UNIVERSITY OF HELSINKI

Course on Computational Intelligence (CI)

Adjunct Prof. Vesa A. Niskanen
University of Helsinki
vesa.a.niskanen@helsinki.fi
www.mv.helsinki.fi/home/niskanen

Agenda

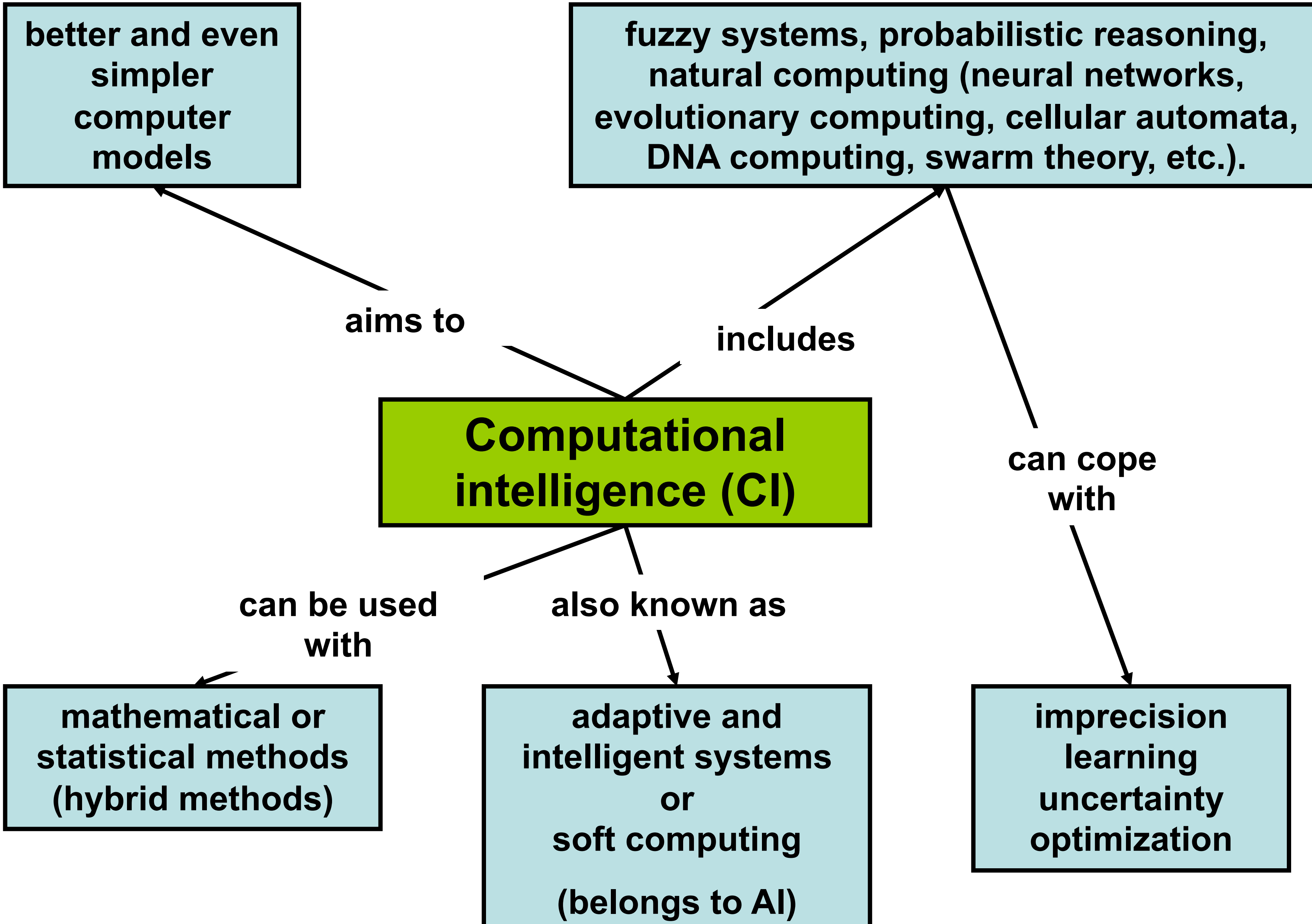
1. Soft computing and fuzzy systems?
2. Fuzzy sets.
3. Model construction with fuzzy language.
4. Typical fuzzy reasoning models.
5. Neuro-fuzzy and genetic-fuzzy systems.
6. Tuning of models with control data.
7. Complex models and cognitive maps.

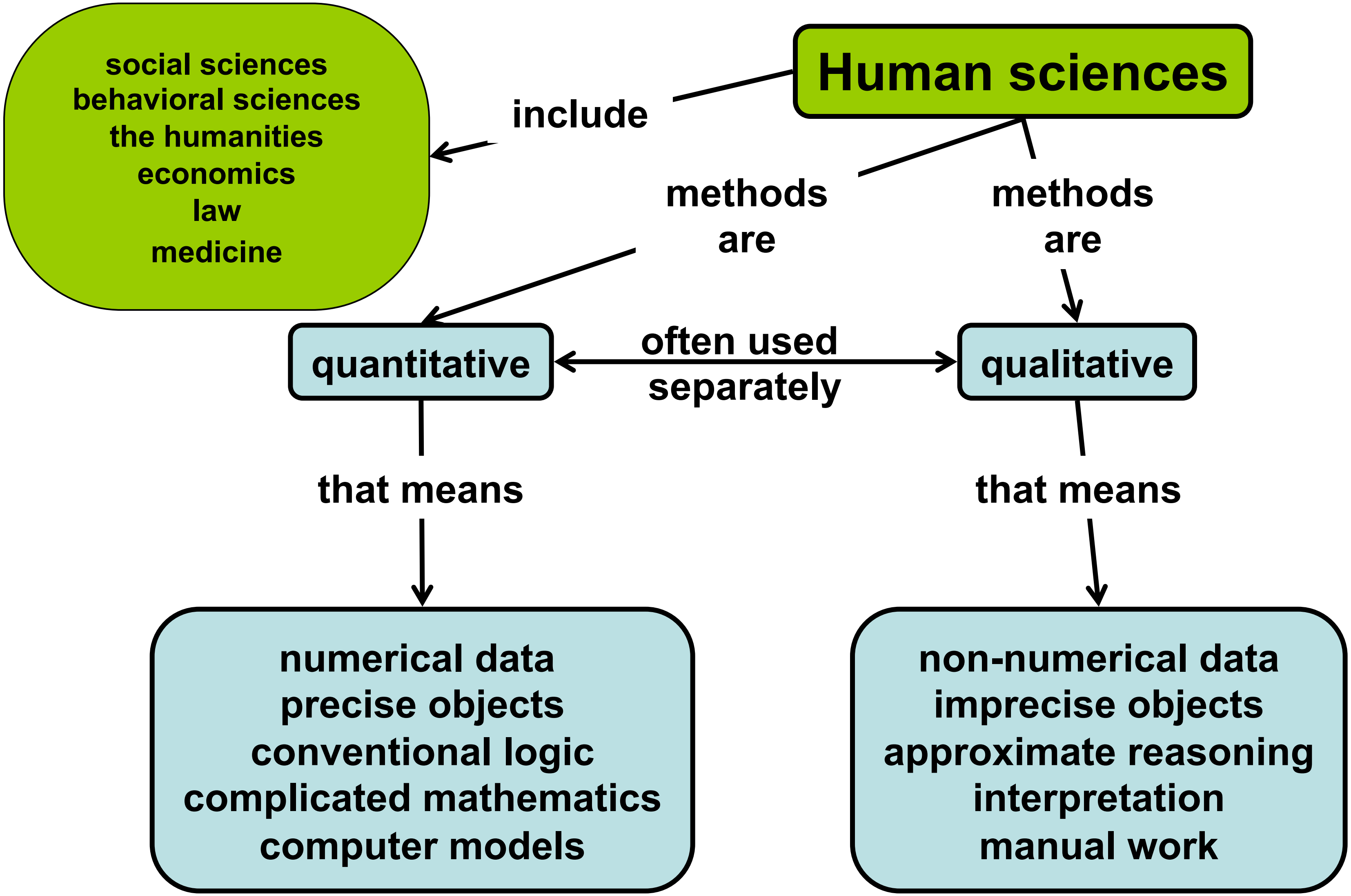
Matlab’s Fuzzy Logic Toolbox

is mainly used in our model constructions.

CI publications in my Box cloud:

https://app.box.com/s/fnroma6w3nb3edesr16dor73tr1vb48u

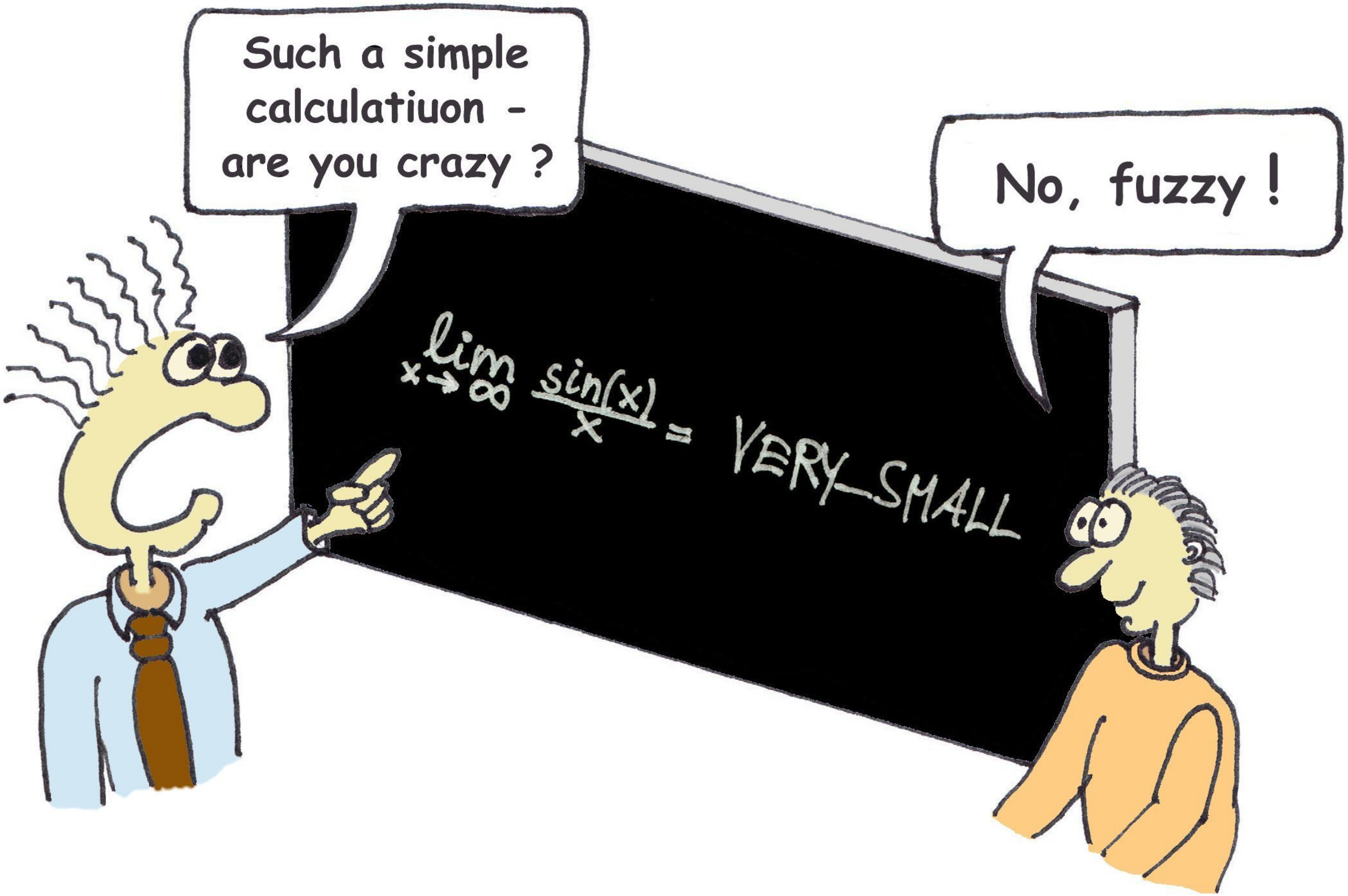

better and even simpler computer models
fuzzy systems, probabilistic reasoning, natural computing (neural networks, evolutionary computing, cellular automata, DNA computing, swarm theory, etc.).
aims to
includes
Computational intelligence (CI)
can cope with
can be used with
also known as
mathematical or statistical methods (hybrid methods)
adaptive and intelligent systems or soft computing
(belongs to AI)
imprecision
learning
uncertainty
optimization



CI a mediator?

Traditional Approaches to Computer Modeling

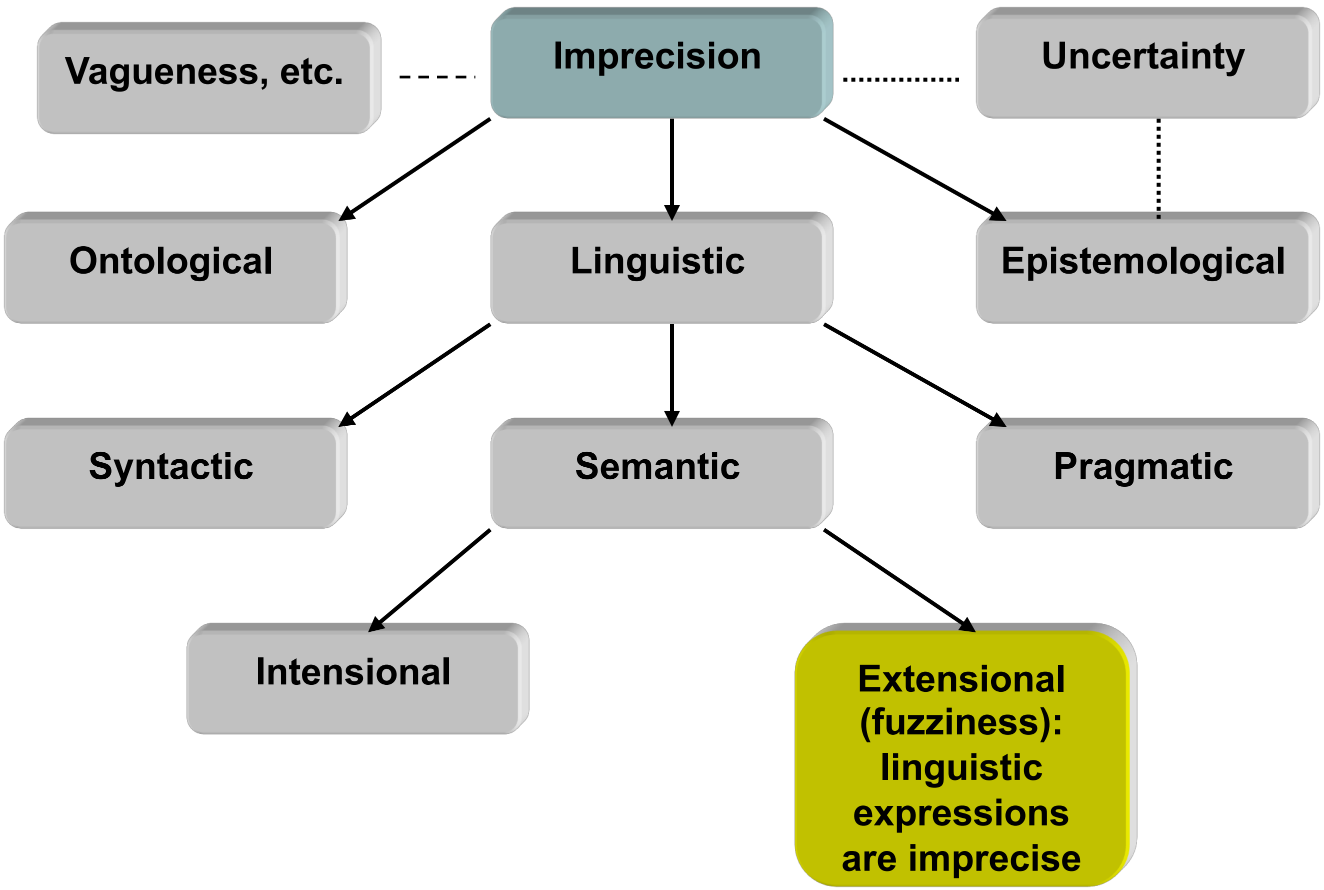
- **Mathematical models:** Complicated, black boxes, number crunching.
- **Rule-based systems (crisp & bivalent):** Large rule bases.


Such a simple calculatiuon - are you crazy ?
No, fuzzy !
lim x→∞ sin(x)/x = VERY_SMALL

According to the theories of the engineering sciences, a bumble bee is unable to fly

?

In the beginning, there was imprecision



Then came fuzziness

- Fuzziness for imprecision, thus **fuzziness = imprecision** in practice (semantics).
- **Probability** for **uncertainty** (epistemology).

- Fuzziness vs. probability:
- John is young (fuzziness)
- Probability(John is 20) = 0.8 (crisp probability)
- Probability(John is 20) = about 0.8 / fairly high (fuzzy probability)
- Probability(John is young) = high (fuzzy probability)

Fuzzy Systems

- Fuzzy systems can cope with linguistic and imprecise entities of a model in a computer environment.
- Invented by Prof. Lotfi Zadeh (1921-2017) at UC Berkeley in the 1960's.
- Stem from novel theories on fuzzy sets and fuzzy logic.

Fuzzy systems

- **Fuzzy set theory**: Also partial memberships to sets are used.
- **Fuzzy logic**: A version of multivalued logic. The truth values may be numeric or linguistic.
- Bivalent logic: either true or false.
- Fuzzy logic: degrees of truth from false to true.

Advantages of Fuzzy Models

- Models aim to mimic real human reasoning.
- Models can be
  - linguistic
  - simple (no number crunching),
  - comprehensible (no black boxes),
  - fast in computing,
  - good in practice.

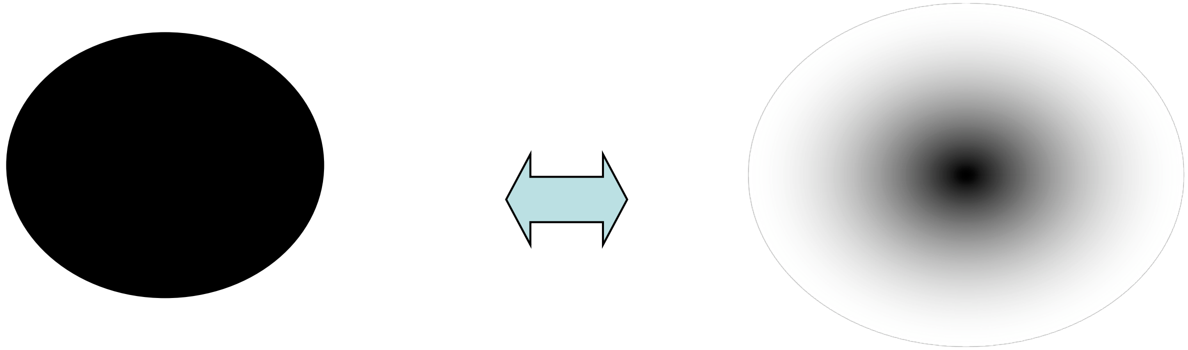
Fuzzy Applications: Control

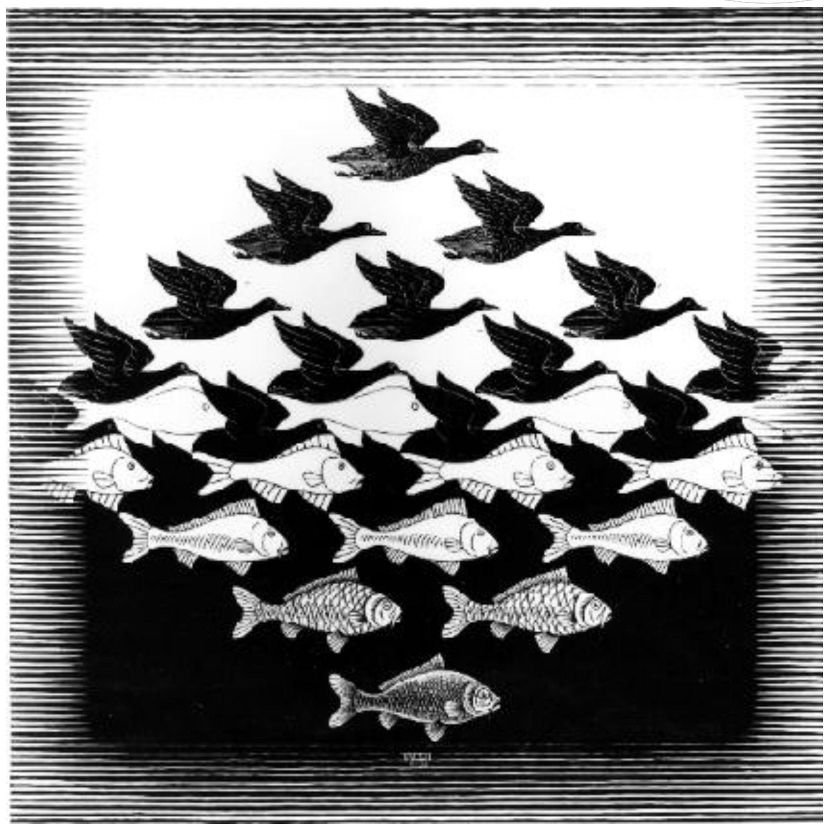
- Heavy industry (Matsushita, Siemens, Stora-Enso, Metso)
- Home appliances (Canon, Sony, Goldstar, Siemens, Whirlpool)
- Automobiles (Nissan, Mitsubishi, Daimler-Benz, Chrysler, BMW, Volkswagen)
- Space crafts (NASA)

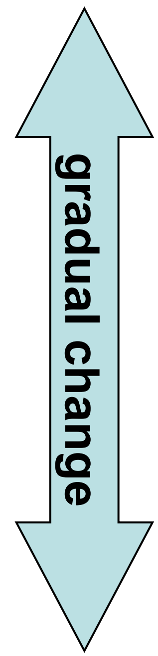
Fuzzy Applications: Decision Making

- Fuzzy scoring for mortgage applicants,
- creditworthiness assessment,
- fuzzy-enhanced score card for lease risk assessment,
- risk profile analysis,
- insurance fraud detection,
- cash supply optimization,
- foreign exchange trading,
- trading surveillance,
- investor classification etc.
- Source: FuzzyTech

Crisp and Fuzzy Sets

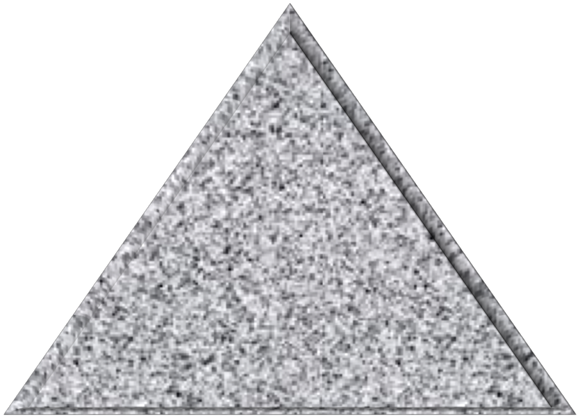






Gradual Change – the Sorites Paradox

Stewing the frog in the pot

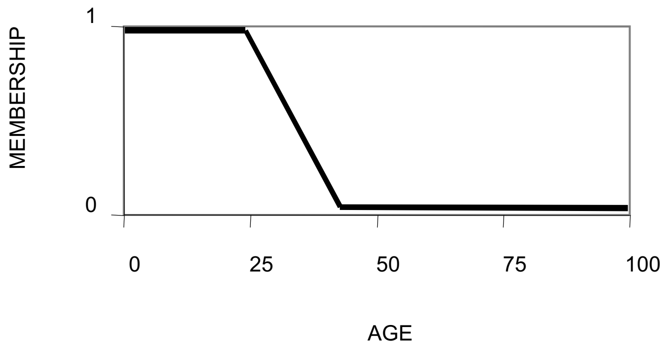


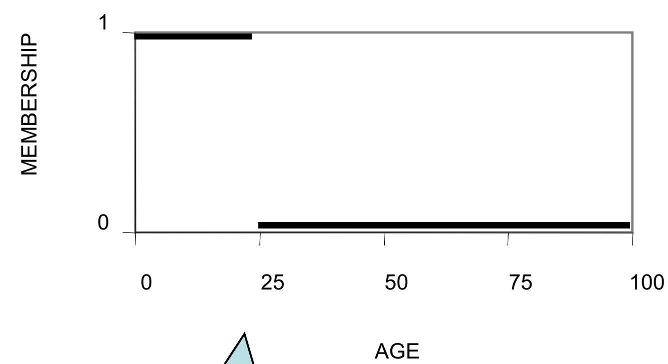
Sorites: How many grains of sand will constitute a heap?

”Quantitative meanings” of linguistic values are fuzzy sets.

E.g. meaning of ”young” is a fuzzy set YOUNG

Fuzzy sets are denoted as functions, membership functions, μ



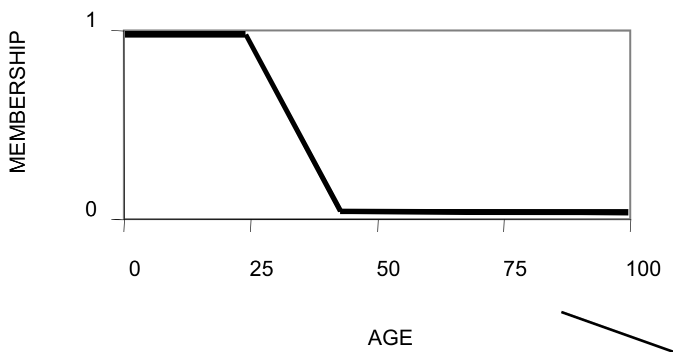


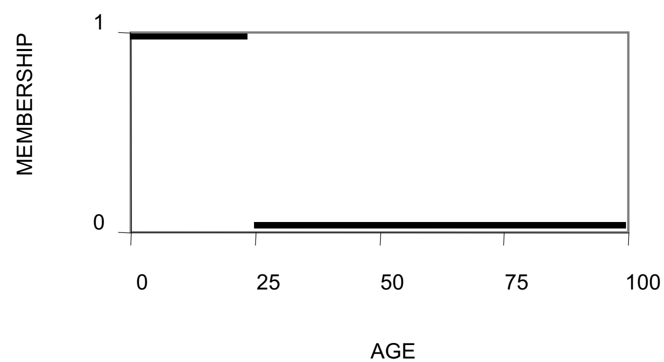
Crisp set YOUNG

Objects can also belong partially to a given fuzzy set.

E.g., given fuzzy set YOUNG and ages of persons,
person aged 10: full membership
person aged 27: almost full
person aged 35: small
person aged 70: no membership

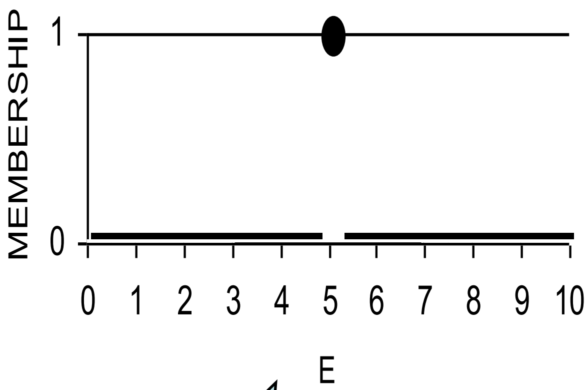
Degrees of membership are denoted as functions, membership functions, $0 \leq \mu(x) \leq 1$

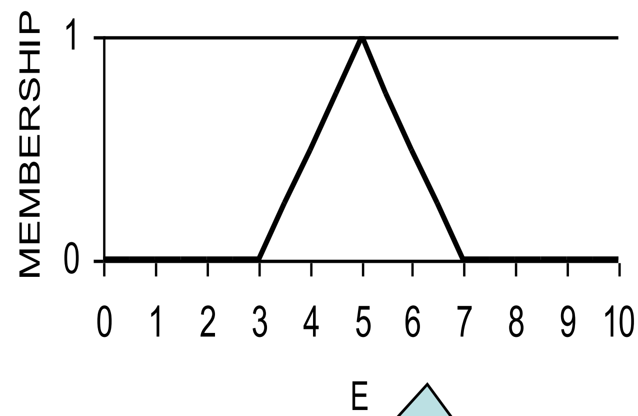




horizontal axis: values of ages, 0 to 100 (reference set, universe of discourse)
vertical axis: degrees of membeship, 0 to 1

Crisp and Fuzzy Membership Functions



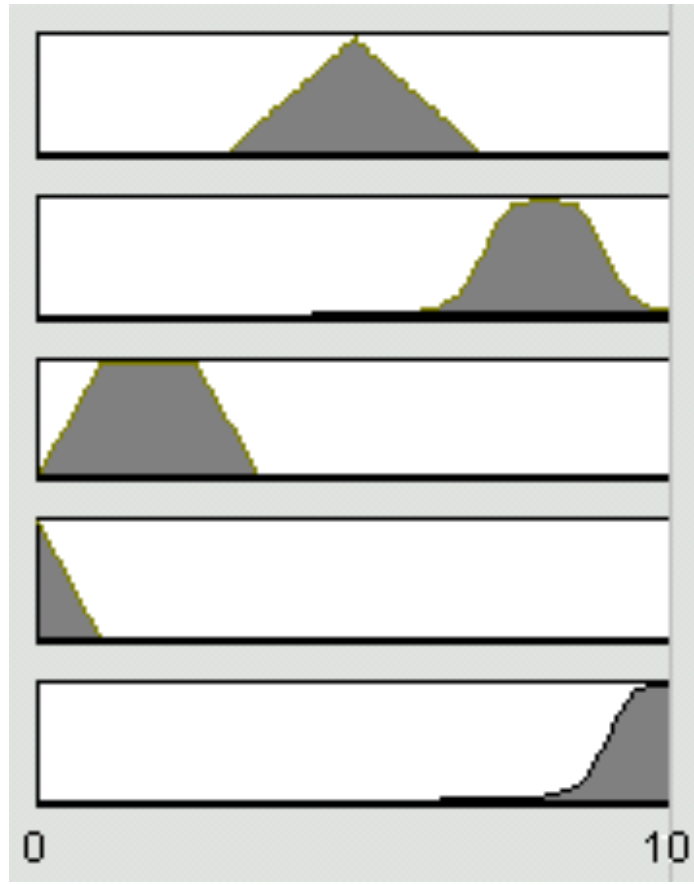


Crisp: Five

Fuzzy: About five
$\{(x,\mu(x)) \mid x \in E,\ \mu(x) \in [0,1]\}$,
In which E is universe of dicourse (reference set).

