Classification of Soil Heterogeneity Indicators

GfKl IFCS 2009

Georg Ruß, Rudolf Kruse, Martin Schneider, Peter Wagner

March 15th, 2009

Precision Agriculture

basic idea

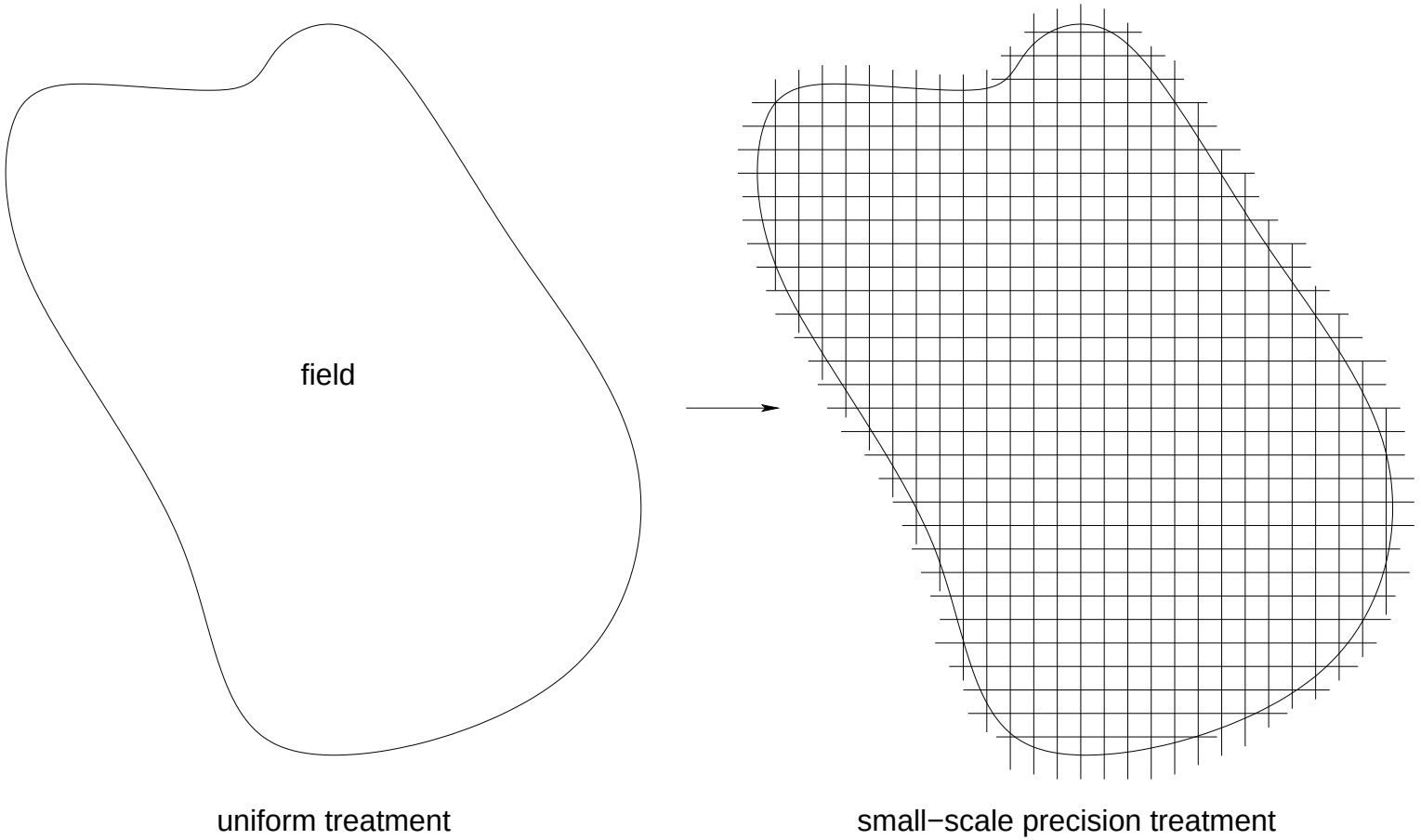

Figure: Basic Idea of Precision Agriculture

Precision Agriculture

some more ideas

- precision agriculture
  - 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
  - identify useful sensors

N Fertilizer and Yield

- Nitrogen fertilizer
  - easy to measure when manuring
  - three points into the growing season where nitrogen fertilizer is applied
  - three attributes: N1, N2, N3
- Yield 2005/2006
  - measure yield when harvesting
  - data from 2005 (previous year) and 2006 (current year)
  - two attributes: Yield05, Yield06

