

# Visualization of Agriculture Data Using Self-Organizing Maps

Georg Ruß, Rudolf Kruse, Martin Schneider, Peter Wagner

December 10th, 2008

# Motivation

- ▶ precision farming
  - ▶ cheap data collection
  - ▶ GPS-based technology
  - ▶ divide field into small-scale parts
  - ▶ treat small parts independently instead of uniformly
- ▶ lots of data (sensors, imagery)
- ▶ use data mining to
  - ▶ improve efficiency
  - ▶ improve yield

## Data Flow Model

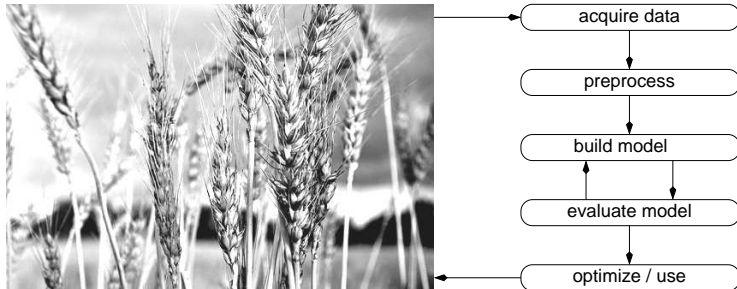


Figure: Data Mining Context

## Nitrogen Fertilizer

- ▶ easy to measure when manuring
- ▶ three points into the growing season where nitrogen fertilizer is applied
- ▶ three attributes: N1, N2, N3

## Vegetation Measuring

- ▶ Red Edge Inflection Point
- ▶ second derivative value along the red edge region
- ▶ aerial photography or tractor-mounted sensor
- ▶ larger value means more vegetation
- ▶ measured before N2 and N3
- ▶ two attributes: REIP32, REIP49

## Electric Conductivity

- ▶ measure apparent conductivity of soil down to 1.5m
- ▶ uses commercial sensors
- ▶ one attribute: EM38

## Yield

- ▶ measure yield when harvesting
- ▶ data from 2005 (previous year) and 2006 (current year)
- ▶ two attributes: Yield05, Yield06

## Data overview, F131 data set

Table: Data overview, F131

<i>F131</i>	<i>min</i>	<i>max</i>	<i>mean</i>	<i>std</i>	<i>Description</i>
YIELD05	1.69	10.68	5.69	0.93	yield in 2005
EM38	51.58	84.08	62.21	8.60	electrical conductivity of soil
N1	47.70	70	64.32	6.02	nitrogen fertilizer applied at first date
N2	14.80	100	51.71	15.67	nitrogen fertilizer applied at second date
N3	0	70	39.65	13.73	nitrogen fertilizer applied at third date
REIP32	719.6	724.4	722.6	0.69	red edge inflection point vegetation index
REIP49	722.3	727.9	725.8	0.95	red edge inflection point vegetation index
YIELD06	1.54	8.83	5.21	0.88	yield in 2006



## Fertilization Strategies

- ▶ uniform treatment
  - ▶ same amount of fertilizer at each part of the field
- ▶ neural network predictor
  - ▶ uses collected past data to predict yield
  - ▶ this prediction is used to optimize fertilizer usage economically
  - ▶ different amounts of fertilizer are used for each part of the field

## Data split

**Table:** Overview on available data sets for specific fertilization strategies for different fields

F131-all	YIELD05, EM38, N1, REIP32, N2, REIP49, N3, YIELD06, <i>fert. strategy</i>
F131-net	subset of F131-all where fertilization strategy is <i>neural network</i>
F330-all	YIELD05, EM38, N1, REIP32, N2, REIP49, N3, YIELD06, <i>fert. strategy</i>
F330-net	subset of F330-all where fertilization strategy is <i>neural network</i>

## Research Questions

- ▶ How much does *fertilization* influence current-year yield?
- ▶ Correlation between data attributes that influences yield?
- ▶ Similarity/Correlation between different fields?