Typical Fuzzy Sets

- **Triangular,**
- **Bell-shaped,**
- **Trapezoidal.**



- **Normalized: max membership = 1**
- **Convexity:**

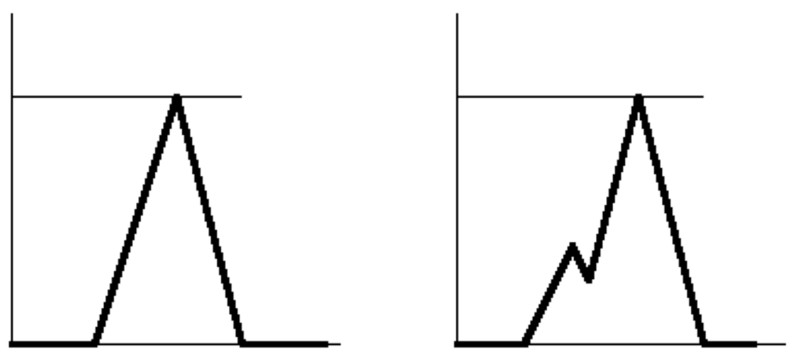
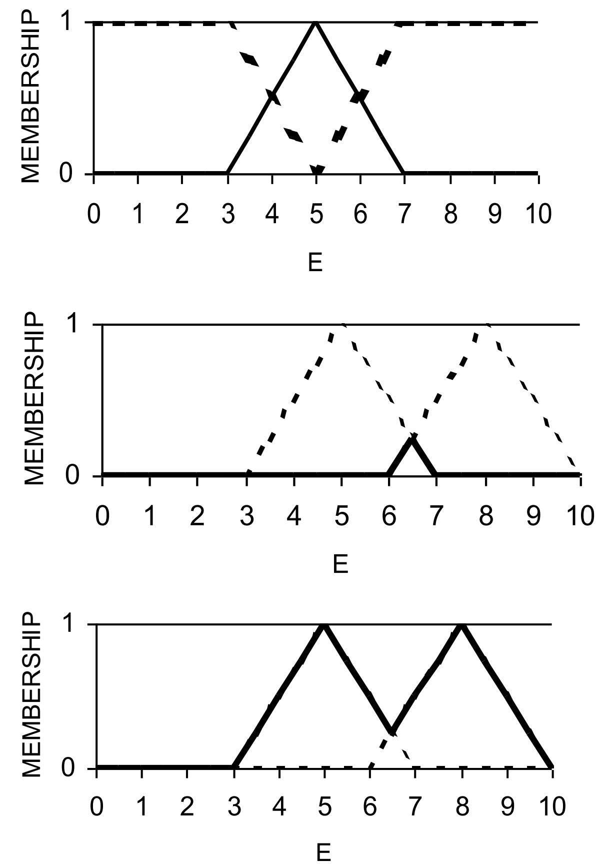


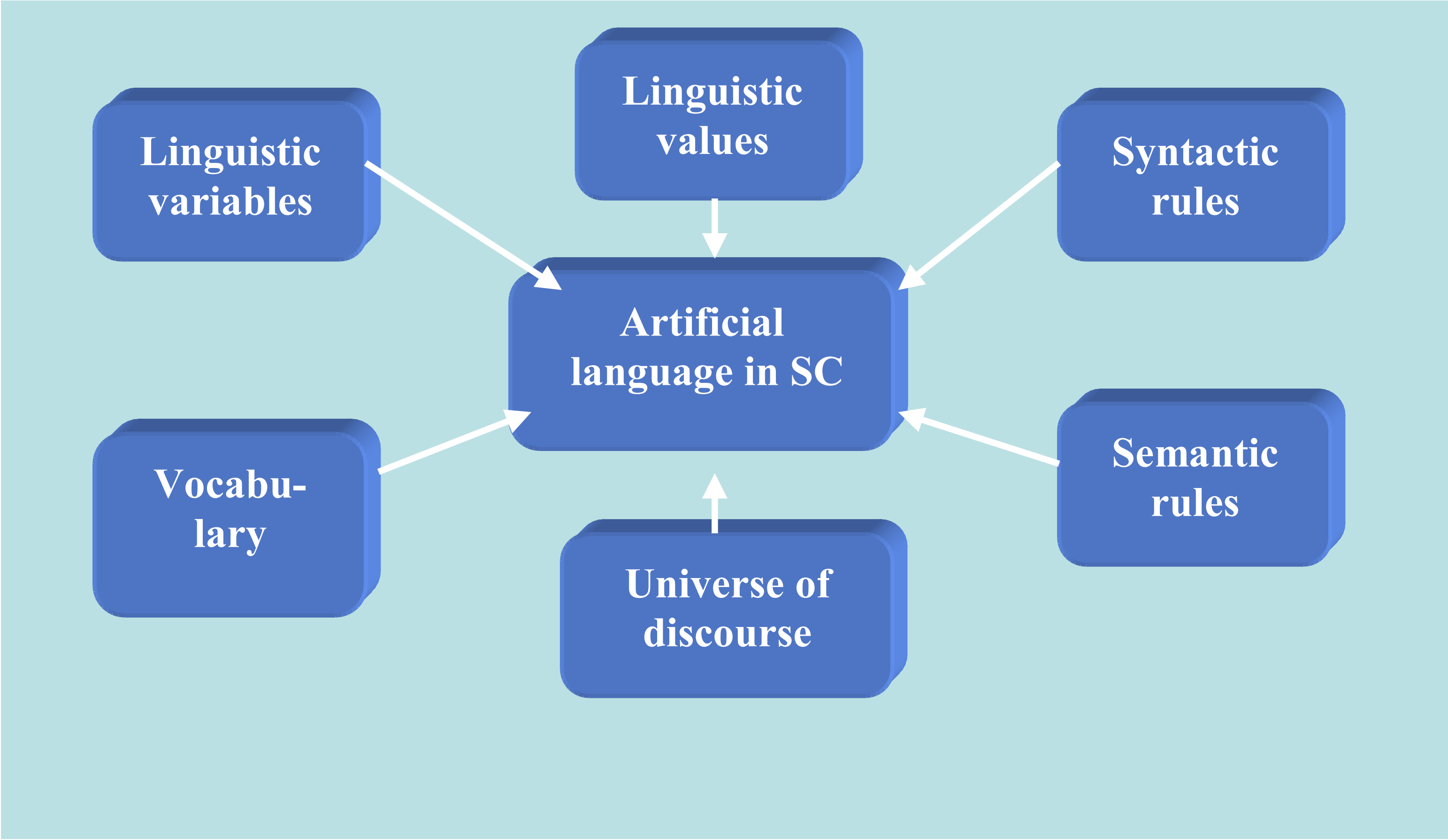
Fig. 1.3. A convex (a) and a nonconvex (b) fuzzy set

Basic Fuzzy Set Operations

- **Complement,**
- **Intersection,**
- **Union.**

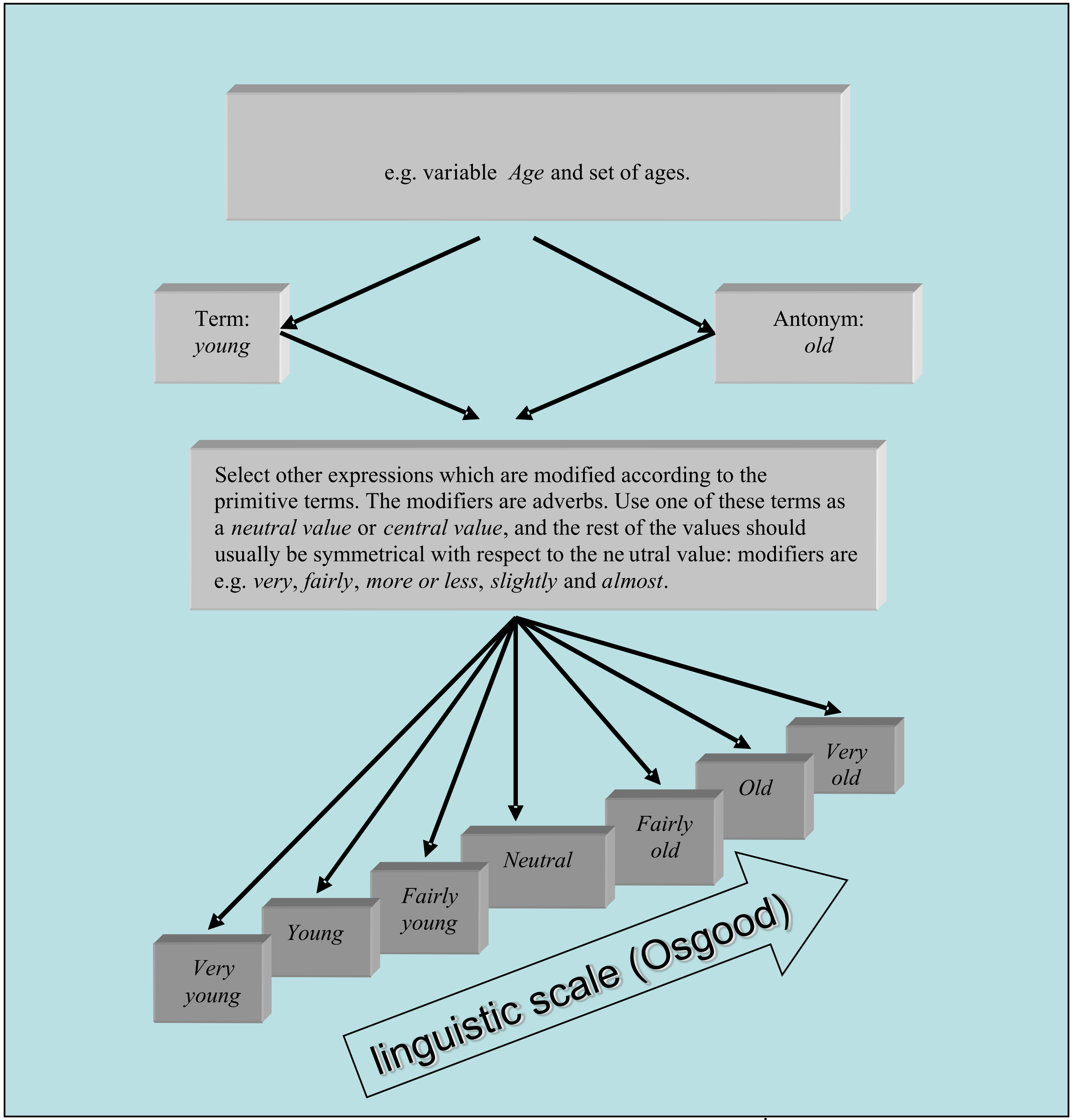


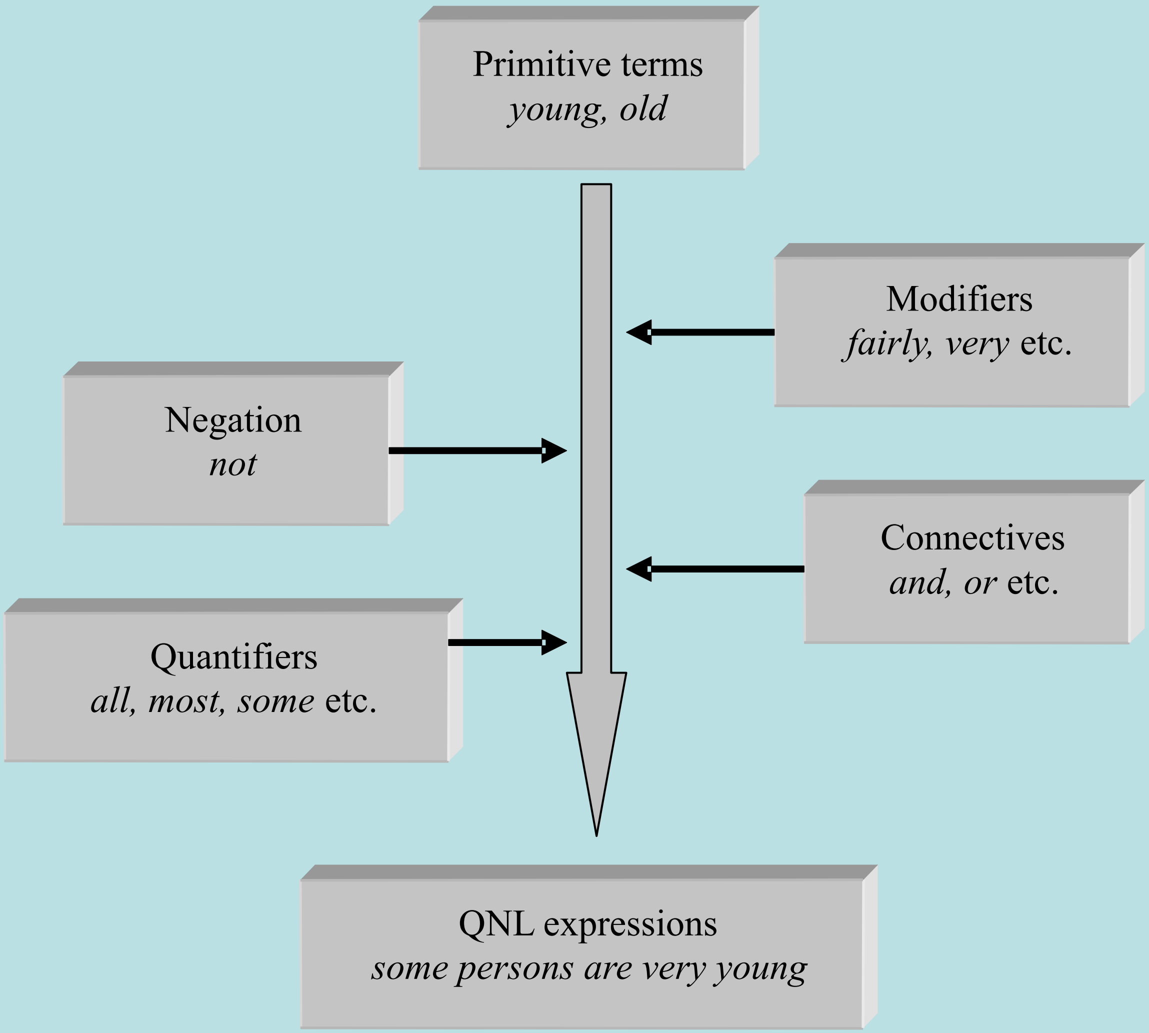
Construction of Fuzzy Language



Possible Values for Variables in Fuzzy Language

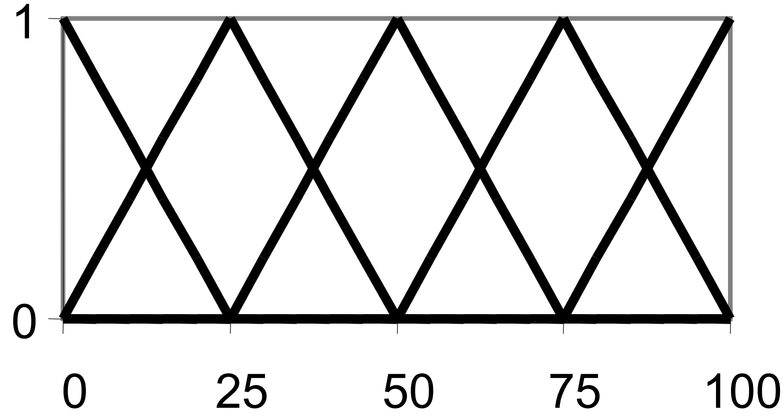
Type of Value	**Examples**
Precise numerical values and intervals	5, 0.5, [4.5,6]
Approximate numerical values and intervals	about 5, about 0.5, about [4.5,6], about from 4.5 to 6
Precise numerical functions and relations	X^2+2y^3+1, x=y
Approximate numerical functions and relations	Approximately x^2+2y^3+1, approximately x=y
Precise and approximate linguistic values and relations	male, negative, small negative, very high, fairly old, not good, young or fairly young, slightly greater than, approximately equal with


e.g. variable Age and set of ages.
Term: young
Antonym: old
Select other expressions which are modified according to the primitive terms. The modifiers are adverbs. Use one of these terms as a neutral value or central value, and the rest of the values should usually be symmetrical with respect to the ne utral value: modifiers are e.g. very, fairly, more or less, slightly and almost.
Very young
Young
Fairly young
Neutral
Fairly old
Old
Very old
linguistic scale (Osgood)



Formation of Fuzzy Language Expressions.

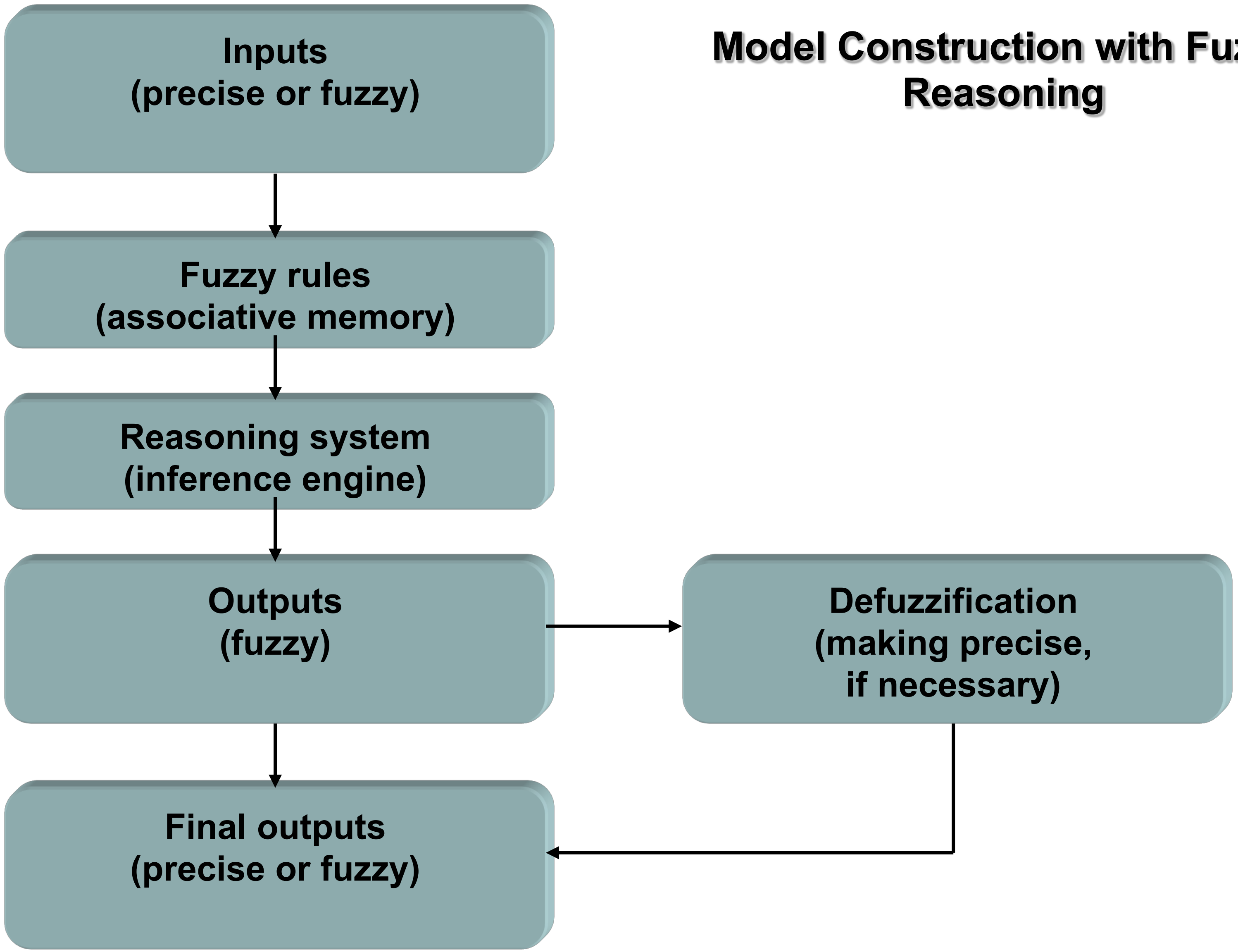
Tentative Fuzzy Sets Denoting Linguistic Values of ”Age”



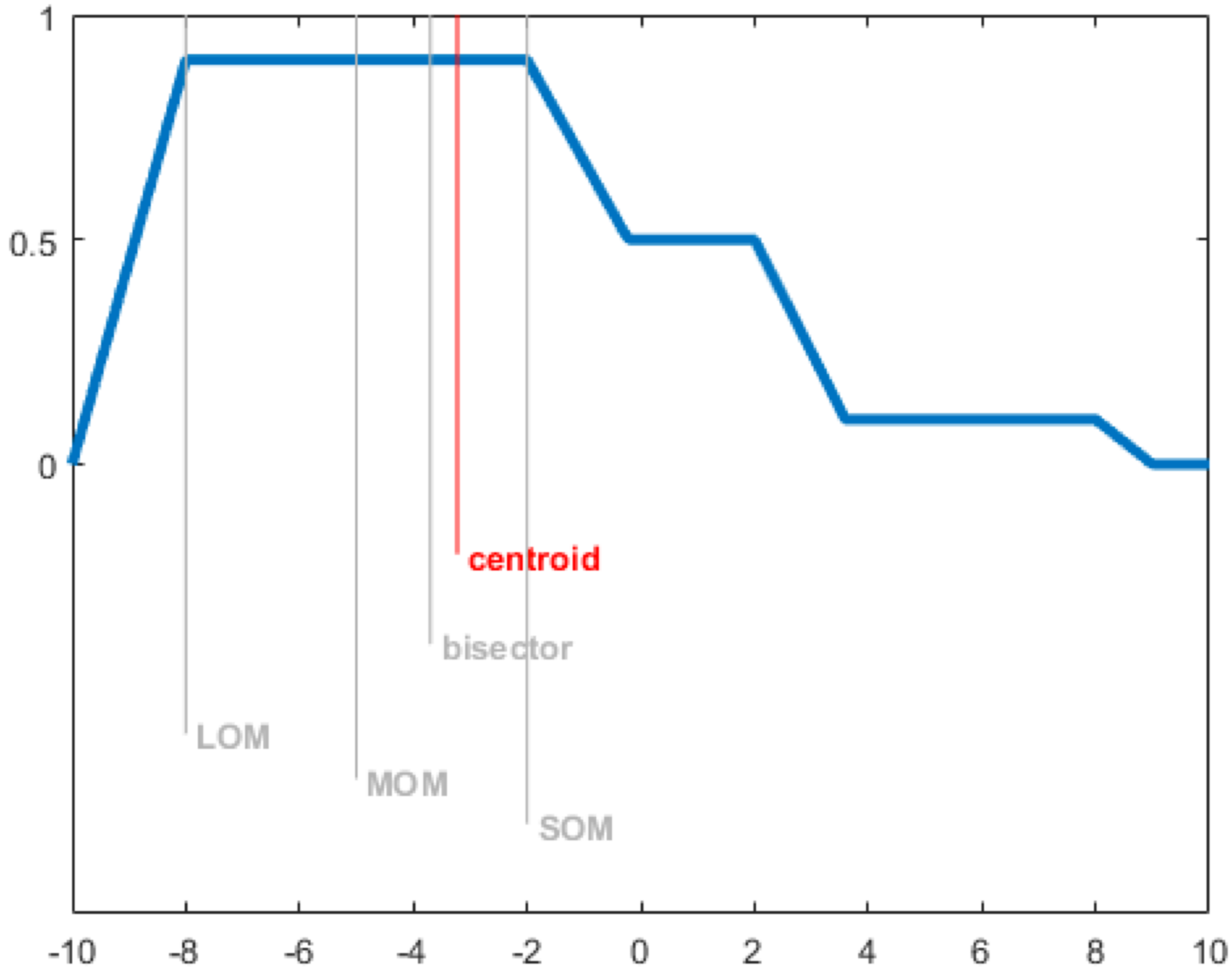
young, fairly young, middle-aged, fairly old, old

Correspondence between Linguistic Expressions and Set-theoretical Operations.	
Expression	**Fuzzy set-theoretical counterpart ("quantitative meaning")**
Primitive terms: young, old	Fuzzy sets: YOUNG, OLD
Modifiers: very, fairly, etc.	Fuzzy sets modified by translation: VERY YOUNG etc.
Negation: not	Modified fuzzy sets: complement, etc.
Compound experessions: and, or, if-then etc.	1.Set-theoretical operations of fuzzy sets: intersection, union, etc. 2.Fuzzy relations: order relation, etc.
Quantifiers: all, most some, etc.	Fuzzy sets: extension principle, etc.

Model Construction with Fuzzy Reasoning



Defuzzification methods (Matlab)



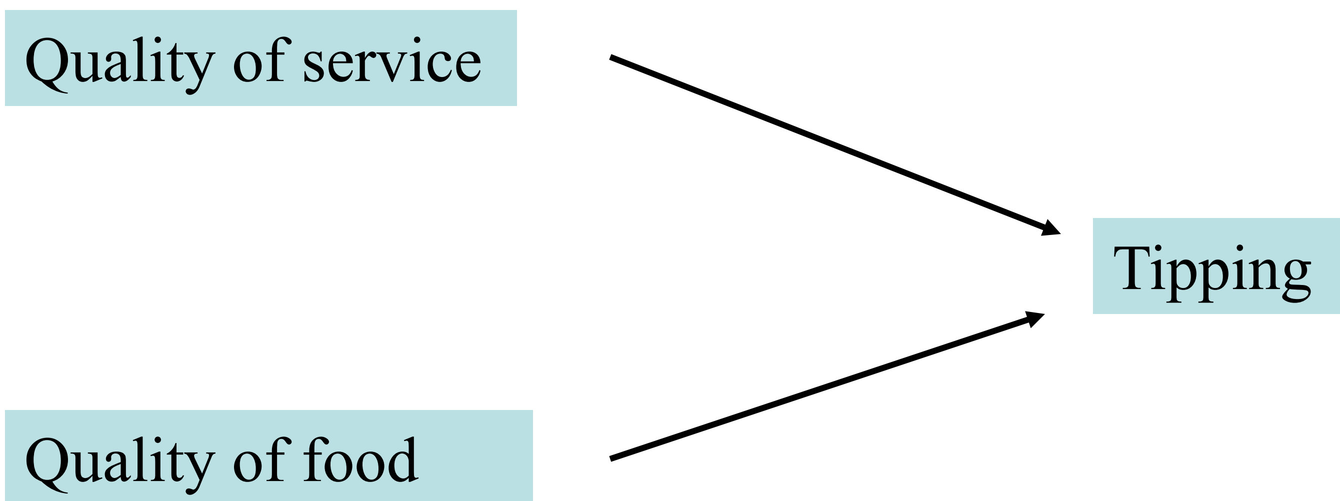
Fuzzy Rule-Based Models

- Types of fuzzy rules:
  1. If height is tall, then weight is fairly heavy.
  2. If height is tall, then weight is 80 kg. (zero-order)
  3. If height is tall, then weight is f(x). (first-order)
  4. If height is tall and body is fat, then weight is _.
  5. If height is tall or body is fat, then weight is _ and risk of heart disease is _.
- Rules have two parts: antecedent (if _) and consequent (then _).

Example of Fuzzy Modeling when Data Unavailable (Mamdani Reasoning)

- **Problem:** How much should I give tip in the restaurant in the USA according to given criteria? (multicriteria decision-making)
- No data, based on expertise.
- Two criteria (inputs):
  - **quality of service (0-10)**
  - **quality of food (0-10)**
- Output:
  - **Tip (%).**

Decision Model (Variables)



Linguistic Values of Variables

- **Service:** poor, good, excellent.
- **Food:** rancid, delicious.
- **Tip:** cheap, average, generous.

Examples of Fuzzy Rules

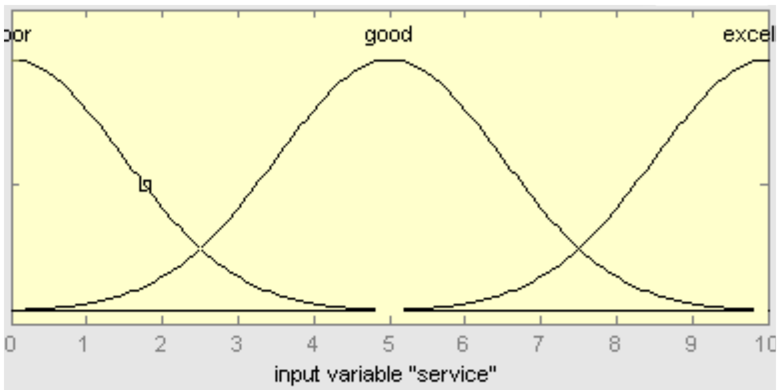
1. If service is poor and food is rancid, then tip is cheap.
2. If service is good and food is delicious, then tip is average.
3. If service is excellent or food is delicious, then tip is generous.

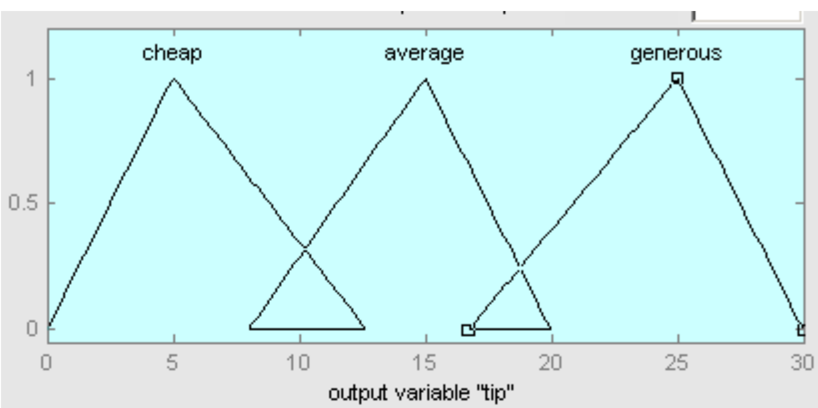
Example of a Fuzzy Decision Table

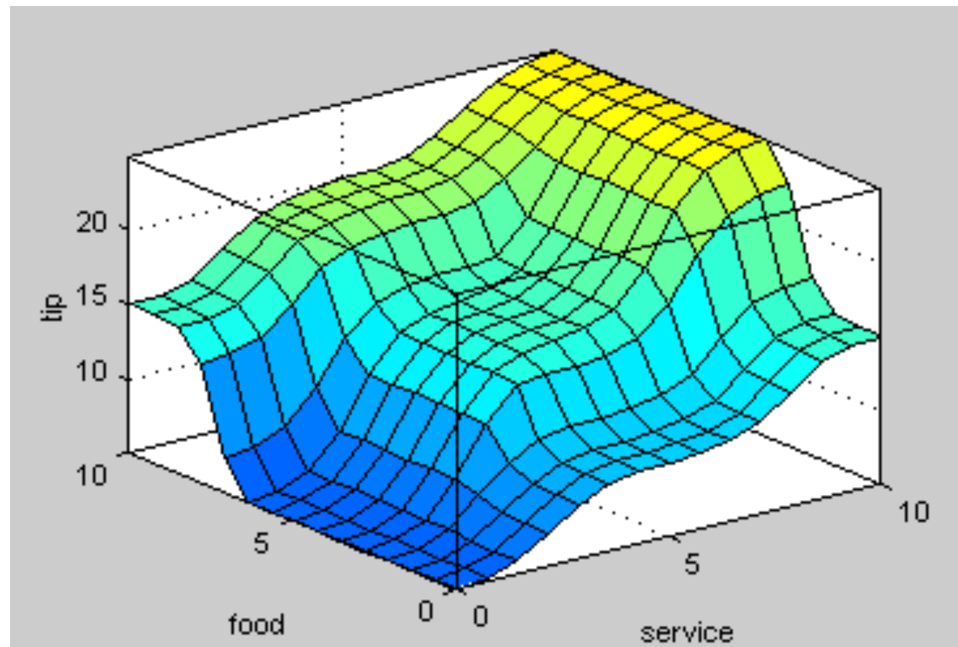
Service	Food Rancid	Delicious
Poor	Tip=?	
Good		Tip=?
Excellent	Tip=?	

E.g. If service is poor and food is rancid, then tip is cheap.

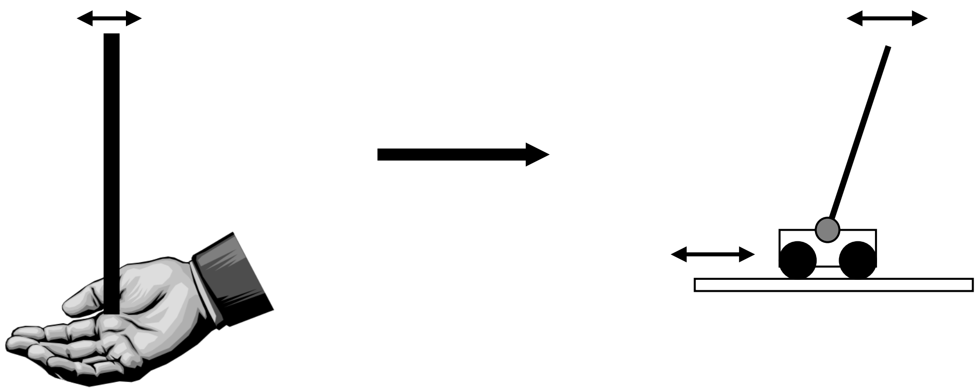
Fuzzy Values and Model

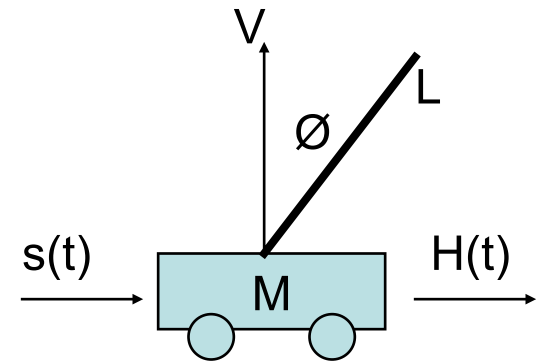






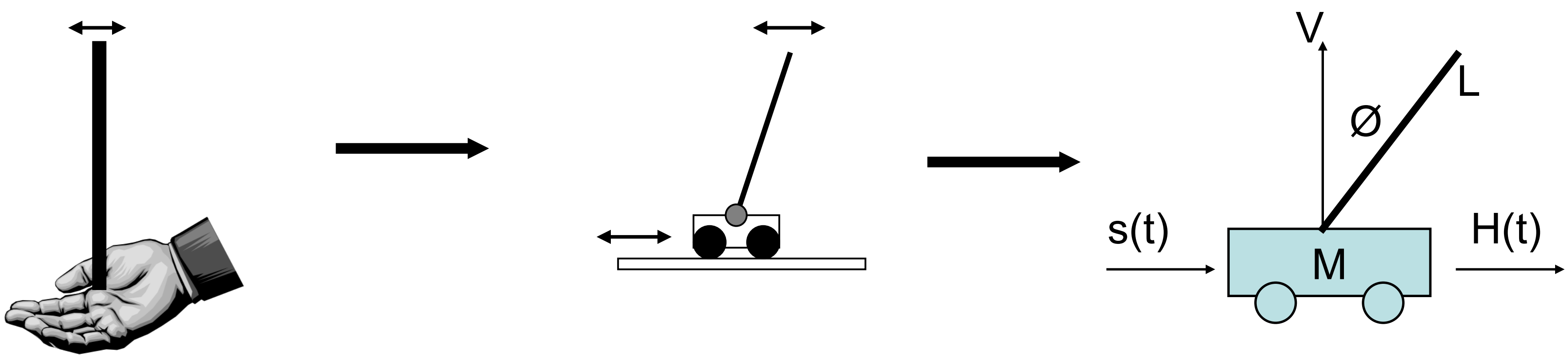
Fuzzy Control, Inverted Pendulum (Omron), Classical Model





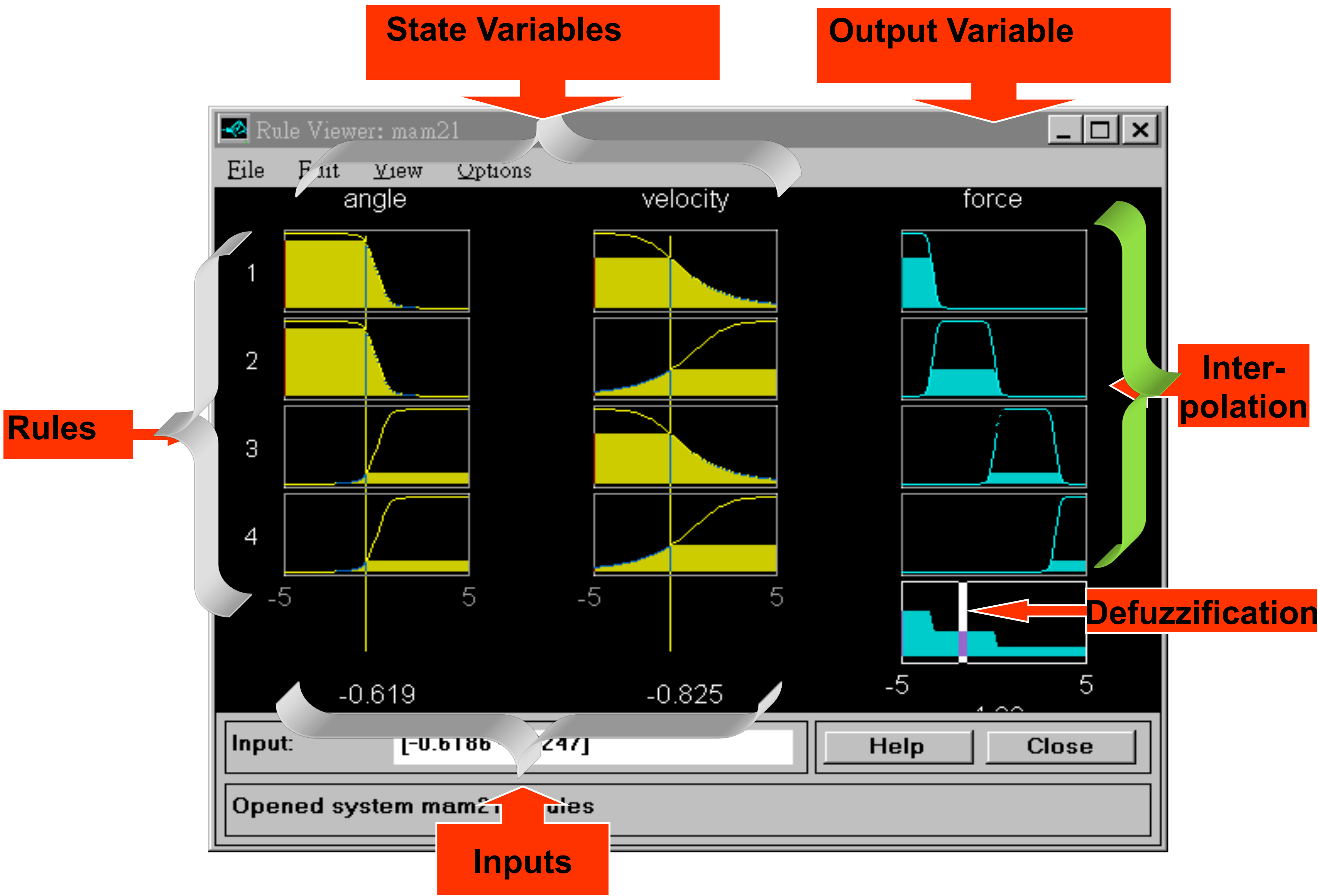
$md^2/dt^2(s(t)+L\cdot \sin ø(t) = H(t)$
$md^2/dt^2(L\cdot \cos ø(t)) = V(t)-m\cdot g$
$Jd^2/dt^2 = (L\cdot V(t)\cdot \sin ø(t)-L\cdot H(t)\cdot \cos ø(t)) = V(t)-m\cdot g$
$Md^2/dt^2\cdot s(t) = \mu(t)-H(t)-Fd/dt\cdot s(t)$

Fuzzy Control, Inverted Pendulum (Omron), Fuzzy Model



- **If error is big negative and angular velocity is zero, then velocity is big negative**
- **If error is small negative and angular velocity is small positive, then velocity is zero**
- **If error is small positive and angular velocity is small positive, then velocity is big positive**
- **etc.**

Fuzzy Logic Control : Inference Method



Two Main Types of Fuzzy Reasoning

- Mamdani (Mamdani-Assilian); no data required
- Takagi-Sugeno (-Kang); data required
- Matlab fuzzy logic toolbox

Comparison of fuzzy reasoning methods

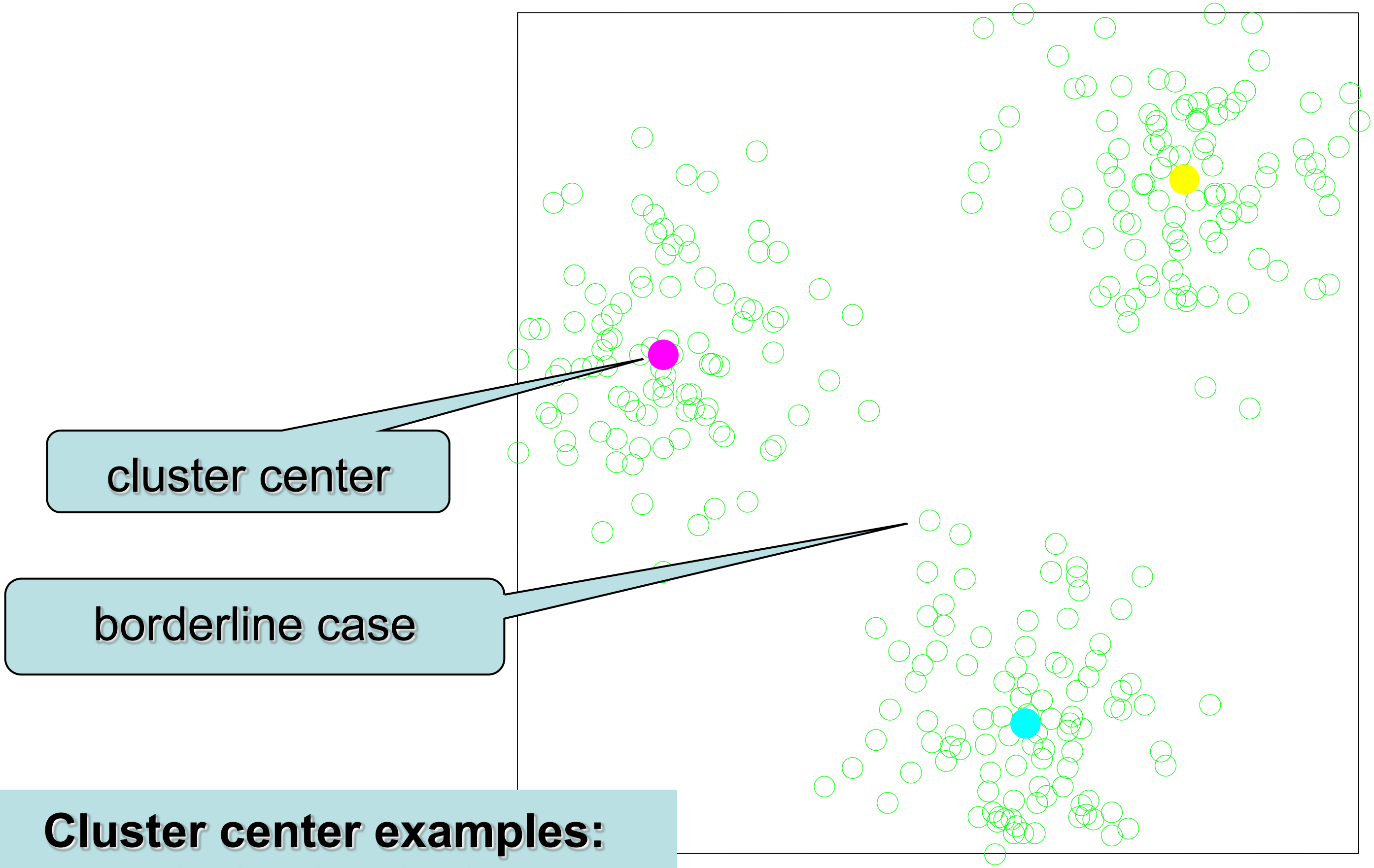
Advantages of the Sugeno Method

- It is computationally efficient.
- It works well with linear techniques (e.g., PID control).
- It works well with optimization and adaptive techniques.
- It has guaranteed continuity of the output surface.
- It is well suited to mathematical analysis.

