Lesson 5:

Examples of Application of MOEAs

António Gaspar-Cunha

Institute for Polymers and Composites/I3N, Dept. of Polymer Engineering,

University of Minho, Guimarães, Portugal

http://www.dep.uminho.pt/agc/

Universidad Carlos III de Madrid
January/February 2012

OUTLINE

- **Design of Shading Devices**
- **Polymer Extrusion**

  (***single and twin-screw extruders***)
- **Scale-up**
- **Feature Selection**

  (***bankruptcy prediction and cardiac SPEC diagnosis***)

Design of shading devices for buildings

(practical architecture problem)

Ecotect: Sustainable Building Design Software

Is a comprehensive concept-to-detail sustainable building design tool

Whole-building energy analysis
Calculate total energy use and carbon emissions of your building model on an annual, monthly, daily, and hourly basis, using a global database of weather information.

Thermal performance
Calculate heating and cooling loads for models and analyze effects of occupancy, internal gains, infiltration, and equipment.

Water usage and cost evaluation
Estimate water use inside and outside the building.

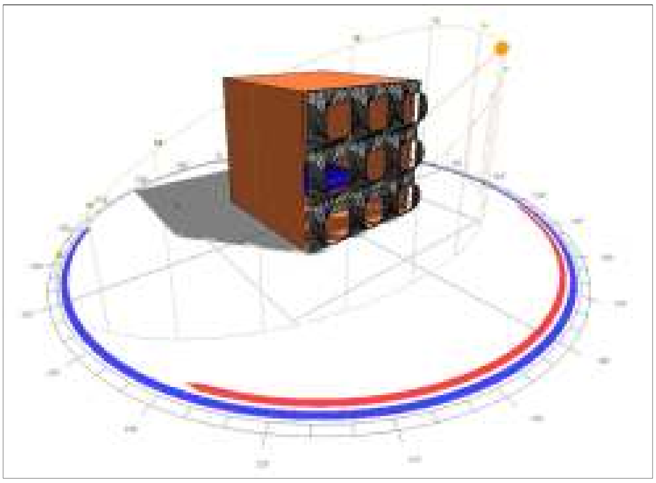
Solar radiation
Visualize incident solar radiation on windows and surfaces, over any period.

Daylighting
Calculate daylight factors and illuminance levels at any point in the model.

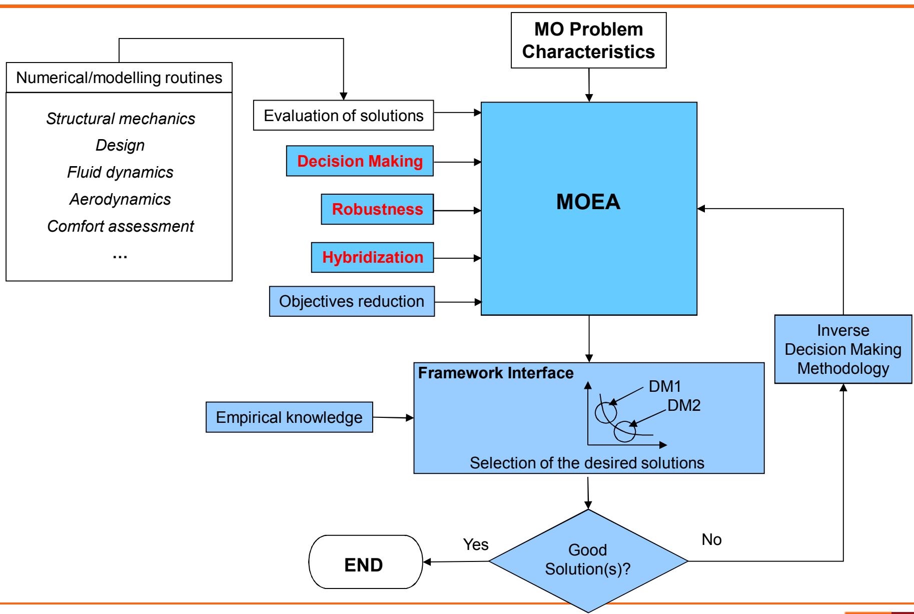
Shadows and reflections
Display the sun's position and path relative to the model at any date, time, and location.

DESIGN OF SHADING DEVICES

ECOTECT SOFTWARE



DESIGN OF SHADING DEVICES



Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

Comfort analysis
Structural Analysis
→ DM based on definition of weights

Aesthetic design
→ DM based on selection of a region on the Pareto frontier

DESIGN OF SHADING DEVICES

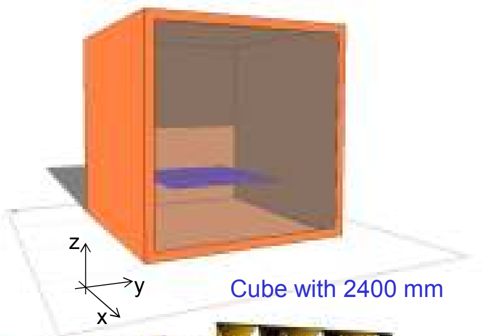
Comfort analysis
Structural Analysis
→ DM based on definition of weights

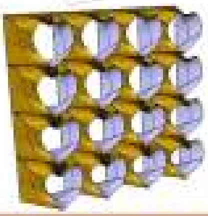
Aesthetic design
→ DM based on selection of a region on the Pareto frontier

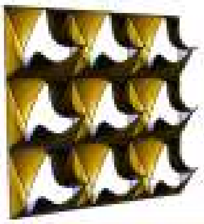
Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

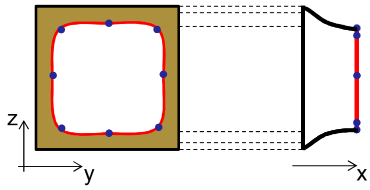
DESIGN OF SHADING DEVICES







Design Variables



Coordinates (x, y, z) of the 8 design points in the figure above:

- x [200, 400] mm
- y [-250, +250] mm
- z [-250, +250] mm

DESIGN OF SHADING DEVICES

Ecotect - Sustainable Building Design Software
analysis and simulation tool

Thermal performance analysis

Direct Solar Gains were calculated on the 21th of June (when the sun is highest) and on the 21th of December (when the sun is lowest). Hourly calculations were performed, and the total direct solar gain in Wh was used as a reference for the performance of the shading device.

The thermal settings for the cube were specified according to standard use of a living room: 60% humidity, 0,30m/s Air Speed, 200 lux Lighting Level, 1 person seated and reading 55W, no Sensible or Latent Gains, Air Change Rate of 0,50 and Wind Sensivity of 0,10 Air changes/hr. Not all of those parameters are important or included in the calculations used for simulation and analysis in this experiment.

Daylighting

The Daylight Factor was calculated over a grid at the height of 600mm and the average value of the results of three nodes, one in de center of the cube and one at a distance of 400mm from the back wall and from the left side wall, and another at the same distances from the back wall but at the right side of the cube, was used as the result.

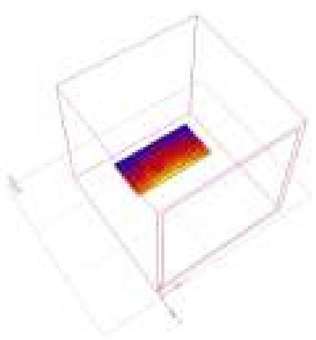
.

DESIGN OF SHADING DEVICES

Objective 1

Maximization of the Average DayLight Factor

(considering 3 Points on the Analysis Grid)



Maximize the daylight quality inside a room.

Ecotect uses the BRE Split Flux Method (DF = Sky Component + Externally Refl. Component + Internally Refl. Component) .

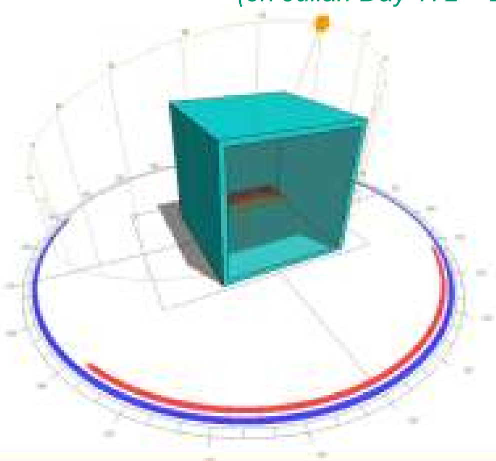
Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

Objective 2

Minimization of Direct Solar Gains in Summer

(on Julian Day 172 – 21th of June)



Minimize the increase in temperature inside a space by blocking the sunlight entering through the window.

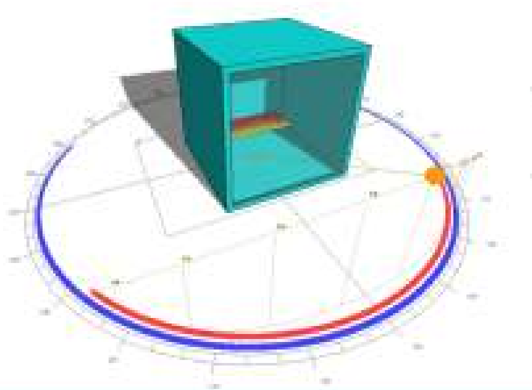
Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

Objective 3

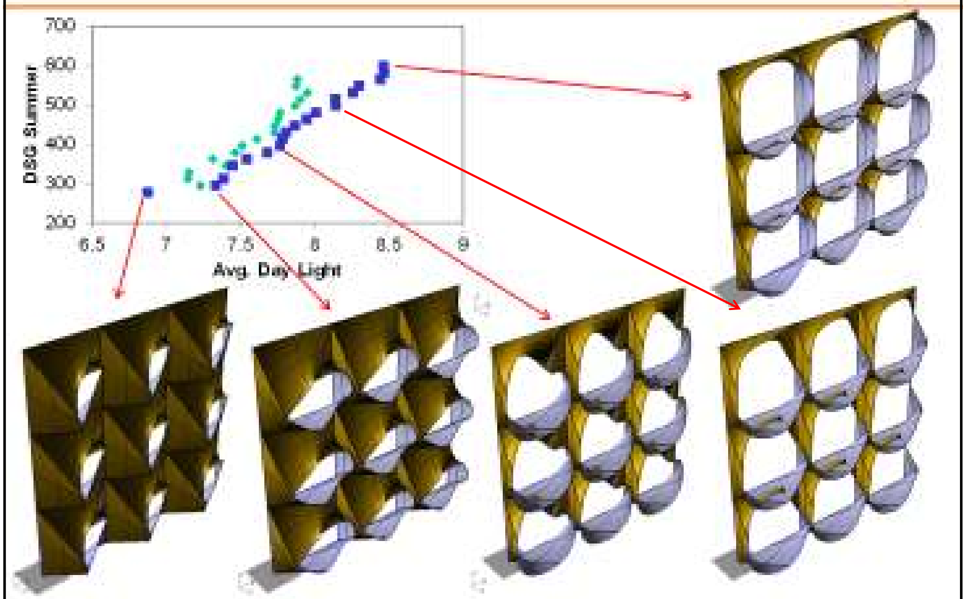
Maximization of Direct Solar Gains in Winter

(on Julian Day 355 – 21th of December)

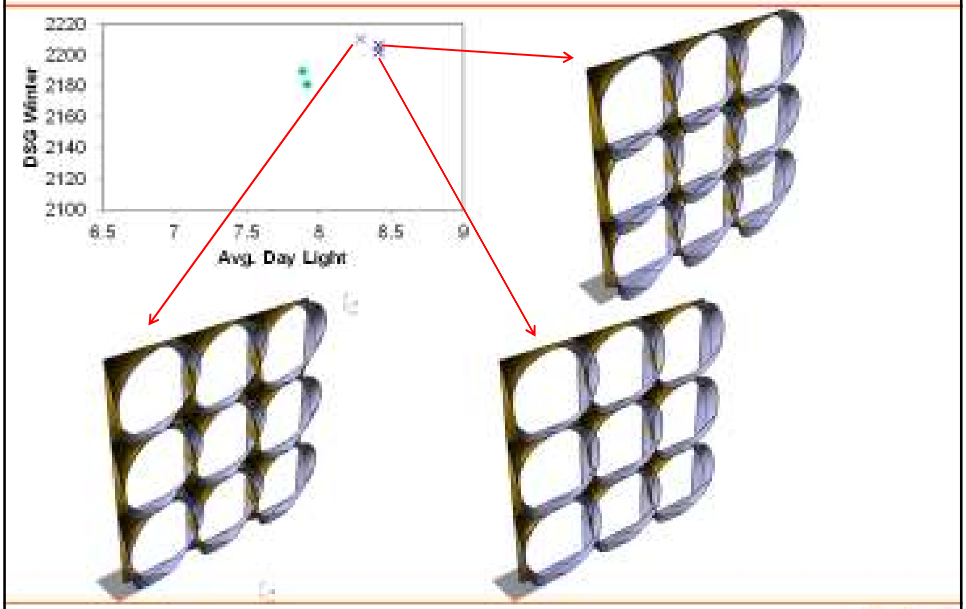


Maximize the increase in temperature inside a space without blocking the sunlight entering through the window.