## Results for F131-all, Labels/U-Matrix

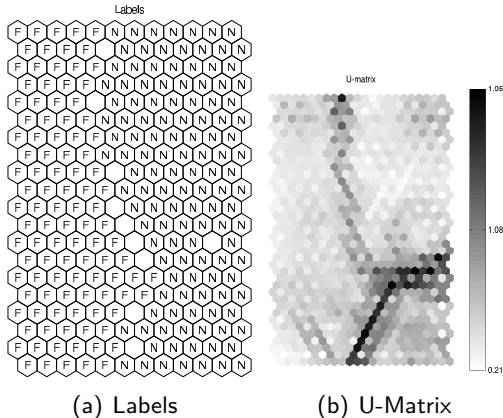


Figure: F131-all, U-Matrix and Labels

## Results for F131-all, Nitrogen

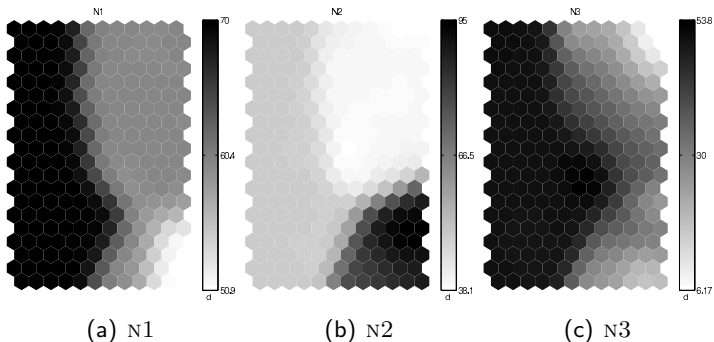


Figure: F131-all, N1, N2, N3

## Results for F131-all, REIP, Yield

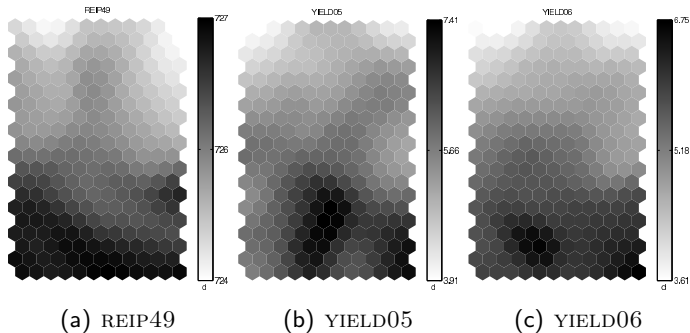


Figure: F131-all, REIP49 vs. YIELD05 vs. YIELD06

## Results for F131-net

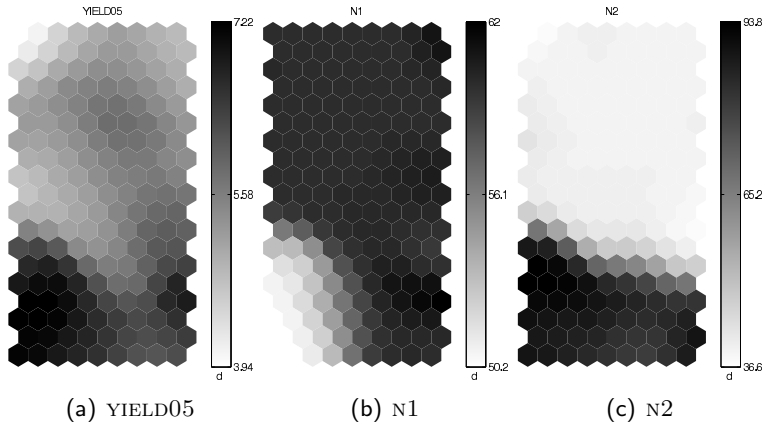


Figure: F131-net, YIELD05, N1, N2

## Results for F131-net

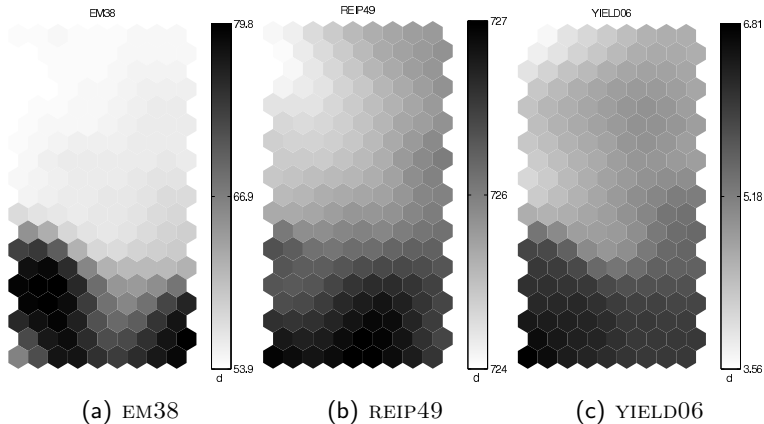


Figure: F131-net, EM38, REIP49, YIELD06



## Conclusion

- ▶ self-organizing maps for data analysis
- ▶ correlations and interdependencies in data visible on maps
- ▶ yield prediction models can partly be understood
- ▶ successful application of data mining in agriculture