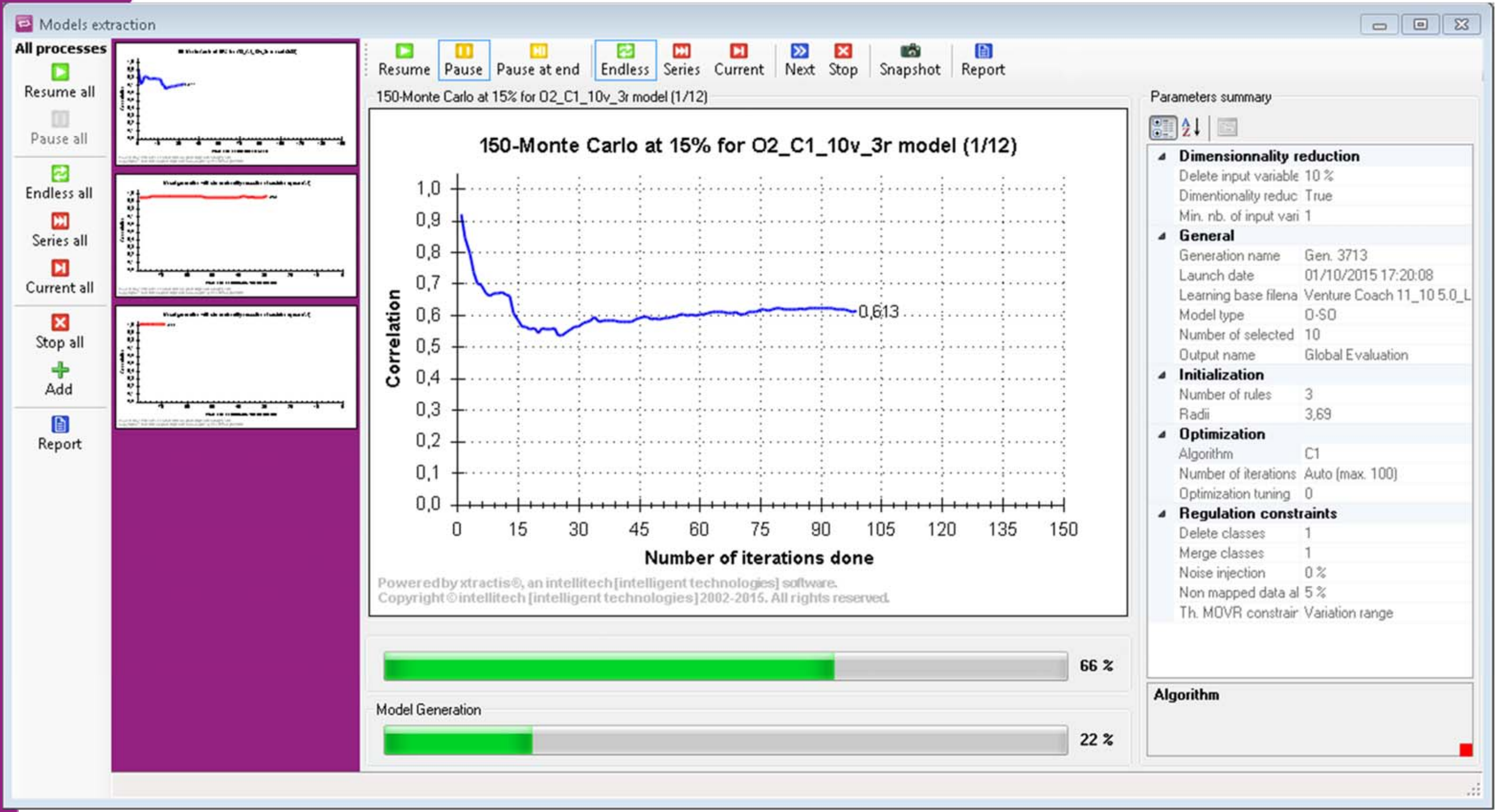
Classification
Scoring

Label	Binary Variable	Nominal Variable	Numerical Variable	…	Var 1 to predict	Var 2 to predict
Observation 1	0	Modality A	1.2558	…	2.35	1
Observation 2	1	Modality C	0.2356	…	1.256	0
Observation 3		Modality D	4.568	…	12.03	1
…	1		125.36	…	1.002	1
Observation n	0	Modality A		…	7.056	0

