Measures of monotone relationships using dominance-based rough set approach (DRSA)

Wojciech Kotłowski[1] Krzysztof Dembczyński[1]
Salvatore Greco[2] Roman Słowiński[1,3]

[1]Institute of Computing Science, Poznań University of Technology

[2]Faculty of Economics, University of Catania

[3]Institute for System Research, Polish Academy of Sciences

Overview

Preliminaries

- X — finite set of objects (observations, records), $X = \{x_1, \ldots, x_\ell\}$
- Each $x \in X$ described by n conditional attributes $Q_1, \ldots, Q_n$ (input variables) and decision attribute (output variable) with finite and ordered domain $T = \{1, \ldots, m\}$ (classes).
- Each $x \in X$ is identified with $(q_1(x), \ldots, q_n(x))$ and decision value (class label) $t(x)$
- Objective: Using the information in X build model for further predictions of value of decision attribute based on the values of conditional attributes.

Monotone relationships

Informal Definition

Increasing (decreasing) the value of conditional attribute will increase or hold the value of decision attribute; attribute with monotone relationship — *criterion*.

Examples

- *the lower maximal temperature in summer, the higher yields* (cost)
- *the larger the market share of a company, the larger its profit* (gain)

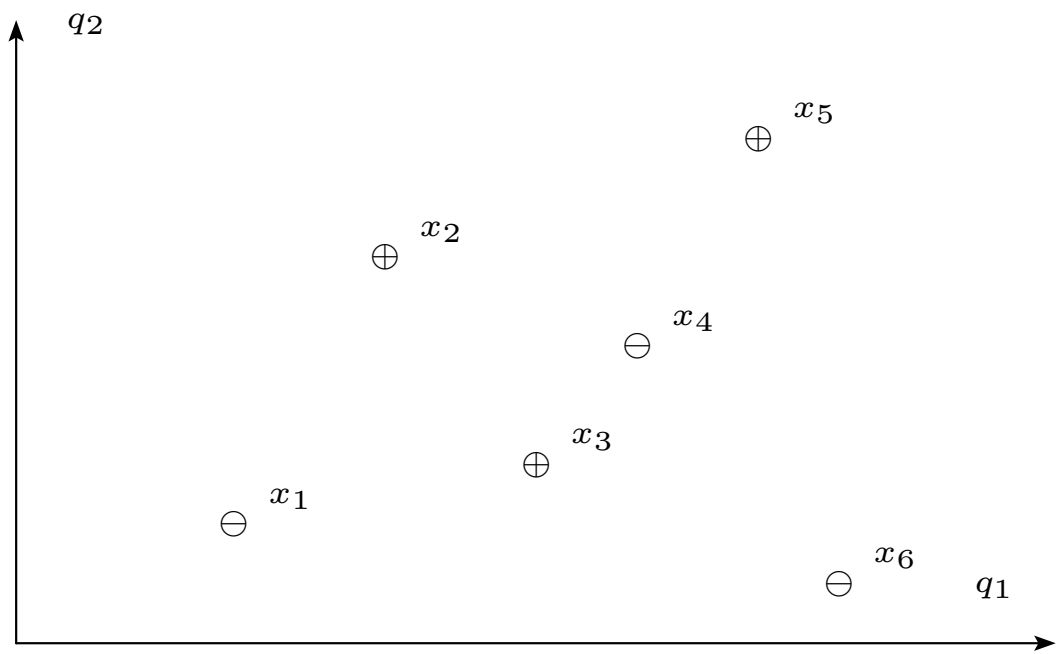
More Formal Definition

The model function from $Q_1 \times \ldots \times Q_n$ to T is monotone in arguments for which monotone relationship holds.