Vegetation Measuring and Electric Conductivity

- Red Edge Inflection Point
  - second derivative value along the spectrum's red edge region
  - aerial photography or tractor-mounted sensor
  - larger value means more vegetation
  - measured before N2 and N3
  - two attributes: REIP32, REIP49
- Electromagnetic Conductivity
  - measure apparent conductivity of soil down to 1.5m
  - uses commercial sensors
  - one attribute: EM38

Data overview, F131 data set

chronological order

Table: Data overview, F131

F131	*min*	*max*	*mean*	*std*	*Description*
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
REIP32	719.6	724.4	722.6	0.69	red edge inflection point vegetation index
N2	14.80	100	51.71	15.67	nitrogen fertilizer applied at second date
REIP49	722.3	727.9	725.8	0.95	red edge inflection point vegetation index
N3	0	70	39.65	13.73	nitrogen fertilizer applied at third date
YIELD06	1.54	8.83	5.21	0.88	yield in 2006

- three additional data sets were used, with more available

Data set sizes

- F04: 5241 records
- F330: 4578 records
- F131: 2278 records
- F131net: 1144 records (subset of F131)
- thereof none with missing values and none with outliers (after preprocessing)

Research Questions

- How well can the current year's yield be predicted?
- How useful are the additional sensor data that were introduced?
  - EM38, REIP32, REIP49 et al.

Yield Prediction from Available Data

Regression Task

- Models used:
  - Artificial Neural Network – Multi-Layer Perceptron
  - Artificial Neural Network – Radial Basis Function
  - Regression Tree
  - Support Vector Regression
- Error Measures (for Cross-Validation)
  - mean absolute error (MAE)
  - root mean squared error (RMSE)

Results for Regression Models

Error Measure / Model	F04		F131	F131net	F330
MAE MLP:	0.3706		0.2468	0.2300	0.3576
RMSE MLP:	0.4784		0.3278	0.3073	0.5020
MAE RBF:	0.3838		0.2466	0.2404	0.3356
RMSE RBF:	0.5031		0.3318	0.3205	**0.4657**
MAE REGTREE:	0.4380		0.2823	0.2530	0.4151
RMSE REGTREE:	0.5724		0.3886	0.3530	0.6014
MAE SVR:	**0.3446**		**0.2237**	**0.2082**	**0.3260**
RMSE SVR:	**0.4508**		**0.3009**	**0.2743**	0.4746

Table: Results of running different models on different data sets. The best predictive model for each data set is marked in **bold** font.

Results for Regression Models

- SVR
  - performs best and takes the least time to compute
  - few parameters to tune
  - but: black-box model, hard to explain *why* it works well
- RegTree
  - worse performance than SVR
  - but: explainable results when considering the tree splits
- MLP, RBF
  - worse performance than SVR
  - longest computation time of models
  - hard to explain, black-box models

Evaluation of Sensor Data

- Question:
  - How much does a particular sensor contribute to the quality of the yield prediction?
- Ideas:
  - use the correlation factor between the sensor and the yield
  - create a standard prediction model and evaluate how much this is improved by adding sensor data
  - further ideas? (work in progress)

Summary

- Yield can be predicted well using standard data and additional sensor data
- One model can be particularly recommended: support vector regression
- Ideas for sensor evaluation have been presented
- overall: successful application of data analysis in agriculture

Bibliography

Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Estimation of neural network parameters for wheat yield prediction.
In Max Bramer, editor, *Artificial Intelligence in Theory and Practice II*, volume 276 of *IFIP International Federation for Information Processing*, pages 109–118. Springer, July 2008.

Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Optimizing wheat yield prediction using different topologies of neural networks.
In José Luis Verdegay, Manuel Ojeda-Aciego, and Luis Magdalena, editors, *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-08)*, pages 576–582. University of Málaga, June 2008.

Georg Ruß, Rudolf Kruse, Martin Schneider, and Peter Wagner.

Visualization of agriculture data using self-organizing maps.
In Tony Allen, Richard Ellis, and Miltos Petridis, editors, *Applications and Innovations in Intelligent Systems*, volume 16 of *Proceedings of AI-2008*, pages 47–60. BCS SGAI, Springer, January 2009.

Georg Ruß, Rudolf Kruse, Peter Wagner, and Martin Schneider.

Data mining with neural networks for wheat yield prediction.
In Petra Perner, editor, *Advances in Data Mining (Proc. ICDM 2008)*, pages 47–56, Berlin, Heidelberg, July 2008. Springer Verlag.

Finally ...

Questions & Answers