Observations

Missing values allowed

Screenshots



150-Monte Carlo at 15% for O2_C1_10v_3r model (1/12)

150-Monte Carlo at 15% for O2_C1_10v_3r model (1/12)

Powered by xtractis®, an intellitech [intelligent technologies] software.
Copyright © intellitech [intelligent technologies] 2002-2015. All rights reserved.

66 %

Model Generation

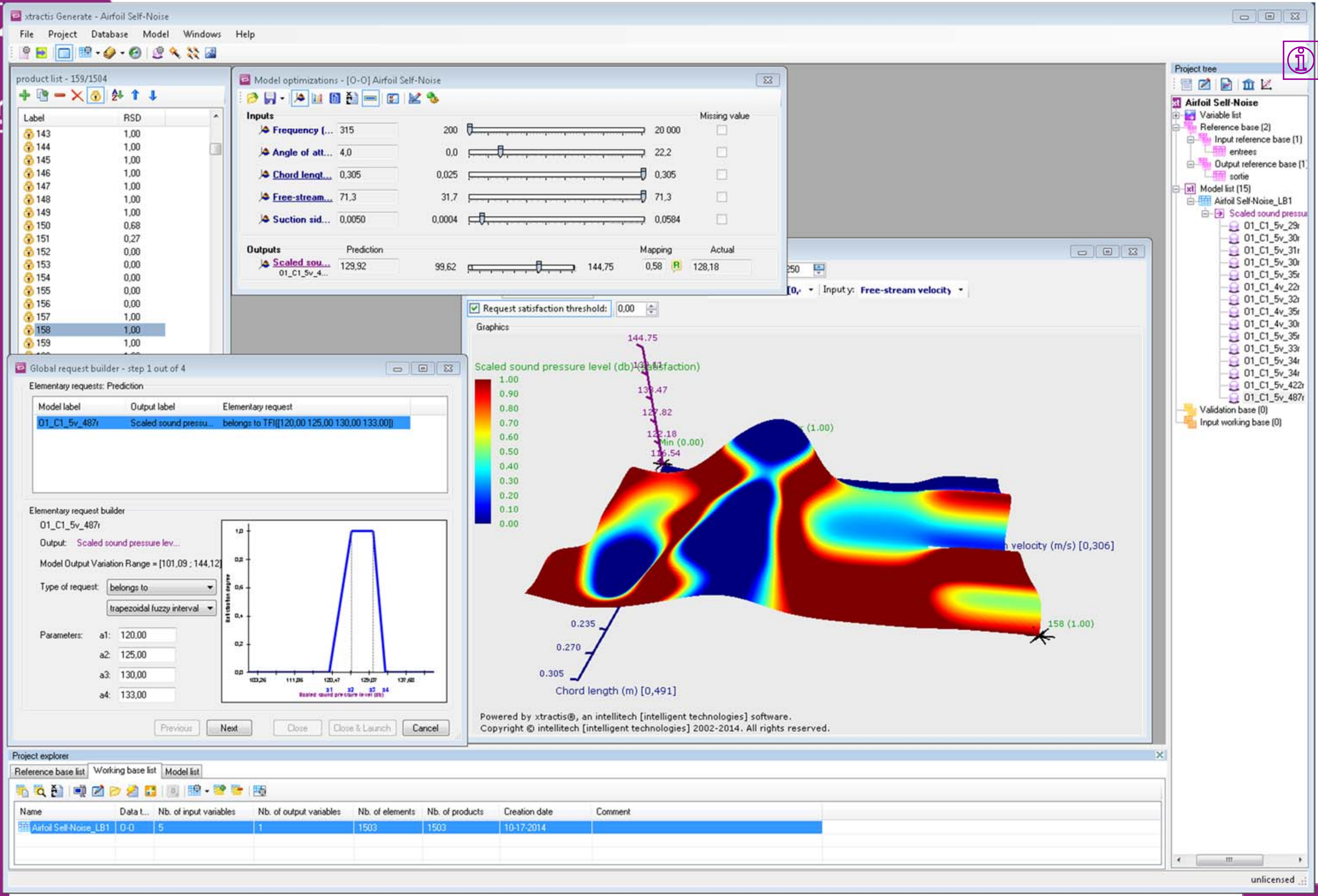
22 %

Parameters summary

Dimensionnality reduction	
Delete input variable	10 %
Dimentionality reduc	True
Min. nb. of input vari	1
General	
Generation name	Gen. 3713
Launch date	01/10/2015 17:20:08
Learning base filena	Venture Coach 11_10 5.0_L
Model type	O-SO
Number of selected	10
Output name	Global Evaluation
Initialization	
Number of rules	3
Radii	3,69
Optimization	
Algorithm	C1
Number of iterations	Auto (max. 100)
Optimization tuning	0
Regulation constraints	
Delete classes	1
Merge classes	1
Noise injection	0 %
Non mapped data al	5 %
Th. MOVR constrair	Variation range

Algorithm

Screenshots



product list - 159/1504

Label	RSD
143	1,00
144	1,00
145	1,00
146	1,00
147	1,00
148	1,00
149	1,00
150	0,68
151	0,27
152	0,00
153	0,00
154	0,00
155	0,00
156	0,00
157	1,00
158	1,00
159	1,00

Model optimizations - [O-O] Airfoil Self-Noise

Inputs				Missing value
Frequency (...	315	200	20 000	☐
Angle of att...	4,0	0,0	22,2	☐
Chord lengt...	0,305	0,025	0,305	☐
Free-stream...	71,3	31,7	71,3	☐
Suction sid...	0,0050	0,0004	0,0584	☐

Outputs	Prediction			Mapping	Actual
Scaled sou... 01_C1_5v_4...	129,92	99,62	144,75	0,58	128,18