Advantages of the Mamdani Method

- It is intuitive.
- It has widespread acceptance.
- It is well suited to human input.

Data Compression when Data: Clusters and Cluster Centers (2-D Data)

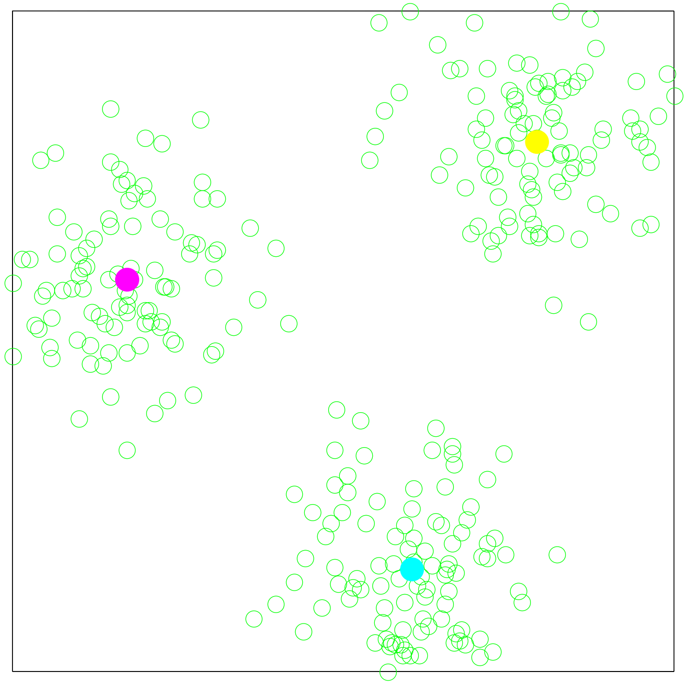


Cluster center examples:
typical points in data
typical persons in data
typical customers
best local centers in area
best central nodes

Clustering methods

- K-means clustering (traditional)
- Fuzzy C-means clustering (fcm, Bezdek)
- Subtractive clustering (Yager, Chiu)
- Best for spherical clusters
- Appropriate number of clusters: methods with **Calinski-Harabasz**, Davies-Bouldin, Gap, etc. (in Matlab: evalclusters)

Fuzzy c-means

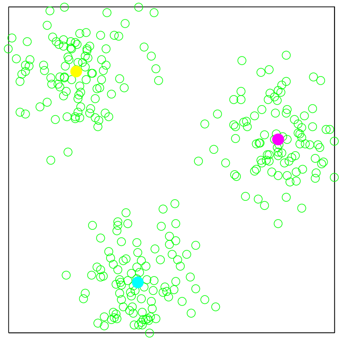


- Nr. of clusters fixed first
- Starts with random centers
- Aims to minimize variance within clusters and maximize the variance between them, step by step.
- The memberships to clusters are used as weights
- ”Theoretical” cluster centers

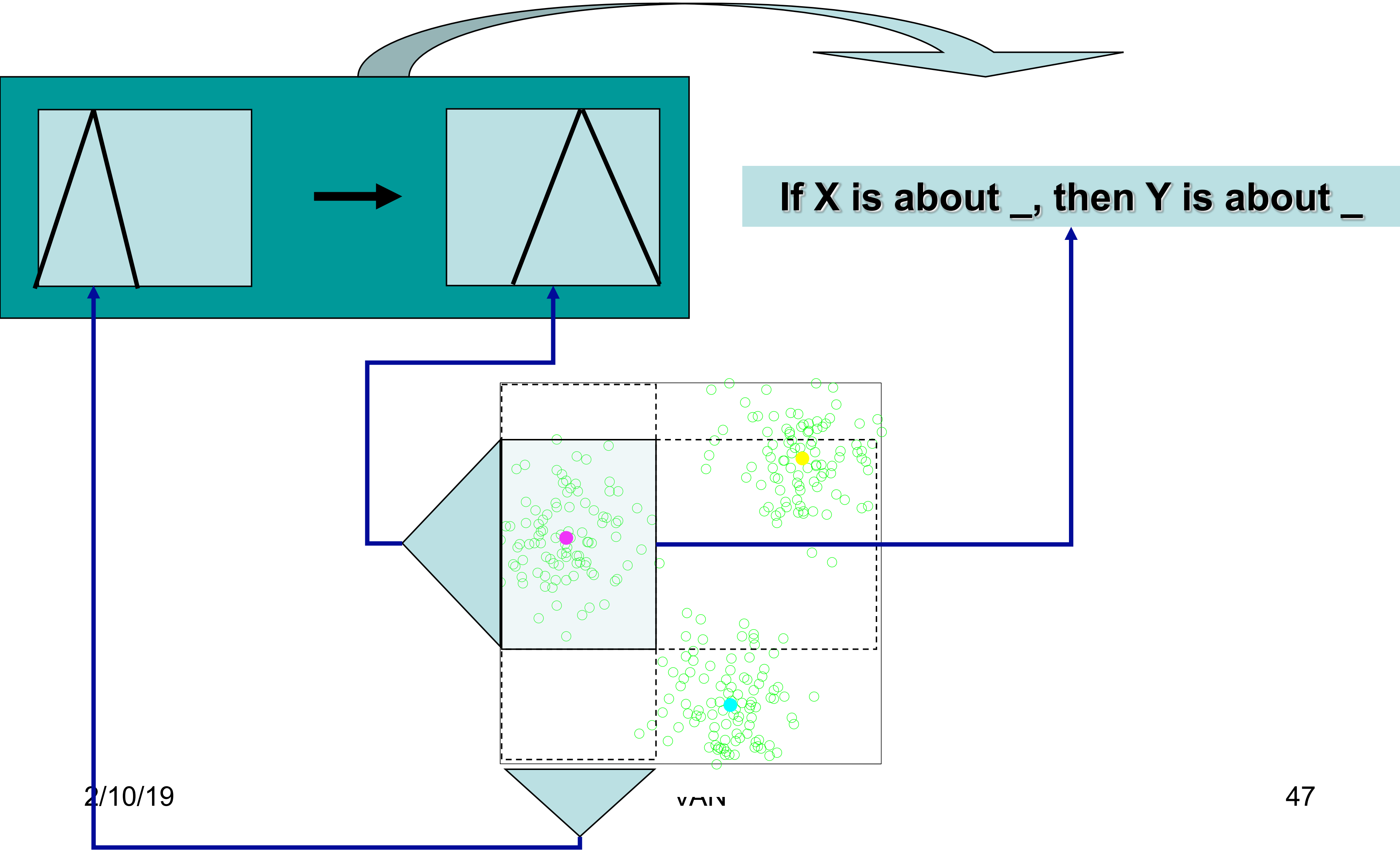
Fuzzy subtractive clustering (mountain clustering)

- Nr. of clusters is based on given radius (range of influence, $0 \leq r \leq 1$)
- The smaller radius, the more clusters
- Subtractive clustering assumes that each data point is a potential cluster center.

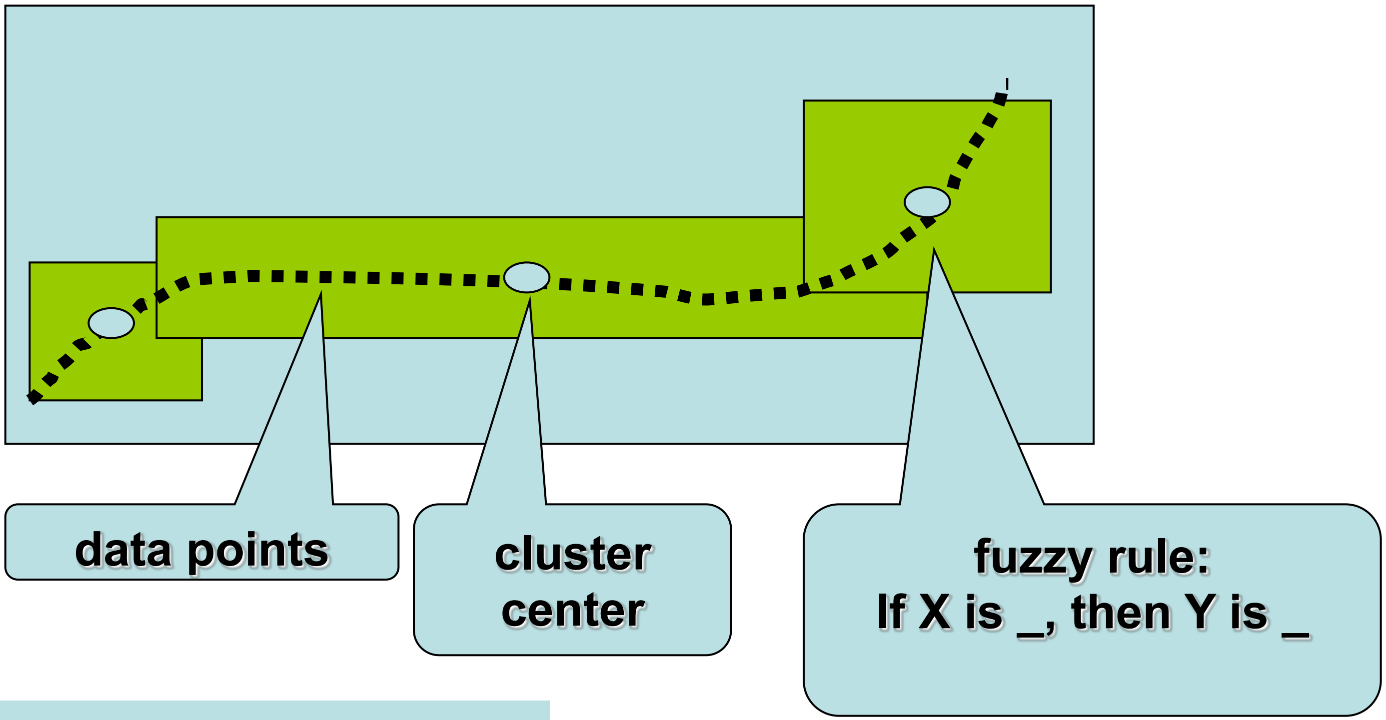
1. Calculate the likelihood that each data point would define a cluster center, based on the density of surrounding data points.
2. Choose the data point with the highest potential to be the first cluster center.
3. Remove all data points near the first cluster center. The vicinity is determined using r.
4. Choose the remaining point with the highest potential as the next cluster center.
5. Repeat steps 3 and 4 until all the data is within the influence range of a cluster center.
6. The subtractive clustering method is an extension of the mountain clustering method.



Reasoning Based on Data: From Clusters to Fuzzy Sets and Rules

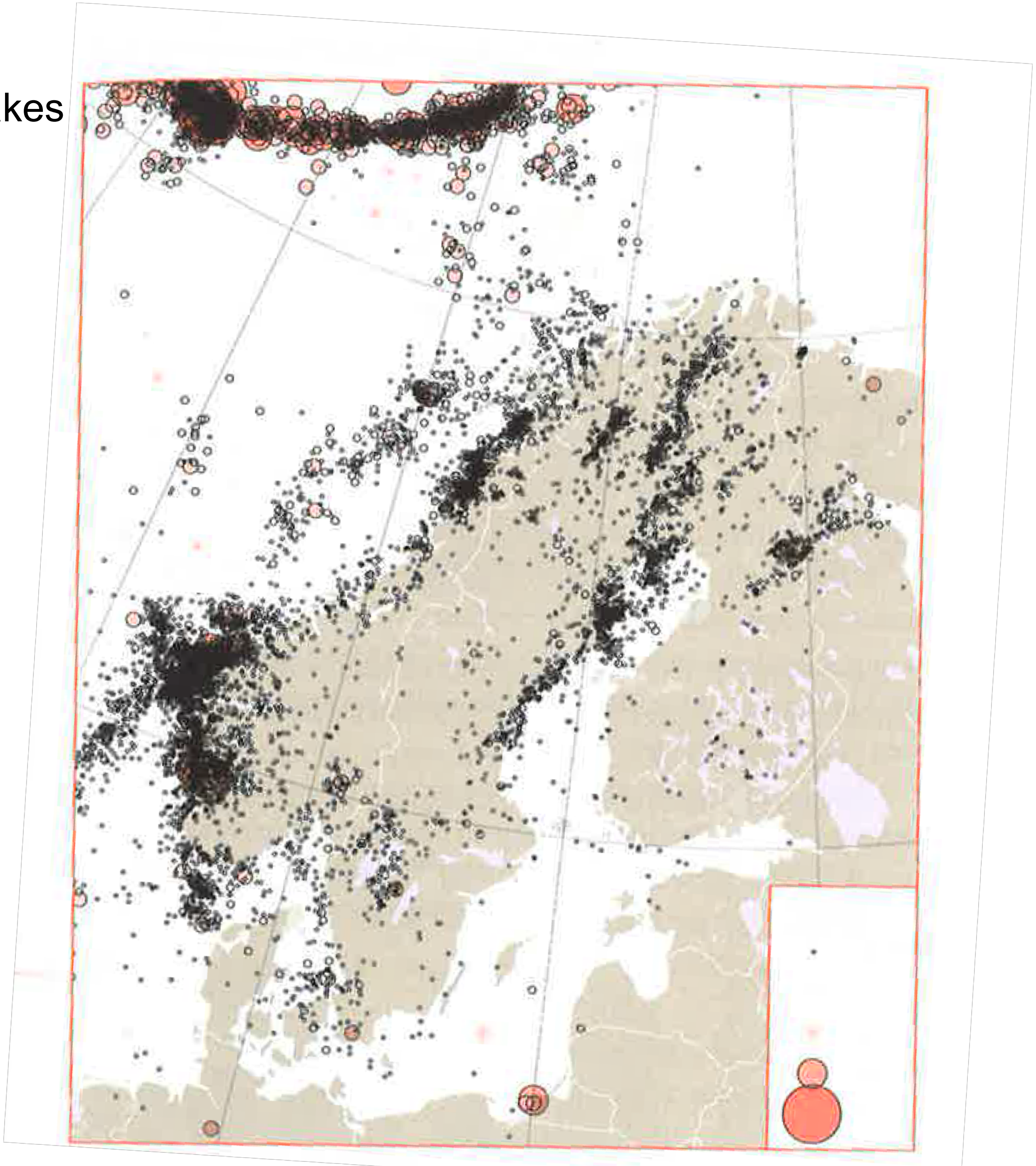


Fuzzy Rules Provide a Basis for Interpolation



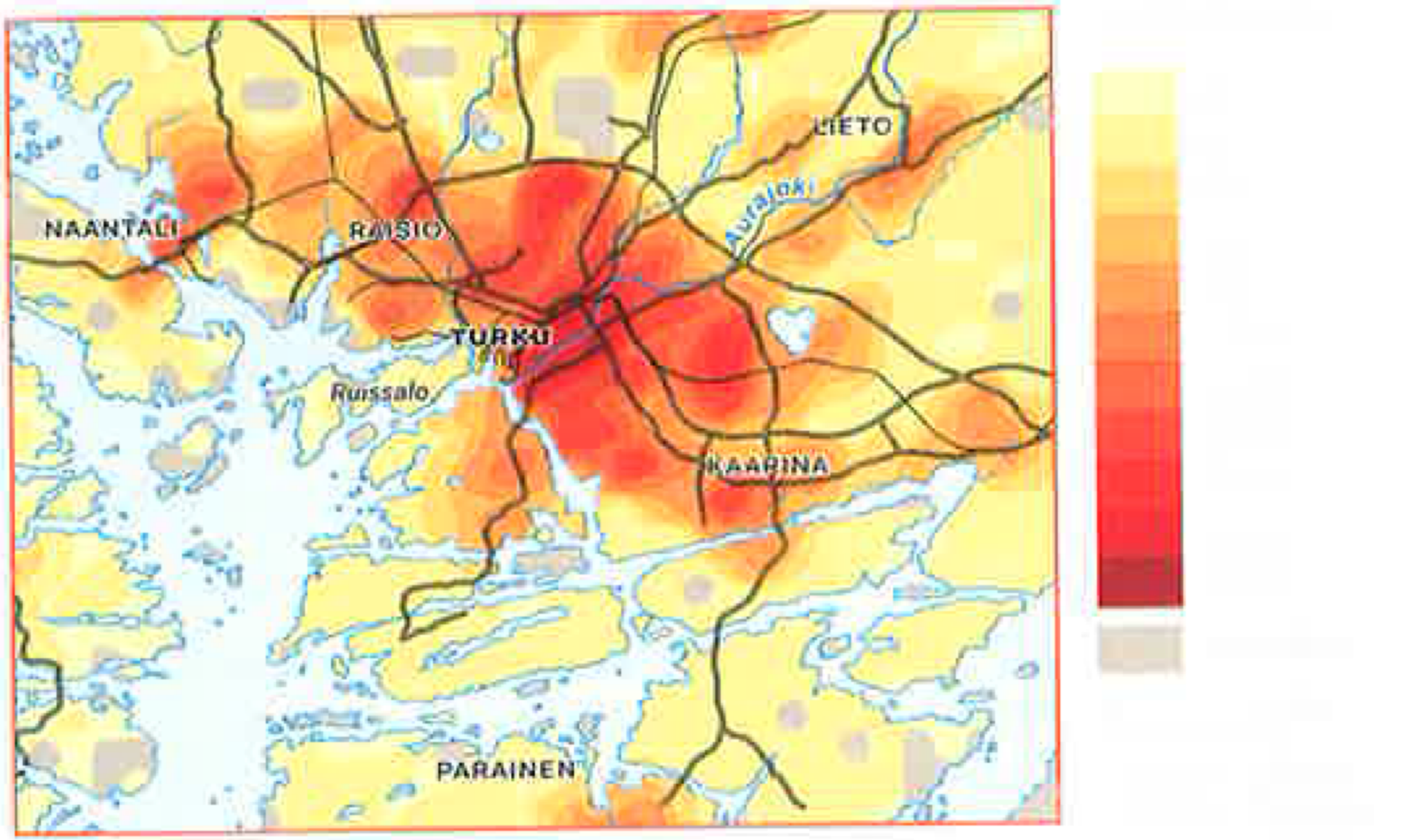
"Fuzzy models are universal approximators"

Earthquakes



CO_2 emission

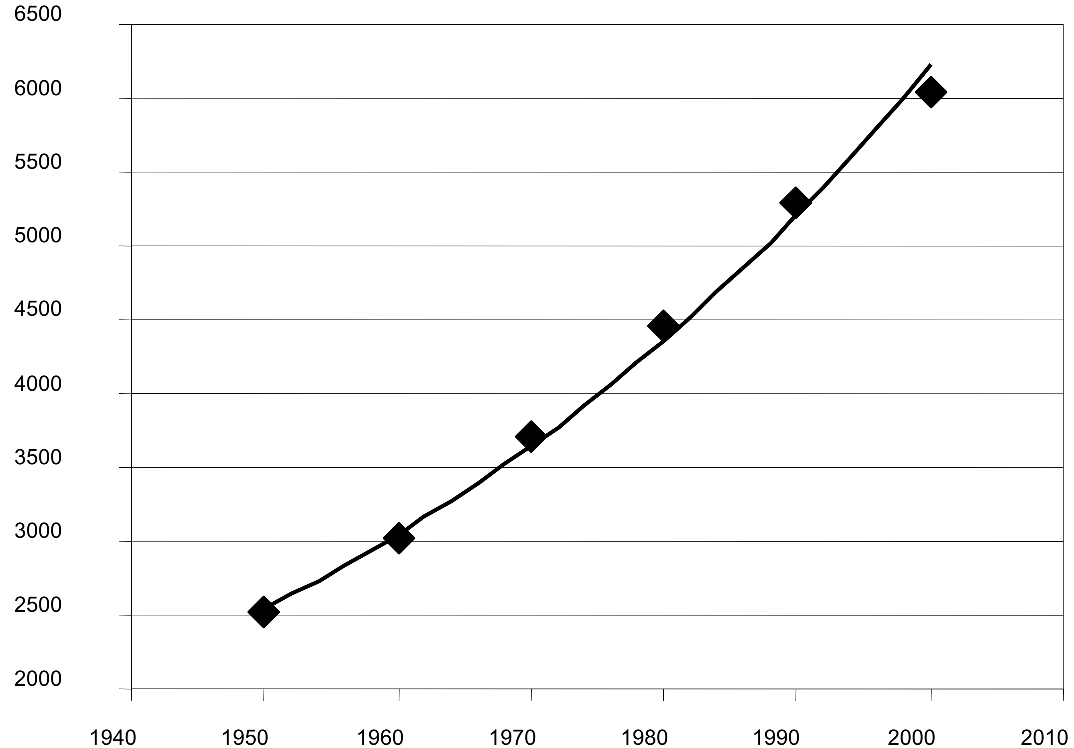
Population / Turku



Example with Data: Population in the Globe

Year	Population (millions)
1950	2515
1960	3019
1970	3698
1980	4448
1990	5292
2000	6045

Model: Year ➔ Population



- **Basic problem in modelling:**
  find a relationship between inputs and outputs.
- **Example of mathematical model:**
  a curve based on function
  $population = 1043{,}06 \cdot 1{,}01803^{(year-1900)}$

Generally: If Three Clusters, Fuzzy Model with Three Rules

1. **If year is about _, then population is about _ million.**
2. **If year is about _, then population is about _ million.**
3. **If year is about _, then population is about _ million.**

Cluster centers

Takagi-Sugeno Model Example (Zero-Order)

1. **If year is about 1950, then population is 2472 million.**
2. **If year is about 1975, then population is 4014 million.**
3. **If year is about 2000, then population is 6075 million.**

These values base on equal intervals (grid technique) or cluster centers

These (precise) numerical values are based on optimization when goal is good model (minimum errors)

Takagi-Sugeno Model (First-Order)

1. If year is about 1950, then population is a1*year+b1 million.
2. If year is about 1975, then population is a2*year+b2 million.
3. If year is about 2000, then population is a3*year+b3 million.

These values base on equal intervals (grid technique) or cluster centers

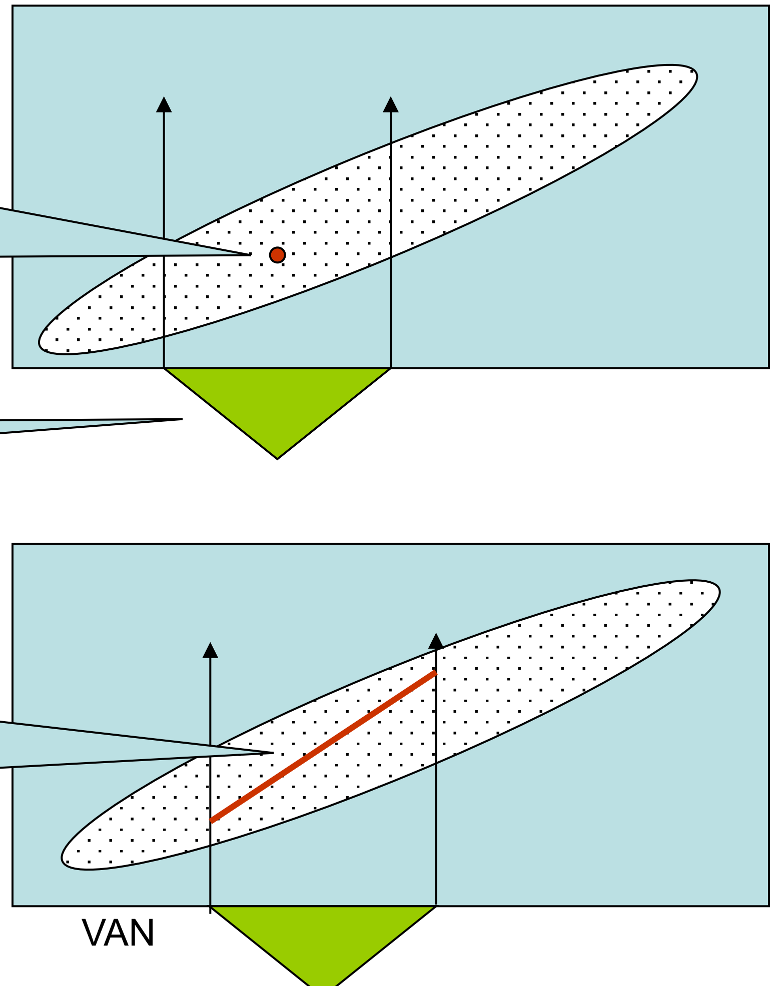
These (precise) numerical values are based on optimization of linear functions when goal is good model (minimum errors)

Takagi-Sugeno: Consequent Calculation Is Always Based on Optimization

Zero-order: Find consequent points that can minimize model errors.

Antecedent

First-order: Find linear functions that can minimize model errors.



Takagi-Sugeno Reasoning with Scatter (clustering) Technique

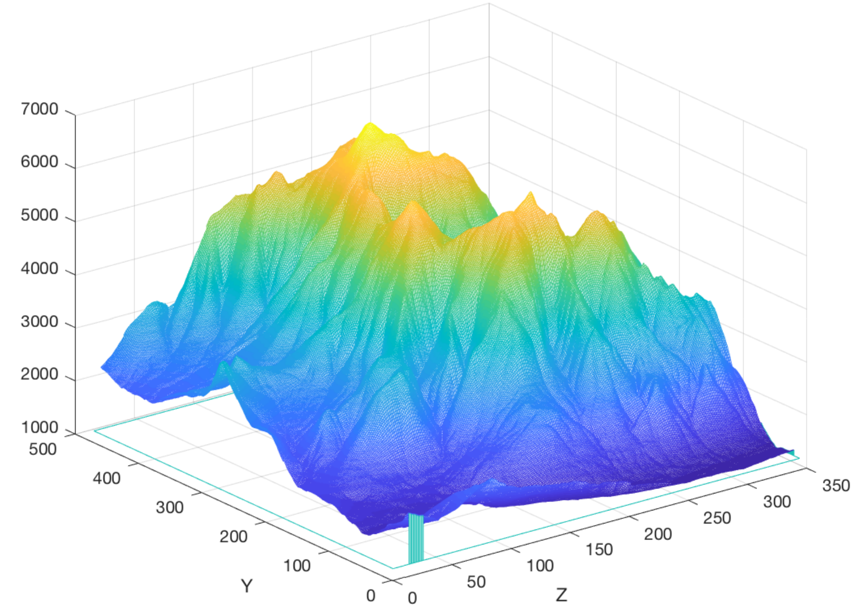
- Antecedents base on fuzzy clustering, consequents with zero- or first-order methods.
- Grid technique: equal intervals for fuzzy sets in inputs.

	Antecedent	
Consequent	Equal intervals	Clusters
Singleton	Zero-order, grid	Zero-order, scatter
Function	First-order, grid	First-order, scatter

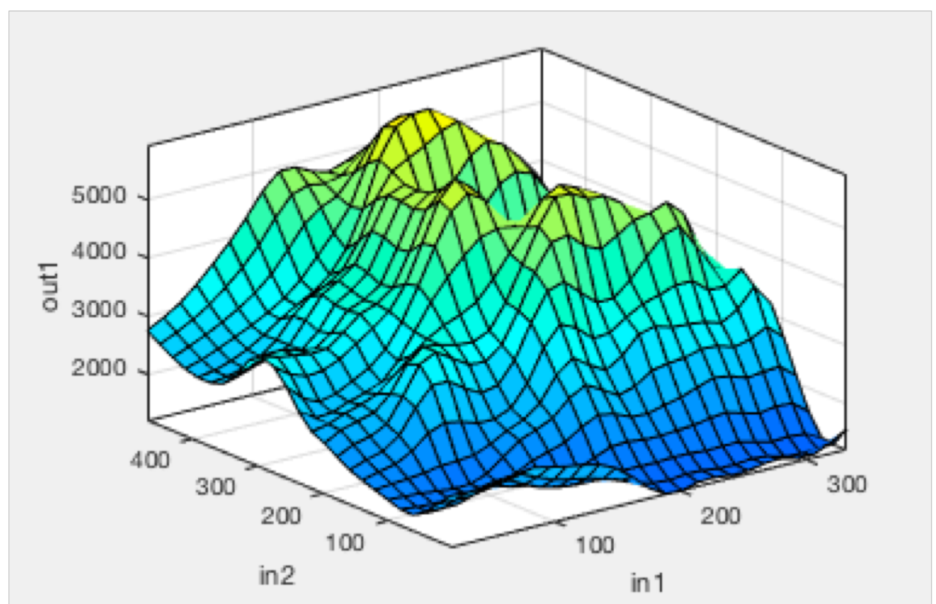
From Data to Fuzzy Model (Mt. Washington)

X	Y	Z (altitude)
16	1	2157
16	16	1904
31	16	1815
46	16	1808
61	16	2026
76	16	2184
91	16	1984
106	16	1877
121	16	1779
136	16	1733
151	16	1656

Data



FUZZY MODEL



Age of buildings in center of Helsinki

Modeling Traffic Patterns (autotrips.mat)

- This example shows traffic patterns in an area based on the area's demographics.
- **The Problem: Understanding Traffic Patterns**
- In this example we attempt to understand the relationship between the **number of automobile trips** generated from an area and the area's demographics. Demographic and trip data were collected from traffic analysis zones in New Castle County, Delaware.
- Five demographic factors are considered: **population, number of dwelling units, vehicle ownership, median household income and total employment.**
- Hereon, the demographic factors will be addressed as inputs and the trips generated will be addressed as output
- Two variables are loaded in the workspace, datin and datout., datout has 1 column representing the 1 output variable.

Goodness of Model

- Graphical presentations.
- Various goodness measures, e.g.
  - root mean square of errors (RMSE), or
  - RMSE/mean of response * 100%

  RMSE:

$$\sqrt{\Sigma(o_i-p_i)^2 /n}$$

- Statistical analysis of errors: mean=0, error distribution, outliers.