DESIGN OF SHADING DEVICES

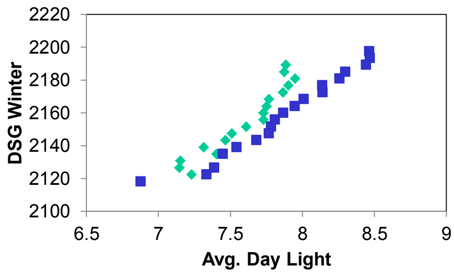


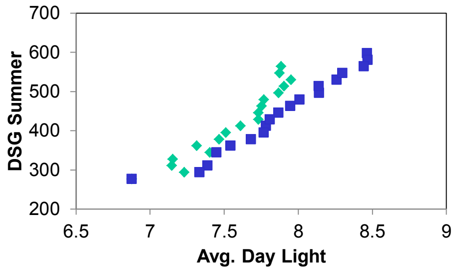
DESIGN OF SHADING DEVICES

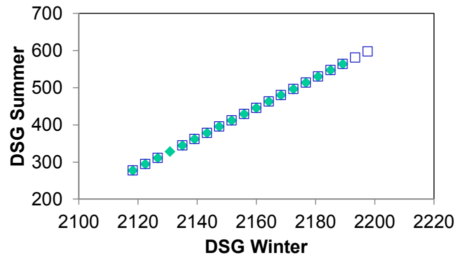


Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

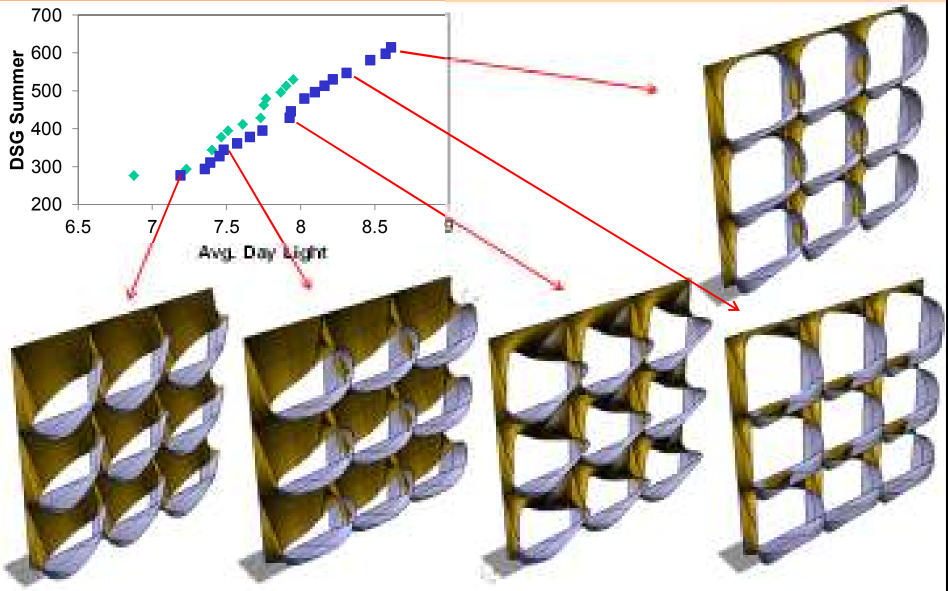






Dept. Polymer Engineering
University of Minho

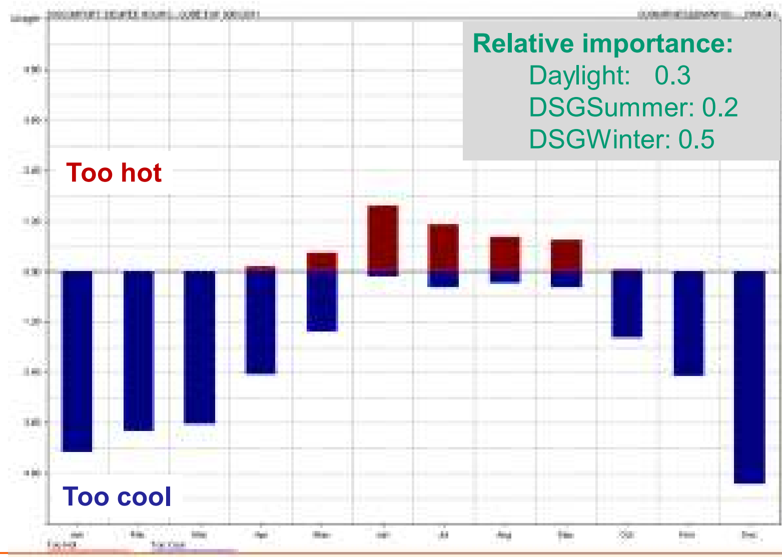
DESIGN OF SHADING DEVICES



Dept. Polymer Engineering
University of Minho

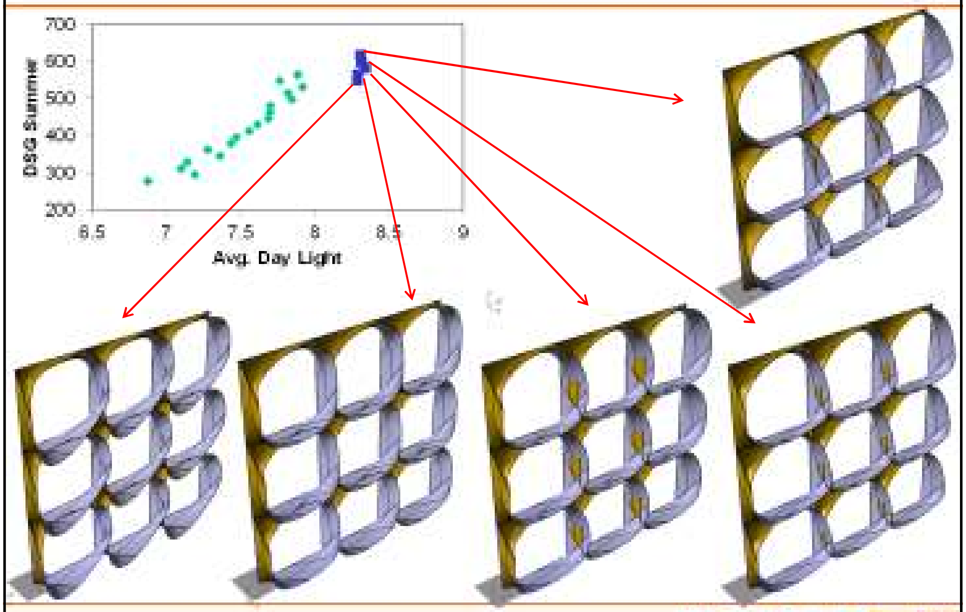
DESIGN OF SHADING DEVICES

Discomfort degree hours (Guimarães)



Dept. Polymer Engineering
University of Minho

DESIGN OF SHADING DEVICES

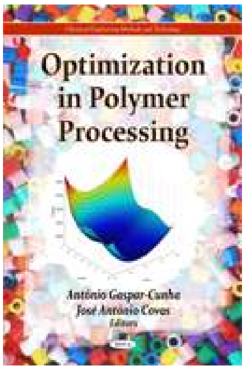


DESIGN OF SHADING DEVICES – Conclusions

- This study has introduced the use of a MOEA in the conceptual phase of the design process.
- The applied strategy for the use of a MOEA allowed for the DM to iteratively control the outcome and steer the process to a personal aesthetical solution.
- The DM can rely less on intuition to solve complicated and conflicting design requirements and concentrate efforts on innovative and aesthetical pleasing results.
- The utility of the design method proposed was demonstrated in the design of an architectural object which can be tested and validated in the real physical world.
- An effort was made in order to prepare the method for general use by less computer literate architects and designers for deployment in real world design processes.

POLYMER EXTRUSION

Application in Polymer Extrusion



https://www.novapublishers.com/catalog/product_info.php?products_id=20015

POLYMER EXTRUSION – *Twin-Screw*

Application of Pareto Local Search,

Multi-Objective Evolutionary Algorithms and

Multi-Objective Ant Colony Algorithms to the

Optimization of Co-Rotating Twin Screw Extruders

POLYMER EXTRUSION – *Twin-Screw*

Polymer extrusion

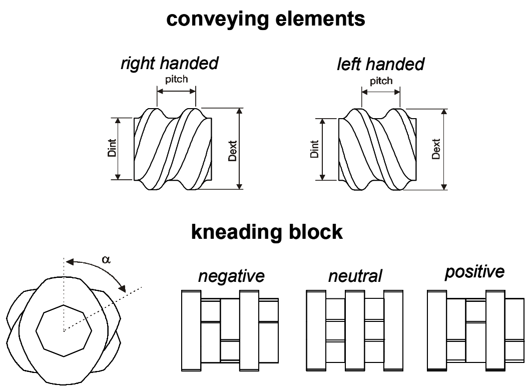
Extruder line

Co-rotating Twin-Screws

Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – *Twin-Screw*

Screw elements



Screws are built connecting screw elements with different geometrical properties.

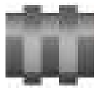
Dept. Polymer Engineering
University of Minho

Screw design for co-rotating twin screw extruders

1. Location of individual screw elements

2. Design of individual screw elements

•Pitch
•Internal screw diameter
•...

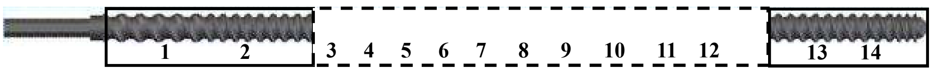


•Angle
•Thickness
•...

3. Both

POLYMER EXTRUSION – Twin-Screw

Twin Screw Configuration Problem (TSCP)

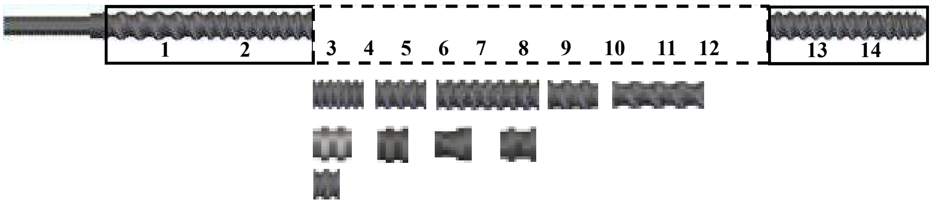


The aim of this work is to define the best location of a pre-defined number of available screw elements in order to maximize the process performance – **Twin Screw Configuration Problem (TSCP)**

Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – Twin-Screw

TSCP as a scheduling problem



TSCP is a scheduling problem where a certain number of resources (screw elements) must be allocated along the screw axis.