Dominance relation

Definition

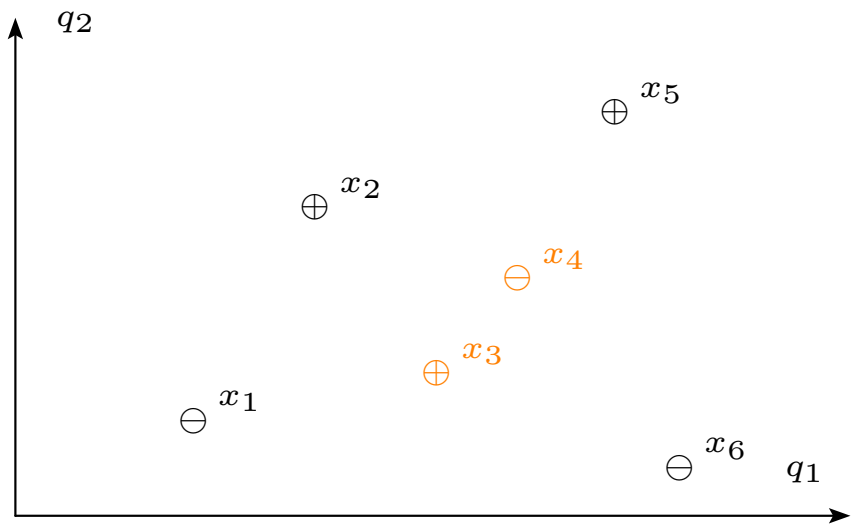
For each $x, y \in X$, x *dominates* y (xDy) if x has better or equal values to y on every criterion, $\forall_{i \in \{1,\ldots,n\}} q_i(x) \geqslant q_i(y)$



Inconsistencies

Definition

- For each $x, y \in X$, x is *inconsistent* with y, if xDy but $t(x) < t(y)$
- Object $x \in X$ is inconsistent, if there exist another object $y \in X$ inconsistent with x



Dominance-based Rough Set Approach (DRSA)

- Handling inconsistencies by the notion of *lower and upper approximations* of classes (certain and possible regions)
- Equivalent (more convenient here) description using *generalized decision function* δ:

$$\delta(x) = \langle l(x), u(x) \rangle$$

where $l(x) = \min\{t(x) \colon yDx, y \in X\}$,
$u(x) = \max\{t(x) \colon xDy, y \in X\}$

- For all $x \in X$, $l(x) \leqslant t(x) \leqslant u(x)$ and if x is consistent, $l(x) = u(x)$.

Quality of Approximation (γ)

Definition

γ measures how well a partition of X into classes $0, \ldots, m$ can be approximated by conditional attributes $Q_1, \ldots, Q_n$ using the dominance relation D. In fact, it measures the overall consistency of the set X.

General Remarks

- $\gamma \in [0, 1]$, $\gamma = 1$ if every $x \in X$ is consistent
- As n grows, γ cannot decrease (monotonicity property)
- γ can be used as a general measure of the strength of monotone relationships

Classical Definition of γ

Definition

$$\gamma = \frac{\#\text{ consistent objects}}{|X|}$$

Pros

- simplicity

Cons

- very restrictive — even one object can boil γ down to 0
- very sensitive to noise

Definition II — Based on the Generalized Decision

Definition

$$\gamma = 1 - \frac{\sum_{x \in X}(u(x) - l(x))}{(n-1)|X|}$$

Pros

- takes into account the strength of inconsistency of each object
- resistant to local inconsistencies

Cons

- sensitive to noise — still one object can boil γ down to 0

Definition III — Based on Objects Reassignment

Definition

$$\gamma = 1 - \frac{L}{|X|}$$

where L is the minimal number of objects that must be reassigned (their decision value is changed) to make X consistent