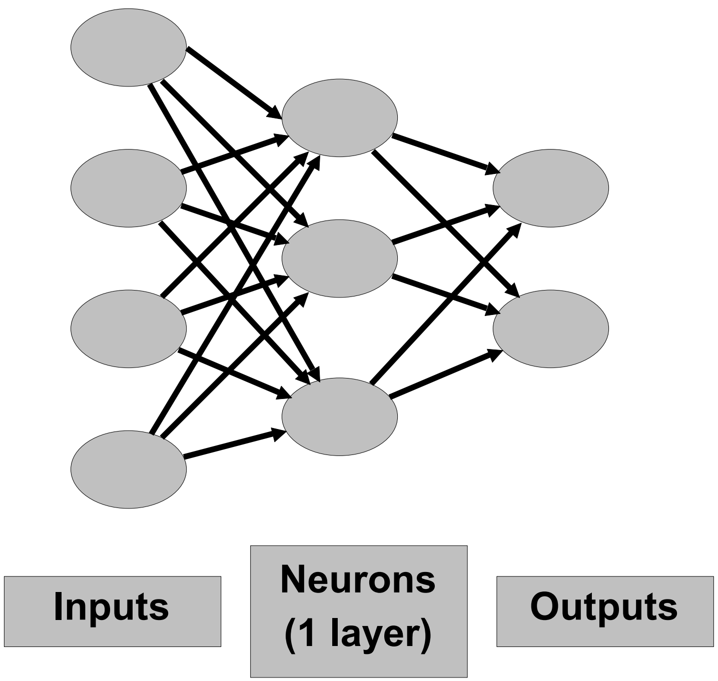
Population in the Globe: Model Errors

Year	Error math (d_i-p_i)	Error square $(d_i-p_i)^2$	Error 2 rules	Error square	Error 3 rules	Error square
1950	-33,81	1143,37	146,70	21520,89	0,00	0,00
1960	-28,50	812,31	-69,60	4844,16	-0,10	0,01
1970	54,24	2942,10	-111,10	12343,21	0,00	0,00
1980	91,32	8339,84	-81,70	6674,89	0,00	0,00
1990	82,92	6875,47	41,60	1730,56	0,00	0,00
2000	-183,26	33585,20	74,10	5490,81	0,00	0,00
Sum		53698,29		52604,52		0,01
Rmse		94,60		93,63		0,04

Neuro-Fuzzy Models

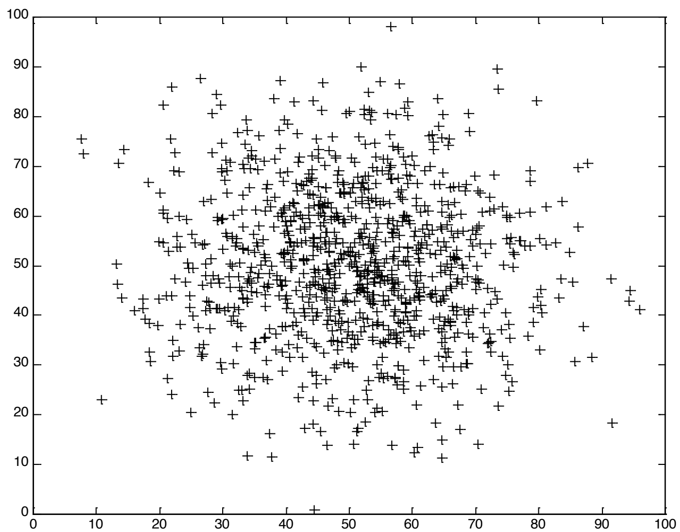
- Fuzzy model is fine-tuned with neural networks.
- In practice locations and/or shapes of fuzzy membership functions are tuned.
- In Matlab ANFIS-algorithm is used.

Neural networks (NN, 1940's)

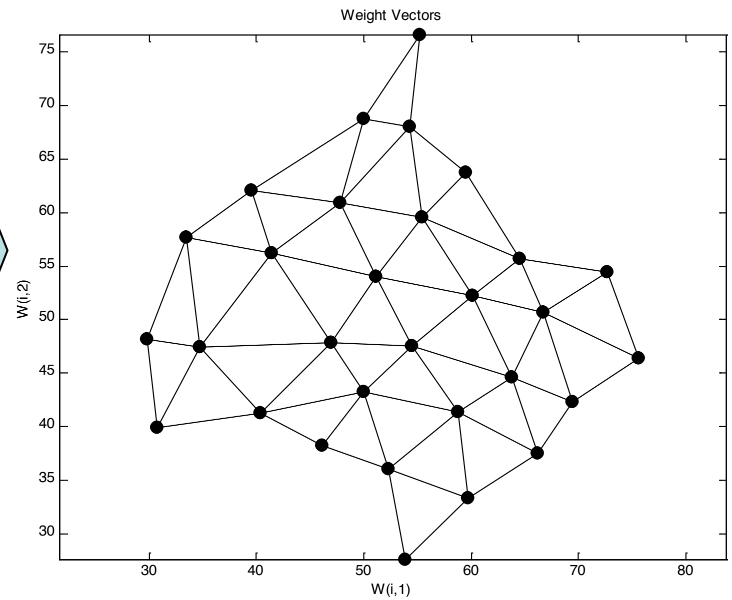


- Neural networks provide a powerful method to explore, classify, and identify patterns in data.
- Today: evolutionary computing replace NN?
- Website of Matlab
- Each neuron: $y=\Sigma w_i x_i$
- **Deep learning**: many layers (Dr. Watson, Facebook)

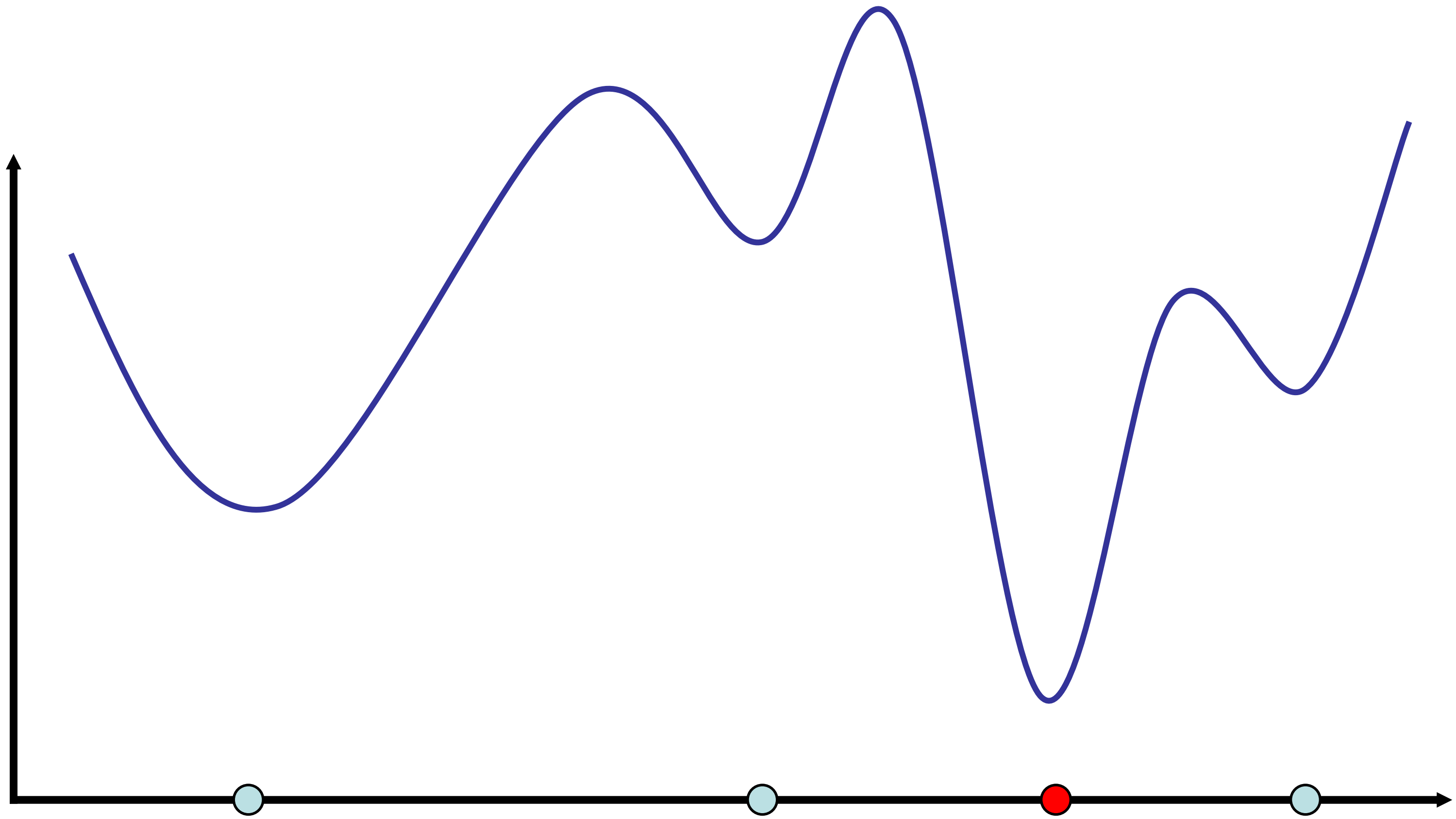
Alternative Clustering with NN: Self-Organized Maps (SOM, Kohonen)



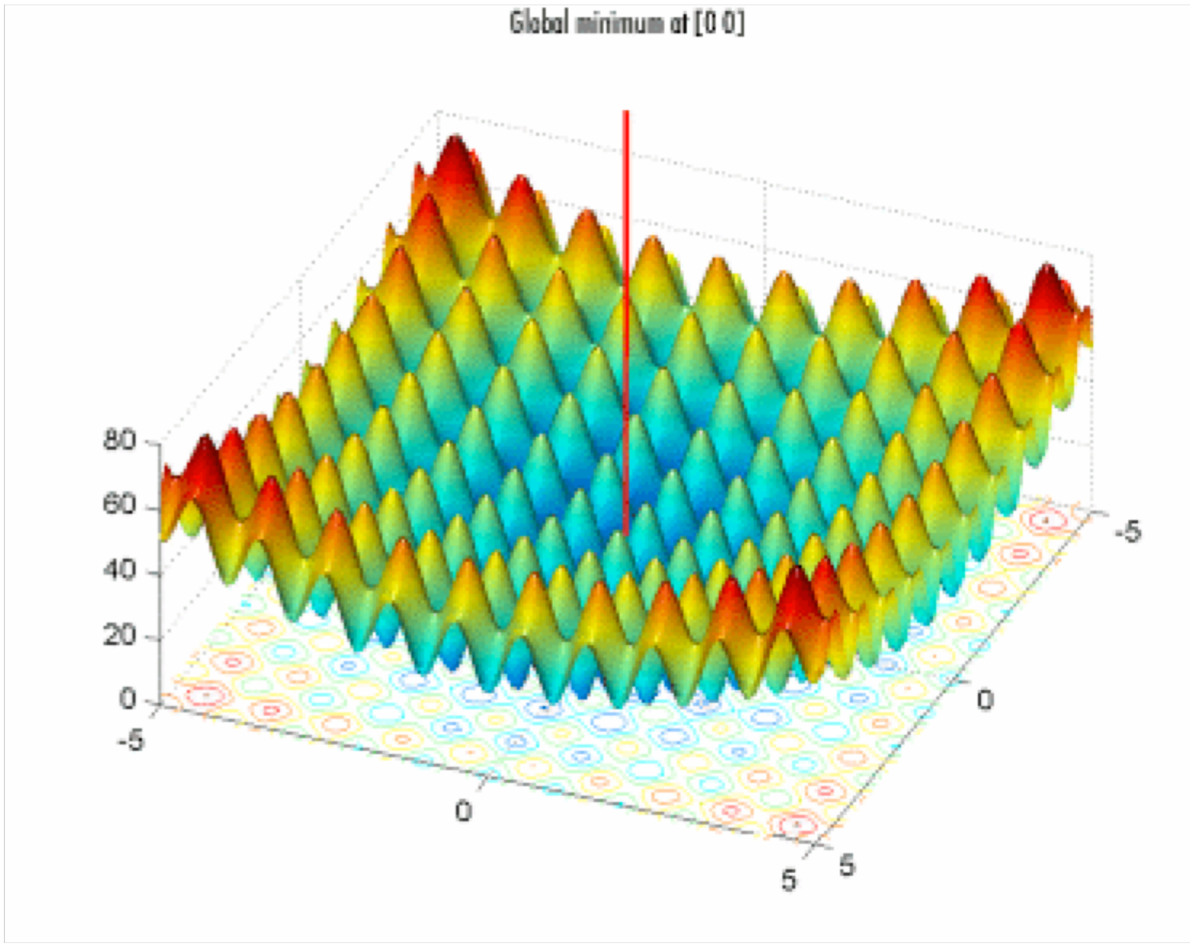
Skeleton with neurons



Local and Global Minima of Error



Finding the global optimum: evolutionary computing



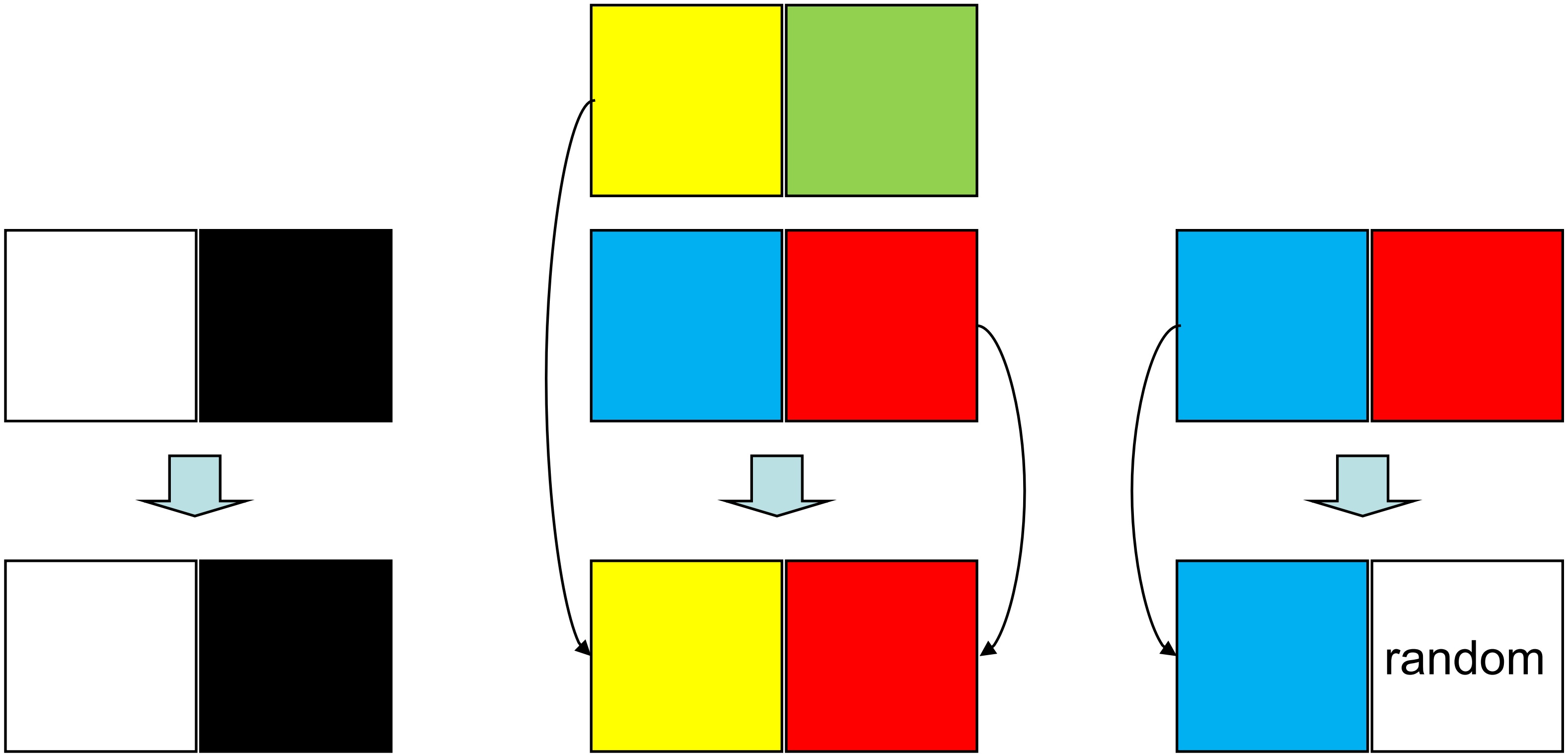
Evolutionary Computing: Genetic Algorithms (Matlab's definition)

- The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution.
- The genetic algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.
- You can apply the genetic algorithm to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, nondifferentiable, stochastic, or highly nonlinear.

Genetic Algorithms 2 (Matlab's definition)

- The genetic algorithm uses three main types of rules at each step to create the next generation from the current population:
  1. Selection rules select the individuals, called **parents**, that contribute to the population at the next generation.
  2. **Crossover** rules combine two parents to form children for the next generation.
  3. **Mutation** rules apply random changes to individual parents to form children.
- **The genetic algorithm differs from a classical, derivative-based, optimization algorithm in two main ways:**
  – Classical Algorithm generates a single point at each iteration. The sequence of points approaches an optimal solution. Selects the next point in the sequence by a deterministic computation.
  – Genetic Algorithm generates a population of points at each iteration. The best point in the population approaches an optimal solution. Selects the next population by computation which uses random number generators.

Genetic Algorithms: Selecting parameters (genes) to chromosomes for the next generation



Elite

Crossover (recombination)

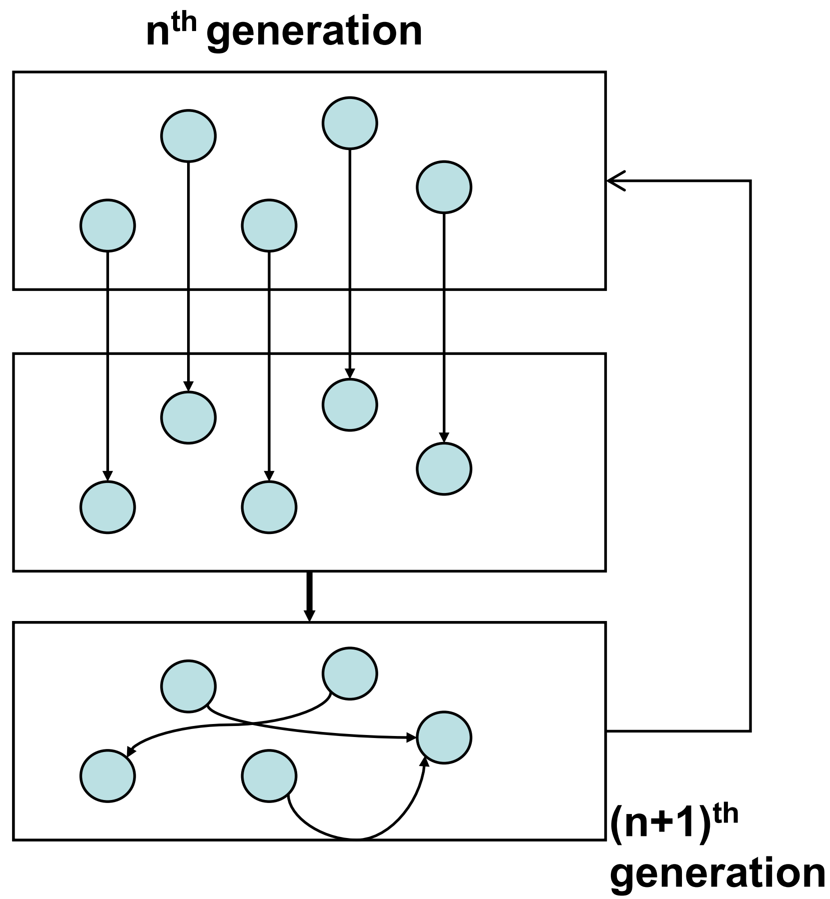
Mutation

More Novel Optimization Methods

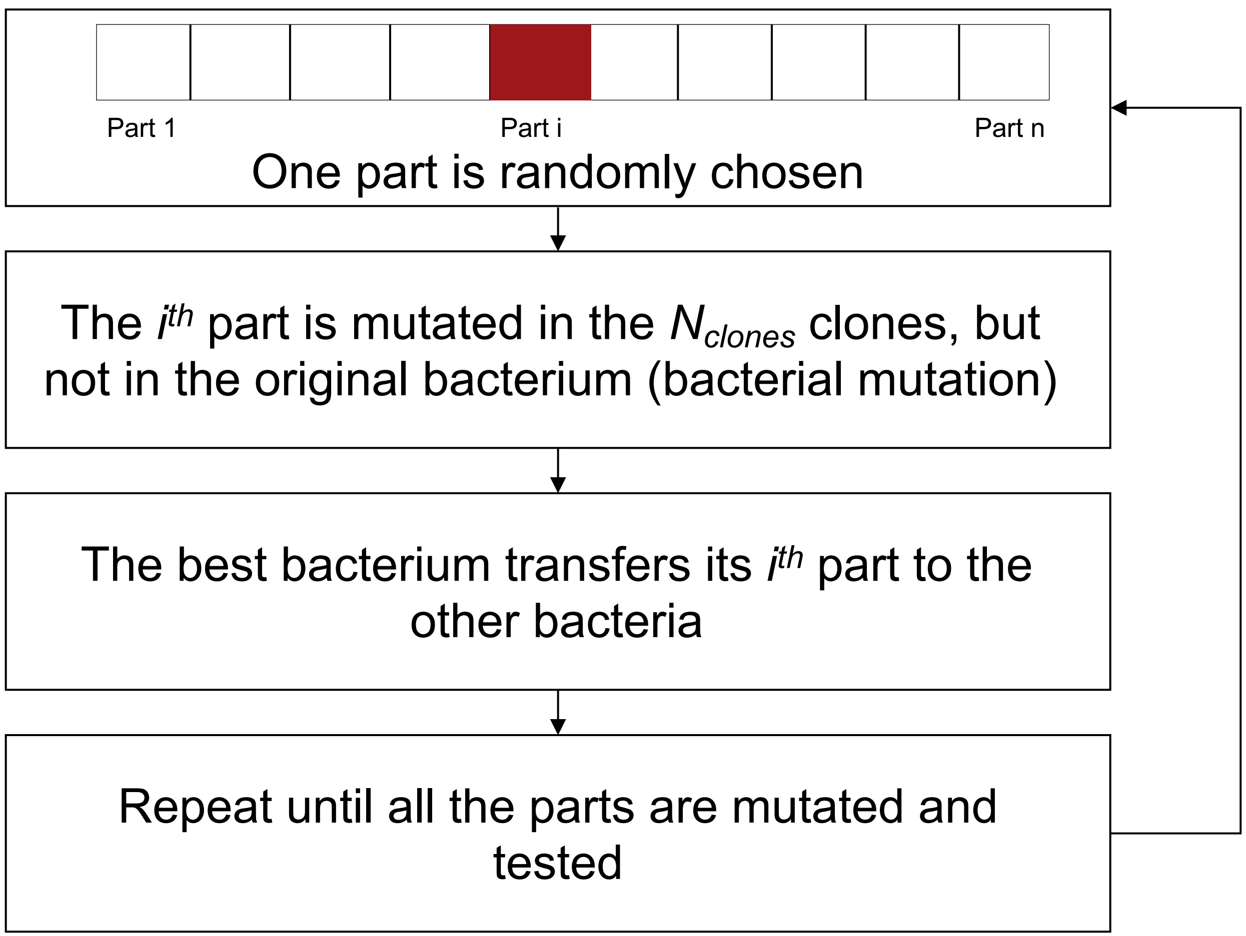
- **Memetic optimization**: evolutionary and traditional optimization in combination.
- **Bacterial evolutionary optimization**: clones of chromosomes and gene transfer are used.
- These may be faster and better than genetic methods alone.

Bacterial evolutionary algorithms (© Koczy)

- Generating the initial population randomly
- Bacterial mutation is applied for each bacterium
- Gene transfer is applied in the population
- If a stopping condition is fullfilled then the algorithm stops, otherwise it continues with the bacterial mutation step



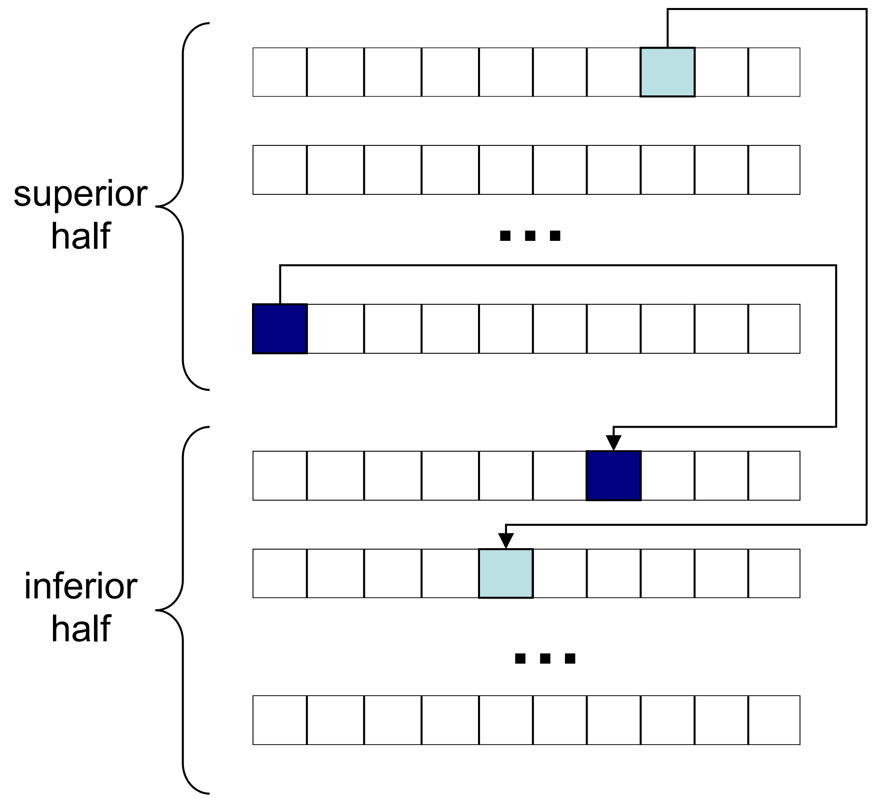
Bacterial mutation for each bacterium (© Koczy)



Gene transfer (© Koczy)

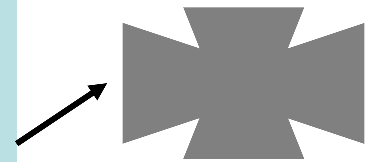
1. The population is divided into two halves
2. One bacterium is randomly chosen from the superior half (source bacterium) and another from the inferior half (destination bacterium)
3. A part from the source bacterium is chosen and this part can overwrite a part of the destination bacterium

This cycle is repeated for N_{inf} times (number of “infections”)



Regression Analysis
(Zimmermann & Zysno Data)

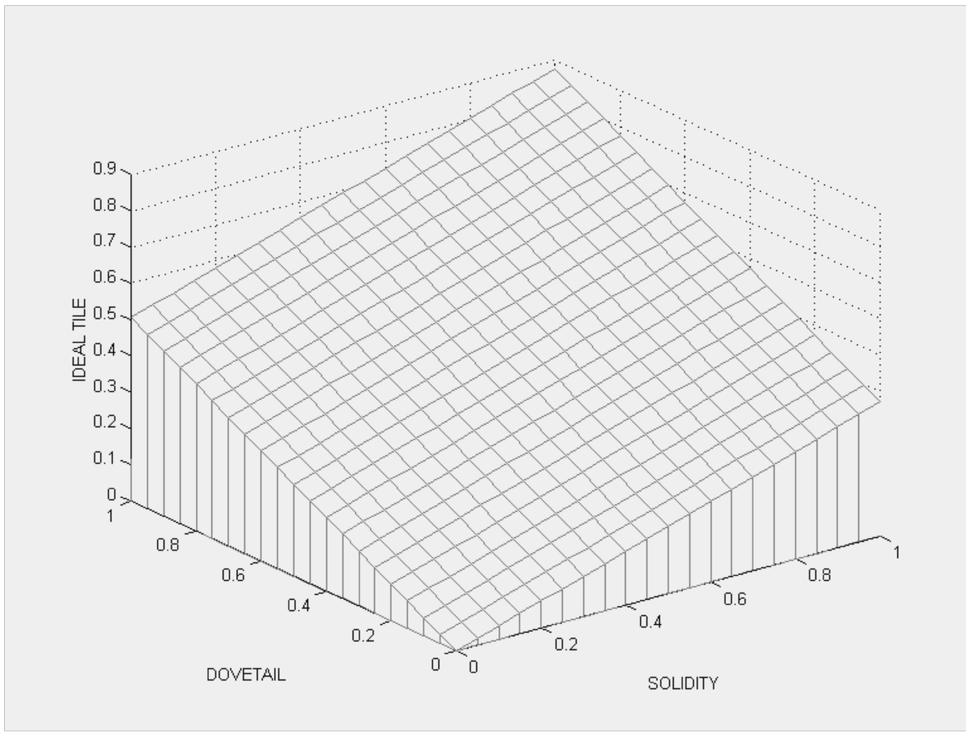
- The objects are tiles. 24 data vectors.
- The degrees of membership of 24 objects with respect to three fuzzy sets.
- Two inputs and one output.
- Inputs:
  - **Good solidity**, as assessed on the basis of the tile's color.
  - **Good dovetailing**, which means that the tiles cling to each other as tightly as possible.
- Output variable: **Ideal tile**, and this feature is assessed on the basis of the input variables.
- Sixty persons performed each of the preceding 24 assessments.



ZZ Data & Linear Regression

Independent:
1. **Solidity**
2. **Dovetailing**

Dependent:
Ideal tile



Obs. nro.	Solidity	Dovetail	Ideal tile
1	0.426	0.241	0.215
2	0.352	0.662	0.427
3	0.109	0.352	0.221
4	0.630	0.052	0.212
5	0.484	0.496	0.486
6	0.000	0.000	0.000
7	0.270	0.403	0.274
8	0.156	0.130	0.119

Ideal tile = 0.370*Solidity+0.507*Dovetailing
Rsquare=0.909

Interlude: Fuzzy Aggregation Operators, e.g. Generalized Mean

$Gmean = (\Sigma_{i=1,\ldots,m} w_i \cdot X_i^p)^{1/p}$, $0 \leq X_i \leq 1$, $\Sigma_{i=1,\ldots,m} w_i = 1$, $-\infty < p < \infty$

- p: degree of compensation
- $p \rightarrow -\infty$: min
- $p = -1$: harmonic mean
- $p \rightarrow 0$: geometric mean
- $p = 1$: arithmetic mean
- $p = 2$: quadratic mean (root mean square)
- $p \rightarrow \infty$: max

Parameter Estimates

Parameter	Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
w1	.448	.026	.395	.502
w2	.552	.025	.499	.604
p	.377	.097	.175	.578

↓

Ideal tile = $(0.448 \cdot Solidity^{0.377} + 0.552 \cdot Dovetailing^{0.377})^{1/0.377}$

Generalized mean (non-linear)

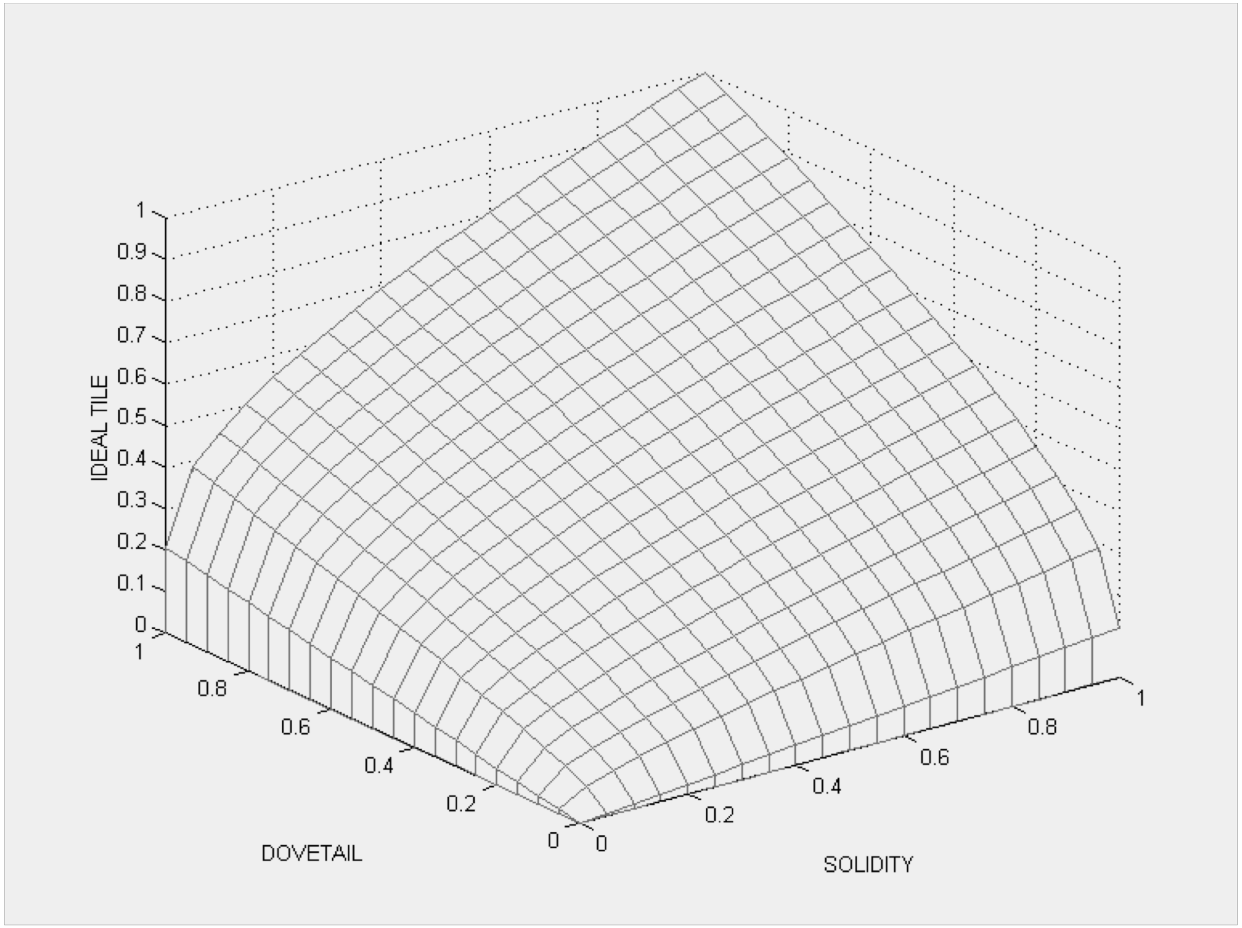
Zimmermann & Zysno:

ideal = $(w_1 \text{solidity}^p + w_2 \text{dovetail}^p)^{1/p}$

$(w_1+w_2=1;\ 0 \leq \text{variable} \leq 1,\ p \neq 0)$

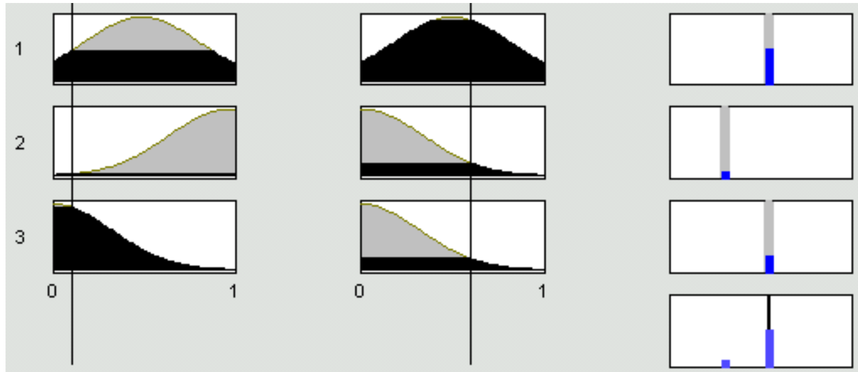
- $w_1=0.45$, $w_2=0.55$, $p=0.38$
- Hence:
  ideal = $(0.45\text{solidity}^{0.38}+0.55\text{dovetail}^{0.38})^{1/0.38}$
- Rmse=0.05
- Gen. means can be neurons in NN.