Dept. Polymer Engineering
University of Minho

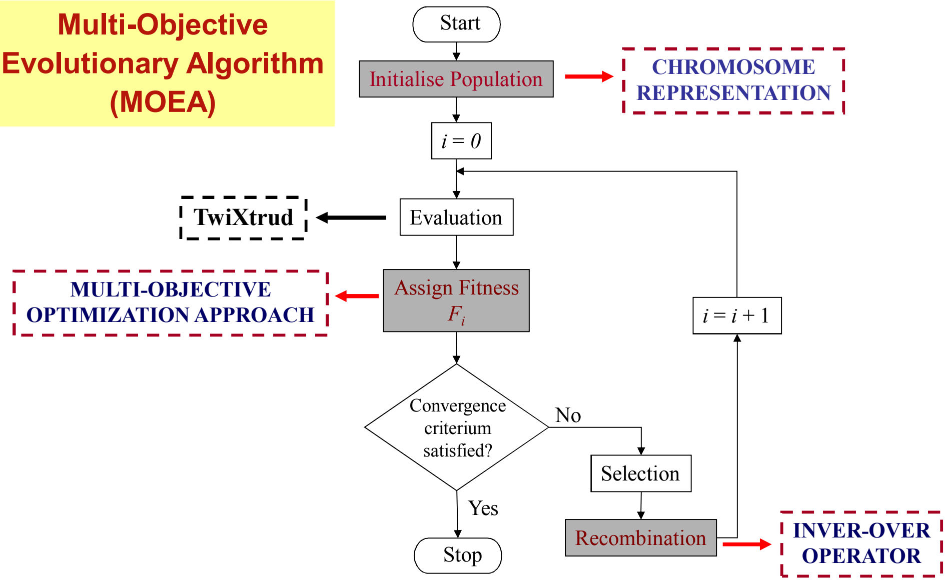
POLYMER EXTRUSION – Twin-Screw

Three alternatives

- **Evolutionary Algorithms (EAs):** Search and optimization metaheuristic based in the principles of natural evolution
- **Stochastic Local Search (SLS):** Optimization algorithm based in the exploration of the neighborhood
- **ANT Colony Optimization (ACO):** Optimization metaheuristic that takes inspiration of real ants foraging behavior

Dept. Polymer Engineering
University of Minho

Multi-Objective Evolutionary Algorithm (MOEA)



Stochastic Local Search (SLS)

- Based in the exploration of the neighborhood of one candidate solution;
- Widely used to solve hard combinatorial optimization problems;
- Can be applied to multi-objective problems changing the acceptance criterion:
    1. Scalarized Acceptance Criterion
    2. Component-wise Acceptance Criterion

POLYMER EXTRUSION – *Twin-Screw*

Two-Phase Local Search (TPLS)

$\Omega = \left(\overrightarrow{w_1}, \ldots, \overrightarrow{w_m}\right)$

$f = w_1 \times f_1 + w_2 \times f_2$

$A = \{\}$

s_t is a randomly generated solution

$s'_1 = SLS(s_1, \overrightarrow{w_1})$

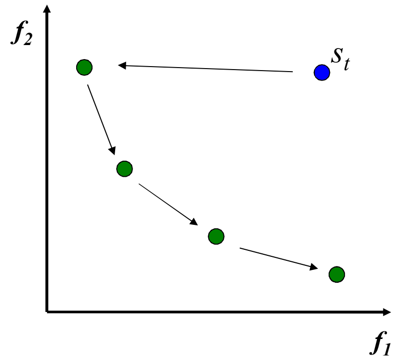
for all $i = 2, \ldots, m$ **do**

  $s'_i = SLS(s'_{i-1}, \overrightarrow{w_i})$

  $A = A \cup s'_i$

$A' = Filter(A)$

return A'



All non-dominated solutions found along the search process are saved into archive

POLYMER EXTRUSION – *Twin-Screw*

Pareto Local Search (PLS)

$t = 0$

s_t is a randomly generated solution

$A = \{s_t\}$

repeat

  for all $s'_t \in N(s_t)$ **do**

    $Evaluate(s'_t)$

    if s'_t is not dominated **then**

      $UpdateArchive(s'_t)$

    end if

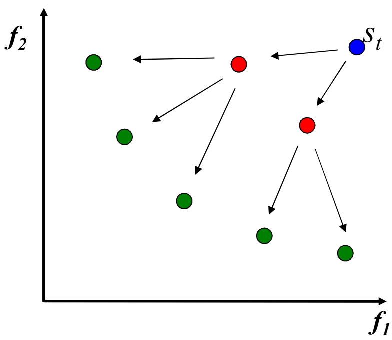
  end for

s_t *defined as visited*

$t = t + 1$

$s_t = PickArchive()$

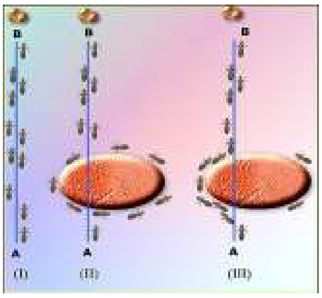
until *all solutions in archive are visited*



Application of an archive bounding technique avoiding the increase of non-dominated solutions

POLYMER EXTRUSION – Twin-Screw

Multi-objective ant colony optimization



(I) Real ants follow a path between nest and food source;

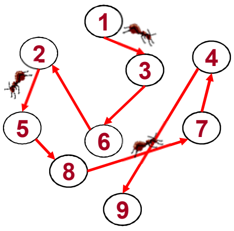
(II) An obstacle appears on the path. Ants choose whether to turn left or right with equal probability. Pheromone is deposited more quickly on the shorter path.

(III) Ants choose the shorter path.

POLYMER EXTRUSION – Twin-Screw

Representation

Path: screw configuration



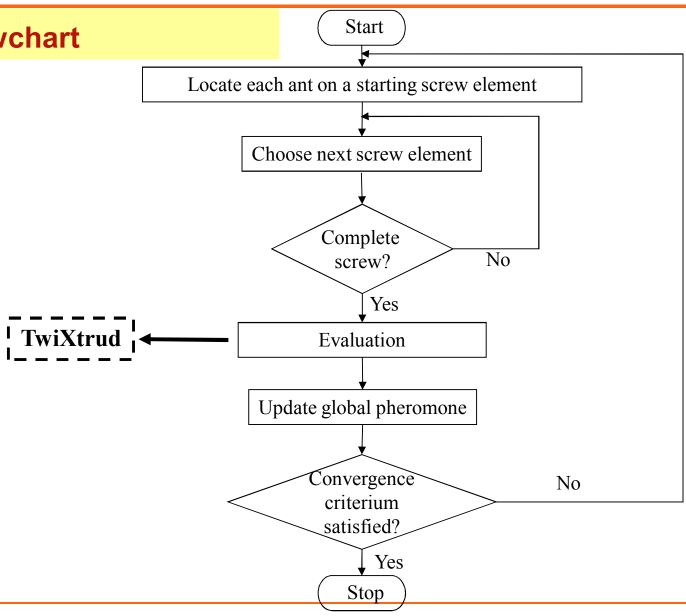
Construction of solutions:
Next screw element to be used is chosen from the screw elements not yet used;
This choice is probabilistic, based on the "intensity of pheromone".

Pheromone update:
Good solutions: positive reinforcement (deposit pheromone)
Bad solutions: negative reinforcement (evaporation)

Pheromone limits → to avoid stagnation

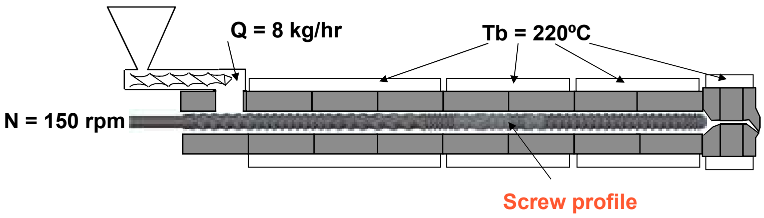
POLYMER EXTRUSION – Twin-Screw

Flowchart

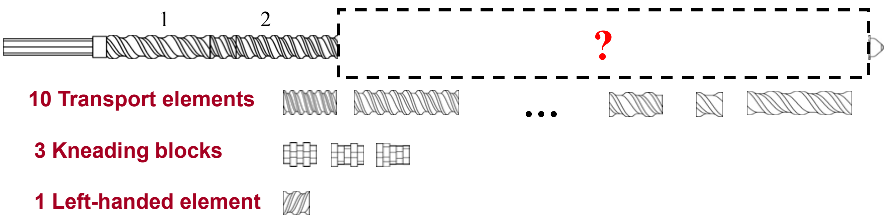


POLYMER EXTRUSION – Twin-Screw

The problem to solve



Screw profile (*location of 14 screw elements*)



POLYMER EXTRUSION – *Twin-Screw*

The problem to solve

1. Polymer: PP *(HB121J, from Borealis)*

2. Extruder: Leistritz LSM 30.34 (***Configuration of the individual screw elements***)

Element N°	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Length	97.5	120	45	60	30	30	30	60	30	120	30	120	37.5	60	60	30
Pitch	45	30	KB -45	30	-20	60	30	20	KB -60	30	30	60	KB -30	45	30	20

3. Optimization criteria, aim of optimization and prescribed range of variation

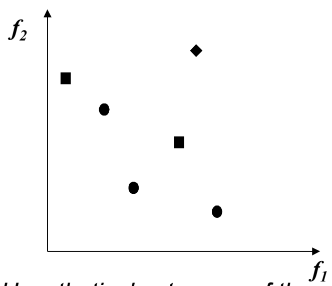
	Criteria	Aim	X_{min}	X_{max}
Case Study 1	Average strain	Maximize	1000	15000
	Specific mechanical energy – SME (MJ/kg)	Minimize	0.5	2
Case Study 2	Average strain	Maximize	1000	15000
	Viscous dissipation – Tmax/Tb	Minimize	0.9	1.5
Case Study 3	Specific mechanical energy – SME (MJ/kg)	Minimize	0.5	2
	Viscous dissipation – Tmax/Tb	Minimize	0.9	1.5

Dept. Polymer Engineering
University of Minho

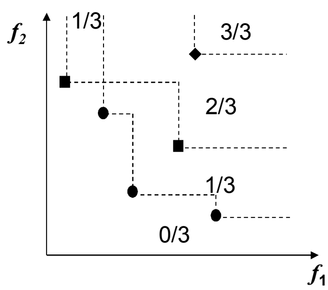
POLYMER EXTRUSION – *Twin-Screw*

Empirical Attainment Function (EAF)

Estimates for each point *z* in the objective space the probability that *z* is attained in one optimization run.



Hypothetical outcomes of three runs in a minimization problem



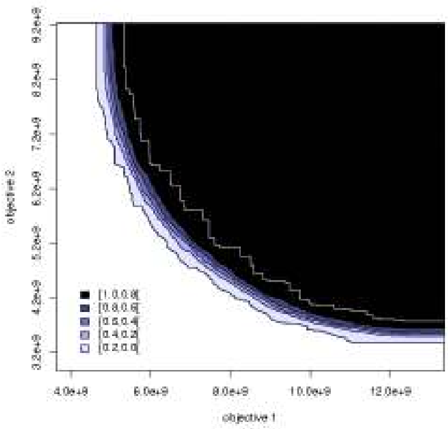
Relative frequencies for the distinct regions in the objective space

38

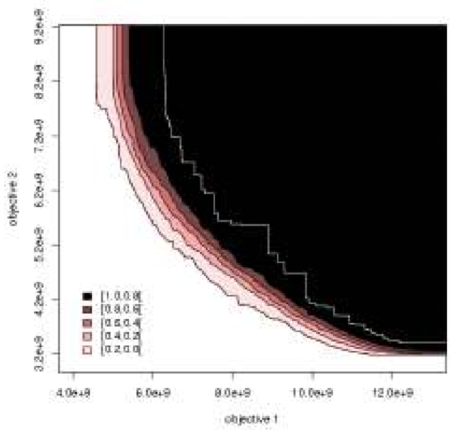
Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – Twin-Screw

example (step 1: compute the EAFs)



EAF_{blue}



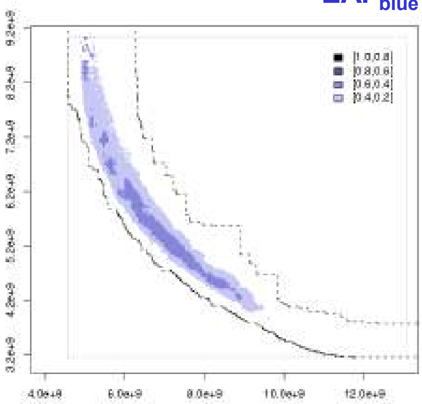
EAF_{red}

Dept. Polymer Engineering
University of Minho

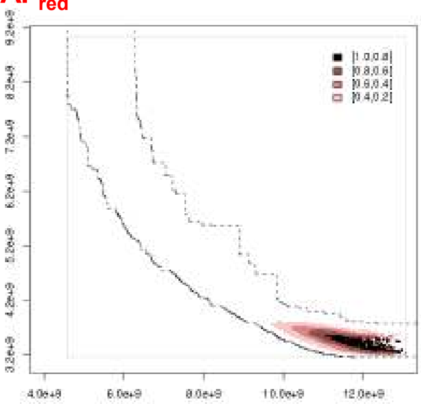
POLYMER EXTRUSION – Twin-Screw

example (step 2: plotting the differences between EAFs)