Input y: Free-stream velocity

☑ Request satisfaction threshold: 0,00

Graphics

Powered by xtractis®, an intellitech [intelligent technologies] software.
Copyright © intellitech [intelligent technologies] 2002-2014. All rights reserved.

Global request builder - step 1 out of 4

Elementary requests: Prediction

Model label	Output label	Elementary request
01_C1_5v_487r	Scaled sound pressu...	belongs to TFI([120,00 125,00 130,00 133,00])

Elementary request builder

01_C1_5v_487r

Output: Scaled sound pressure lev...

Model Output Variation Range = [101,09 ; 144,12]

Type of request: belongs to

trapezoidal fuzzy interval

Parameters:
a1: 120,00
a2: 125,00
a3: 130,00
a4: 133,00

Project tree

- Airfoil Self-Noise
  - Variable list
  - Reference base (2)
    - Input reference base (1)
      - entrees
    - Output reference base (1)
      - sortie
  - Model list (15)
    - Airfoil Self-Noise_LB1
      - Scaled sound pressu
        - 01_C1_5v_29r
        - 01_C1_5v_30r
        - 01_C1_5v_31r
        - 01_C1_5v_30r
        - 01_C1_5v_35r
        - 01_C1_4v_22r
        - 01_C1_5v_32r
        - 01_C1_4v_35r
        - 01_C1_4v_30r
        - 01_C1_5v_35r
        - 01_C1_5v_33r
        - 01_C1_5v_34r
        - 01_C1_5v_34r
        - 01_C1_5v_422r
        - 01_C1_5v_487r
  - Validation base (0)
  - Input working base (0)

Project explorer

Name	Data t...	Nb. of input variables	Nb. of output variables	Nb. of elements	Nb. of products	Creation date	Comment
Airfoil Self-Noise_LB1	0-0	5	1	1503	1503	10-17-2014	

unlicensed