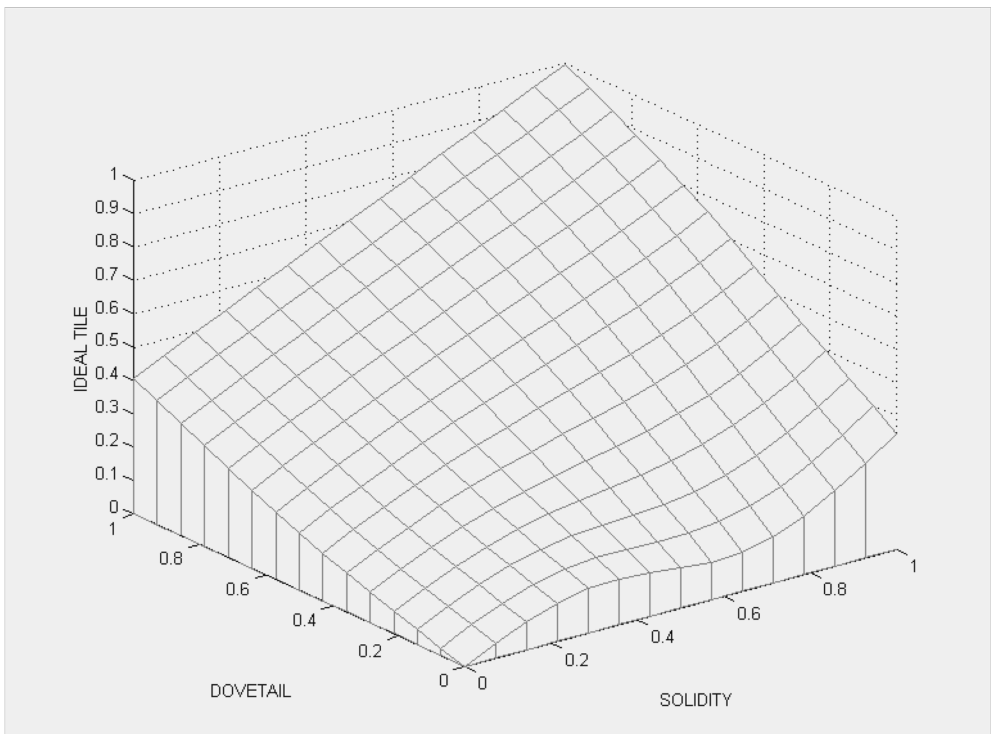
ZZ Model with Generalized Mean



Ideal tile = $(0.448 \cdot \text{Solidity}^{0.377} + 0.552 \cdot \text{Dovetailing}^{0.377})^{1/0.377}$

Fuzzy ZZ Model, Subtractive Clustering





Fuzzy ZZ Model, Interpretation

Initial rules:

1. **If *Solidity* is about 0.484 and *Dovetailing* is about 0.496, then the degree of *Ideal tile* is about 0.486.**
2. **If *Solidity* is about 0.949 and *Dovetailing* is about 0.020, then the degree of *Ideal tile* is about 0.247.**
3. **If *Solidity* is about 0.000 and *Dovetailing* is about 0.000, then the degree of *Ideal tile* is about 0.000.**

Final rules:

1. **If *Solidity* is about 0.484 and *Dovetailing* is about 0.496, then the degree of *Ideal tile* is 0.588**Solidity*+0.467**Dovetailing*-0.053.**
2. **If *Solidity* is about 0.949 and *Dovetailing* is about 0.020, then the degree of *Ideal tile* is 1.063**Solidity*+0.631**Dovetailing*-0.734.**
3. **If *Solidity* is about 0.000 and *Dovetailing* is about 0.000, then the degree of *Ideal tile* is 0.779**Solidity*+0.315**Dovetailing*+0.014.**

Goodness Evaluation of ZZ Models

Table 6.2.6.1.6. Statistics of Residuals of Data. Tile Assessment.					
Statistic	**LINREG0**	**Gmean**	**ANFIS**	**Tuned ANFIS**	**FMT**
Mean	-0.010	-0.019	0.001	0.000	0.001
Median	0.001	-0.009	-0.006	0.003	0.007
Standard Deviation	0.068	0.065	0.047	0.045	0.052
Sample Variance	0.005	0.004	0.002	0.002	0.003
Rmse	0.067	0.067	0.046	0.044	0.051
Kurtosis	-0.159	-0.263	-0.346	-0.423	-0.927
Skewness	-0.056	0.035	0.344	-0.501	-0.250
Range	0.282	0.262	0.183	0.165	0.177
Minimum	-0.142	-0.135	-0.084	-0.097	-0.092
Maximum	0.140	0.127	0.099	0.068	0.085
Sum	-0.245	-0.457	0.032	0.002	0.035
Count	24	24	24	24	24
t-test: mean=0	yes	yes	yes	yes	yes
Shapiro-Wilk test: normally distributed	yes	yes	yes	yes	yes

Definition 3–20 [Zimmermann and Zysno 1980]

The "*compensatory and*" *operator* is defined as follows:

$$\mu_{\tilde{A},comp}(x) = \left(\prod_{i=1}^{m} \mu_i(x)\right)^{(1-\gamma)} \left(1 - \prod_{i=1}^{m} (1 - \mu_i(x))\right)^{\gamma}, \qquad x \in X, 0 \leq \gamma \leq 1$$

$y = prod(x)^{(1-p)} * (1-prod(1-x))^{p}$

$0 \leq p \leq 1$

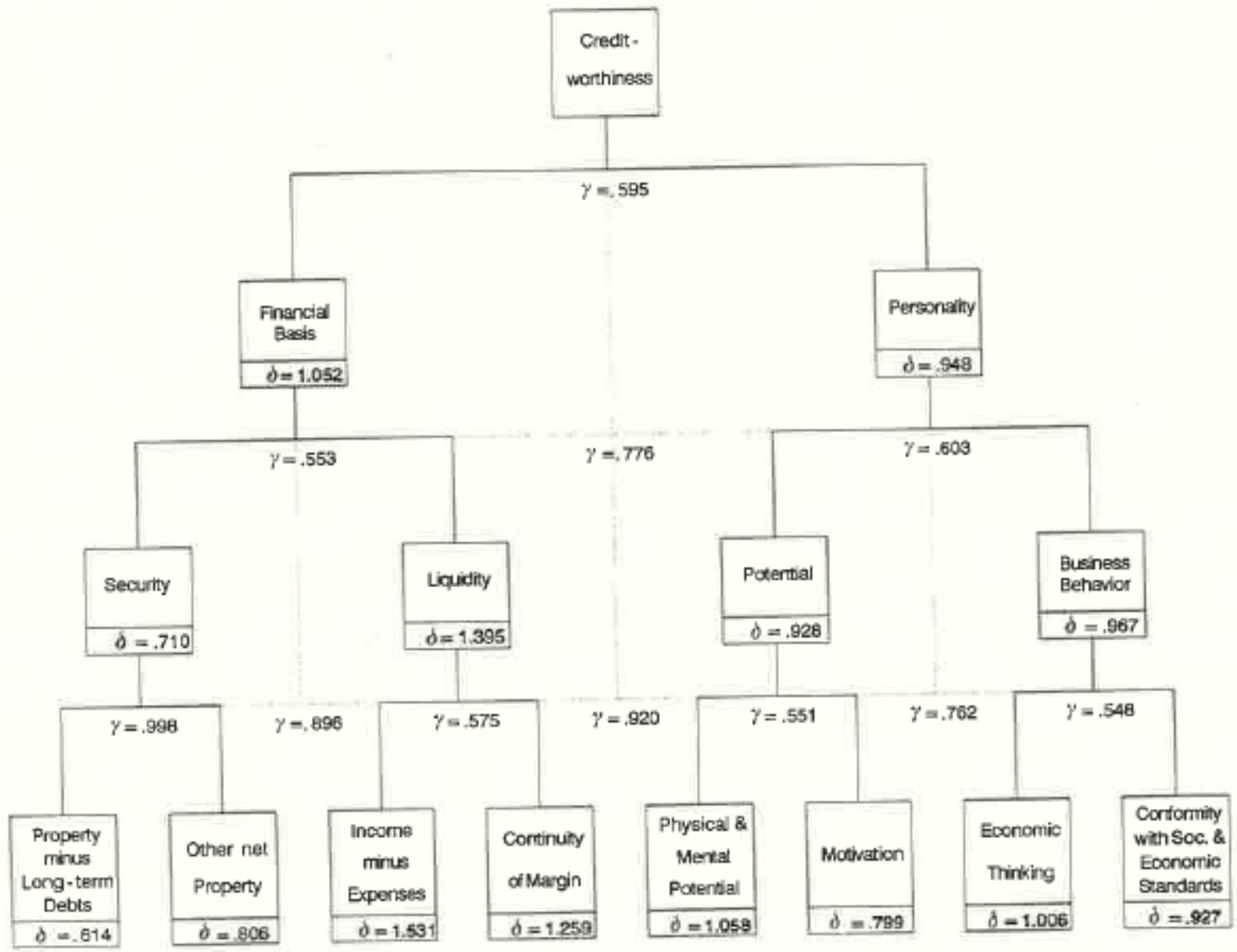
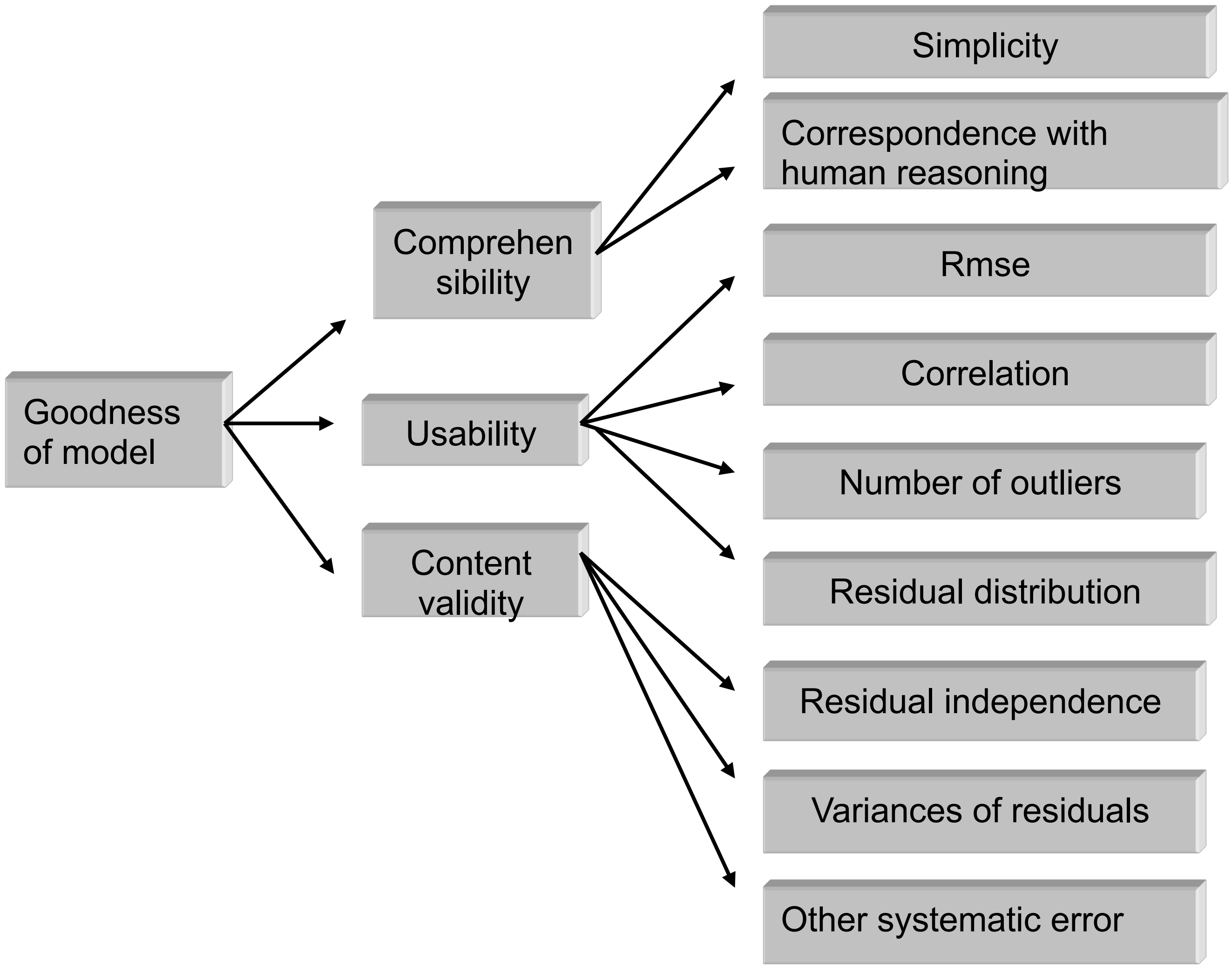


Figure 14–16. Concept hierarchy of creditworthiness together with individual weights δ and γ-values for each level of aggregation.

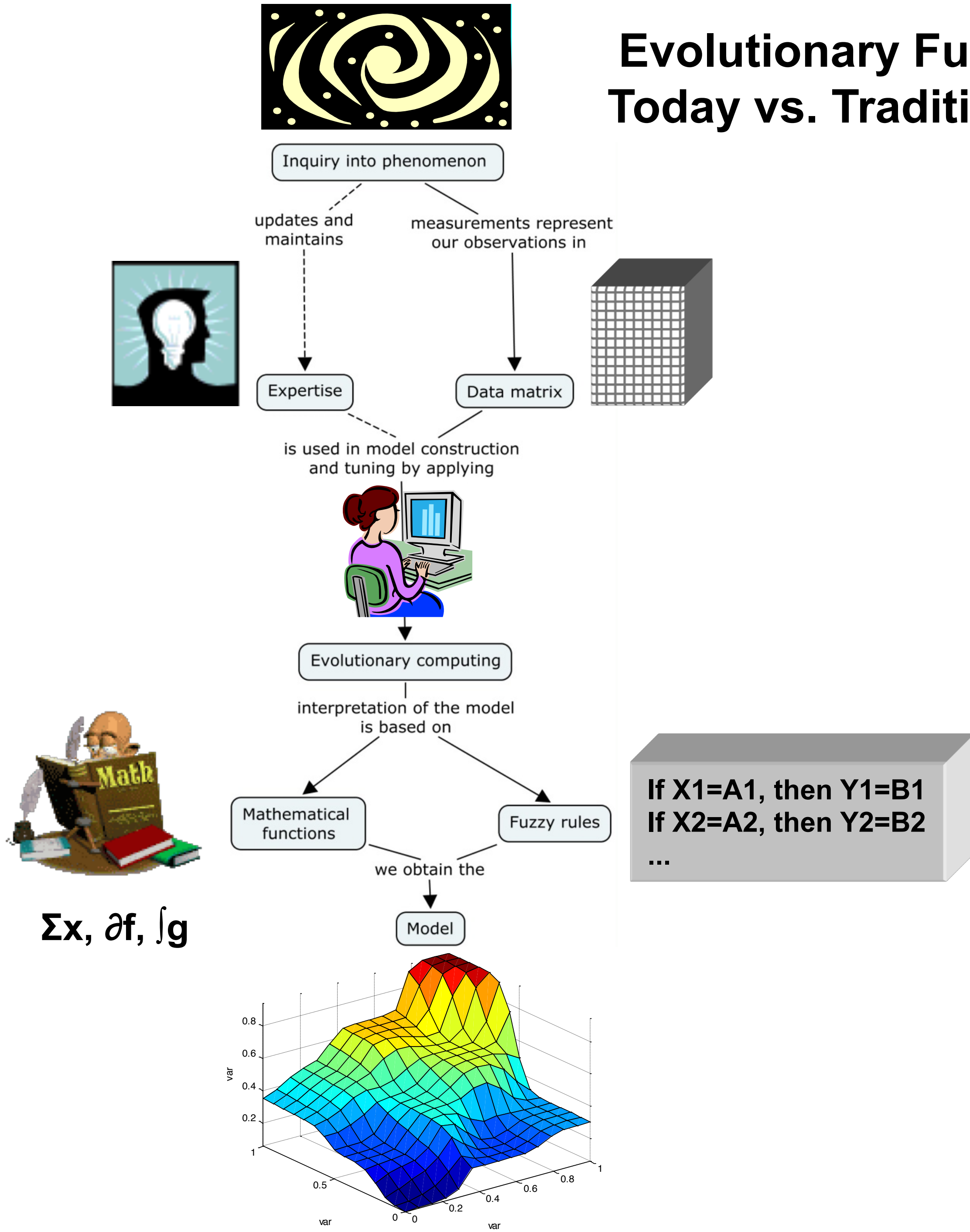
Goodness Evaluation



Goodness Evaluation: Residuals

Mean
Median
Standard Deviation
Sample Variance
Rmse
Kurtosis
Skewness
Range
Minimum
Maximum
t-test: mean=0?
Shapiro-Wilk or Kolmogorov-Smirnov test: normally distributed?
t-test or Wilcoxon test: 2 related samples (samples are sets of residuals)

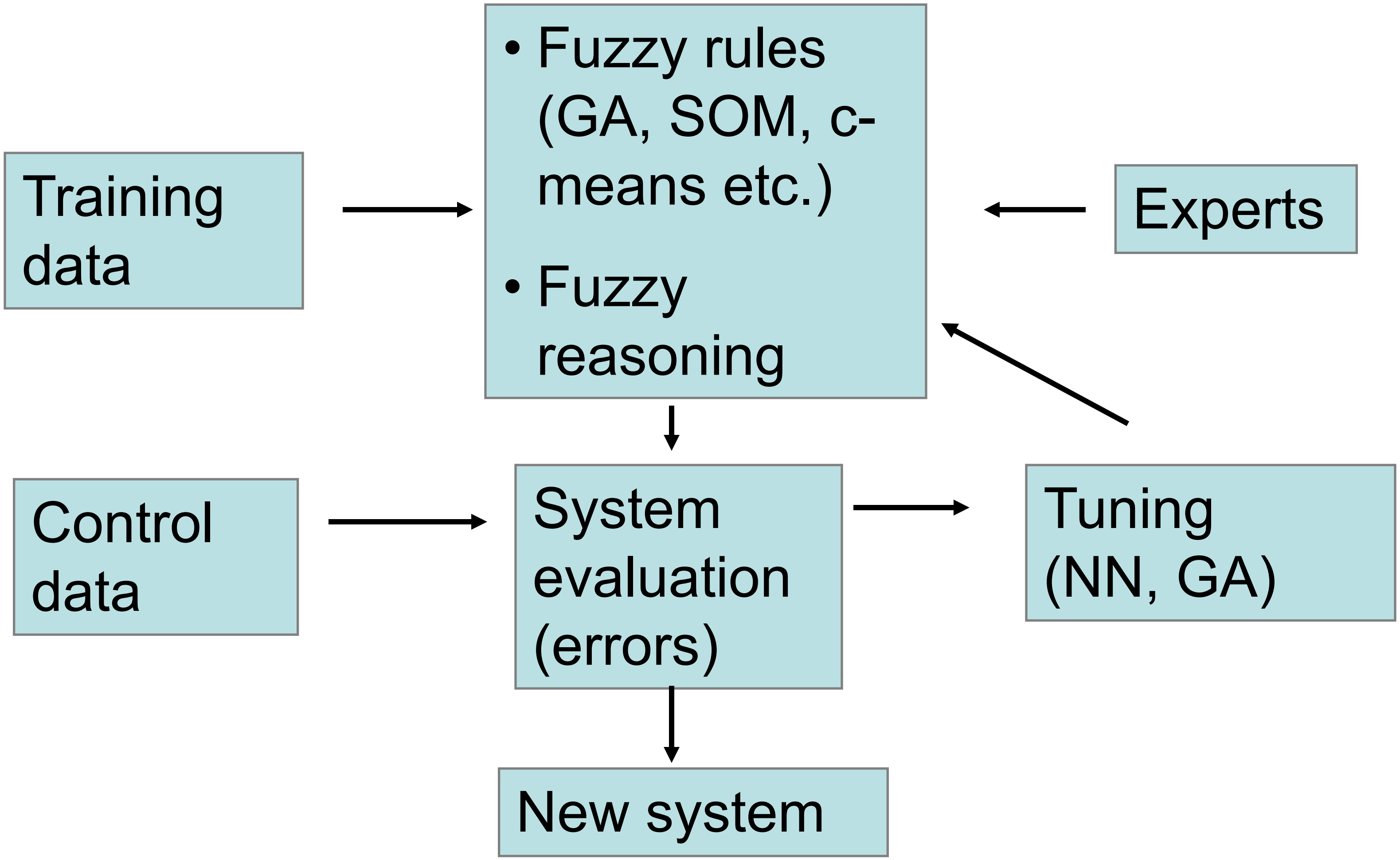
Evolutionary Fuzzy Modeling Today vs. Traditional Modeling

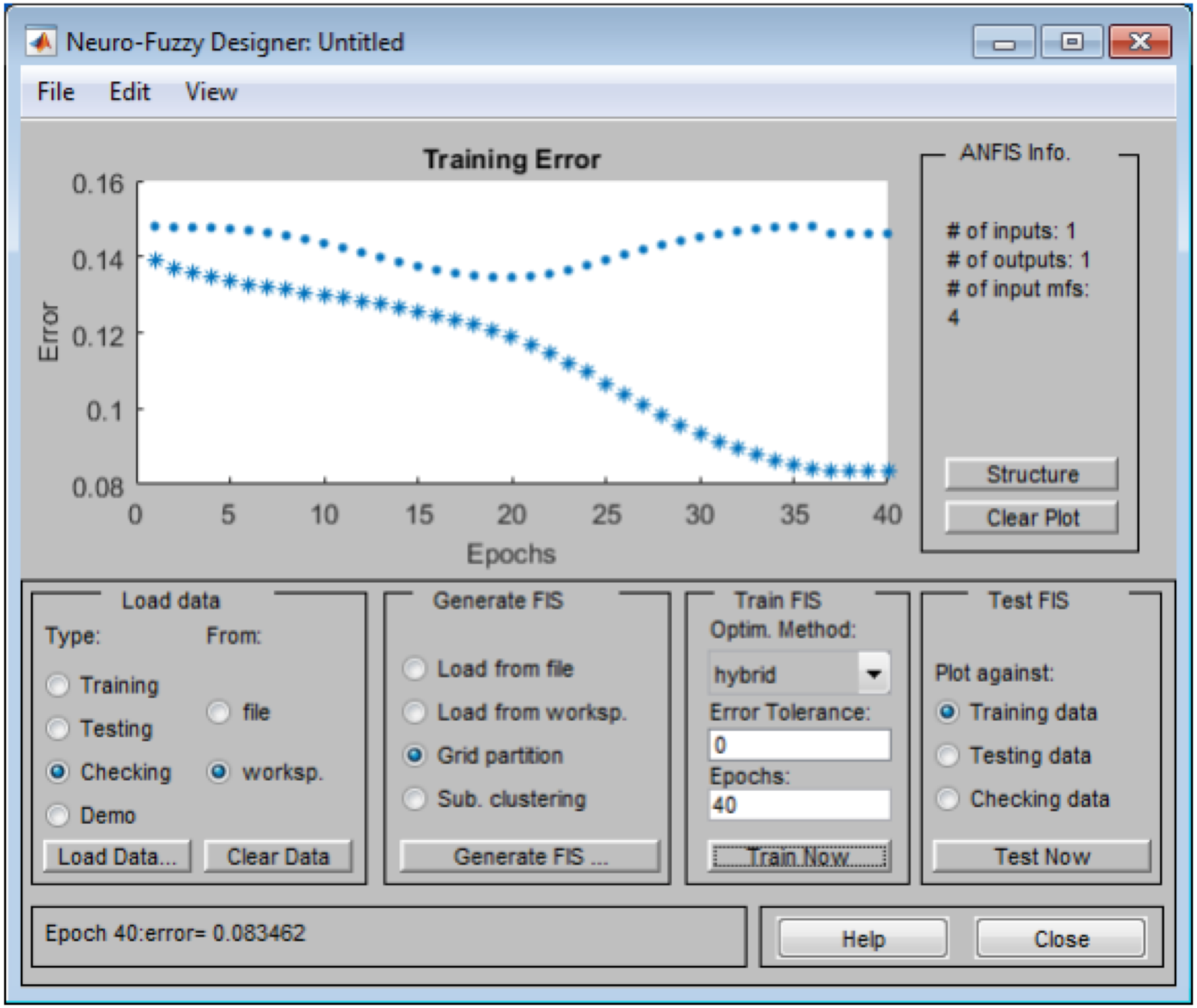


Goodness Evaluation of Model with Control Data (cross-validation)

- We can divide data into two parts: **training** and **control** (validation) data.
- We construct model with training data but evaluate its goodness with control data.
- This method shows whether we can generalize our model to any data set in the same population.

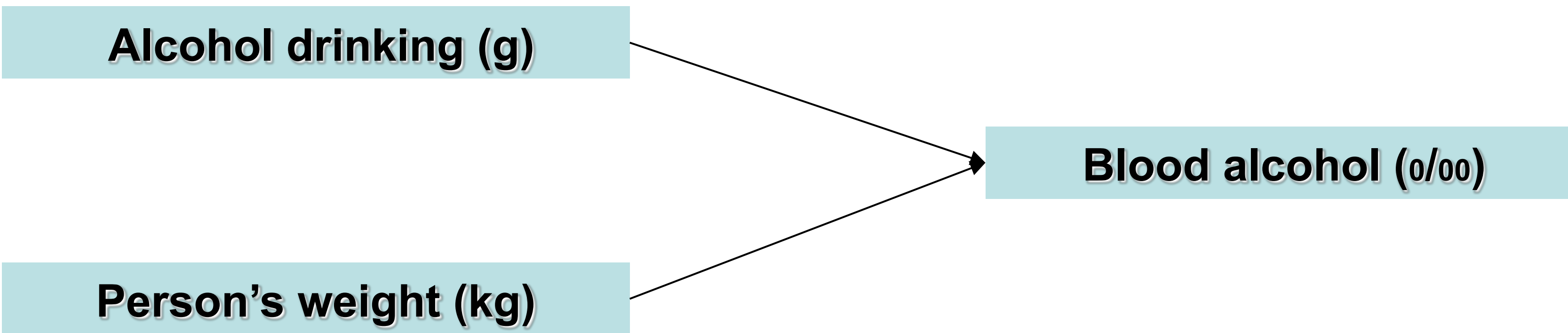
Model construction with Data and Expertise





The plot shows the checking error as ♦ ♦ on the top . The training error appears as * * on the bottom. The checking error decreases up to a certain point in the training, and then it increases. This increase represents the point of model overfitting. `anfis` chooses the model parameters associated with the minimum checking error (just prior to this jump point). This example shows why the checking data option of `anfis` is useful.

Example: Measuring Blood Alcohol Concentration among Males

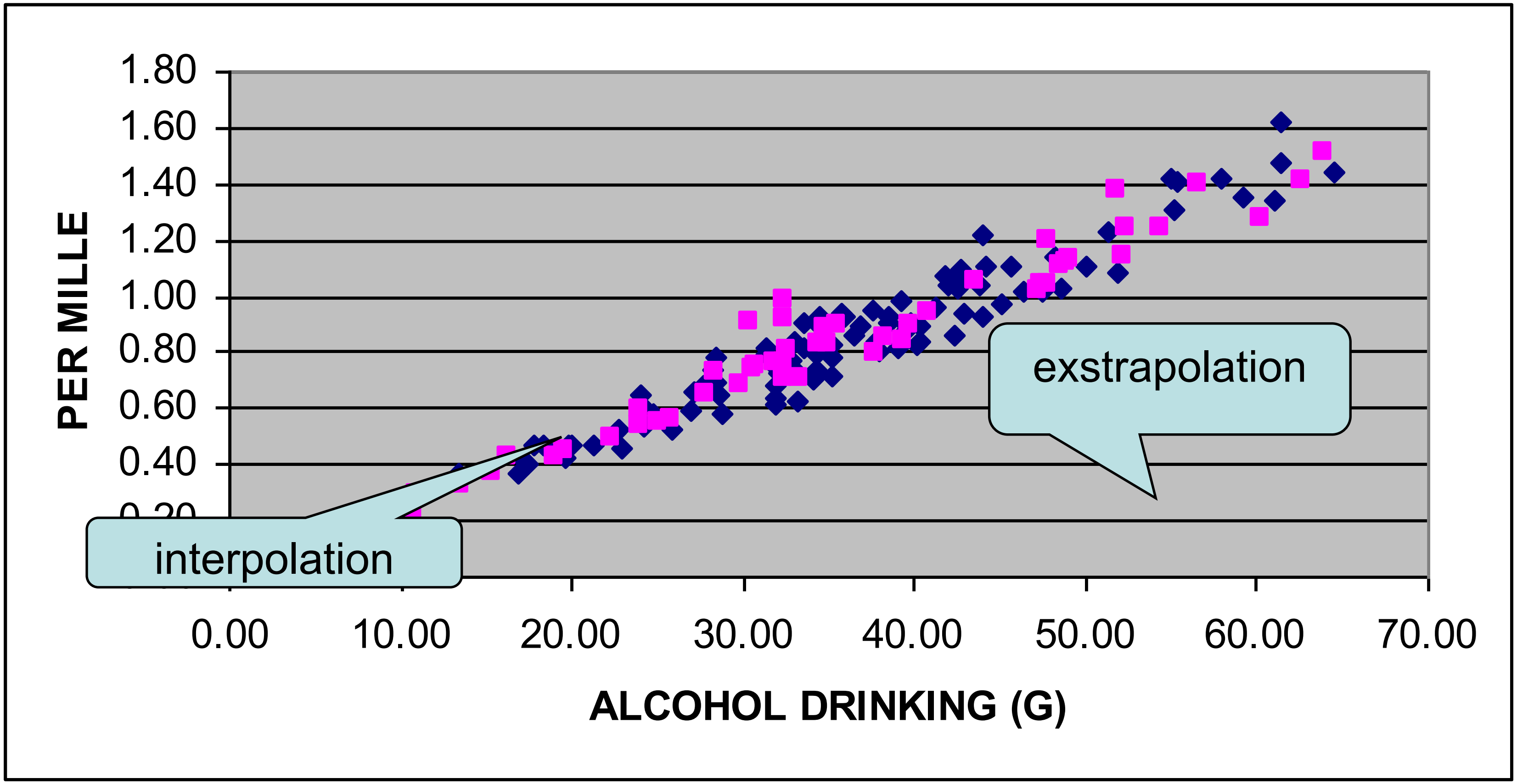


Training data: 100 men

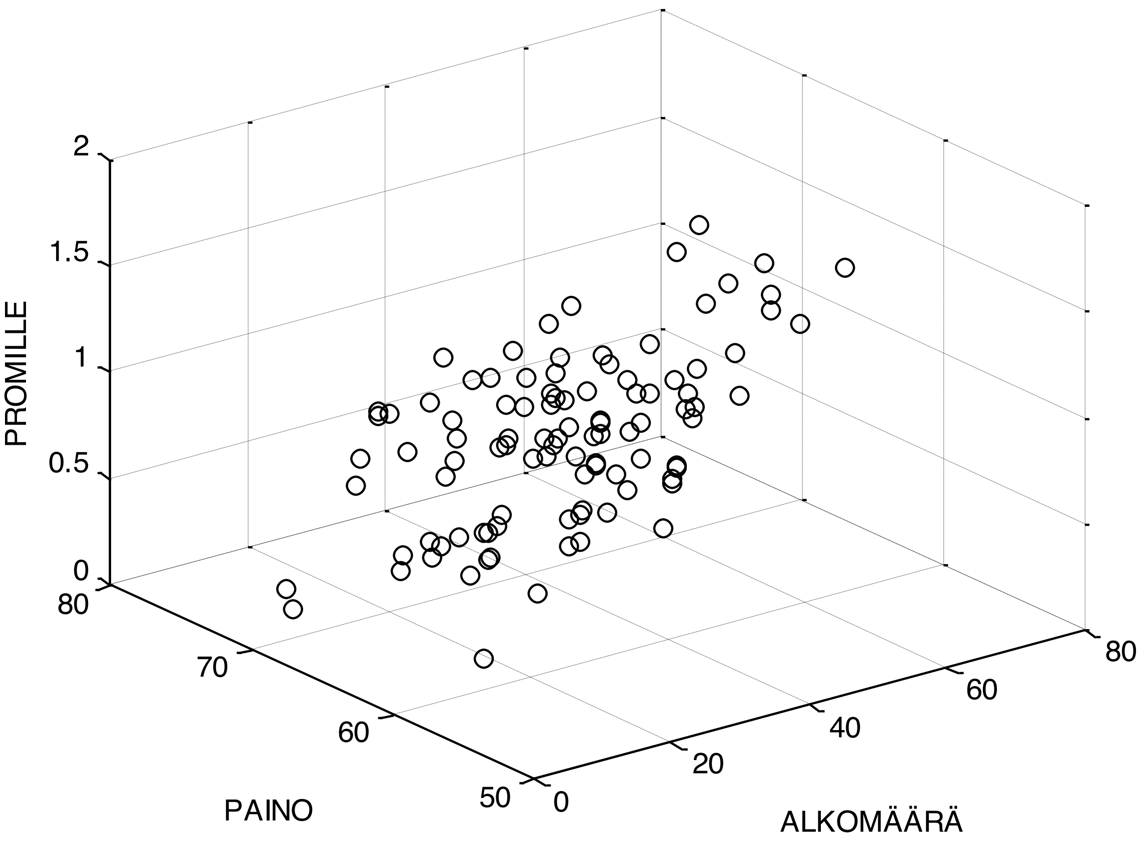
Control data: 50 men

One bottle of beer contains 16 g alcohol

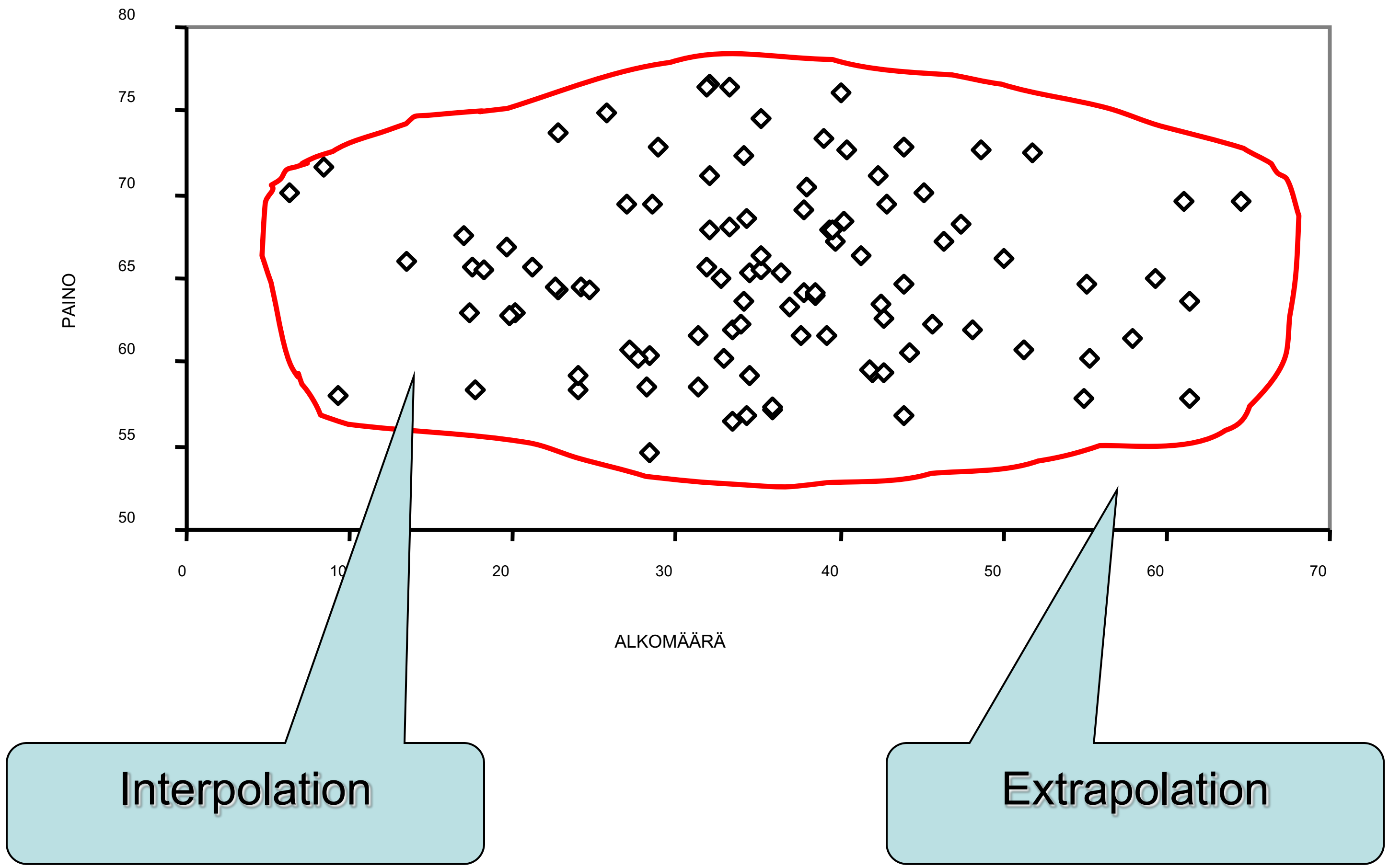
Training and Control Data for One-Input Alcohol Model