EAF_{blue} - EAF_{red}



differences are positive

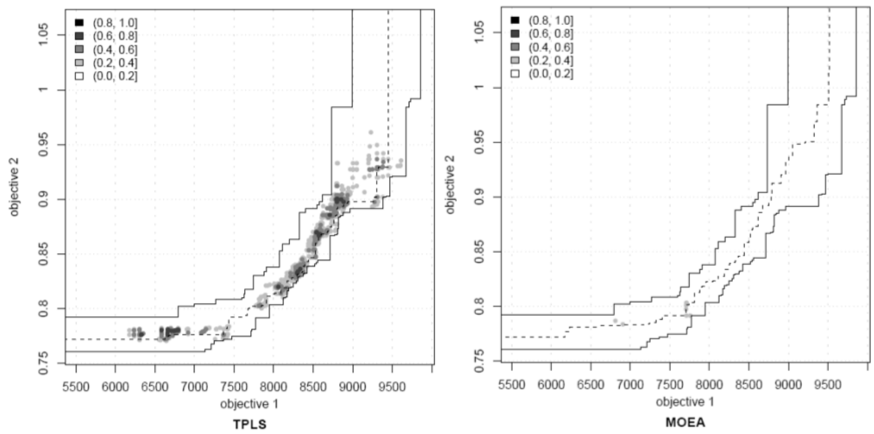


differences are negative

Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – *Twin-Screw*

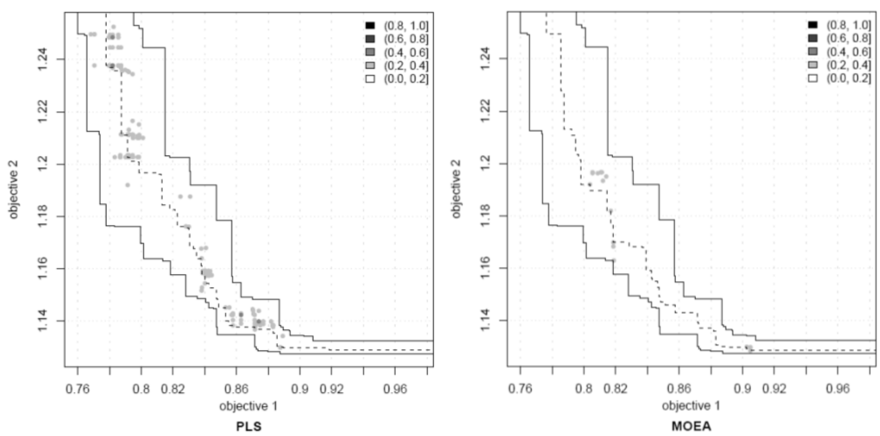
TPLS *vs* MOEA (case study 1)



Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – *Twin-Screw*

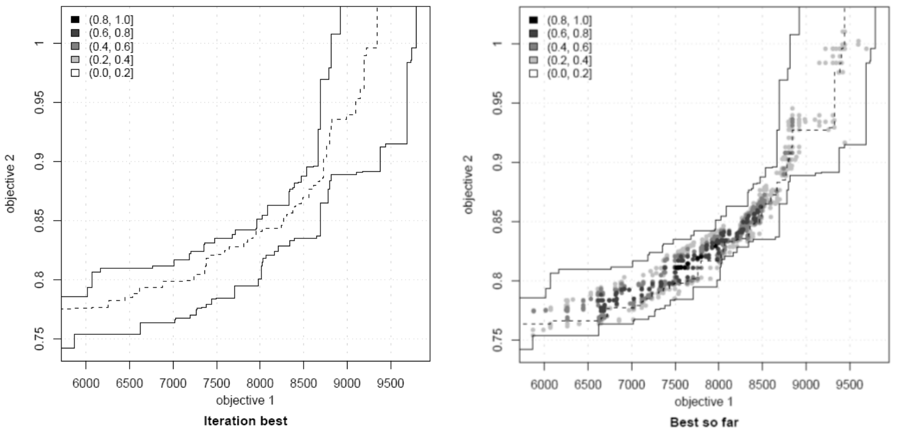
PLS *vs* MOEA (case study 3)



Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – *Twin-Screw*

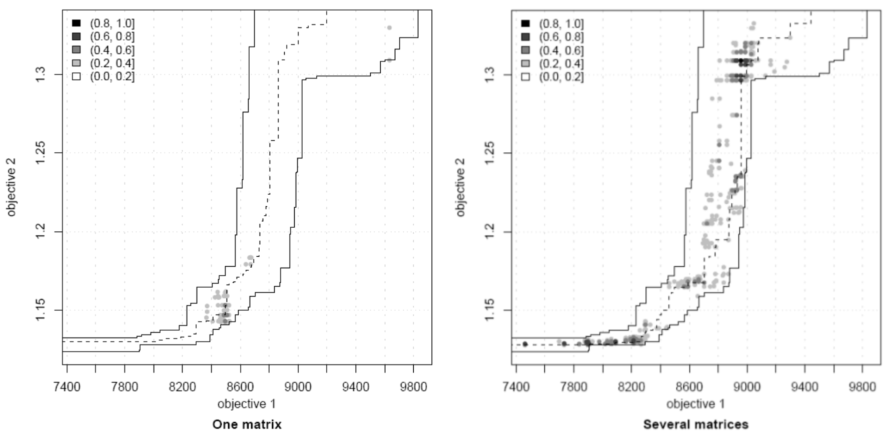
MO-ACO: Iteration best *vs* best so far strategies (case study 1)



Dept. Polymer Engineering
University of Minho

POLYMER EXTRUSION – *Twin-Screw*

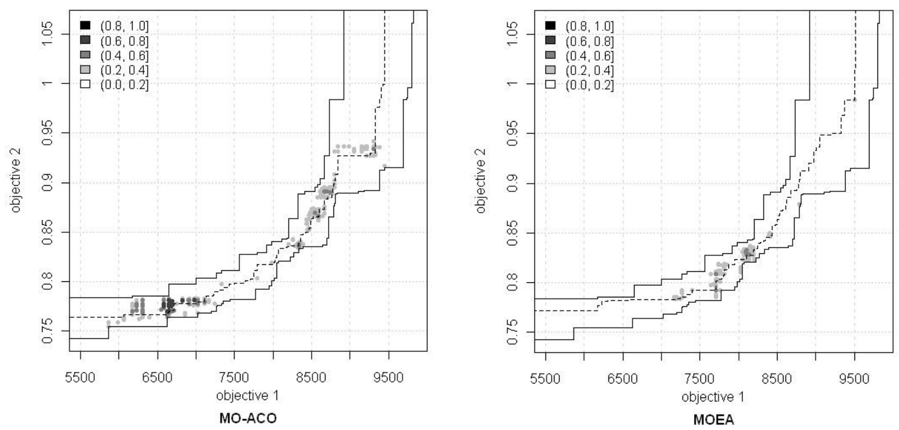
MO-ACO: Pheromone information (case study 2)



Dept. Polymer Engineering
University of Minho

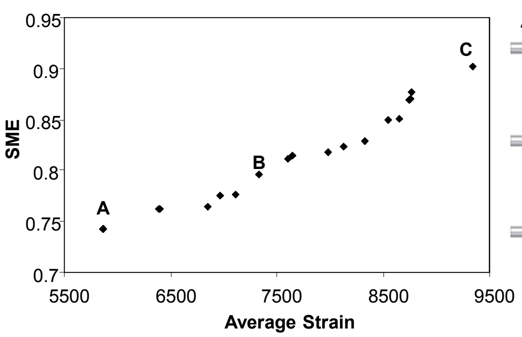
POLYMER EXTRUSION – *Twin-Screw*

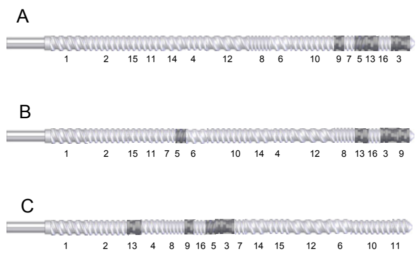
MO-ACO *vs* MOEA (case study 1)



POLYMER EXTRUSION – *Twin-Screw*

screw configurations





Screw configurations obtained in the optimization process are in agreement with the physical knowledge of the process

POLYMER EXTRUSION – *Twin-Screw*

- Simple SLS Algorithms and MO-ACO were applied with success for solving the TSCP.
- The solutions obtained are in agreement with the knowledge about the process and have physical meaning.
- The good performance obtained with the simple SLS algorithms indicates that the development of a hybrid algorithm through the incorporation of iterative improvement methods can be a good way to further improve the performance.

SCALE-UP

Scale-up

SCALE-UP

- In science and engineering, researchers are often challenged with the need to replicate the innovative results obtained in an equipment of a given size in another equipment of a different dimension.
- In practice, this often involves passing from laboratory or prototype dimensions to industrial level.
- The process is known as scale-up and consists in ensuring that the values of the criteria that describe the process characteristics at a given scale are preserved at different scales.

SCALE-UP

- Available scale-up rules are often based on oversimplified process analyses and generate unsatisfactory results.
- Scale-up can also be understood as an optimization process where various objectives are to be satisfied simultaneously, i.e., the performance at the two scales must be as similar as possible.
- For that purpose, the adoption of a multi-objective optimization algorithm is proposed.

SCALE-UP

Industrially-oriented R&D

Problem identified in the production plant (*e.g. material degradation*) → Studies performed in lab processing equipment → Problem solved in the lab → Results exported to production plant

New polymer system developed in the lab (*e.g. nanocomposite, compatibilized blend*) → Screw profile and operating conditions optimized in the lab → Results exported to production plant

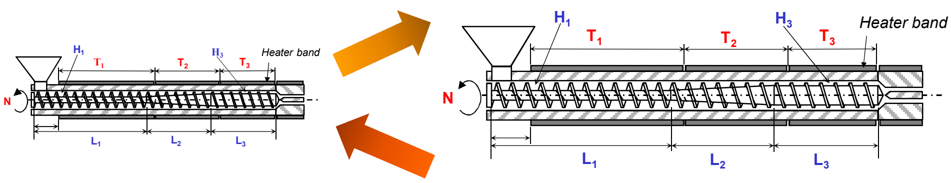
Successful production operation → New equipment with higher output

SCALE-UP

To define the *screw geometry and/or the operating conditions* that induce the same thermo-mechanical environment in machines of different sizes processing the same material.

↓

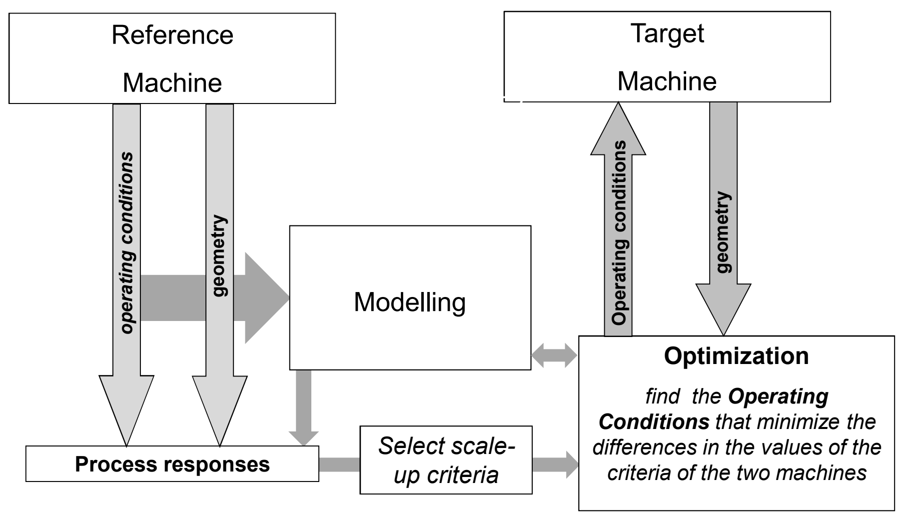
Extrusion scale-up (*scale-down*) rules



SCALE-UP

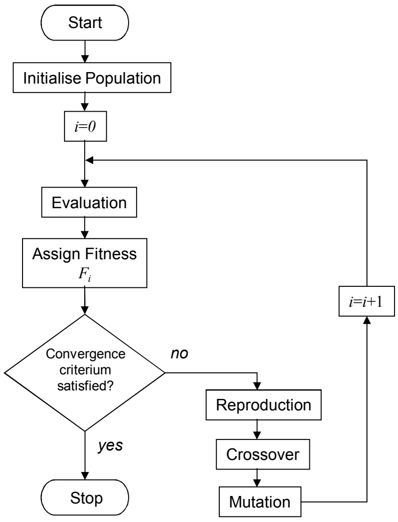
SCALE-UP

Methodology Proposed



SCALE-UP

Multi-Objective Evolutionary Algorithm (*MOEA*)



Population: *set of individuals*

Initialization of population: *random definition of all individuals of the population*

Evaluation: *calculation of the values of the criteria using the modelling routine*

Fitness: *calculation of a single value identifying the performance of the individuals*

Reproduction: *selection of the best individuals for crossover and/or mutation*

Crossover/Mutation: *methods to obtain new individuals for the next generation (i+1)*

SCALE-UP

Type of Objective Functions

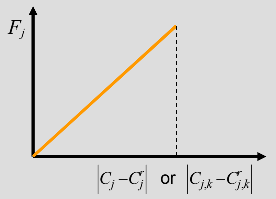
F_j

Single value

$$F_j = \frac{|C_j - C_j^r|}{C_j^r}$$

C_j

SME
Mixing index (e.g. WATS)
Total strain
Power consumption
Avg. Shear Rate
Avg. Shear Stress
Viscous Dissipation
Adimensional nos.