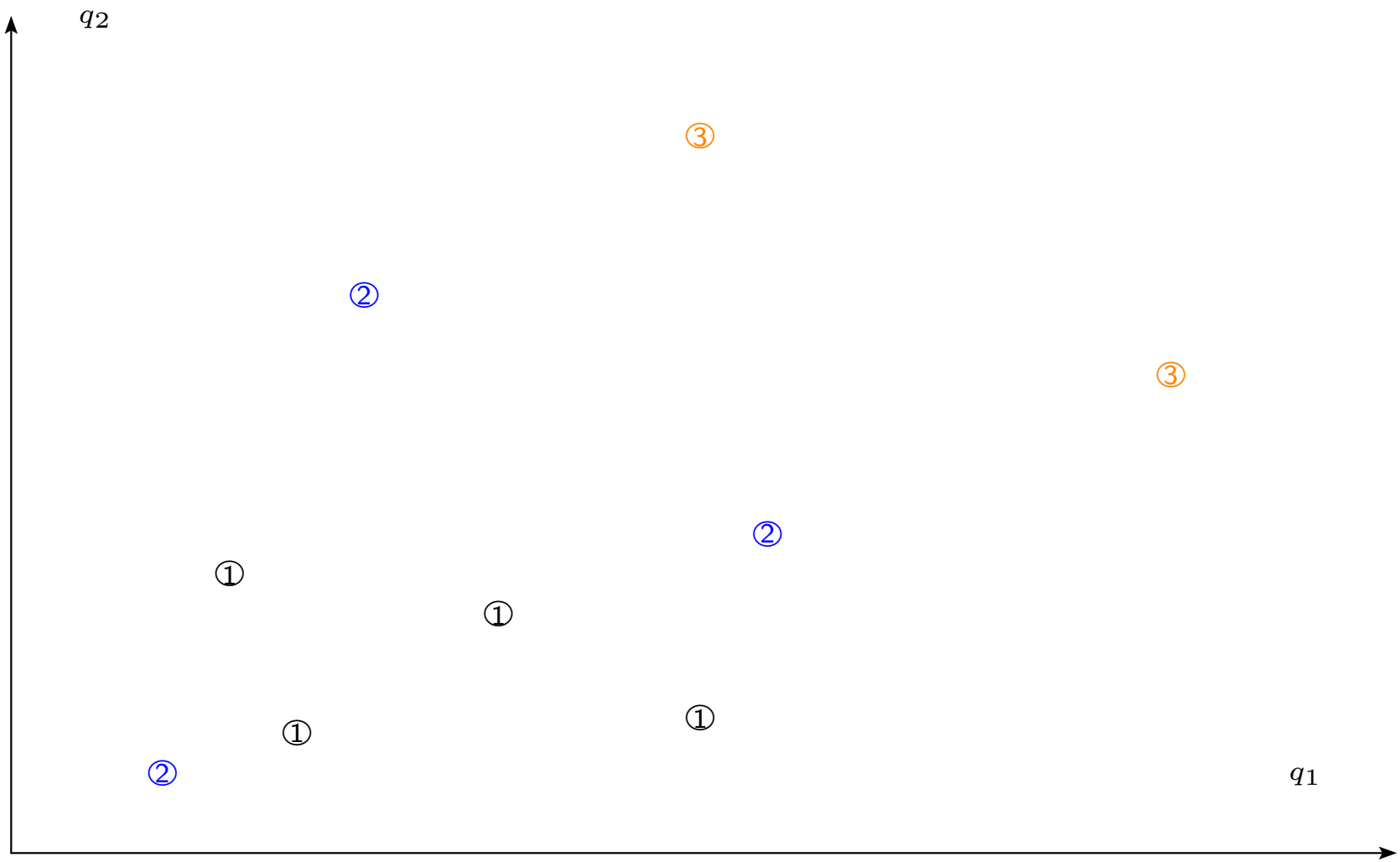
Pros

- resistant to noise
- indicates possible errors in X

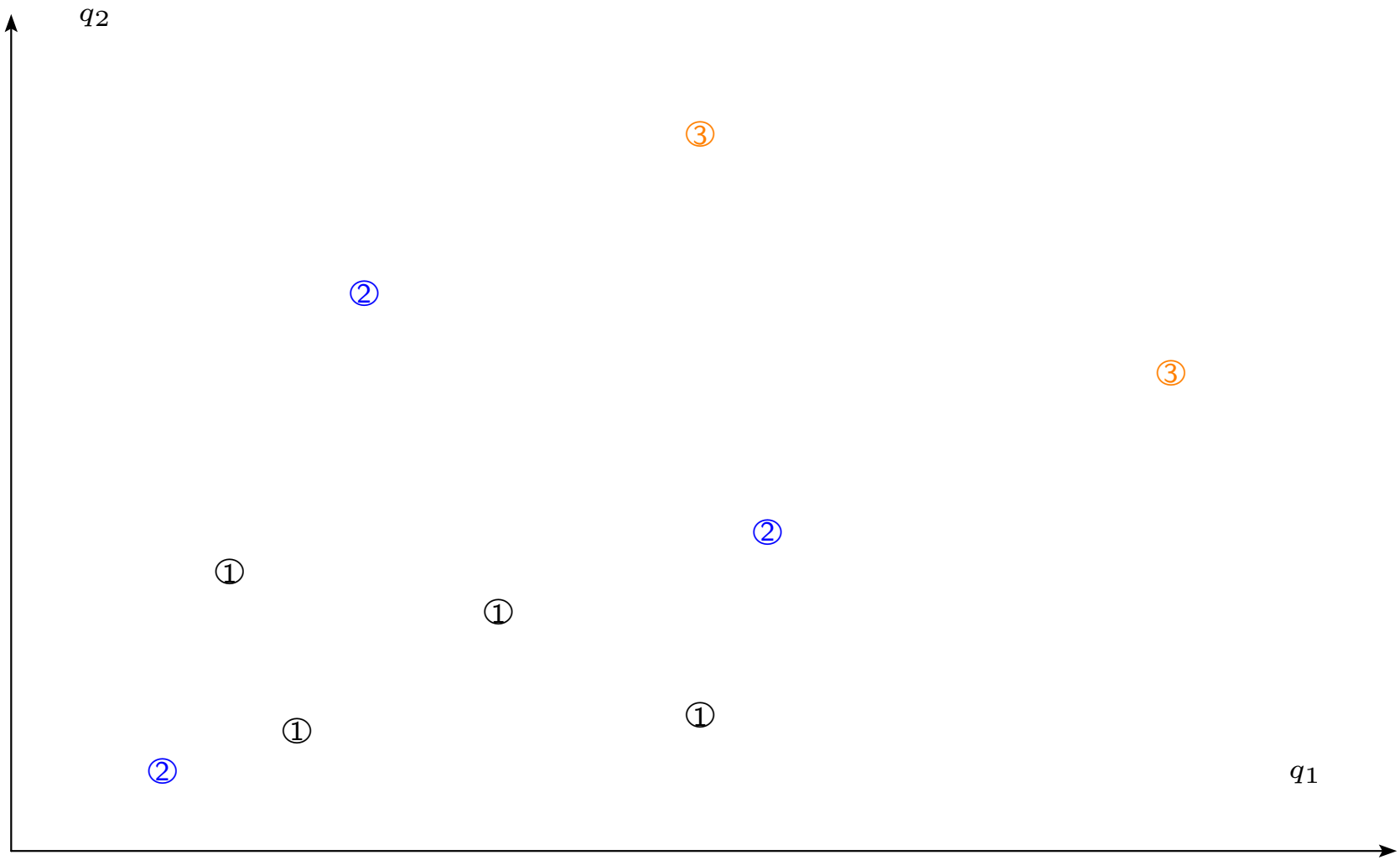
Cons

- combinatorial definition requires an algorithm to solve.

Example



Example: $\gamma_1 = \frac{4}{9}$, $\gamma_2 = \frac{13}{18}$, $\gamma_3 = \frac{8}{9}$



How to Calculate Minimal Reassignment?