Training Data for Two-Input Alcohol Model



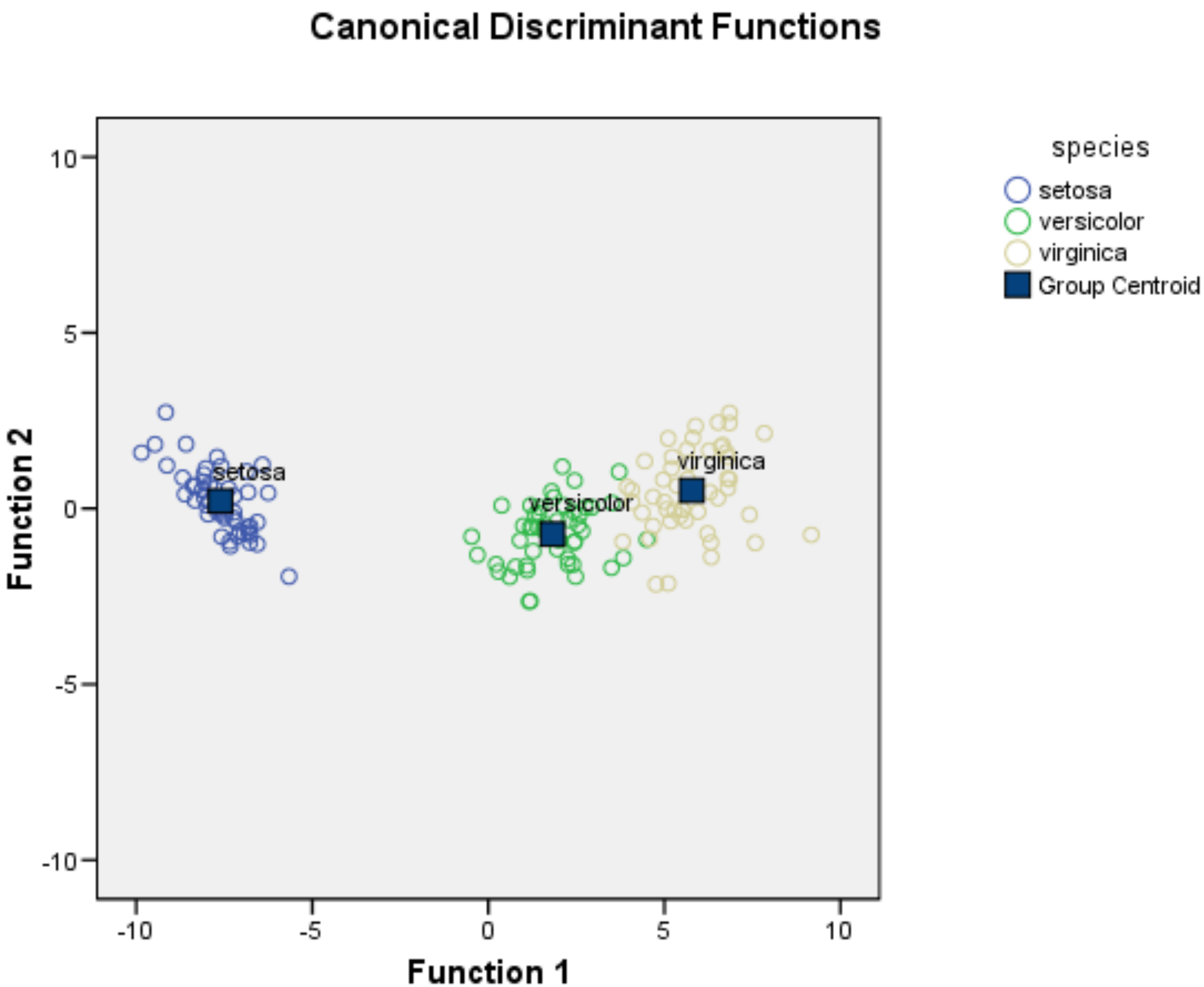
Blood Alcohol: Input Space for Training Data in Two-Inputs Model (Control Data Points Must Locate in Same Area)



Alcohol: Linear Regression Analysis

- Blood alcohol =
  -0.01*drinking + 0.02*weight + 0.82
- Rsquare (training) = 0.99
- Rmse (training) = 0.03

Conventional Discriminant Analysis with Iris Data



Iris Data: Fuzzy Model (in NN: Learning Vector Quantization, LVQ)

Supervised learning

Independent:
- **Sepal length**
- **Sepal width**
- **Petal length**
- **Petal width**

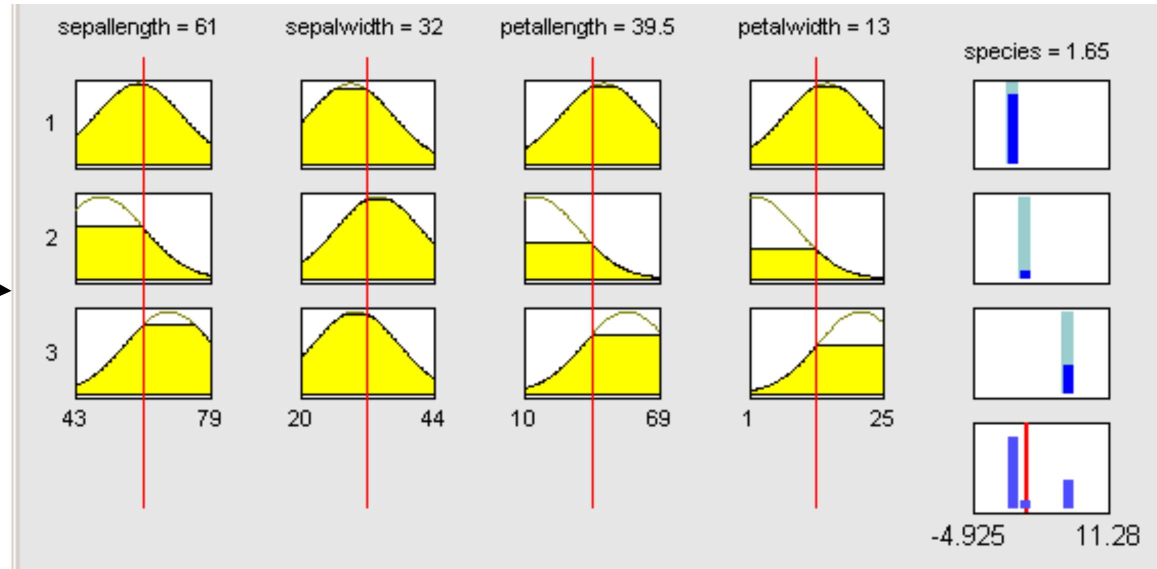
→

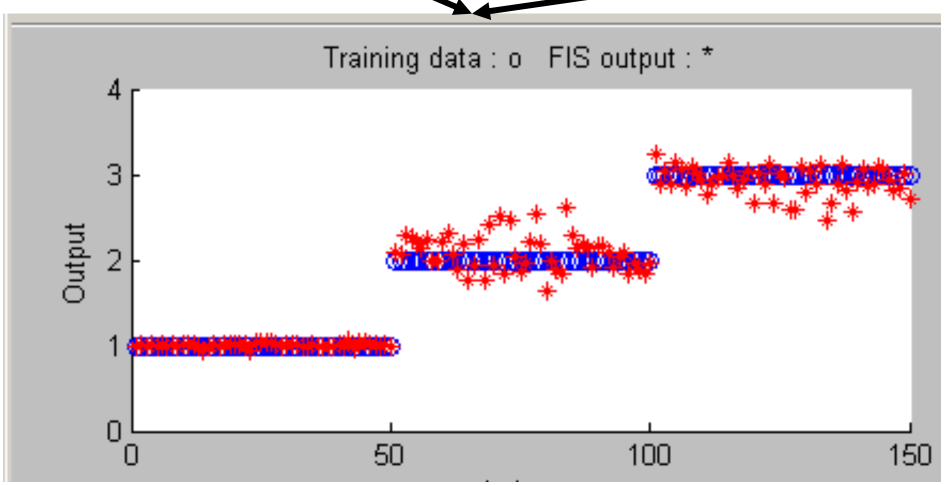
Dependent:
Species (setosa, versicolor, virginica)

Iris Data: Fuzzy Model with Three Rules (Subclust, 0.9)

Initial rules:

SL	SW	PL	PW →	S
60	29	45	15	2
50	34	15	2	1
68	30	55	21	3





Iris Data: Classification Results in Discriminant Analysis, SPSS (top), Fuzzy (bottom)

Classification Results[a]

			Predicted Group Membership			
		species	setosa	versicolor	virginica	Total
Original	Count	setosa	50	0	0	50
		versicolor	0	48	2	50
		virginica	0	1	49	50
	%	setosa	100.0	.0	.0	100.0
		versicolor	.0	96.0	4.0	100.0
		virginica	.0	2.0	98.0	100.0

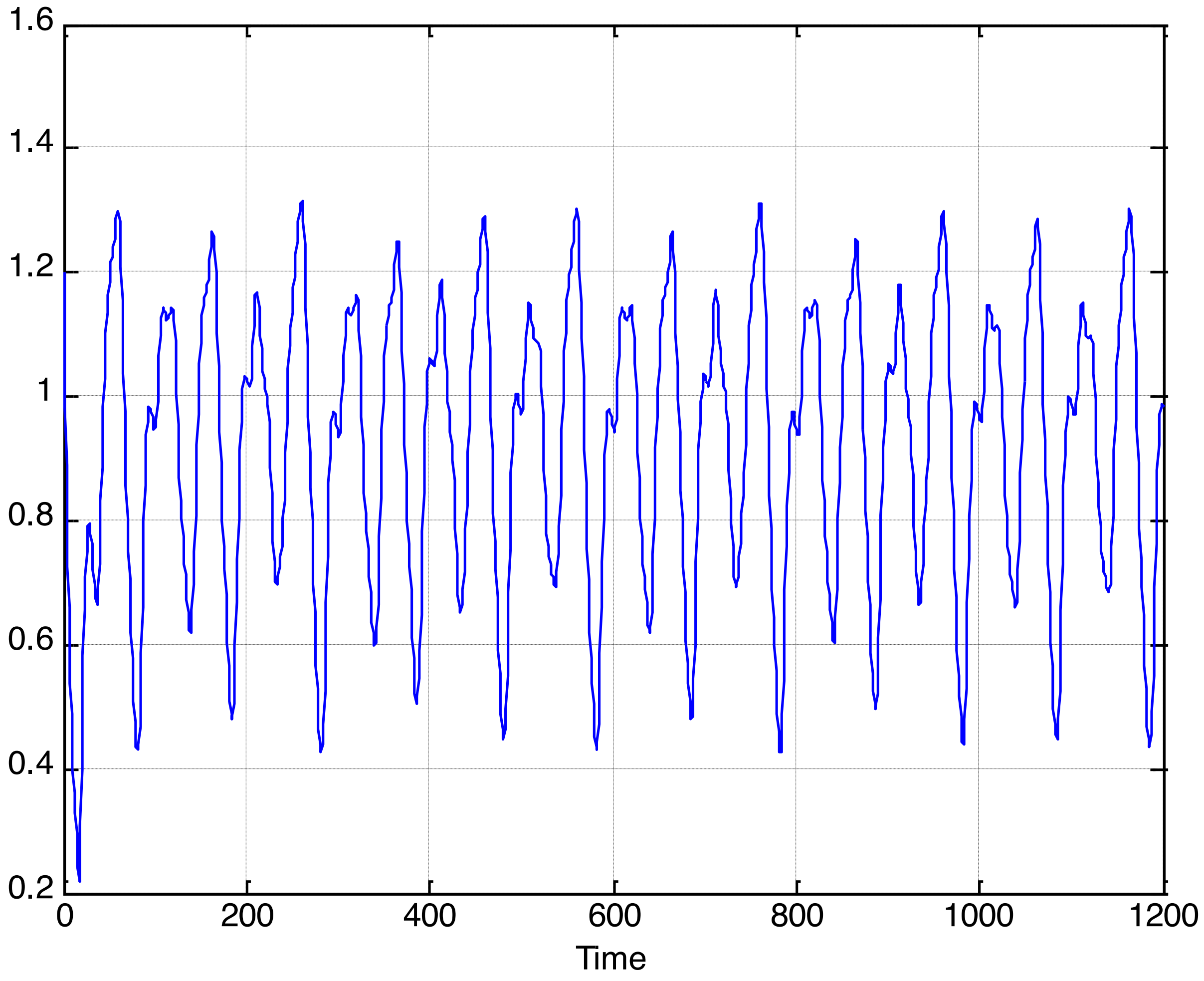
a. 98.0% of original grouped cases correctly classified.

species * predicted Crosstabulation

Count

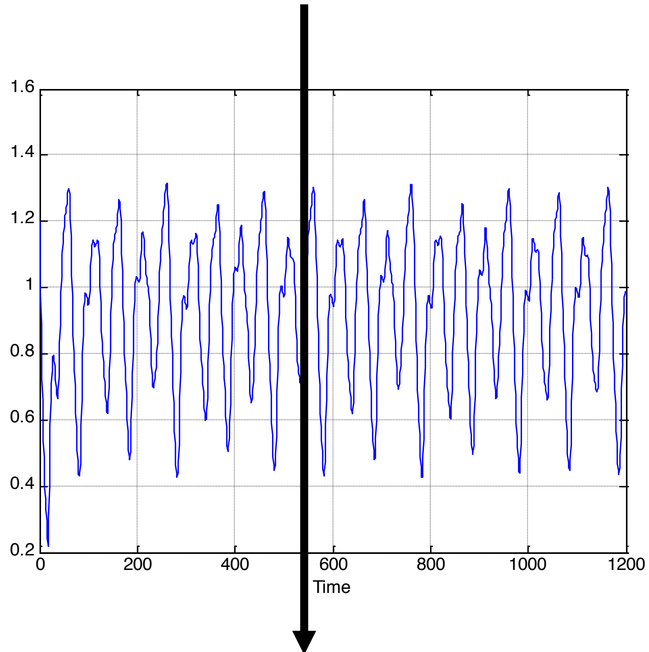
		predicted			
		1.00	2.00	3.00	Total
species	setosa	50	0	0	50
	versicolor	0	47	3	50
	virginica	0	1	49	50
Total		50	48	52	150

Time Series: Mackey-Glass Chaotic Time Series



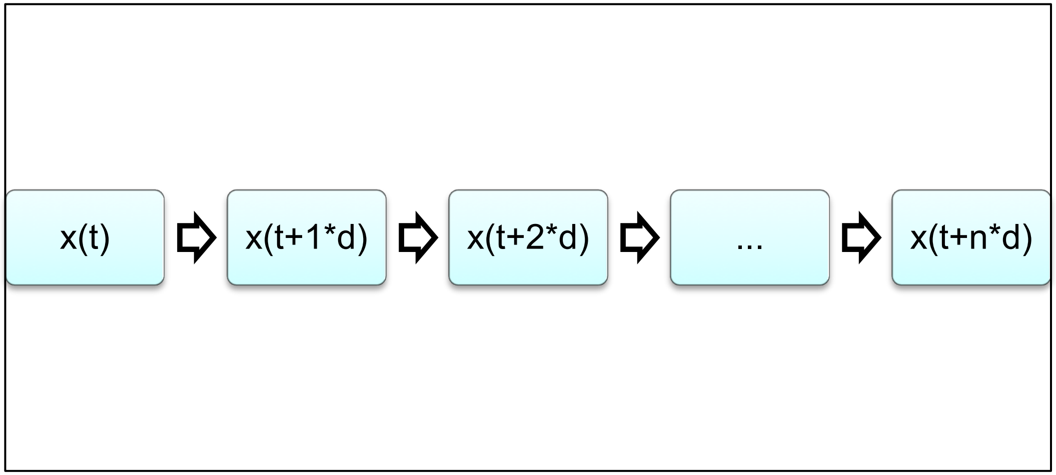
Mackey-Glass Chaotic Time Series

Assign
- starting point,
- step,
- nr. of inputs



training data

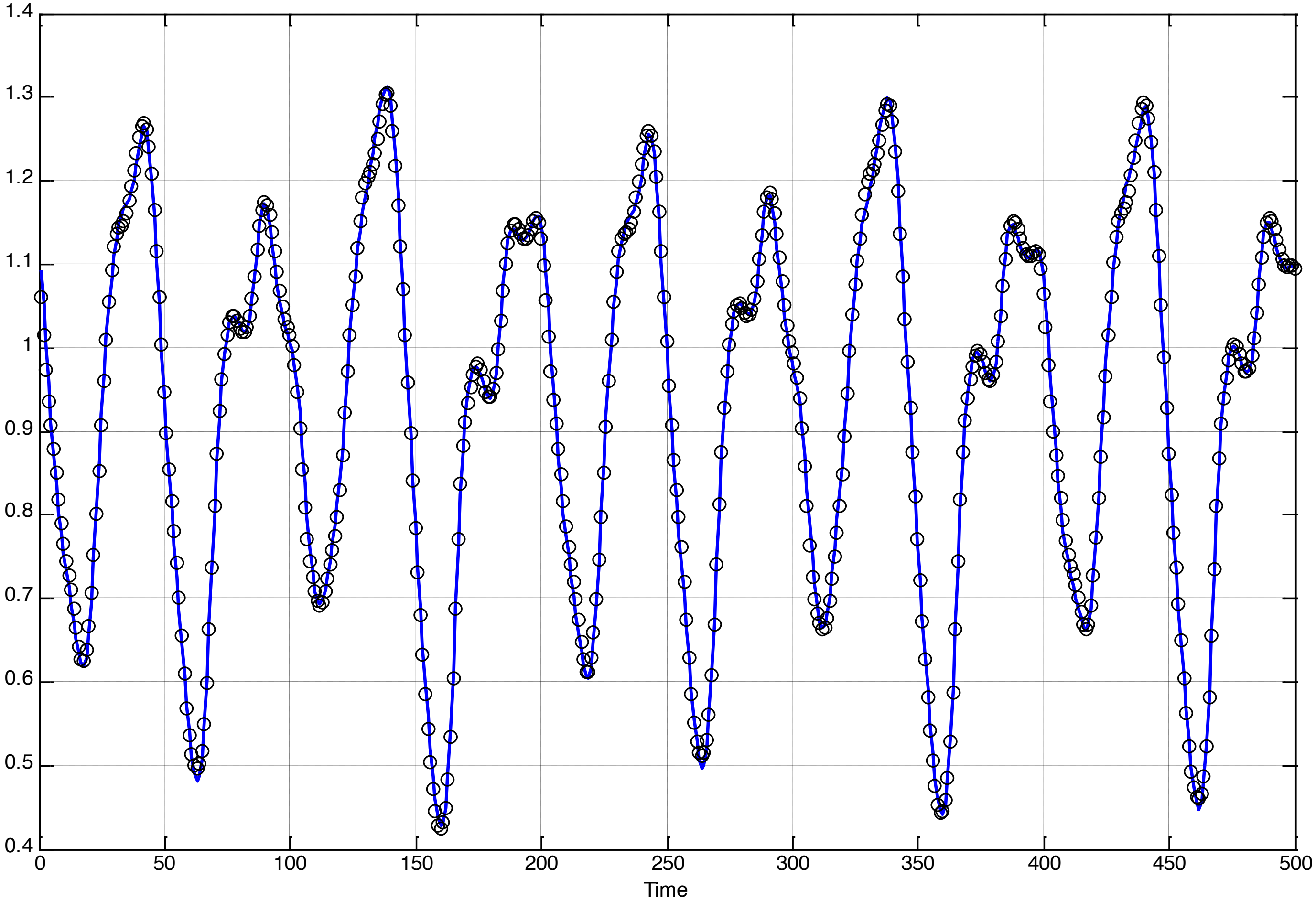
control data



data vector when d=6, n=4:
4 inputs, 1 output

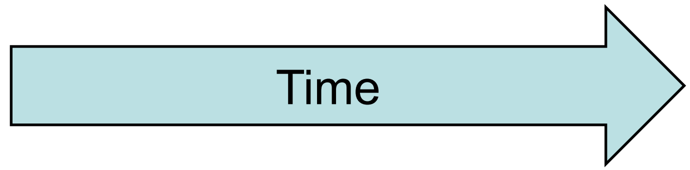
[x(t), x(t+6), x(t+12), x(t+18), x(t+24)]

Mackey-Glass Chaotic Time Series: Desired and Predicted (o) Values in Control Data



Simple Time Series Analysis (Pedrycz)

Original data vector: (x1,x2,x3,x4,x5,x6,x7,x8, ..., xn)



Create matrix which contains in 3 columns (2 inputs, 1 output)
- original values (from x_2),
- differences x_k-x_{k-1},
- original values (from x_3)

Data matrix for fuzzy system (today & today-yesterday => tomorrow):
x2, x2-x1, x3
x3, x3-x2, x4
x4, x4-x3, x5
...

Future Trends: How to Construct Simple Models?

Reduce variables (eg. when intercorrelated)

Factor analysis
Item analysis

Fuzzy clustering of variables

Adjusted rsquare

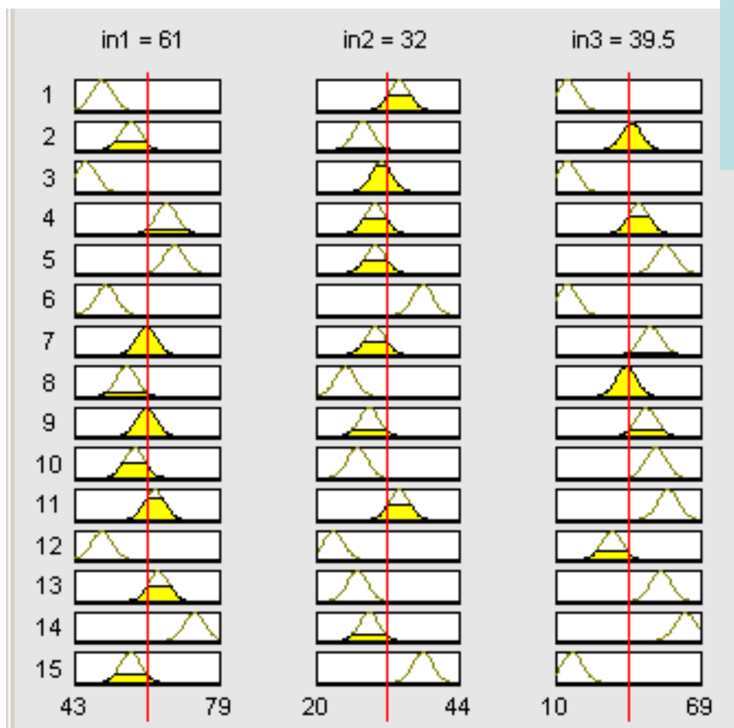
Fuzzy clustering of cases

Fuzzy rules

Reduce clusters (i.e. rules)

Reduce values of variables

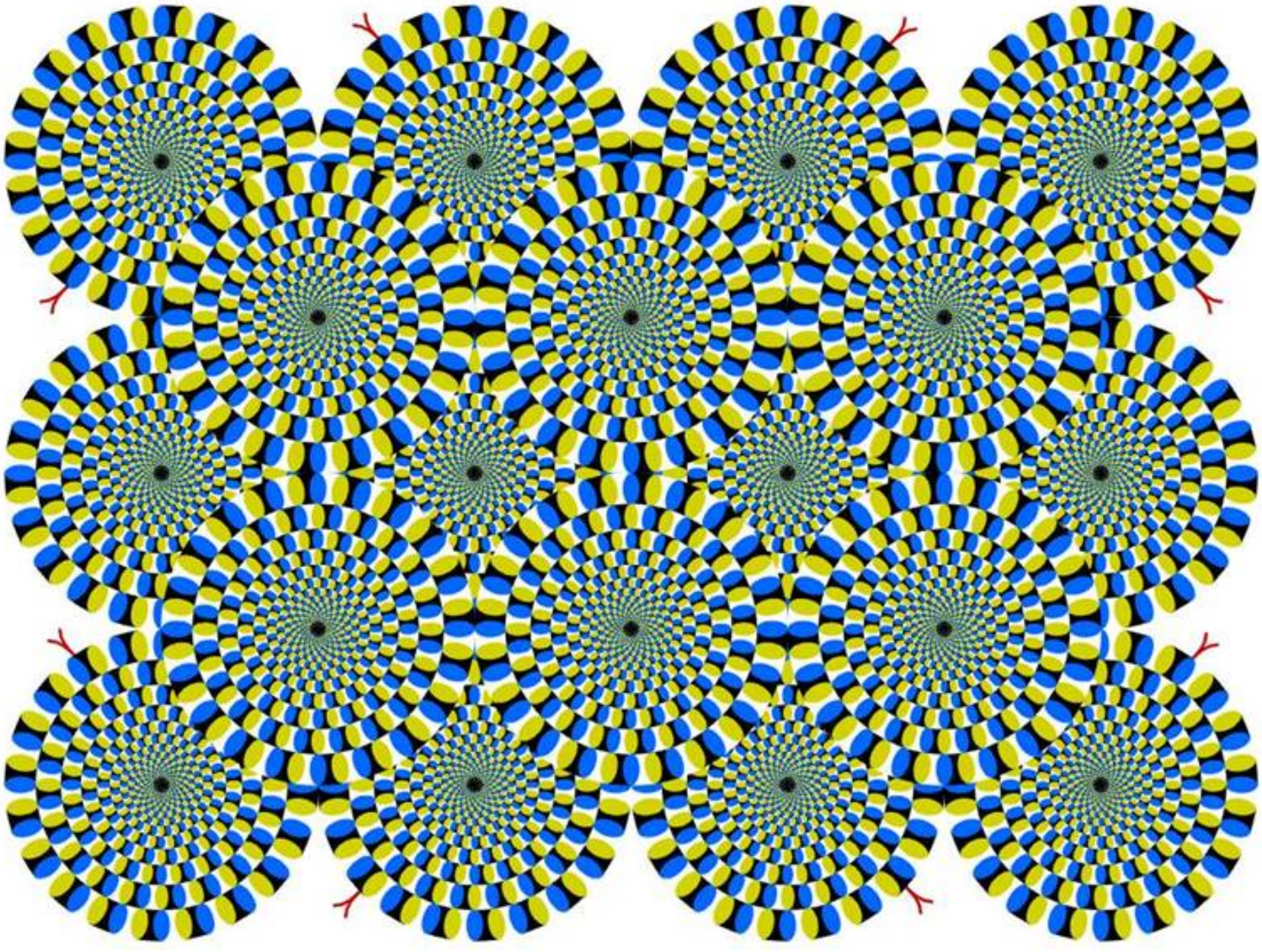
Use archetype values



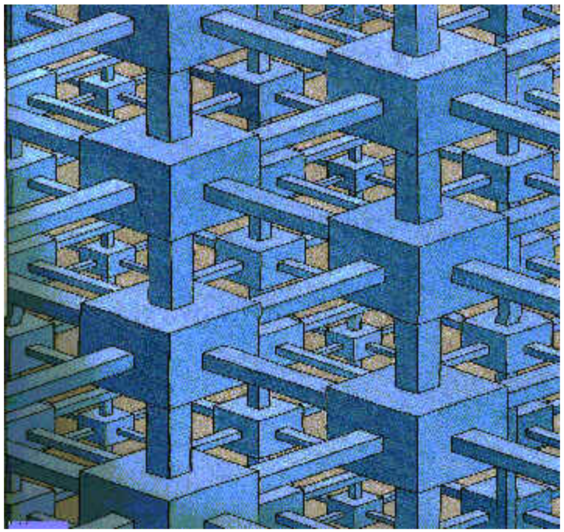
Simple dimension reduction for creating composite variables (Niskanen)

- Cluster analysis (CA), k-means, fcm or subclust.
- Originally CA finds clusters of objects.
- If data matrix transposed, CA can find clusters of variables. Cluster centers will be the composite variables.
- Traditionally: principal component and factor analysis, or multidimensional scaling (in SPSS: proxcal).

"Everything Depends on Everything Else"
Modeling of Complicated Phenomena with Concept and Cognitive Maps



Computer Models of Complicated Phenomena

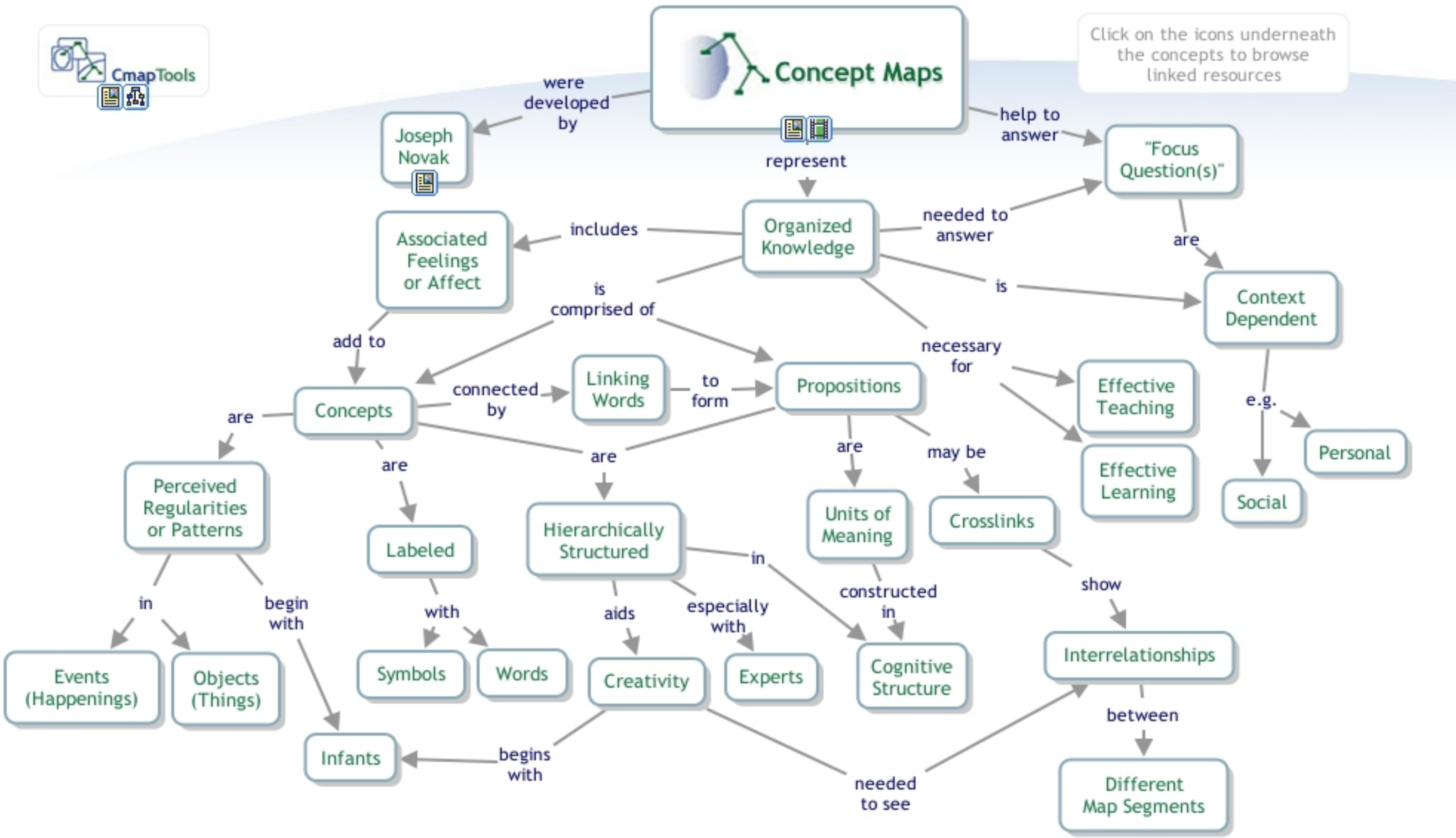


- Networks of variables and their interrelationships.
- CI models can also cope with these phenomena.
- We can use these CI models in both qualitative and quantitative research.
- Typical conventional quantitative models apply path analysis or structural equation modeling (Lisrel, Amos, Mplus).
- We apply concept and cognitive maps in this context.