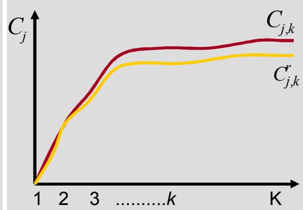


Function

$$F_j = \frac{\sum_{k=1}^{K} \frac{|C_{j,k} - C_{j,k}^r|}{C_{j,k}^r}}{K}$$

$C_{j,k}$

Pressure profile
Power consump. profile
Temperature profile
Shear rate profile
Viscosity profile
Shear stress profile



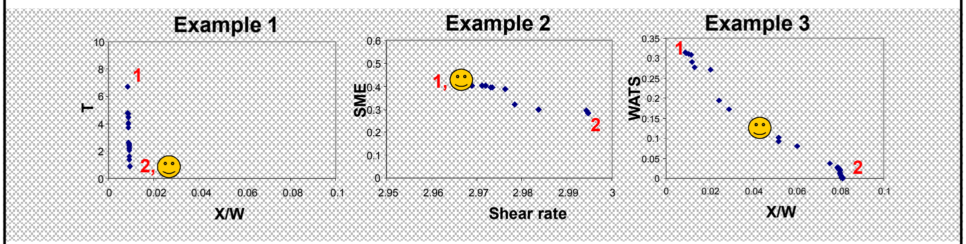
SCALE-UP

Illustrative Results

Scaling-up - operating conditions (*fixed geometry*) (N, T_{b1}, T_{b2}, T_{b3})

Criterion	Symbol	Type
Melting profile	*X/W*	Function
Melt temperature profile	T_{melt}	Function
Shear rate profile	$\dot{\gamma}$	Function
Specific mechanical energy	*SME*	Single value
Degree of mixing	*WATS*	Single value

Example	Criteria
1	*X/W*, T_{melt}
2	$\dot{\gamma}$, SME
3	*X/W*, WATS



SCALE-UP - Conclusions

- A general scale-up methodology, based in the use of MOEA coupled with modelling and function objective calculation routines, was proposed.
- This approach is able to take into account the character multi-objective of the problem and, due to its modularity, can be used in a wide range of applications.
- The capabilities of the approach were shown with an example of scale-up in the field of single screw extrusion.

FEATURE SELECTION – Bankruptcy

Multi-Objective Evolutionary Algorithms for

Feature Selection:

Application in Bankruptcy Prediction

FEATURE SELECTION – Bankruptcy

- The high importance of financial bankruptcy prediction for banks, insurance companies, creditors and investors.
- One of the most important threats for business is the credit risk.
- The rate of bankruptcies have increased in recent years.
- The companies develop sophisticated schemes to hide their real situation.

↓

The ability to discriminate between “good” customers from potential bad ones is of crucial importance.

Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

The Problem

The bankruptcy prediction problem can be stated as follows: given a set of financial statements from a company over one, or several years, predict the probability that it will become distressed over a given period, normally the next year or two ahead.

FEATURE SELECTION – Bankruptcy

Objective

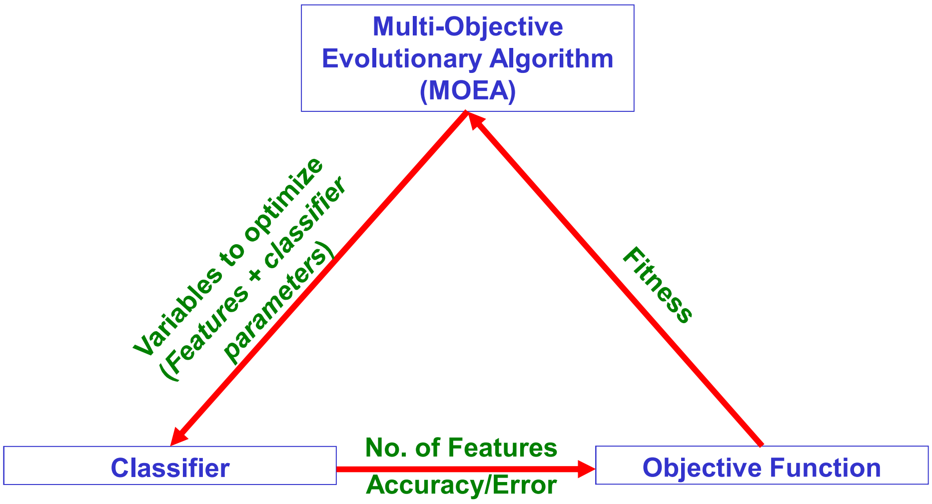
To apply a Multi-Objective Evolutionary Algorithm (MOEA) for feature selection in the problem of bankruptcy prediction.

Objectives to be accomplished simultaneously:

- the minimization of the number of features used and
- the maximization of the accuracy of the classifier and/or the error accomplished by the classifier.

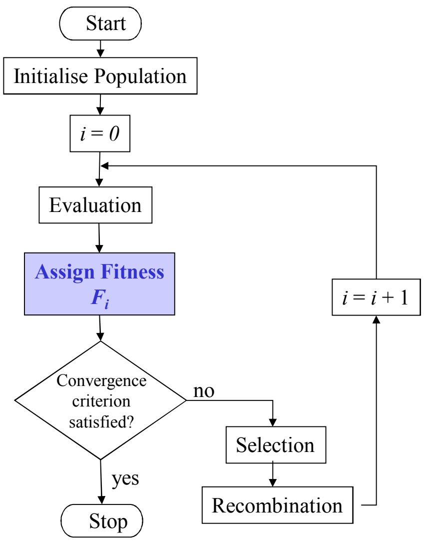
Dept. Polymer Engineering
University of Minho

Global Methodology



FEATURE SELECTION – Bankruptcy

Reduced Pareto Set Genetic Algorithm (RPSGAe)



RPSGAe sorts the population individuals in a number of pre-defined ranks using a clustering technique, in order to reduce the number of solutions on the efficient frontier.

Details:

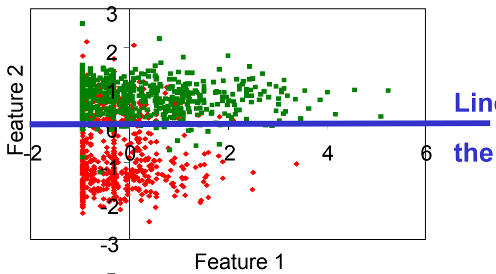
A. Gaspar-Cunha, J.A. Covas, - RPSGAe - A Multiobjective Genetic Algorithm with Elitism: Application to Polymer Extrusion, in Metaheuristics for Multiobjective Optimisation, Lecture Notes in Economics and Mathematical Systems, Gandibleux, X.; Sevaux, M.; Sörensen, K.; T'kindt, V. (Eds.), Springer, pp. 221-249, 2004.

Concept of Classification

Two classes:

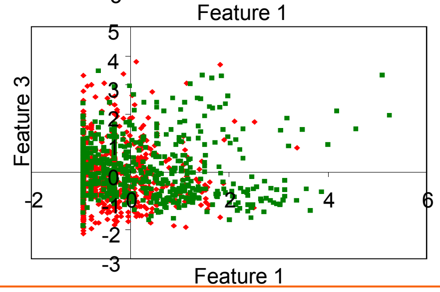
Red: Bankruptcy

Green: Healthy



Line separating the classes

It is not possible to separate the classes



Classification measures

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$R = \frac{TP}{TP + FN}$$

$$P = \frac{TP}{TP + FP}$$

$$e_I = \frac{FP}{TN + FP}$$

$$e_{II} = \frac{FN}{FN + TP}$$

$$F_m = \frac{2.P.R}{P + R}$$

prediction outcome		actual value P	actual value N
	P'	TP	FP
	N'	FN	TN

TP – Positives correctly classified
TN – Negatives correctly classified
FP – Positives incorrectly classified
FN – Negatives incorrectly classified
R – Recall or sensitivity
P – Precision
e_I, e_{II} – Classification errors

FEATURE SELECTION – Bankruptcy

Classification measures

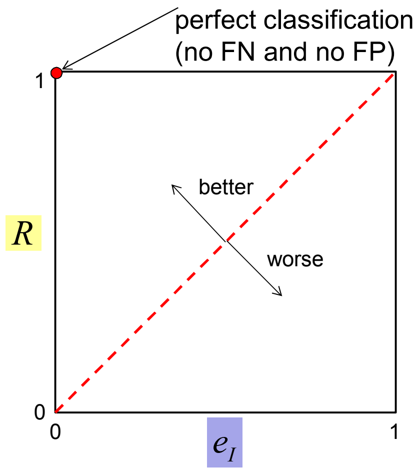
Receiver Operating Characteristic,

ROC curve: $FP_{rate} = f(TP_{rate})$

$e_I = f(R)$

$$e_I = \frac{FP}{TN + FP}$$

$$R = \frac{TP}{TP + FN}$$



e_I – Companies that test "well" but actually are in "bankruptcy" **[LOSSES]**

R – Companies that test "well" and are in reality "healthy" **[PROFITS]**

FEATURE SELECTION – Bankruptcy

METHODS: Characteristics and Parameters

Logistic regression (LG):

Method of the gradient descent

Parameters Studied:

- Training method
- Learning rate
- Training fraction

Objectives:

- NF + Accuracy

Support Vector Machines (SVM):

Method: ν-SVC and C-SVC

Kernel: Radial Basis Function (RBF)

Parameters Studied:

- Training method
- Regularization parameter (ν, C)
- Kernel parameter (γ)

Objectives:

- NF + Accuracy
- NF + Other measures

FEATURE SELECTION – *Bankruptcy*

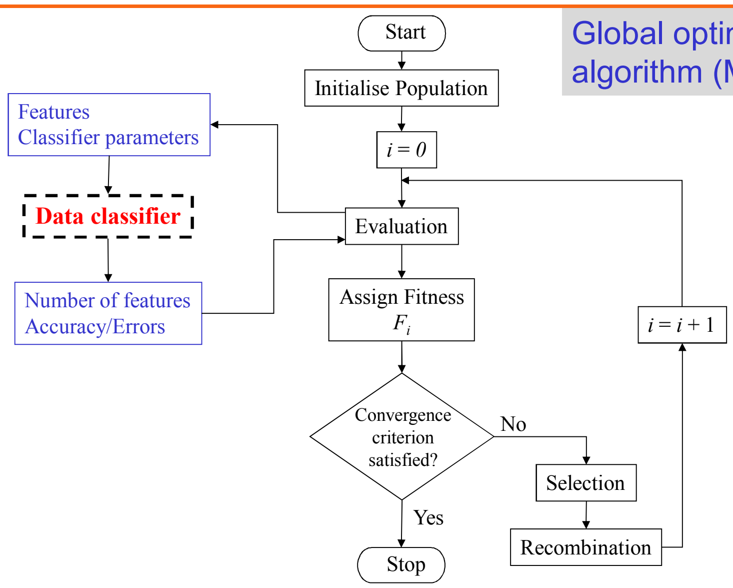
Training Approaches

Holdout method
- A given fraction (*training fraction*) is used to train
- The remaining is used to evaluate (accuracy)

K-fold validation
- The database is divided in *k-folds*
- In each k iteration:
  - (N-k)/N are used to train
  - k/N are used to evaluate
- The accuracy is the average of all folds
- Additional data is used to evaluate the accuracy

FEATURE SELECTION – *Bankruptcy*

Global optimization algorithm (MOEA)



Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

The problem to solve

A database containing financial statements of 1200 small and medium size private French companies during 2006.

Each company is characterized by a set of 30 features measuring its most important ratios, from profitability to debt.