- Problem can be formulated using linear integer programming.
- Notation: $t(x_i)$ — initial decision values; $t'(x_i)$ — decision values after reassignment
- For each $x_i \in X$ introduce m binary variables d_{ij} $i \in \{1, \ldots, n\}$, $j \in \{1, \ldots, m\}$, with interpretation $d_{ij} = 1$ iff $t'(x_i) \geqslant j$
- Penalty for each $x_i \in X$ is $L(x_i) = (1 - d_{i,t(x_i)}) + d_{i,t(x_i)+1}$
- Goal function is $L = \sum_{x_i \in X} L(x_i)$
- Constrains are:

$$d_{ij} \geqslant d_{kj} \qquad \forall i, k\colon \; x_i D x_k \quad 1 \leqslant j \leqslant n$$

$$d_{ij'} \leqslant d_{ij} \qquad 1 \leqslant i \leqslant m, \quad 1 \leqslant j < j' \leqslant n$$

- Final assignment for each $x_i \in X$ is $t'(x_i) = \max_{d_{ij}=1}\{j\}$

Solution of the Problem

- For two class problem we end up with *isotonic separation* (Chandrasekaran et al.).
- Totally unimodular matrix allows to relax integer constraints or solve the dual problem as a network flow problem, O(n^3).
- Strong reduction of the problem due to the rough set theory (DRSA)

Theorem

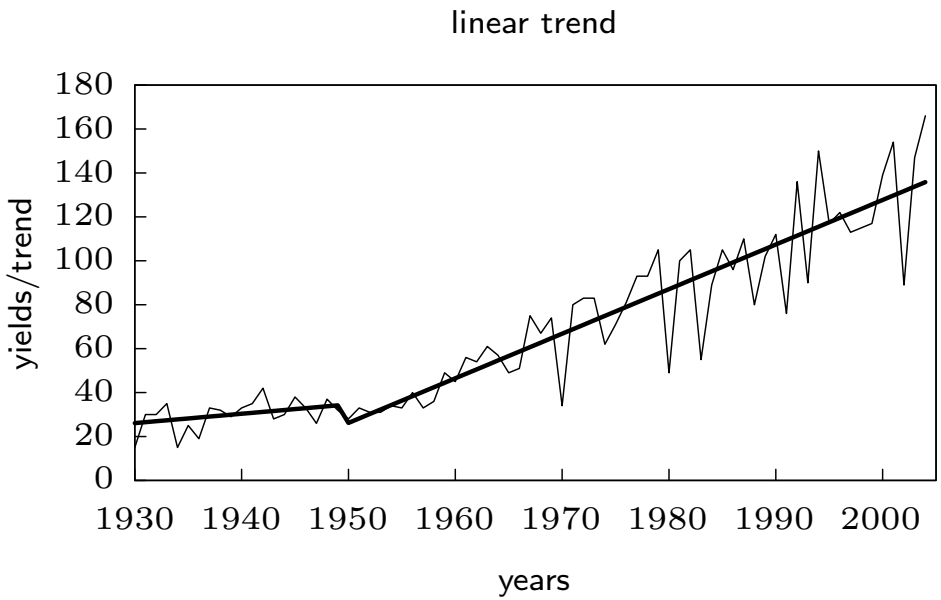
There always exists an optimal solution for which the following condition holds: $l(x_i) \leqslant t'(x_i) \leqslant u(x_i)$