What Is a Concept Map? (Novak, Åhlberg)

- Concept mapping is a technique for representing knowledge in graphs.
- Knowledge graphs are networks of concepts. Networks consist of nodes (points/vertices) and links (arcs/edges).
- Nodes represent concepts and links represent the relations between concepts. Concepts and links may be divided in categories such as causal or temporal relations.
- Concept mapping can be done for for several purposes:
  - to generate ideas (brain storming, etc.);
  - to design a complex structure (long texts, hypermedia, large web sites, etc.);
  - to communicate complex ideas;
  - to aid learning by explicitly integrating new and old knowledge;
  - to assess understanding or diagnose misunderstanding.

Concept Map about Concept Maps (Cmap Tools™)



Fuzzy Cognitive Maps (FCM, Kosko & al.)

- **Originally Axelrod & al.**
- **FCMs are fuzzy directed graphs with possible feedback.**
- **The nodes are causal concepts.**
- **The nodes have numeric or linguistic interrelationships.**
- **They can model events, actions, values, goals, stories etc.**
- **More applicable than e.g. Bayesian networks because feedback (loops) are also allowed.**
- **We can use these in both qualitative and quantitative research.**
- **Internet applications can be financially very profitable.**

Numerical FCM: Public City Health Model (Lee & al.): Variables (Nodes)

N1: Number of people in a city

N2: Migration into city

N3: Modernization

N5: Sanitation facilities

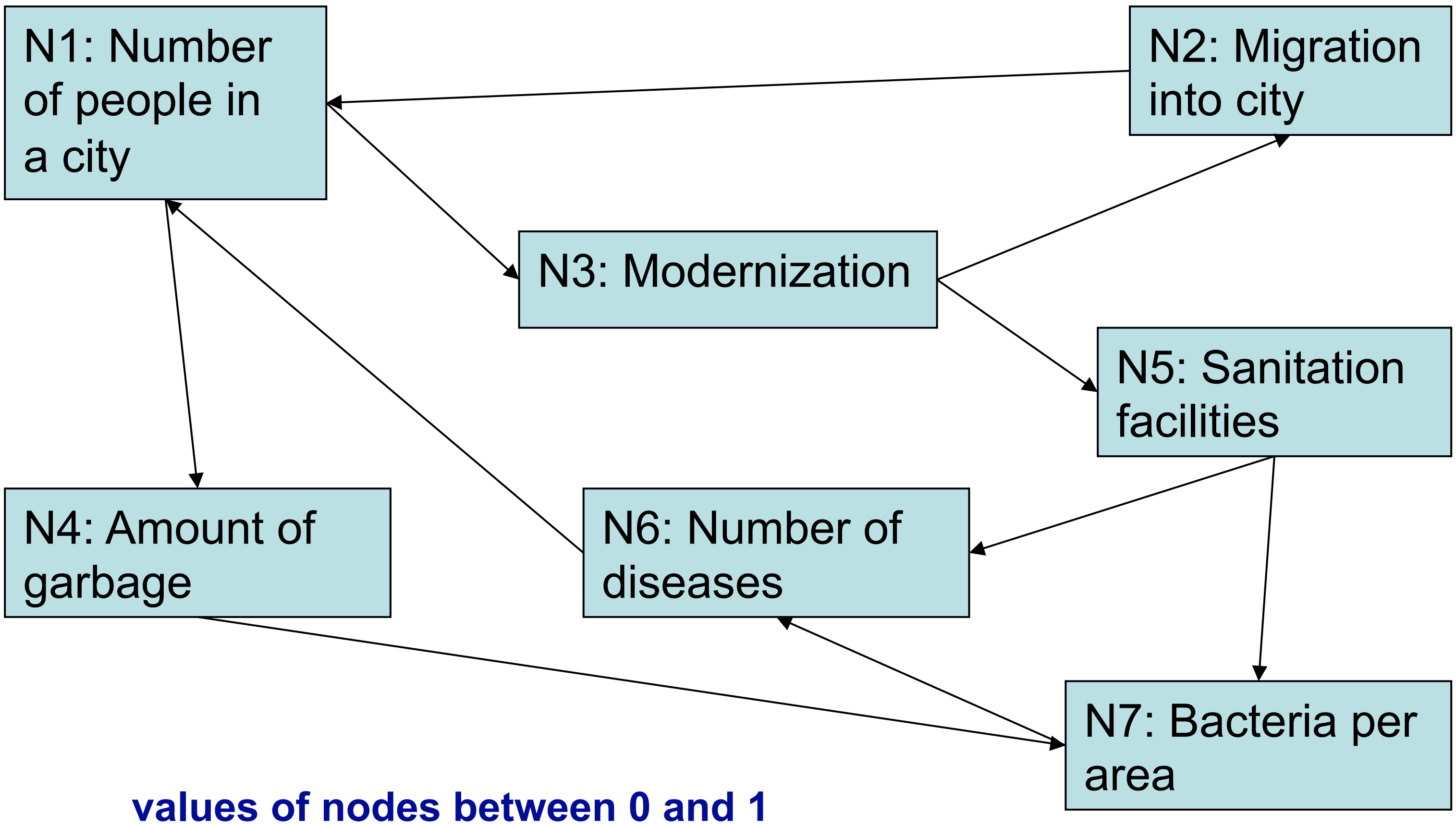
N4: Amount of garbage

N6: Number of diseases

N7: Bacteria per area

values of nodes usually between 0 and 1

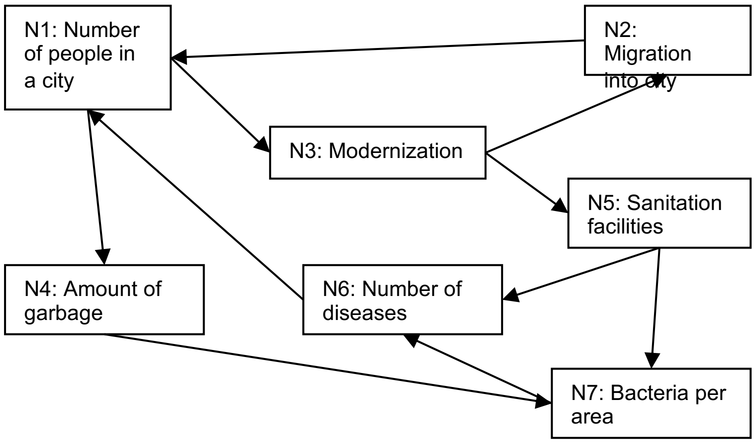
Numerical FCM: Public City Health Model (Lee & al.): drivers and targets

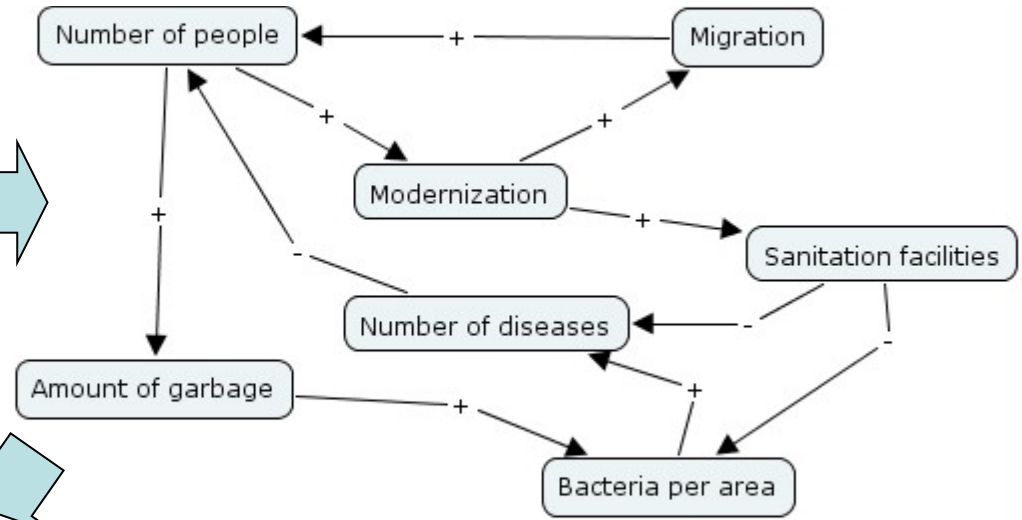


values of nodes between 0 and 1

intensity of (monotonic) relationship usually between -1 and 1

Numerical FCM: Public City Health Model





targets → drivers ↓	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7
Node 1	0	0	0.6	0.9	0	0	0
Node 2	0.5	0	0	0	0	0	0
Node 3	0	0.6	0	0	0.8	0	0
Node 4	0	0	0	0	0	0	1
Node 5	0	0	0	0	0	-0.8	-0.9
Node 6	-0.3	0	0	0	0	0	0
Node 7	0	0	0	0	0	0.8	0

Connection matrix: Intensities of relationship (between -1 and 1)

Numerical FCM: Public City Health Model in Iteration: matrix product of inputs and connection matrix

Nodes at t

Nodes at t+1

input

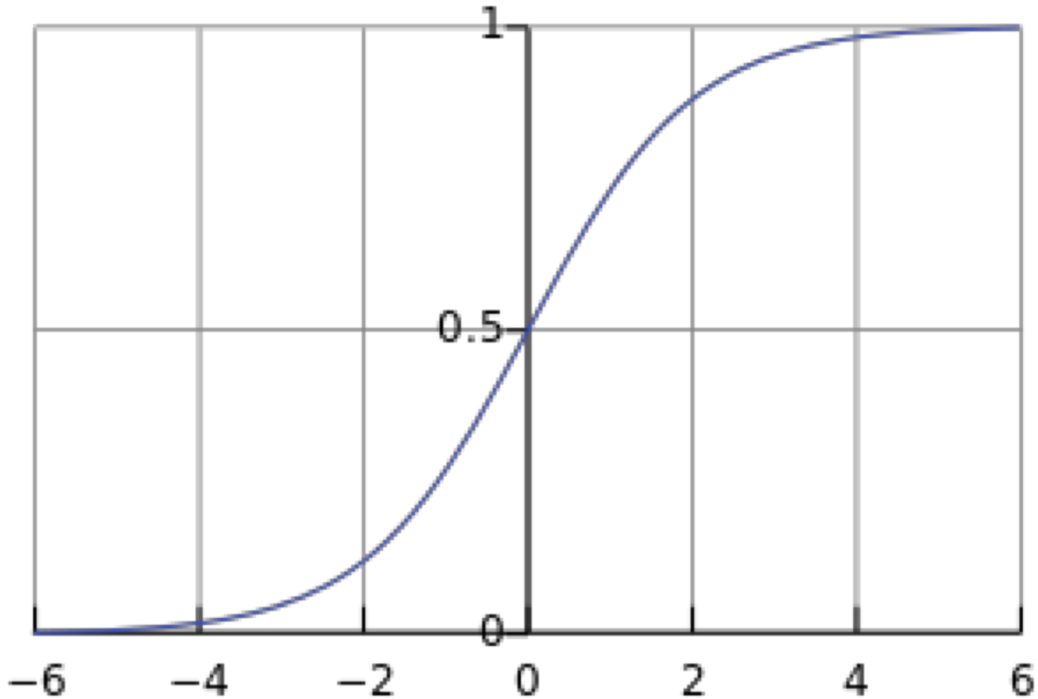
	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7
Node 1	0	0	0.6	0.9	0	0	0
Node 2	0.5	0	0	0	0	0	0
Node 3	0	0.6	0	0	0.8	0	0
Node 4	0	0	0	0	0	0	1
Node 5	0	0	0	0	0	-0.8	-0.9
Node 6	-0.3	0	0	0	0	0	0
Node 7	0	0	0	0	0	0.8	0

output

transformation (squashing) into interval 0 to 1

Intensity of relationship (between -1 and 1)

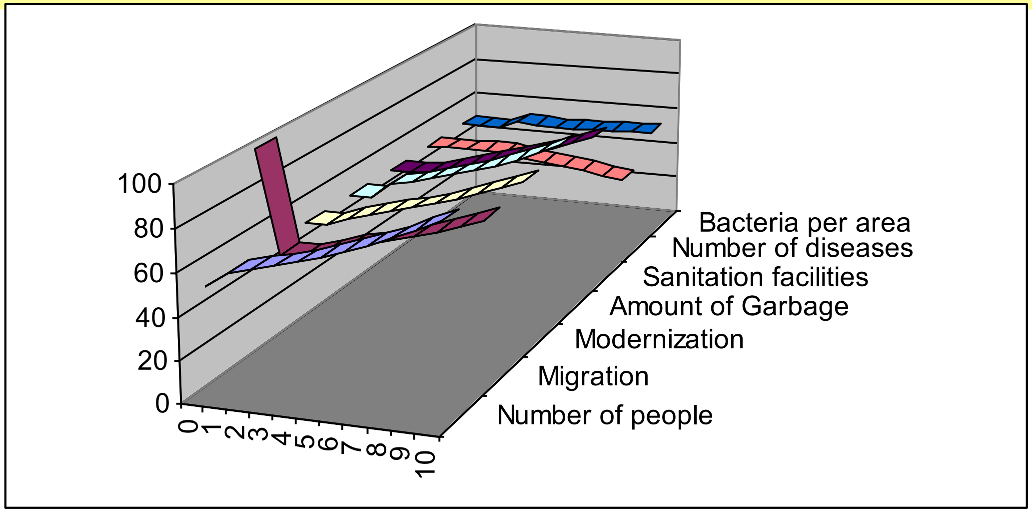
Transformation function
Y=1/(1+exp(-lambda*X))
1 ≤ lambda ≤ 5



X = matrix product

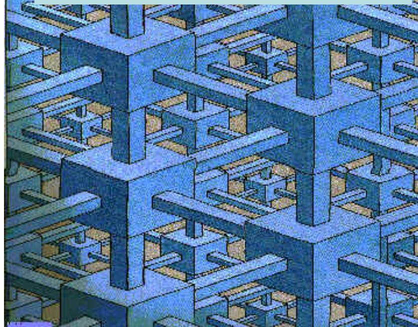
FCM: Public City Health History Data after Iterations

Time	Number of people	Migr ation	Moderniz ation	Amount of Garbage	Sanitation facilities	Number of diseases	Bacteria per area
0	0.90	0.20	0.20	0.10	0.20	0.20	0.20
1	0.55	0.65	0.94	0.98	0.69	0.50	0.39
2	0.70	0.94	0.84	0.92	0.98	0.23	0.79
3	0.88	0.93	0.89	0.96	0.97	0.32	0.44
4	0.86	0.94	0.93	0.98	0.97	0.11	0.49
5	0.90	0.94	0.93	0.98	0.98	0.13	0.51
6	0.90	0.94	0.94	0.98	0.98	0.13	0.50
7	0.90	0.94	0.94	0.98	0.98	0.13	0.51
8	0.90	0.94	0.94	0.98	0.98	0.13	0.51
9	0.90	0.94	0.94	0.98	0.98	0.13	0.51
10	0.90	0.94	0.94	0.98	0.98	0.13	0.51



FCM construction with history data

Real world



Human expertise

NN, Evolutionary computing & History data

TIME

Node 1	Node 2	Node 3	Node3
2	3.5	2.1	1
5.1	10	4.7	9.5
8.2	8.1	2.1	3.8
1.4	6.9	9.3	7.7
5.2	6.3	0.4	0

FCM model

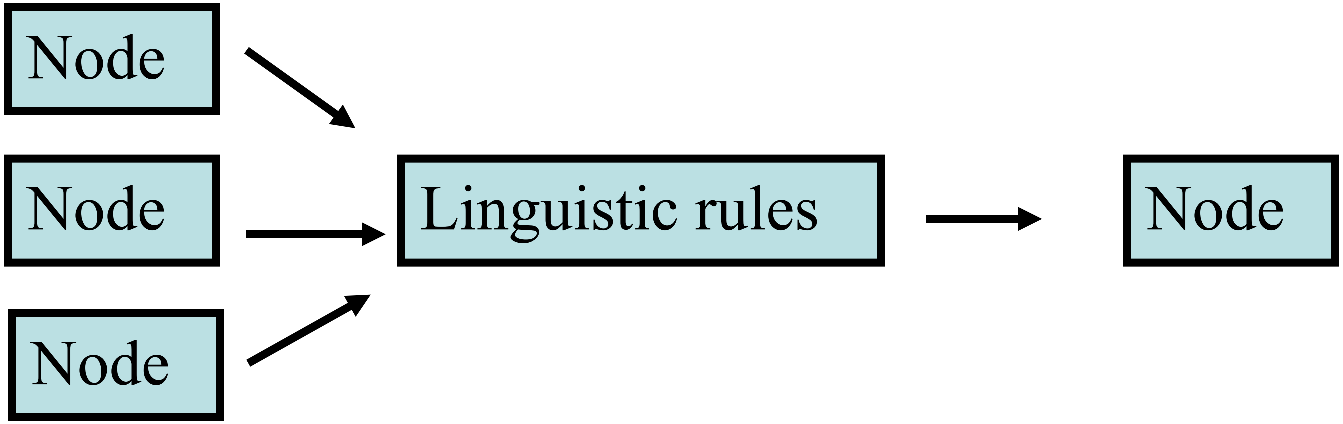
From history data to connection matrix with evolutionary computing

Time	Number of people	Migr ation	Moderniz ation	Amount of Garbage	Sanitation facilities	Number of diseases	Bacteria per area
0	0.90	0.20	0.20	0.10	0.20	0.20	0.20
1	0.55	0.65	0.94	0.98	0.69	0.50	0.39
2	0.70	0.94	0.84	0.92	0.98	0.23	0.79
3	0.88	0.93	0.89	0.96	0.97	0.32	0.44
4	0.86	0.94	0.93	0.98	0.97	0.11	0.49
5	0.90	0.94	0.93	0.98	0.98	0.13	0.51
6	0.90	0.94	0.94	0.98	0.98	0.13	0.50
7	0.90	0.94	0.94	0.98	0.98	0.13	0.51
8	0.90	0.94	0.94	0.98	0.98	0.13	0.51
9	0.90	0.94					
10	0.90	0.94					

	Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7
Node 1	0	0	0.6	0.9	0	0	0
Node 2	0.5	0	0	0	0	0	0
Node 3	0	0.6	0	0	0.8	0	0
Node 4	0	0	0	0	0	0	1
Node 5	0	0	0	0	0	-0.8	-0.9
Node 6	-0.3	0	0	0	0	0	0
Node 7	0	0	0	0	0	0.8	0

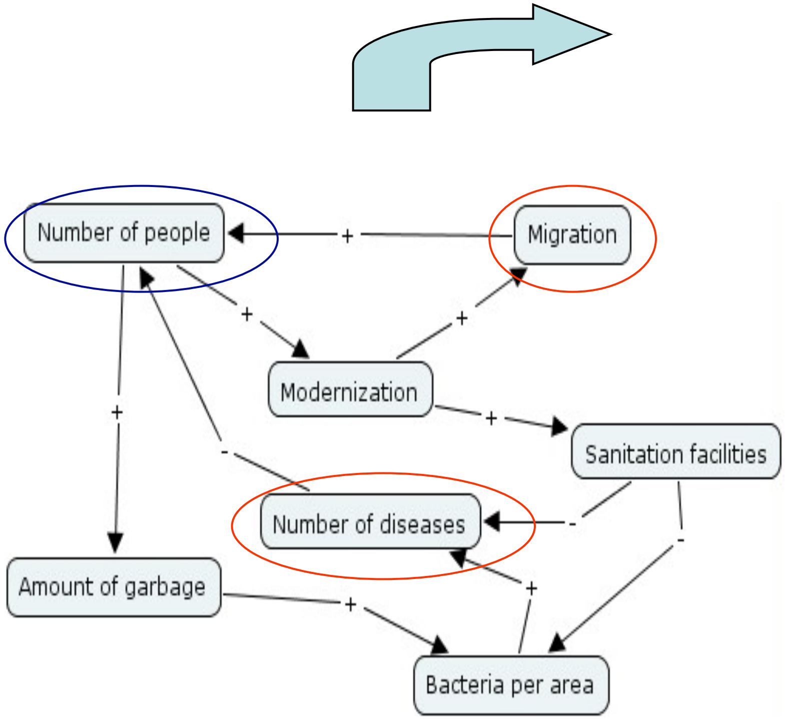
intensities are parameters in optimization

Linguistic FCM



- We can use linguistic inputs and outputs.
- The relationships between nodes are assigned with linguistic rules.

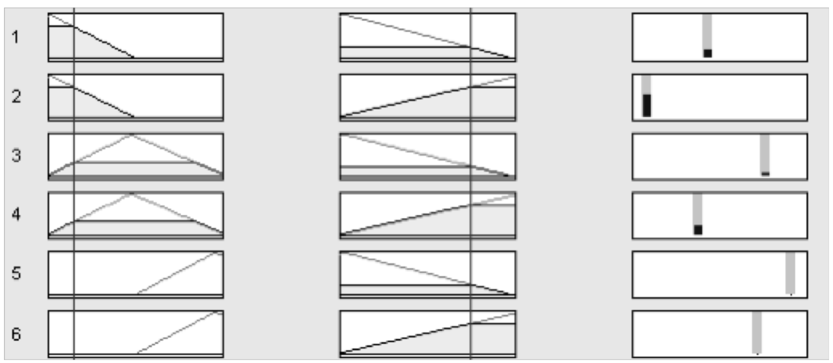
Linguistic Public City-Health Model



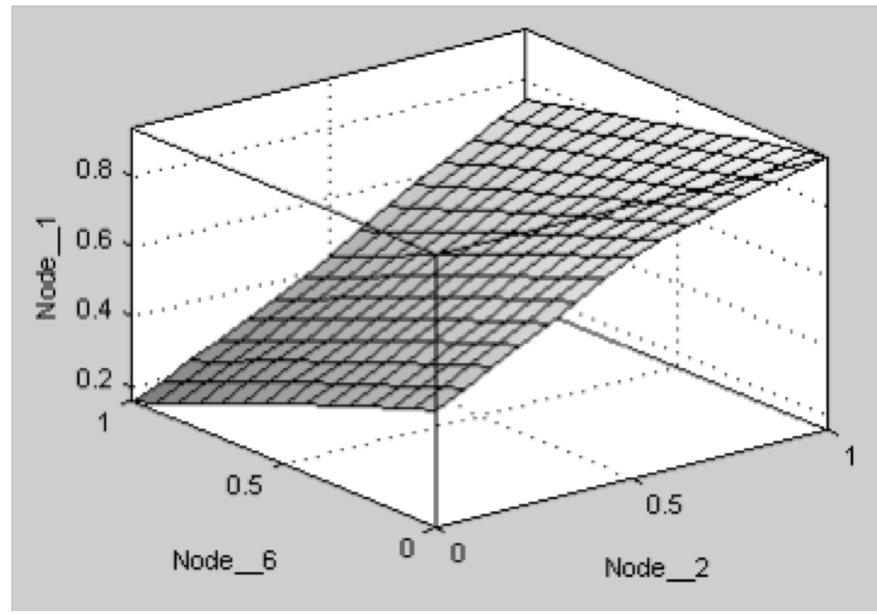
Fuzzy rules:

1. If the migration into city is low and the number of diseases is small, then the increase in the number of people in the city is medium.
2. If the migration into city is low and the number of diseases is large, then the increase in the number of people in the city is fairly small.
3. If the migration into city is average and the number of diseases is small, then the increase in the number of people in the city is fairly large.
4. If the migration into city is average and the number of diseases is large, then the increase in the number of people in the city is medium.
5. If the migration into city is high and the number of diseases is small, then the increase in the number of people in the city is large.
6. If the migration into city is high and the number of diseases is large, then the increase in the number of people in the city is fairly large.

From Fuzzy Rules to an FCM Model



Node 2 **Node 6** **Node 1**



Simple FCM control model (Papageorgiou, Stach, Kurgan, Pedrycz)

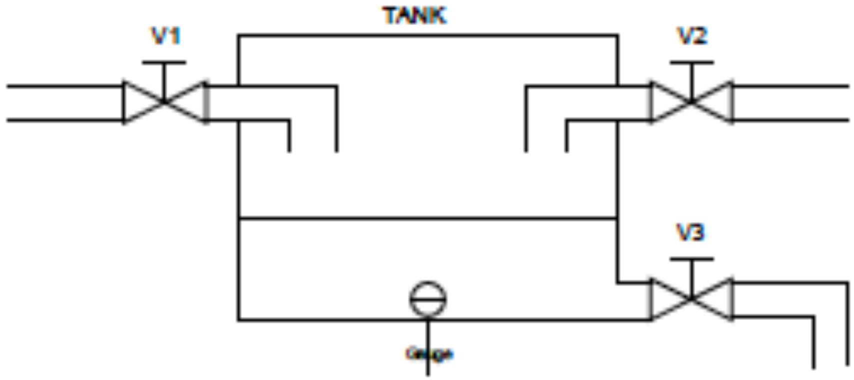
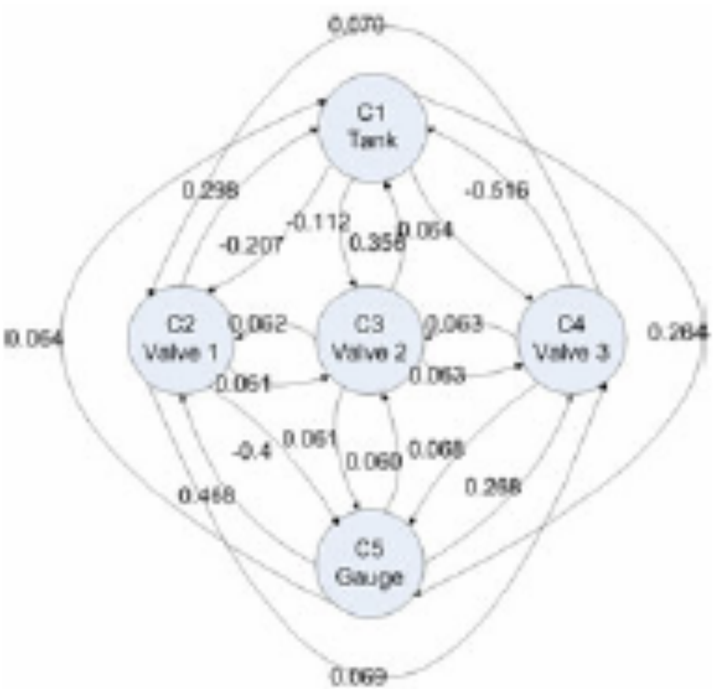


Fig. 1 Process control problem

Figure 1 shows an example process control problem which was discussed by Papageorgiou et al. (Papageorgiou et al. 2003). Two valves, *valve 1* (*V1*) and *valve 2* (*V2*), supply two different liquids into the tank. The liquids are mixed and a chemical reaction takes place. The control objective is to maintain the desired level of liquid and its specific gravity. *Valve 3* (*V3*) is used to drain liquid from the tank.



	C1	C2	C3	C4	C5
C1	0	-0.207	-0.112	0.064	0.264
C2	0.298	0	0.061	0.069	0.067
C3	0.356	0.062	0	0.063	0.061
C4	-0.516	0.070	0.063	0	0.068
C5	0.064	0.468	0.060	0.268	0

Fig. 2 FCM graph along with its connection matrix for the process control problem

Goals:
0.68 ≤ Tank (C1) ≤ 0.70
0.74 ≤ Gauge (C5) ≤ 0.80

***Introducing Fuzzy Cognitive Maps for decision making in precision agriculture*.**

The FCM model developed consists of nodes which describe soil properties and cotton yield and of the weighted relationships between these nodes. The nodes of the FCM model represent the main factors influencing cotton crop production i.e. essential soil properties such as texture, pH, OM, K, and P.
The proposed FCM model addresses the problem of crop development and spatial variability of cotton yield, taking into consideration the spatial distribution of all the important factors affecting yield.

Available from:
https://www.researchgate.net/publication/237007624_Introducing_Fuzzy_Cognitive_Maps_for_decision_making_in_precision_agriculture

Table 1. Soil parameters that affect cotton yield.

Concept	Description
C1: EC	Soil shallow electrical conductivity Veris (mS/m)
C2: Mg	Magnesium (ppm)
C3: Ca	The measured calcium in the soil in depth 0–30 cm (ppm)
C4: Na	The measured Na (Sodium) in the soil in depth 0–30 cm (ppm)
C5: K	The measured Potassium in the soil in depth 0–30 cm (ppm)
C6: P	The measured Phosphorus in the soil in depth 0–30 cm (ppm)
C7: N	The measured NO_3 in the soil profile of 0–30 cm (ppm)
C8: OM	The percent organic matter content in soil profile in depth 0–30 cm
C9: Ph	The pH of the soil in depth 0–30 cm
C10: S	The percent of the sand in the soil samples in depth 0–30 cm
C11: Cl	The percent of the clay in samples in depth 0–30 cm
C12: Y	Seed cotton yield from 1st picking measured by yield monitor ($t\ ha^{-1}$)

Table 1. Concepts of the FCM: Type of values.

C1: ShallowEC (mS/m)	C2: Mg (ppm)	C3: Ca (ppm)	C4: Na (ppm)
Five Fuzzy	*Five Fuzzy*	*Five Fuzzy*	*Five Fuzzy*
0 – 10 Very Low	< 60 Very Low	< 400 Very Low	< 25 Very Low
10 – 20 Low	60 – 180 Low	400 – 1000 Low	25 – 70 Low
20 – 30 Medium	181 – 360 Medium	1001 – 2000 Medium	71 - 160 Medium
30 – 40 High	361 - 950 High	2001 – 4000 High	161 – 460 High
> 40 Very High	> 950 Very High	> 4000 Very High	> 460 Very High
C5: K (ppm)	C6: P (ppm)	C7: N (ppm)	C8: OM (ppm)
Five Fuzzy	*Five Fuzzy*	*Five Fuzzy*	*Three Fuzzy*
< 40 Very Low	< 5 Very Low	< 3 Very Low	< 1.0 Low
40 – 120 Low	5 – 15 Low	3 – 10 Low	1.0 – 2.0 Medium
121 – 240 Medium	16 – 25 Medium	11 – 20 Medium	> 2.0 High
241 – 470 High	26 – 45 High	21 – 40 High	
> 470 Very High	> 45 Very High	> 40 Very High	
C9: Ph	C10: Sand %	C11: Clay %	C12: Yield (tons/ha)
Seven Fuzzy	Four Fuzzy	Three Fuzzy	Three Fuzzy
<4.5 Very Low	< 20 Low	< 15 Low	< 2.5 Low
4.6 – 5.5 Low	20 – 70 Medium	15 – 37 Medium	2.5 - 3.5 Medium
5.6 – 6.5 Slightly Low	71 – 80 High	Texture	>3.5 High
6.6 – 7.5 Neutral	> 80 Very High	> 37 High	
7.6 – 8.5 Slightly High			
8.6 - 9.5 High			
> 9.5 Very High			

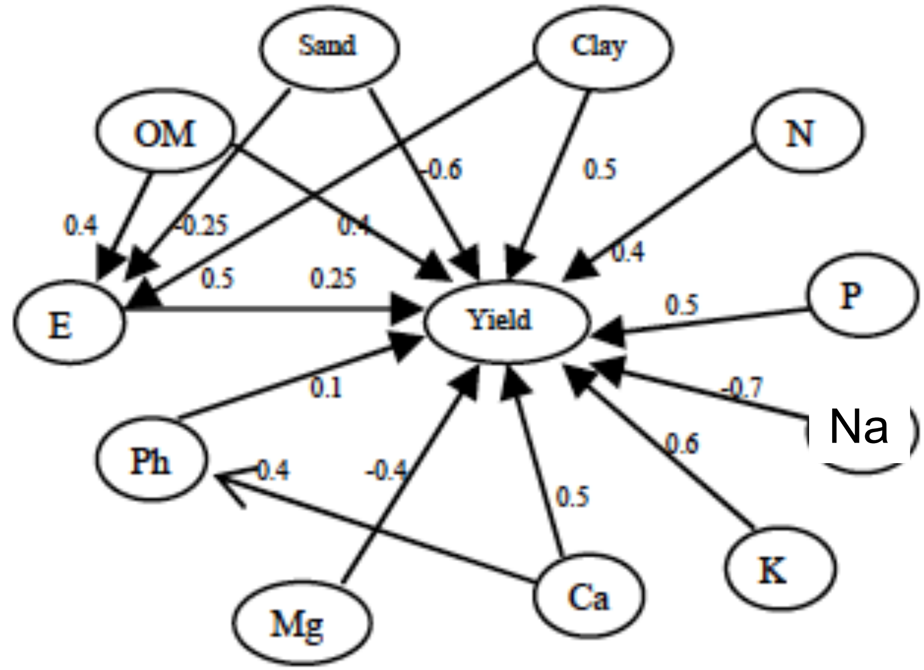


Figure 4. The FCM model for describing the final cotton yield.

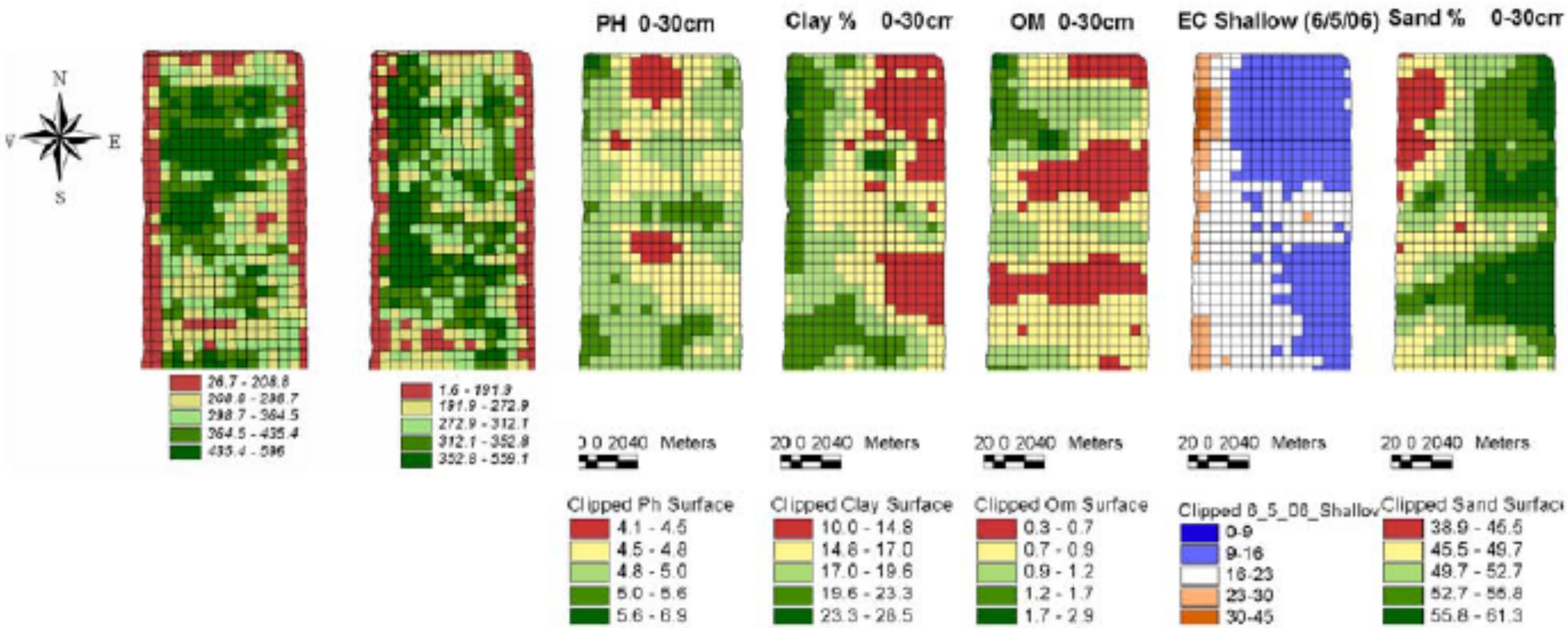


Figure 5. Two of the yield (years 2001 and 2003) and some of the soil properties maps.