FEATURE SELECTION – Bankruptcy

Feature	Designation
F1	Number of employees
F2	Capital Employed / Fixed Assets
F3	Financ. Debt / Capital Employed (%)
F4	Depreciation of Tangible Assets (%)
F5	Working capital / current assets
F6	Current ratio
F7	Liquidity ratio
F8	Stock Turnover days
F9	Collection period
F10	Credit Period
F11	Turnover per Employee (thousands euros)
F12	Interest / Turnover
F13	Debt Period days
F14	Financial Debt / Equity (%)
F15	Financial Debt / Cashflow

Feature	Designation
F16	Cashflow / Turnover (%)
F17	Working Capital / Turnover (days)
F18	Net Current Assets/Turnover (days)
F19	Working Capital Needs / Turnover (%)
F20	Export (%)
F21	Value added per employee
F22	Total Assets / Turnover
F23	Operating Profit Margin (%)
F24	Net Profit Margin (%)
F25	Added Value Margin (%)
F26	Part of Employees (%)
F27	Return on Capital Employed (%)
F28	Return on Total Assets (%)
F29	EBIT Margin (%)
F30	EBITDA Margin (%)

FEATURE SELECTION – Bankruptcy

RPSGA parameters

Due to the stochastic nature of the initial population, 16 runs of each experiment were performed:

- Number of ranks (N_{ranks}) = 30
- Indifference limits of the clustering algorithm = 0.1
- Size of the internal population (N) = 100
- Size of the external population (N_e) = 100
- Number of generations = 100
- Crossover rate = 0.8
- Mutation rate = 0.05

Details:

A. Gaspar-Cunha, J.A. Covas, - RPSGAe - A Multiobjective Genetic Algorithm with Elitism: Application to Polymer Extrusion, in Metaheuristics for Multiobjective Optimisation, Lecture Notes in Economics and Mathematical Systems, Gandibleux, X.; Sevaux, M.; Sörensen, K.; T'kindt, V. (Eds.), Springer, pp. 221-249, 2004.

FEATURE SELECTION – Bankruptcy

Logistic regression (NF + Accuracy):

Experiment	Training Method	Learning Rate	Training Fraction
Log1	H	0.001	2/3
Log2	H	0.01	2/3
Log3	H	0.02	2/3
Log4	H	0.1	2/3
Log5	H	0.01	0.5
Log6	H	0.01	0.8
Log11	K (10)	0.001	NA
Log12	K (10)	0.01	NA
Log13	K (10)	0.02	NA
Log14	K (10)	0.1	NA
Log15	K (5)	0.01	NA
Log20	H	[0.001, 0.1]	[0.2, 0.9]
Log21	K (10)	[0.001, 0.1]	NA

NA: Not Applicable

FEATURE SELECTION – Bankruptcy

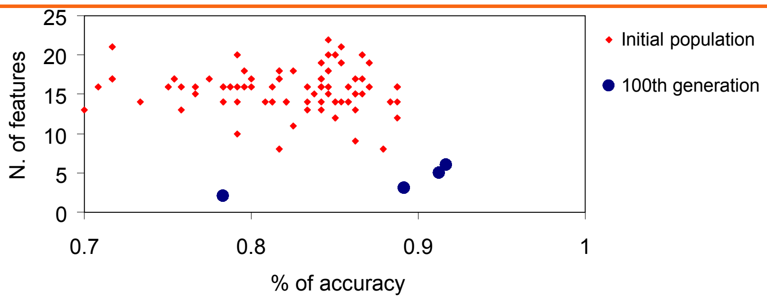
Logistic regression (NF + Accuracy):

Experiment	Training Method	Learning Rate	Training Fraction
Log1	H	**0.001**	2/3
Log2	H	**0.01**	2/3
Log3	H	**0.02**	2/3
Log4	H	**0.1**	2/3
Log5	H	0.01	**0.5**
Log6	H	0.01	**0.8**
Log11	K (10)	**0.001**	NA
Log12	K (10)	**0.01**	NA
Log13	K (10)	**0.02**	NA
Log14	K (10)	**0.1**	NA
Log15	**K (5)**	0.01	NA
Log20	**H**	**[0.001, 0.1]**	**[0.2, 0.9]**
Log21	**K (10)**	**[0.001, 0.1]**	**NA**

NA: Not Applicable

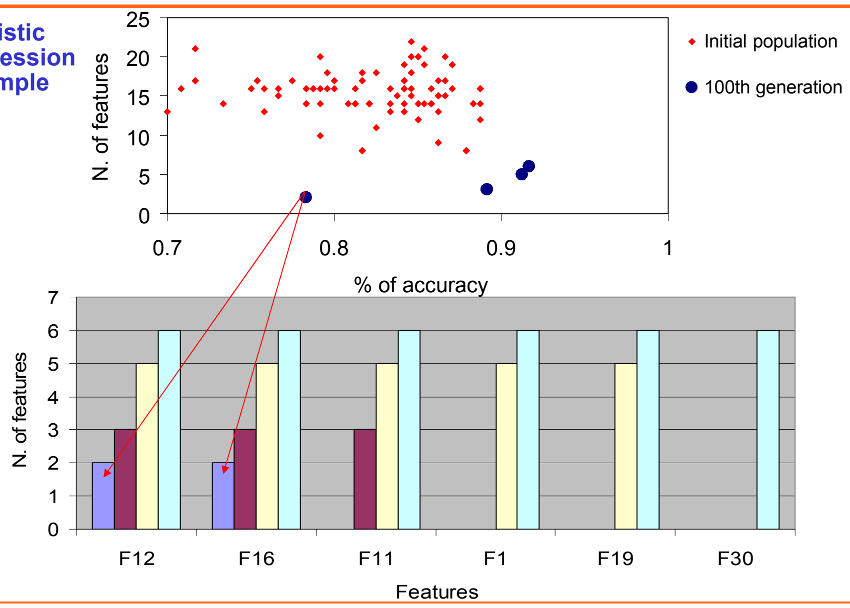
FEATURE SELECTION – Bankruptcy

Logistic regression example



FEATURE SELECTION – Bankruptcy

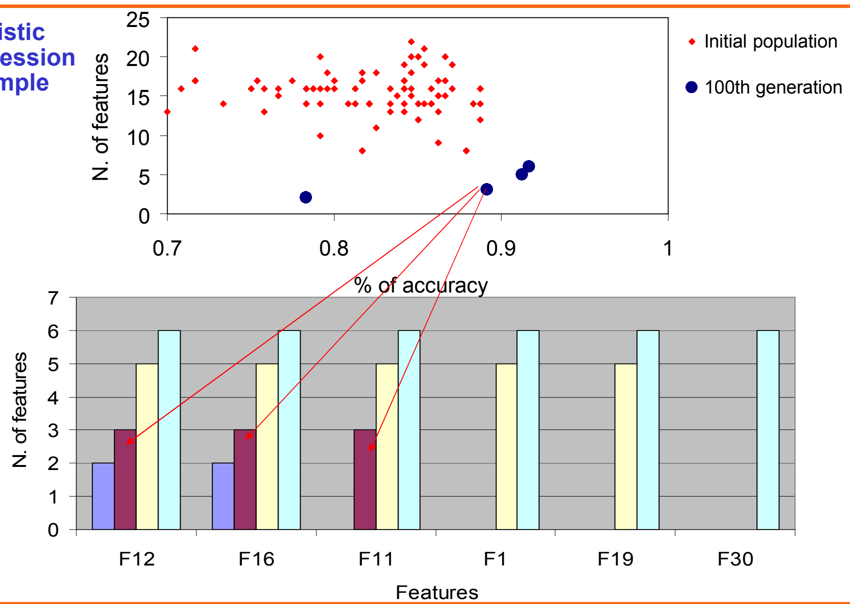
Logistic regression example



Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

Logistic regression example

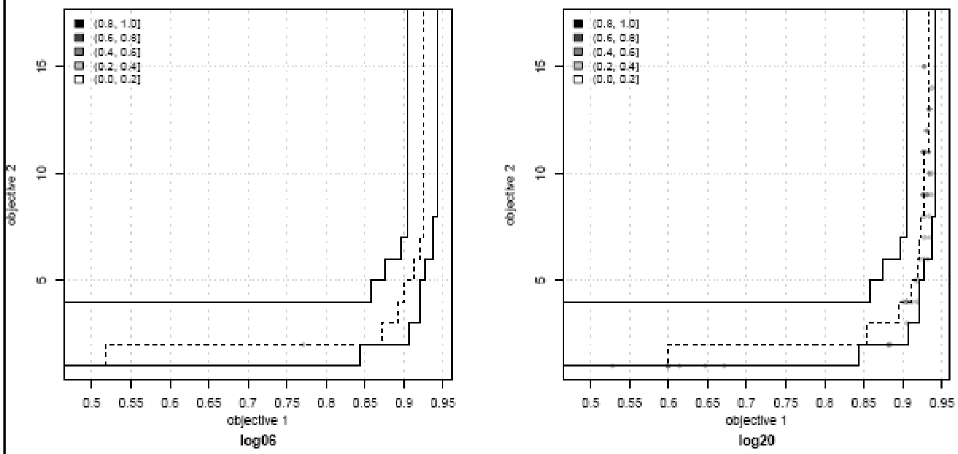


Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

Comparison between experiments Log06 and Log20

logistic regression



More results: http://www.dep.uminho.pt/agc/results

Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

Logistic regression (NF + Accuracy):

Run	Training Method	Learning Rate	Training Fraction
Log1	H	0.001	2/3
Log2	H	0.01	2/3
Log3	H	0.02	2/3
Log4	H	0.1	2/3
Log5	H	0.01	0.5
Log6	H	0.01	0.8
Log11	K (10)	0.001	NA
Log12	K (10)	0.01	NA
Log13	K (10)	0.02	NA
Log14	K (10)	0.1	NA
Log15	K (5)	0.01	NA
Log20	H	[0.001, 0.1]	[0.2, 0.9]
Log21	K (10)	[0.001, 0.1]	NA

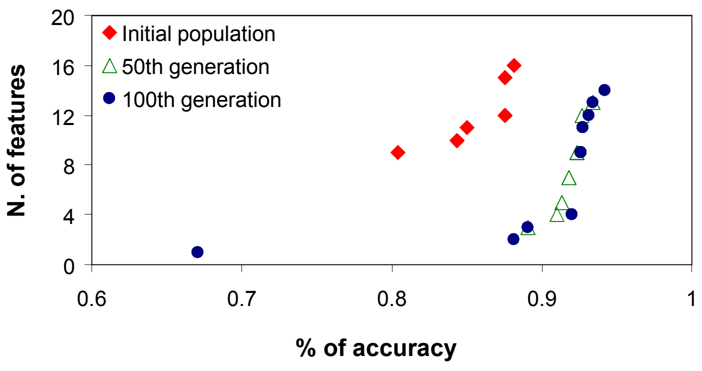
NA: Not Applicable

Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

Pareto frontiers: experiment Log20

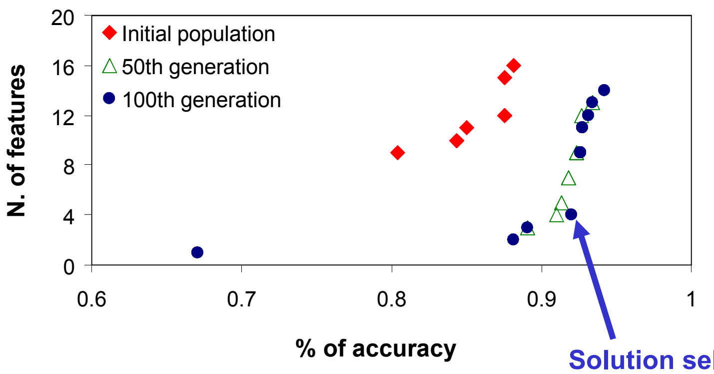
Logistic regression



FEATURE SELECTION – Bankruptcy

Pareto frontiers: experiment Log20

Logistic regression

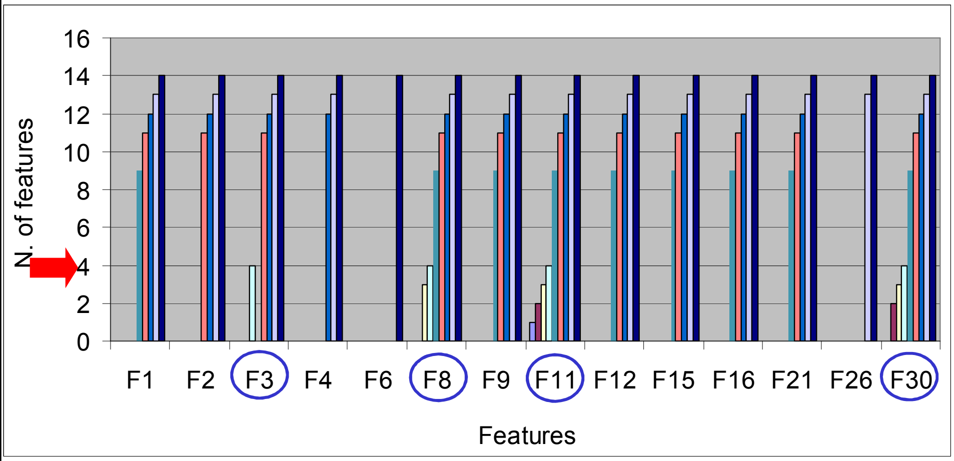


Solution selected:
- 4 features
- 92% of accuracy

FEATURE SELECTION – *Bankruptcy*

Features selected: experiment Log20

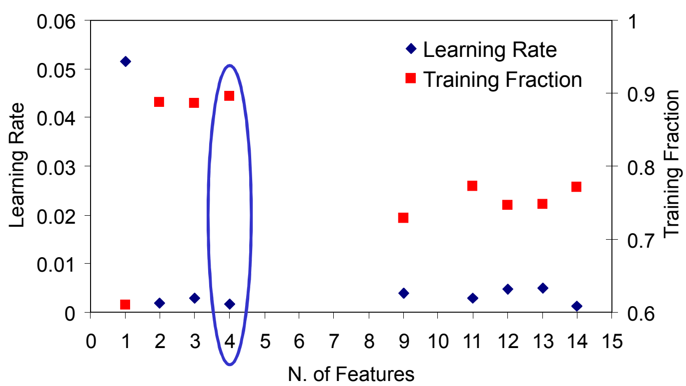
Logistic regression



FEATURE SELECTION – *Bankruptcy*

Parameters: experiment Log20

Logistic regression



FEATURE SELECTION – Bankruptcy

Features selected: experiment Log20 corresponding to 92% of accuracy

F3: Financial Debt / Capital Employed (%). This measures the capital structure ratio, i.e., the amount of financial debt in relation to the total amount of capital invested in the firm. The higher the ratio the closest to failure the company is. Several author used this ratio as a good predictor of bankruptcy (Alfaro-Cid & Castillo, 2008);