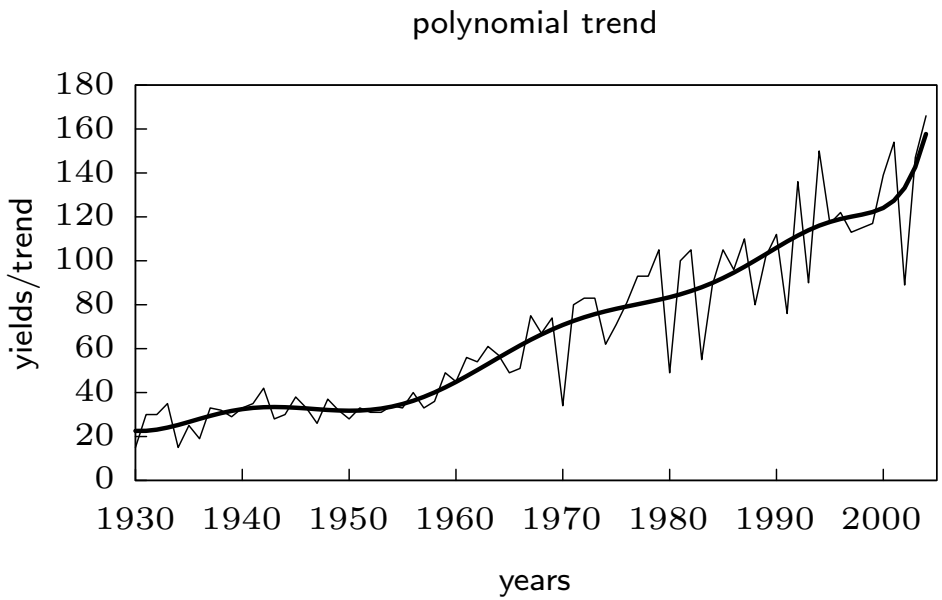
Case Study

- Investigating the impact of weather on crop yields in USA
- Weather data collected by counties, monthly; period of time: about 1930-2004:
  - precipitation
  - maximal, minimal and average temperature
- Yield data collect yearly, for each county in various periods (from 1930-1950 till 2004):
  - maize (Iowa, Illinois, Indiana)
  - winter wheat (North Carolina, South Carolina, Virginia)

Trend Identification

- The goal was to remove impact of non-weather factors (long term variations included in the general trend)
- Two types of trend considered: piecewise linear and polynomial





Data Preprocessing

- Changing the value of detrended yields into two classes *low* and *high* (below the trend and above the trend)
- Removing objects (observations) with yield values [-0.1,0.1] around the trend
- Aggregation of monthly weather conditions into seasonal (winter, spring, sumer, fall) by taking the average value (in case of precipitation and average temperature) or extreme value (in case of minimal and maximal temperature)
- Monotonicity directions were chosen using Kendall's τ coefficients for each of the attributes; if τ exceeded some threshold value, the attribute was treated as monotonically (positively or negatively) related to the decision attribute

Classification

- The algorithm used was the *ensemble of decision rules*
- Possibility of using both criteria and non-monotone attributes
- Analysis for several τ thresholds (different number of criteria) — for each threshold calculation of γ measures, their significance and accuracy.
- Significance of γ: $s = \frac{\gamma - \gamma^{rev}}{\gamma + \gamma^{rev}}$.

Results: Maize

thresh.	**# criteria**	γ_1	**sign.**	γ_3	**sign.**	**accuracy**
1	0	-	-	-	-	83.55%
0.2	3	0.08	0.86	0.55	0.66	84.2%
0.15	4	0.16	0.88	0.64	0.63	84.24%
0.1	6	0.47	0.79	0.8	0.57	83.56%
0	16	0.99	0.08	0.99	0.06	82.23%

Results: Wheat

thresh.	**# criteria**	γ_1	**sign.**	γ_3	**sign.**	**accuracy**
1	0	-	-	-	-	82.54%
0.2	3	0.09	0.88	0.5	0.65	82.46%
0.15	4	0.15	0.87	0.59	0.64	82.28%
0.1	5	0.35	0.82	0.73	0.63	82.25%
0	16	0.99	0.07	0.99	0.05	79.67%

Conclusions

- Low strength of monotone relationships in the data probably due to "noise" (non-weather factors).
- Imposing monotone constraints decrease the prediction accuracy.
- Possible improvements: different detrending methods, outliers analysis, introducing non-weather factors.