Mannerheimintie	Mäkelänkatu2	Kallio2	Vartiokylä	Leppävaara4	Tikkurila3	Luukki	Lohja	Kerava(M5)	Lentoasema(M4)	Mechelininkatu(M1)	Olari(M2)	Rekola(M3)
84	72	108	79	111	164	39	51	74	65	35	81	118
86	74	77	62	82	83	52	32	64	60	32	54	71
64	50	53	42	57	51	47	29	51	50	35	46	53
59	40	48	43	50	53	32	31	39	44	56	42	38
53	35	43	39	56	52	25	31	30	41	49	45	44
42	31	40	36	51	47	24	30	26	43	43	36	33
41	30	41	38	46	50	18	28	19	46	51	38	44
33	33	40	37	38	46	25	26	30	44	30	35	43
20	30	35	33	44	35	25	22	32	34	19	38	26
36	39	36	31	37	44	26	21	30	42	31	37	37
36	45	38	31	48	50	22	29	32	40	37	38	38
36	34	41	31	38	52	16	29	35	43	43	33	47
28	40	33	33	40	45	17	25	24	45	50	33	51

Koczy & al.
Waste management

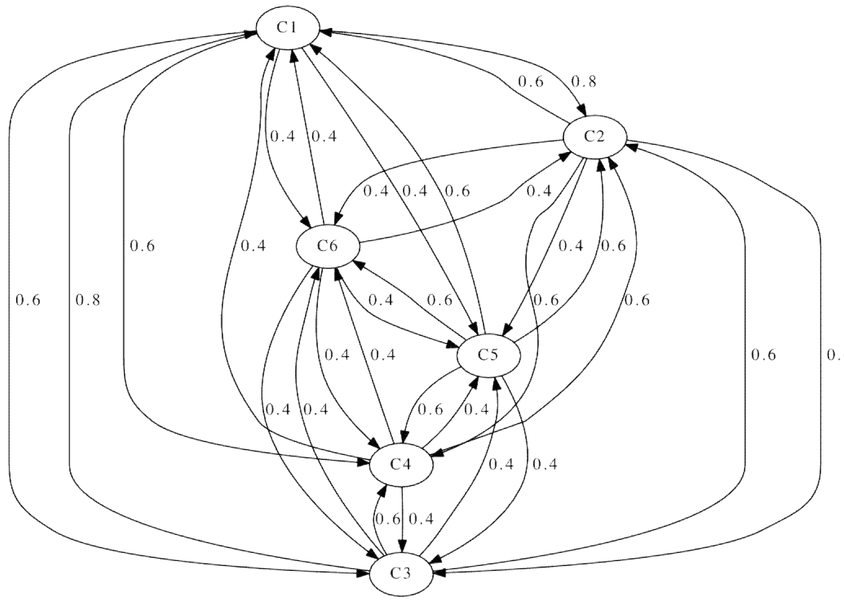


Fig. 6. The initial Fuzzy Cognitive Maps

The factors in the matrix are represented as follows:

- C1: technical factor (collection, transport, treatment methods, etc.)
- C2: environmental factor (emission of pollution, depletion of resources, human toxicity, etc.)
- C3: economic factor (subsidies, efficiency at system/subsystem level, economic sound and continuous operation, coverage of all aftercase expenses, etc.)
- C4: social factor (involving local need and requirements, minimizing public health risk, providing employment, etc.)
- C5: legal factor (EU packaging directive, EU landfill directive, waste hierarchy, national, regional and local regulations)
- C6: institutional factor (involvement of stakeholders, existence of feedback mechanisms of citizens, organisational structure, etc.)

FCM (Carlsson & Fuller)

Strategic Management is defined as a system of action programs which form sustainable competitive advantages for a corporation, its divisions and its business units in a strategic planning period. A research team of the IAMSR institute developed a support system for strategic management, called the *Woodstrat*, in two major Finnish forest industry corporations in 1992-96. The system is modular and is built around the actual business logic of strategic management in the two corporations, i.e. the main modules cover the

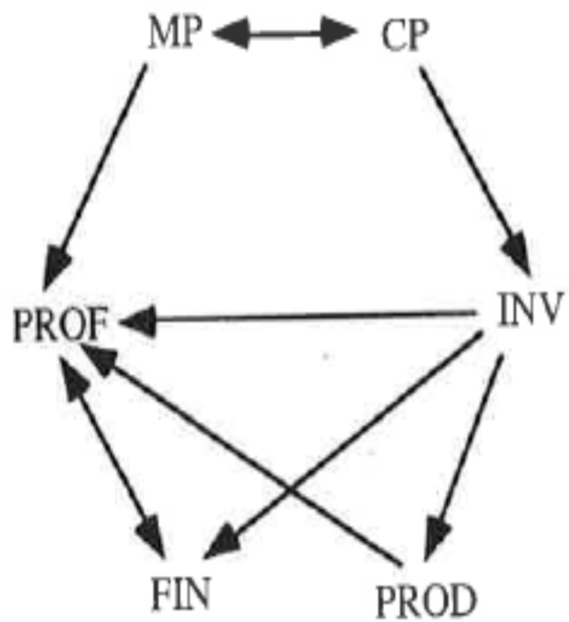
market position	(MP),
competitive position	(CP),
productivity position	(PROD),
profitability	(PROF),
investments	(INV)
financing of investments	(FIN).

Woodstrat

Each arrow in Fig. 7.1 defines a fuzzy rule. We weight these rules or arrows with a number from the interval $[-1, 1]$, or alternatively we could use *word weights* like *little*, or *somewhat*, or *more or less*. The states or nodes are fuzzy too. Each state can fire to some degree from 0% to 100%. In the crisp case the nodes of the network are *on* or *off*. In a real FCM the nodes are fuzzy and fire more as more causal juice flows into them.

Adaptive fuzzy cognitive maps can learn the weights from historical data. Once the FCM is trained it lets us play what-if games (e.g. *What if demand goes up and prices remain stable? - i.e. we improve our MP*) and can predict the future.

Woodstrat



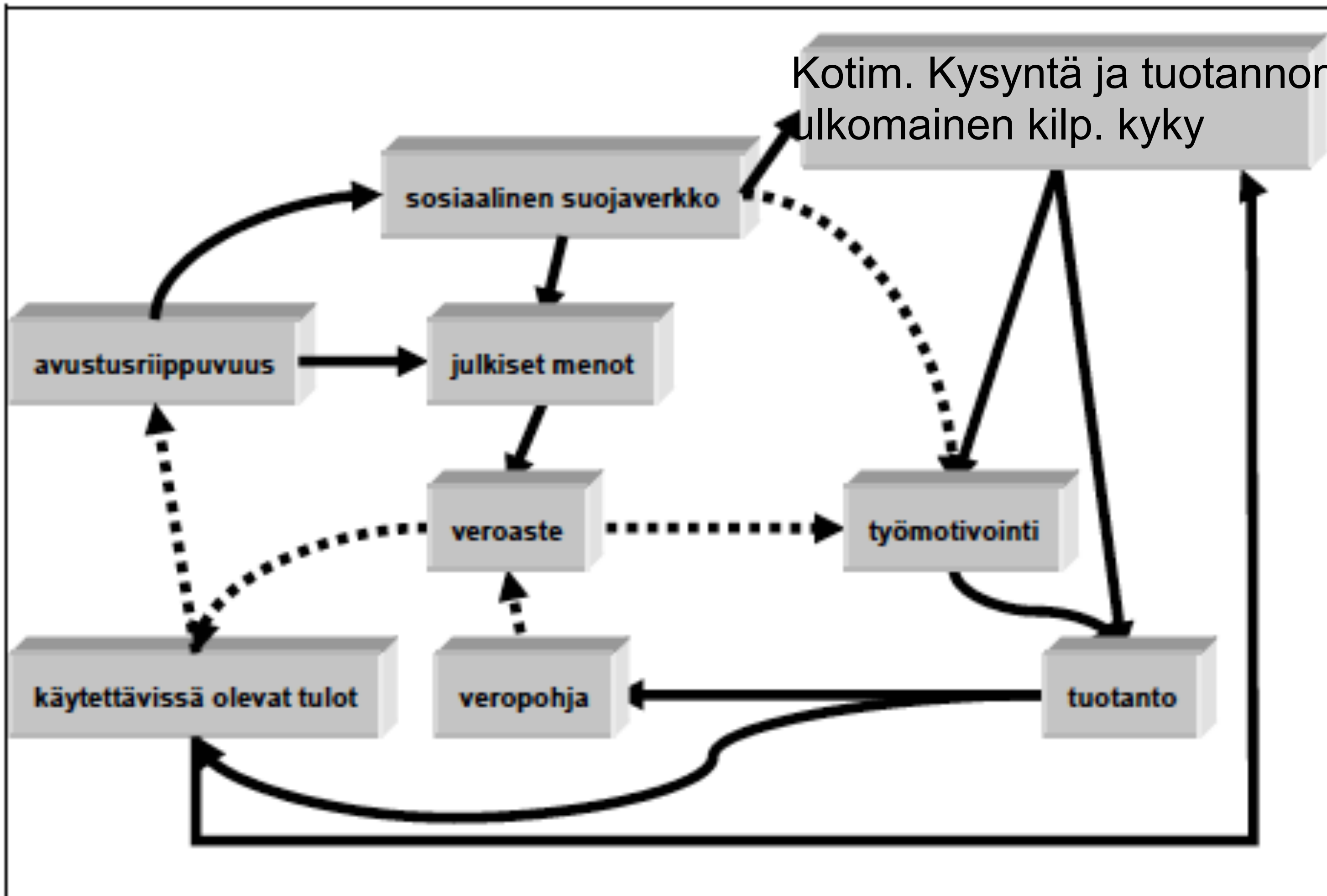
$$W = \begin{pmatrix} 0 & w_{12} & 0 & 0 & 0 & 0 \\ w_{21} & 0 & 0 & 0 & 0 & 0 \\ w_{31} & 0 & 0 & w_{34} & w_{35} & w_{36} \\ 0 & w_{42} & 0 & 0 & 0 & 0 \\ 0 & 0 & w_{53} & w_{54} & 0 & 0 \\ 0 & 0 & 0 & w_{64} & 0 & 0 \end{pmatrix}$$

Figure 7.1: Essential elements of the strategy building process.

Woodstrat history data

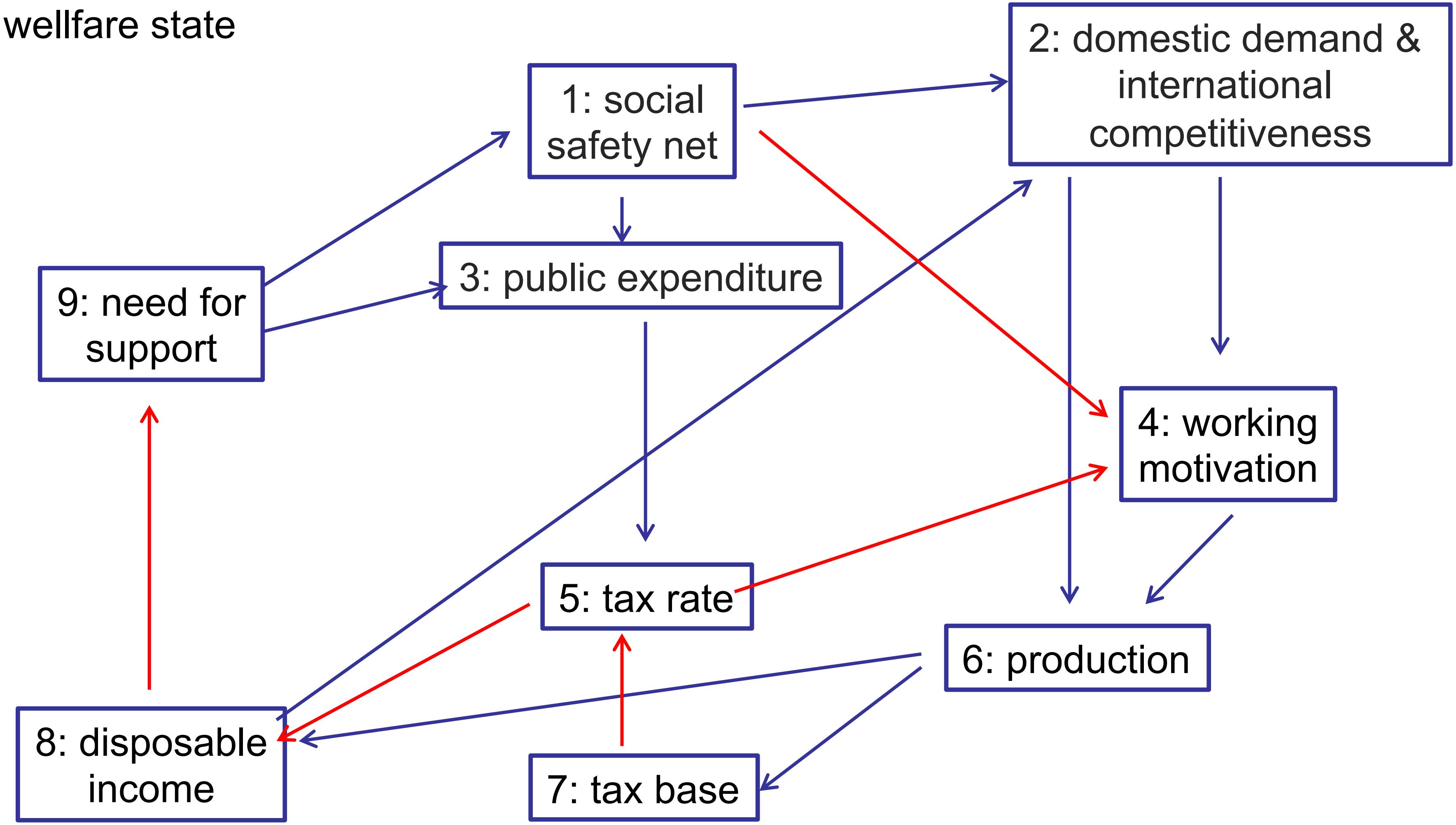
	MP	CP	PROF	INV	FIN	PROD
1.	3	3	3	3	3	3
2.	4	3.5	3.5	3	4	3
3.	4	4	3.5	4	5	3.5
4.	3	4	3.5	4	4	3.5
5.	3	3.5	4	4	3	4
6.	2	3	4	4	2	4
7.	3	2.5	4	5	1	4
8.	3	3	4	5	2	3.5
9.	4	3	4	5	3	3.5
10.	3	3.5	4	5	4	3

Table 5. A training set.



Kuva 13.9. Sumea kognitiivinen kartta Suomen hyvinvointivaltion osatekijöistä (kiinteä viiva = lisää, katkoviiva = vähentää).

Constituents of the Finnish wellfare state



- positive, - negative

Supervisory FCM

- An FCM which supervises or controls the operation of another FCM
- In a sense, a meta-level FCM

Pros and Cons of Cognitive Maps

Numerical	Linguistic
1. Good for modeling complicated phenomena.	1. Good for modeling complicated phenomena.
2. Quite simple systems from the mathematical standpoint.	2. Quite simple systems from the mathematical standpoint.
3. A great number of nodes/variables can be used.	3. Only a limited number of nodes/variables.
4. Can only establish monotonic causal interrelationships between the nodes.	4. Various relationships can be used.
5. Only numerical values and models can be used, thus less user-friendly.	5. Linguistic values and relationships are more user-friendly and human-like.
6. Allow us to use feedback or loops effortlessly.	6. Feedback or loops require more work.
7. The structure can be changed effortlessly.	7. The change of the structure can be laborious.
8. Most are still a priori maps.	8. Most are still a priori maps.
9. Time delays are problematic.	9. Time delays are problematic.
10. Interpretation of nodes and relationships can be problematic.	10. Interpretation of nodes and relationships can be problematic.

Example: Carvalho & Tomè

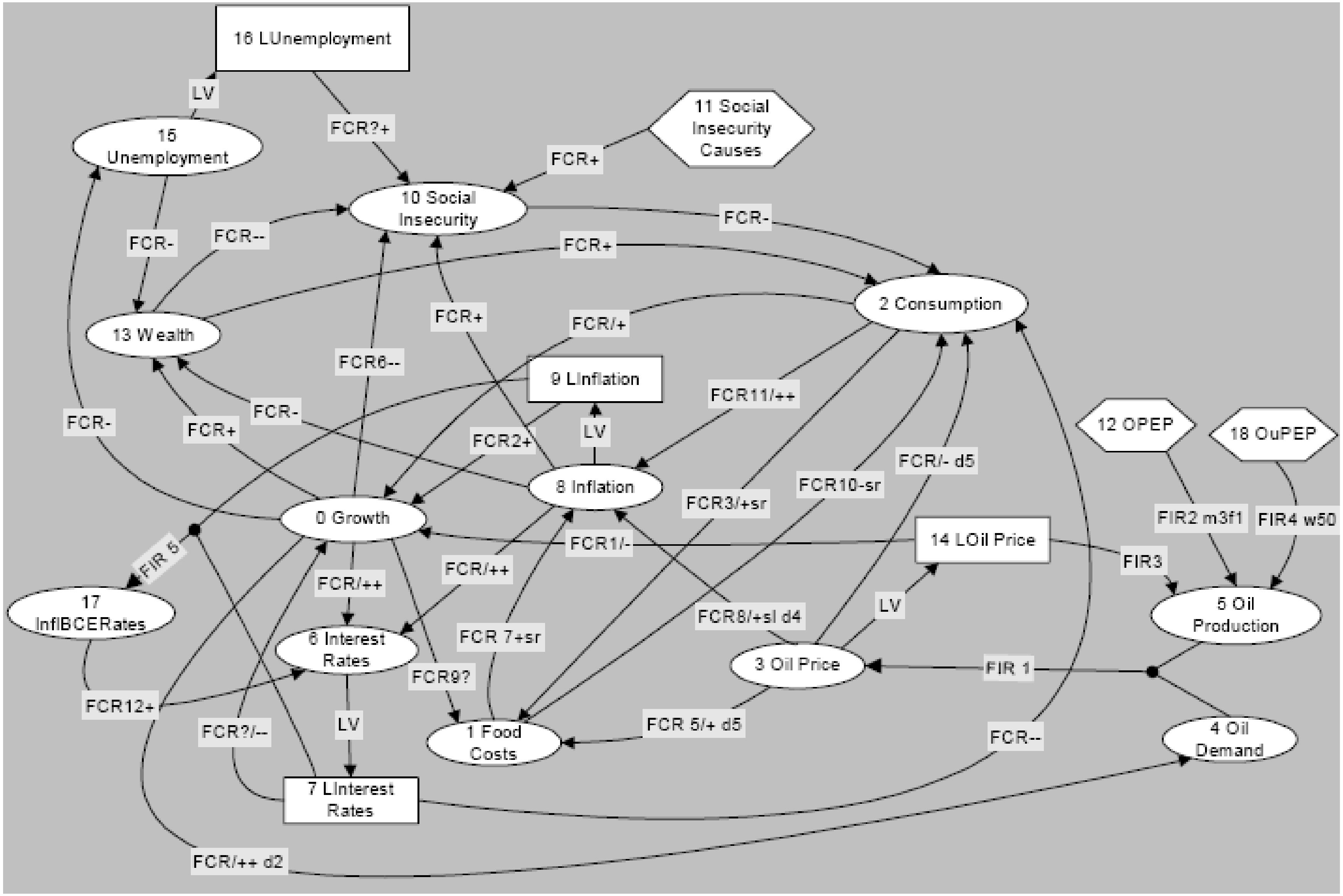
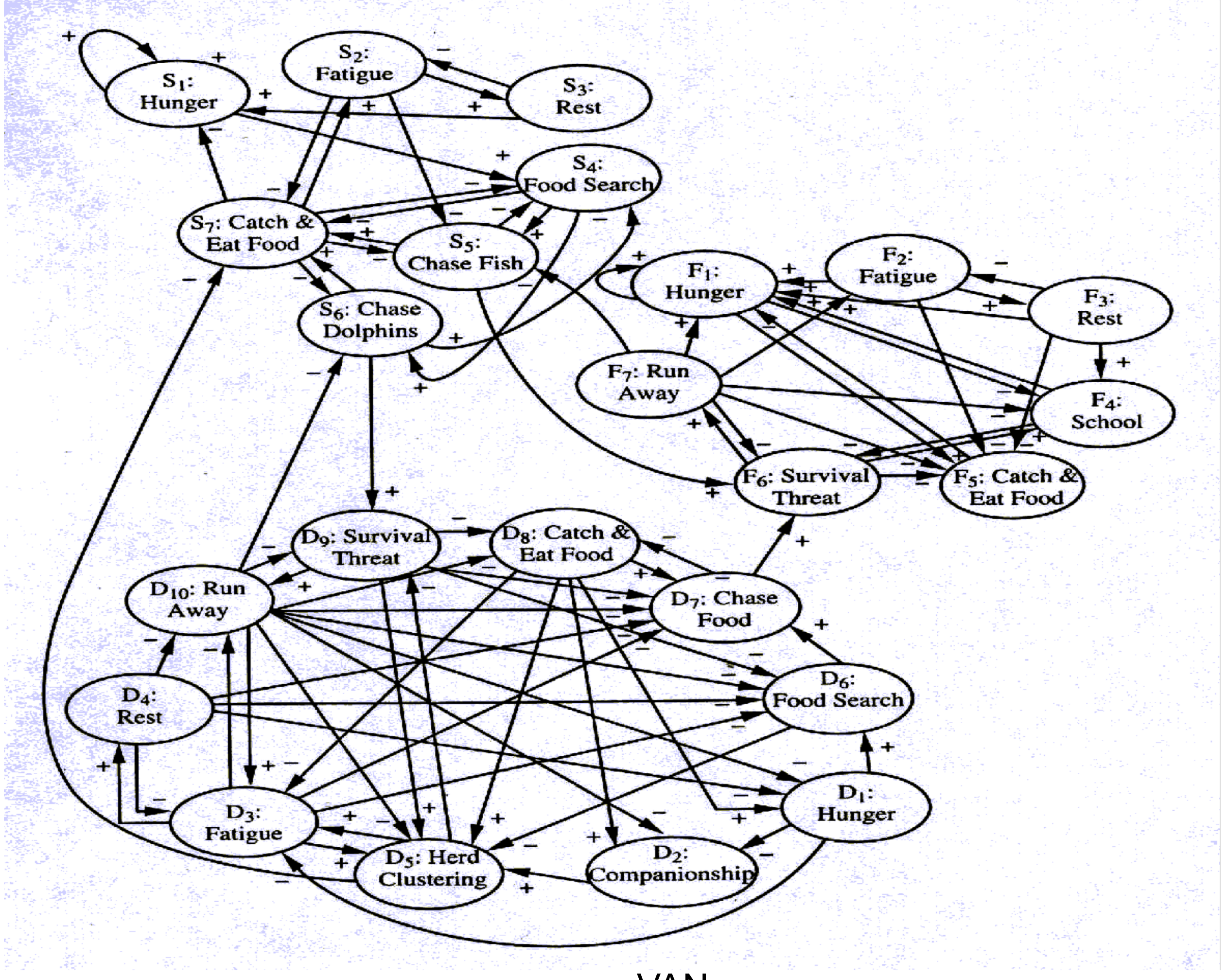


Fig. 8. RB-FCM: A qualitative model of economy. Concept 17, representing a simple FISS was added later

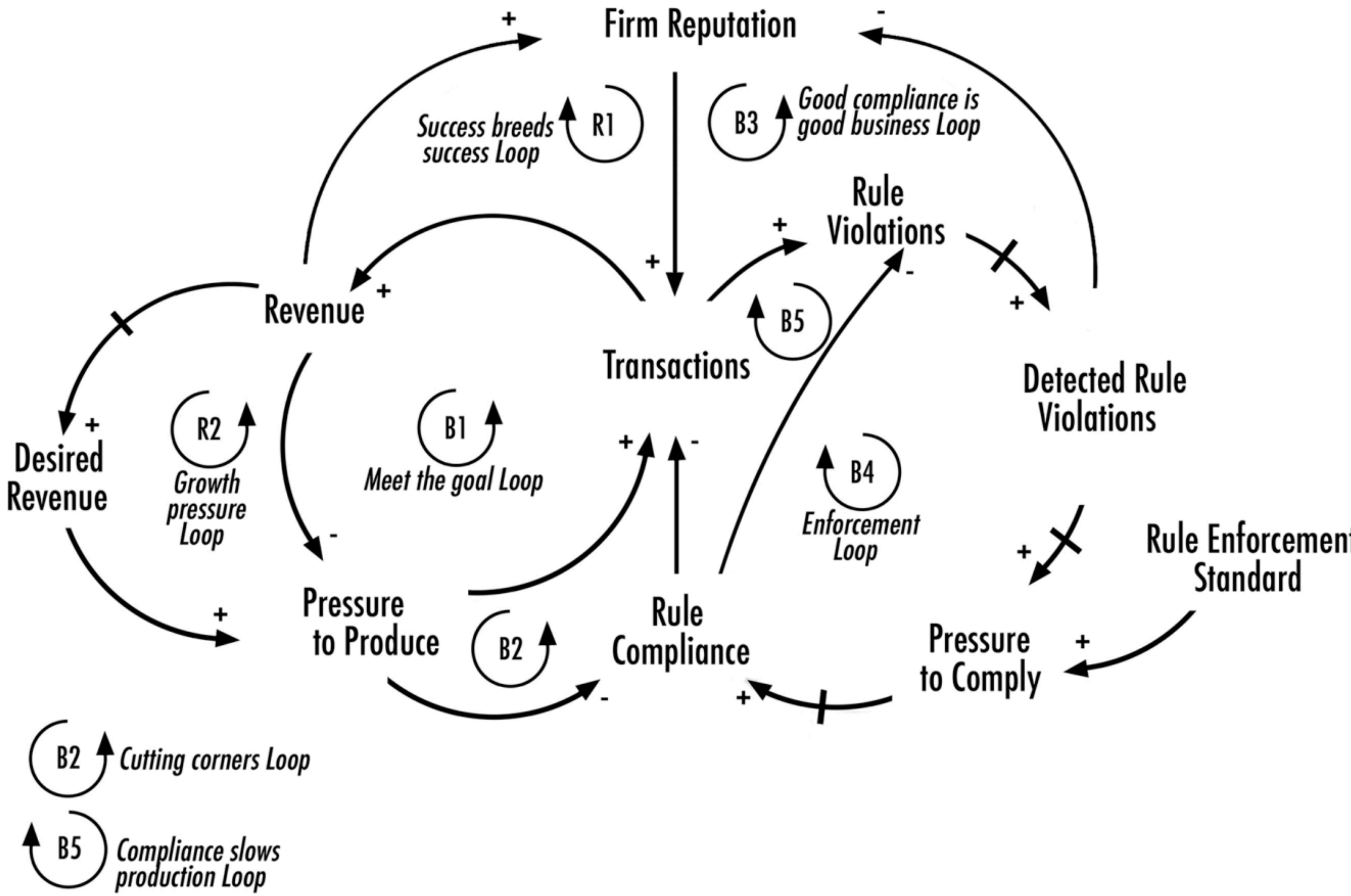
Virtual Undersea World (Kosko)

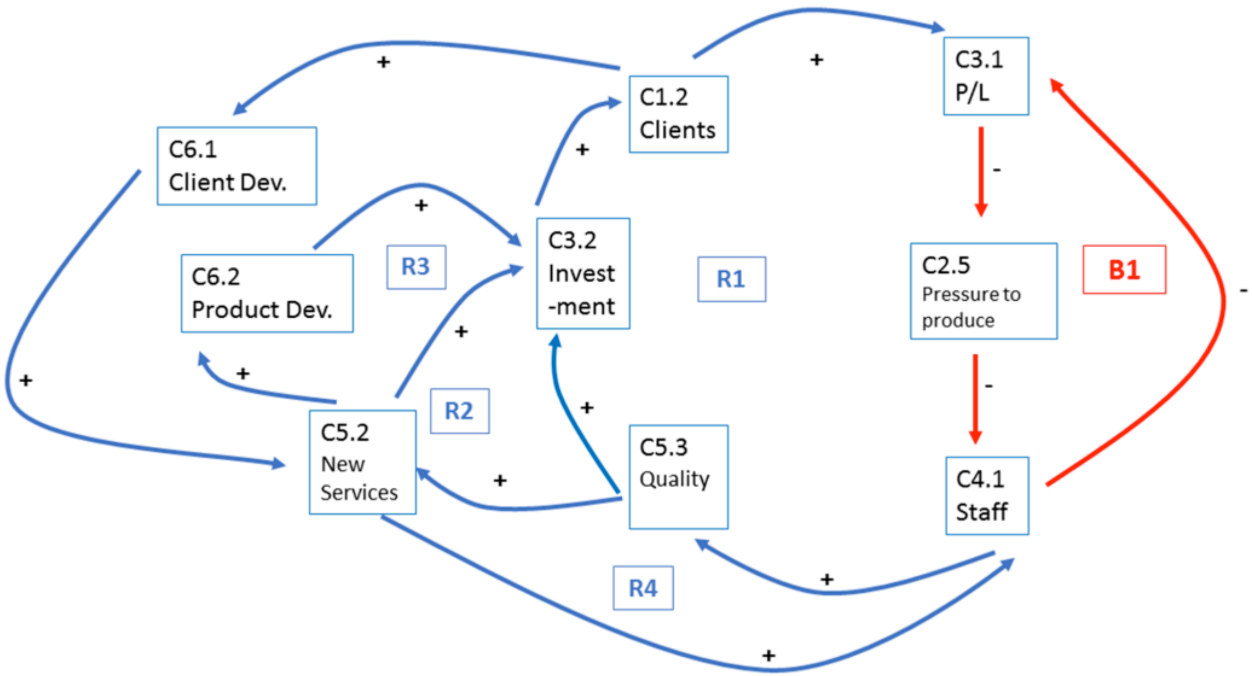
shark



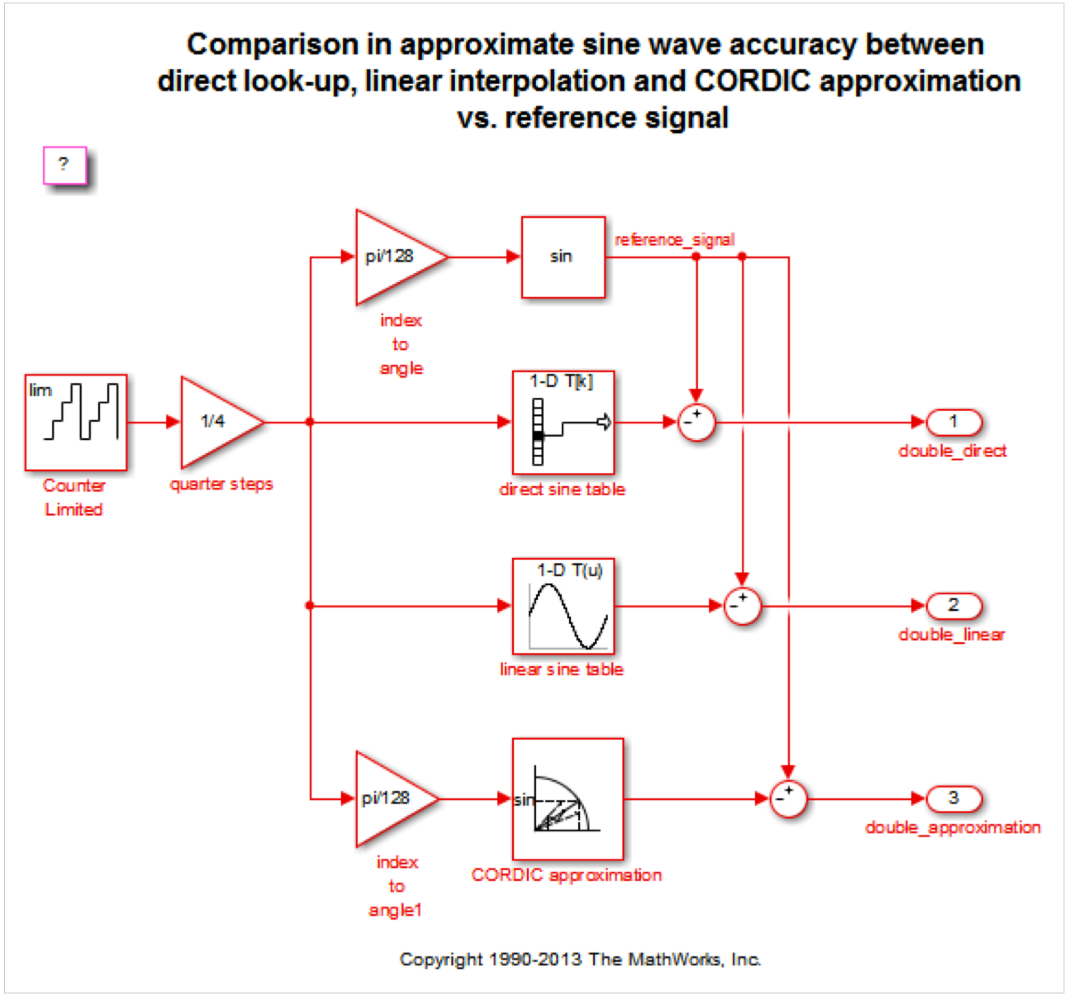
fish school

dolphin herd


Firm Reputation
Success breeds success Loop
R1
B3
Good compliance is good business Loop
Rule Violations
Revenue
Transactions
B5
Detected Rule Violations
Desired Revenue
R2
Growth pressure Loop
B1
Meet the goal Loop
B4
Enforcement Loop
Rule Enforcement Standard
Pressure to Produce
B2
Rule Compliance
Pressure to Comply
B2
Cutting corners Loop
B5
Compliance slows production Loop


C6.1
Client Dev.
C1.2
Clients
C3.1
P/L
C6.2
Product Dev.
R3
C3.2
Invest
-ment
R1
C2.5
Pressure to
produce
B1
R2
C5.2
New
Services
C5.3
Quality
C4.1
Staff
R4
+
-

In Matlab Simulink also usable for FCM construction



Examples of Urban Research

- http://www.tkk.fi/Yksikot/YTK/koulutus/metodikortti/Kognkart.html
- http://www.spacesyntax.com/

London Pedestrian Routemap http://www.spacesyntax.com/