F8: Stock Turnover days. This is a ratio that measures the number of days invested in inventories. It measures the efficiency of the firm in the conversion of inventories into revenues. Companies with a low ratio may denote difficulty in selling their stock in comparison with other of the same industry. Inventories are part of the total assets and some author may prefer to use sales/total assets ratio such as Atiya (2001) and Altman & al. (1968, 1977);

F11: Turnover per Employee (thousands Euros) is a measure of employee profitability. If all the other ratios are constant, a higher productivity decreases the probability of bankruptcy. This has been used as an indicator for bankruptcy prediction in several studies (Neves & Vieira, 2006);

F30: EBITDA Margin (%) measures the Earnings Before Interest, Taxes, Depreciation and Amortization for the total revenues. This is a measure of operational profitability of the firm. This ratio is commonly used by financial analysts and investors to benchmark profitability within a given industry and to understand the effects of competition in operating profitability.

FEATURE SELECTION – Bankruptcy

SVM (NF + Accuracy):

Experiment	γ	C	Training Method	Training Fraction
Ref. values	0.01	1	-	2/3
C-svc01	0.01	1	H	-
C-svc02	0.1	-	H	-
C-svc03	1.0	-	H	-
C-svc04	10	-	H	-
C-svc07	-	10	H	-
C-svc08	-	100	H	-
C-svc09	-	1000	H	-
C-svc21	0.01	1	K	NA
C-svc22	0.1	-	K	NA
C-svc23	1.0	-	K	NA
C-svc24	10	-	K	NA
C-svc27	-	10	K	NA
C-svc28	-	100	K	NA
C-svc29	-	1000	K	NA
C-svc50	-	-	H	[0.2, 0.9]
C-svc51	-	-	K	NA
C-svc52	[0.005, 10]	[1, 1000]	H	[0.2, 0.9]
C-svc53	[0.005, 10]	[1, 1000]	K	NA

NA: Not Applicable

FEATURE SELECTION – Bankruptcy

SVM (NF + Accuracy):

Experiment	γ	ν	Training Method	Training Fraction
Ref. values	0.01	0.05	-	2/3
ν -svc01	0.01	-	H	-
ν -svc02	0.1	-	H	-
ν -svc03	1.0	-	H	-
ν -svc04	10	-	H	-
ν -svc07	-	0.01	H	-
ν -svc08	-	0.1	H	-
ν -svc09	-	0.5	H	-
ν -svc21	0.01	-	K	NA
ν -svc22	0.1	-	K	NA
ν -svc23	1.0	-	K	NA
ν -svc24	10	-	K	NA
ν -svc27	-	0.01	K	NA
ν -svc28	-	0.1	K	NA
ν -svc29	-	0.5	K	NA
ν -svc50	-	-	H	[0.2, 0.9]
ν -svc51	-	-	K	NA
ν -svc52	[0.005, 10]	[0.01, 0.5]	H	[0.2, 0.9]
ν -svc53	[0.005, 10]	[0.01, 0.5]	K	NA

NA: Not Applicable

FEATURE SELECTION – Bankruptcy

C-SVC: results for experiment c-svc53

N. Features	Accuracy (%)	γ	C	Features
1	62.0	7.2	875	F4
2	84.5	7.7	905	F4, F30
3	93.0	10.0	995	F4, F7, F30
4	98.9	10.0	977	F4, F7, F22, F30
5	100.0	9.3	959	F4, F7, F11, F22, F30

ν-SVC: results for experiment ν-svc53

N. Features	Accuracy (%)	γ	ν	Features
1	59.3	7.91	0.492	F2
2	66.2	7.55	0.486	F2, F22
3	91.4	9.91	0.232	F2, F16, F22
4	99.3	9.88	0.079	F2, F8, F16, F22
5	100.0	6.63	0.043	F1, F2, F8, F16, F22

FEATURE SELECTION – Bankruptcy

SVM (NF + other objectives)

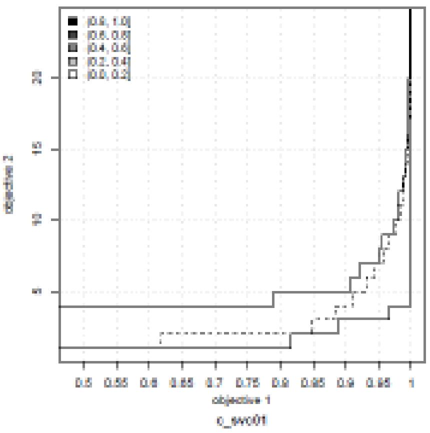
Experiment	C	γ	Objectives
c-svc1	1	10	$NF + F_m$
c-svc2	1	10	$NF + e_I$
c-svc3	1	10	$NF + e_{II}$
c-svc4	1	10	$NF + F_m + e_I$
c-svc11	[1,10]	[0.001,1]	$NF + F_m$
c-svc12	[1,10]	[0.001,1]	$NF + e_I$
c-svc13	[1,10]	[0.001,1]	$NF + e_{II}$
c-svc14	[1,10]	[0.001,1]	$NF + F_m + e_I$
c-svc15	[1,10]	[0.001,1]	$NF + R + e_I$
c-svc16	[1,10]	[0.001,1]	$R + e_I$

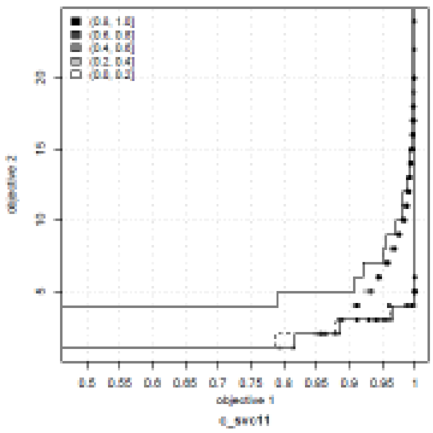
Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

SVM (NF + other objectives)

Comparison using EAF



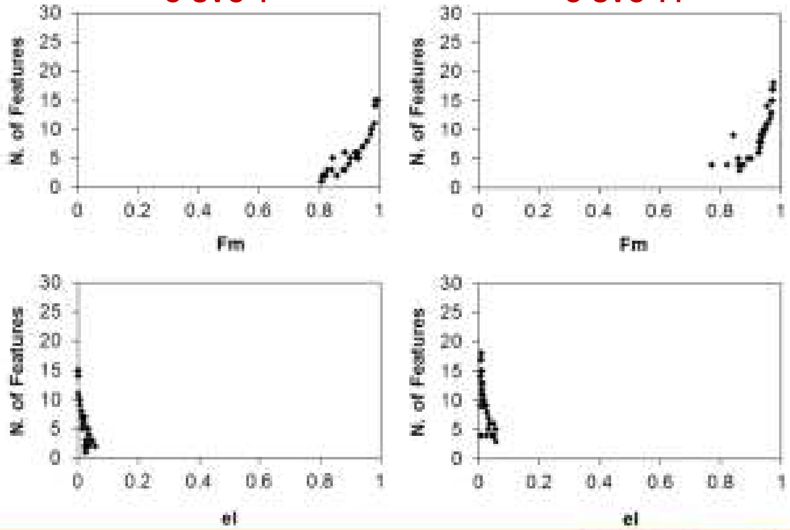


Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Bankruptcy

SVM (NF + other objectives)

C-SVC 4 | C-SVC 14



FEATURE SELECTION – Bankruptcy

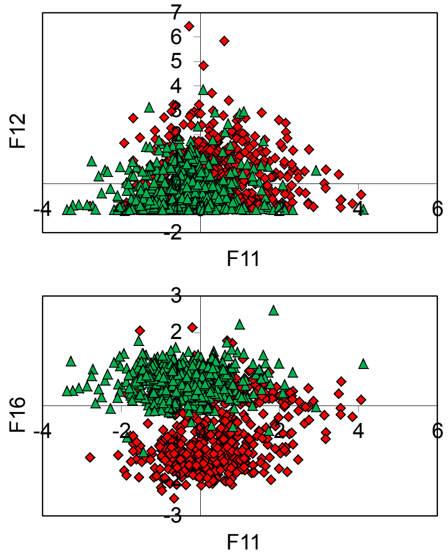
SVM (NF + other objectives)

Results for a single run of case C-SVC14

N. Features	F_m	e_I	γ	C	Features
3	0.864	0.058	0.780	7.26	F11, F12, F16
4	0.770	0.007	0.052	2.42	F 3, F 5, F15, F16
4	0.823	0.025	0.739	4.85	F12, F15, F16, F24
4	0.862	0.043	0.803	6.39	F11, F12, F15, F16
4	0.876	0.052	0.742	7.45	F11, F12, F16, F24
5	0.860	0.032	0.924	8.79	F 3, F12, F15, F16, F28
5	0.889	0.033	0.915	8.87	F11, F12, F15, F16, F23

FEATURE SELECTION – *Bankruptcy*

SVM (NF + other objectives)



Results for a single run of case C-SVC14

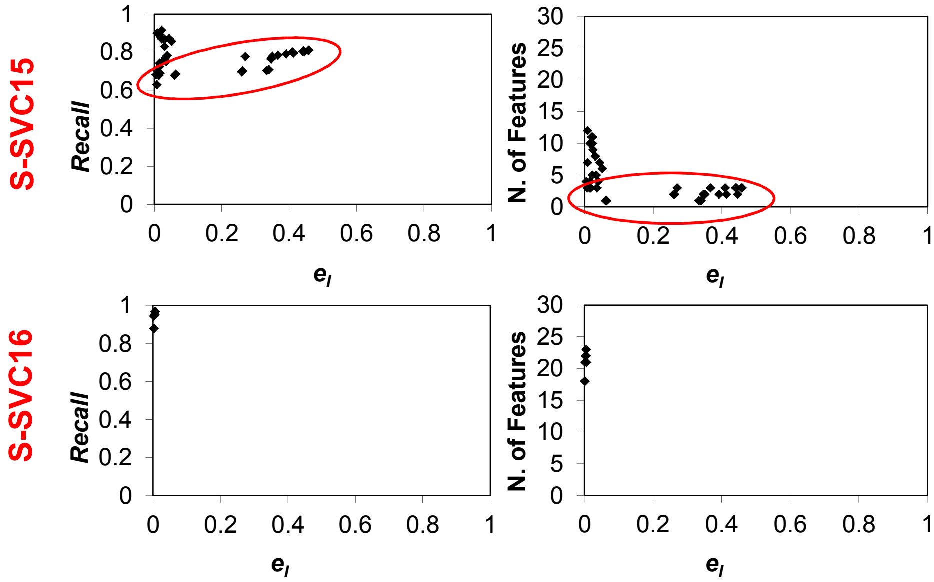
FEATURE SELECTION – *Bankruptcy*

SVM (NF + other objectives)

Experiment	*C*	γ	Objectives
c-svc1	1	10	$NF + F_m$
c-svc2	1	10	$NF + e_I$
c-svc3	1	10	$NF + e_{II}$
c-svc4	1	10	$NF + F_m + e_I$
c-svc11	[1,10]	[0.001,1]	$NF + F_m$
c-svc12	[1,10]	[0.001,1]	$NF + e_I$
c-svc13	[1,10]	[0.001,1]	$NF + e_{II}$
c-svc14	[1,10]	[0.001,1]	$NF + F_m + e_I$
c-svc15	[1,10]	[0.001,1]	$NF + R + e_I$
c-svc16	[1,10]	[0.001,1]	$R + e_I$

Dept. Polymer Engineering
University of Minho

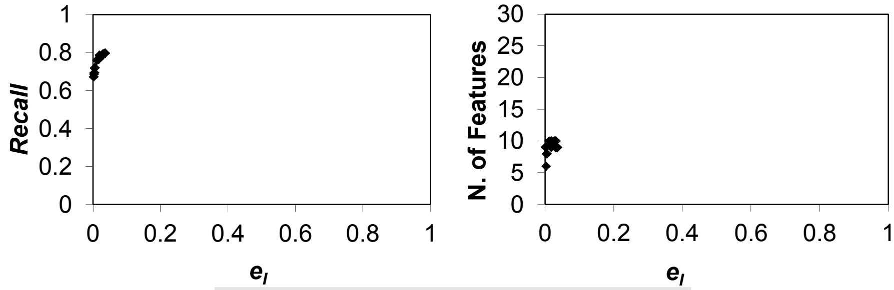
FEATURE SELECTION – Bankruptcy SVM (NF + e_I + R)



FEATURE SELECTION – Bankruptcy SVM (e_I + R)

To obtain solutions with lower number of features, low e_I and high R:

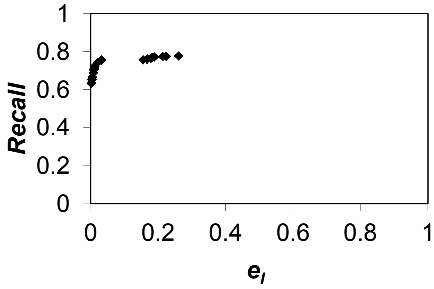
Perform optimizations with the maximum number of features fixed.



Initial population:	NF fixed in 5
During optimization:	NF fixed in 10

FEATURE SELECTION – *Bankruptcy*

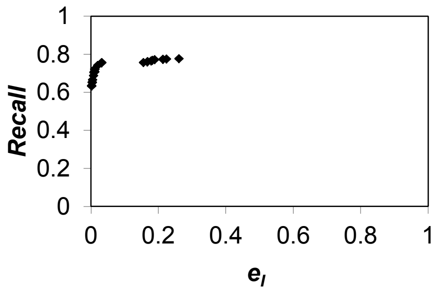
SVM (e_I + *R*)



Initial population:	NF fixed in 5
During optimization:	NF fixed in 5

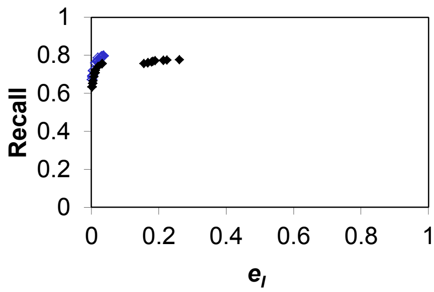
FEATURE SELECTION – *Bankruptcy*

SVM (e_I + *R*)

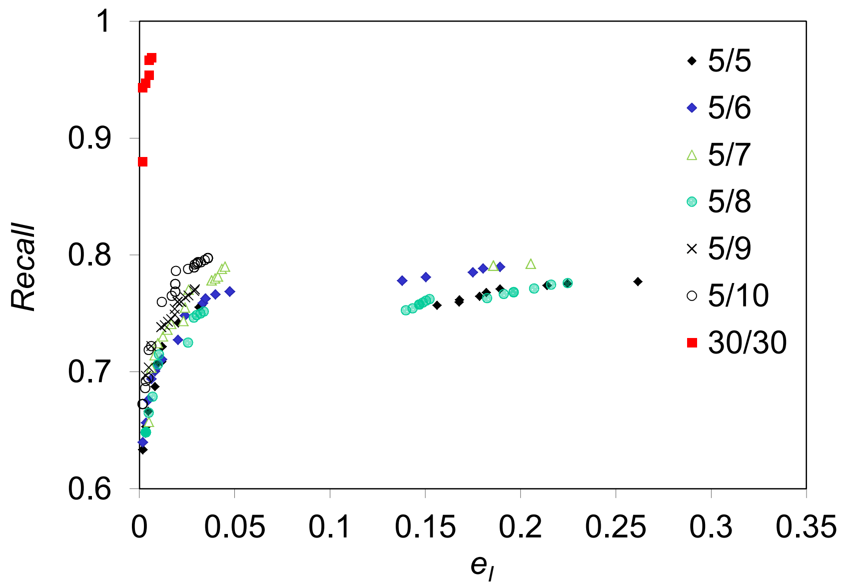


Initial population:	NF fixed in 5
During optimization:	NF fixed in 5

Comparison with the previous case:



FEATURE SELECTION – *Bankruptcy* SVM (e_I + R)



FEATURE SELECTION – *Bankruptcy* Conclusions

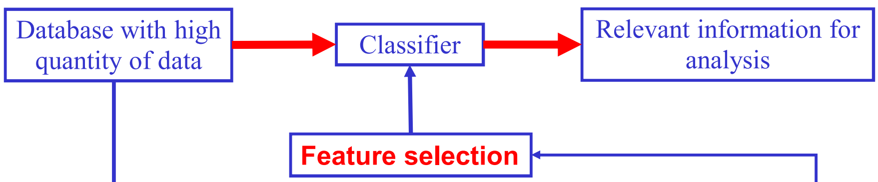
- In the present study a MOEA was applied in the problem of feature selection for bankruptcy prediction using two complementary classifier methods, LR and SVM;
- The proposed methodology is able to provide good solutions, not only by reducing the number of features but also making available to the decision maker useful information, concerning with both, the best features to be used and the best parameters of the classifier;
- An important characteristic of the MOEA strategy is the possibility of the decision maker to have multiple Pareto optimal solutions to perform the final analysis;
- The best performance is obtained when the classifier parameters are optimized simultaneously with the features.
- The approach followed here showed good potentialities in obtaining a good approximation to the ROC curve.

Multi-Objective Evolutionary Algorithms for Feature Selection:

Application to Cardiac SPECT Diagnosis

FEATURE SELECTION – Cardiac SPECT Diagnosis

Problem characteristics



Feature selection is of crucial importance when dealing with problems with high amount of data:

i) the processing of all features available can be computational infeasible;

ii) the existence of high number of variables for small number of available data points can invalidate the resolution of the problem;

iii) an high number of features can be redundant or irrelevant for the classification problem under study.

Dept. Polymer Engineering
University of Minho

FEATURE SELECTION – Cardiac SPECT Diagnosis

Diagnostic problem of cardiac Single Proton Emission Computed Tomography (SPECT) images (UCI Machine Learning Repository):

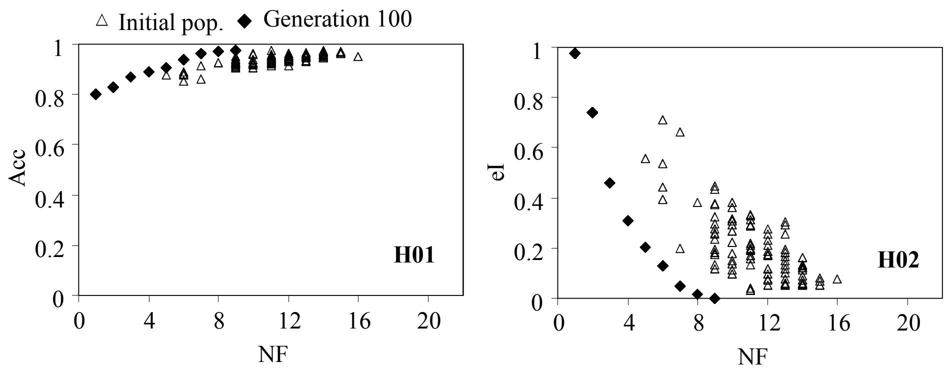
- Each of the patients is classified into two categories: normal and abnormal;
- 22 binary feature patterns.

Exp.	γ	C	TM	LR	TF	Objectives
H01	10	1	K(10)	0.01	*	NA + PA
H02	10	1	K(10)	0.01	*	NA + e_I
H03	10	1	K(10)	0.01	*	NA + e_{II}
H04	10	1	K(10)	0.01	*	NA + F_m
H05	10	1	K(10)	0.01	*	NA + e_I + F_m
H06	[0.01,10]	[1,150]	K(10)	[0.001,0.1]	*	NA + F_m
H07	[0.01,10]	[1,150]	K(10)	[0.001,0.1]	*	NA + e_I + F_m
H08	10	1	H	0.01	0.7	NA + F_m
H09	[0.01,10]	[1,150]	H	[0.001,0.1]	[0.2,0.9]	NA + F_m
H10	[0.01,10]	[1,150]	H	[0.001,0.1]	[0.2,0.9]	NA + e_I + F_m
H11	10	1	K(10)	0.01	*	NA + e_I + R
H12	[0.01,10]	[1,150]	K(10)	[0.001,0.1]	*	NA + e_I + R
H13	[0.01,10]	[1,150]	K(10)	[0.001,0.1]	*	NA + e_I + R + F_m

*: Not Applicable

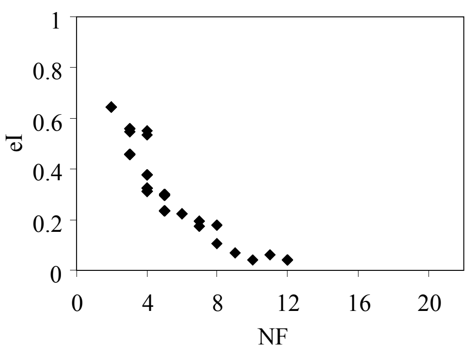
FEATURE SELECTION – Cardiac SPECT Diagnosis

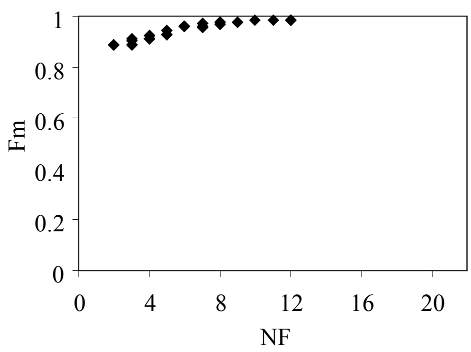
Experiments H01 and H02: run 1



FEATURE SELECTION – Cardiac SPECT Diagnosis

Experiment H05: run 1

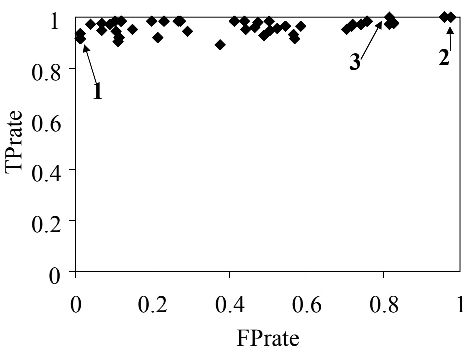


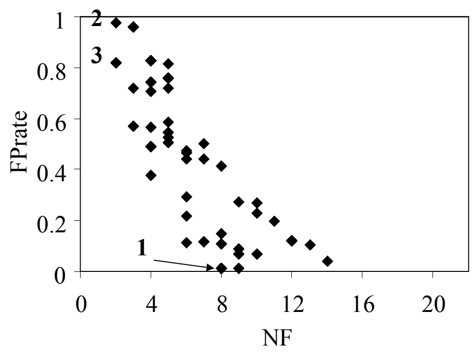


FEATURE SELECTION – Cardiac SPECT Diagnosis

Results using ROC curves

Experiment H13: run 1





FEATURE SELECTION – *Cardiac SPECT Diagnosis*

Results using ROC curves

Experiment H13: run 1

Sol.	Decision Variables				Objectives		
	Features selected	γ	*C*	*NF*	*FPrate*	*TPrate*	F_m
1	F3,F4,F11,F13,F14,F16,F18,F22	0.078	0.17	8	0.013	0.91	0.951
2	F4, F11	0.040	0.43	2	0.975	1.00	0.886
3	F11, F13	0.043	0.46	2	0.818	0.97	0.892

FEATURE SELECTION – *Cardiac SPECT Diagnosis*

- In this work a MOEA was used for feature selection in data mining problems using a Support Vector Machines classifier.
- The methodology proposed was able not only to propose solutions with a few number of features necessary but is able also to provide relevant information to the decision maker, not only the best features to be used but, also, the best parameters of the classifier and the trade-off between the different objectives used.
- Finally, the approach followed here showed good potentialities in obtaining a good approximation to the ROC curves.

Dept. Polymer Engineering
University of Minho