Dealing with uncertainty in a real-time knowledge process

Monica Wachowicz and Gary J. Hunter

ALTERRA
WAGENINGEN UR

Outline

- Knowledge Discovery in Disaster Management
- Real Time and Uncertainty
- Our Approach
- Results

Knowledge Discovery in Environmental Sciences

Approaches, tools, environment and facilities for

- Exploring anomalies in geophysical data,
- Scaling up the current analysis techniques.

Behnke et al. 1999

Knowledge Discovery in Disaster Management

- How to predict storm tracks?
- How to detect fires for a rapid dissemination of warnings?
- How to predict changes in intensity as a storm approaches land?

Knowledge Discovery in Disaster Management

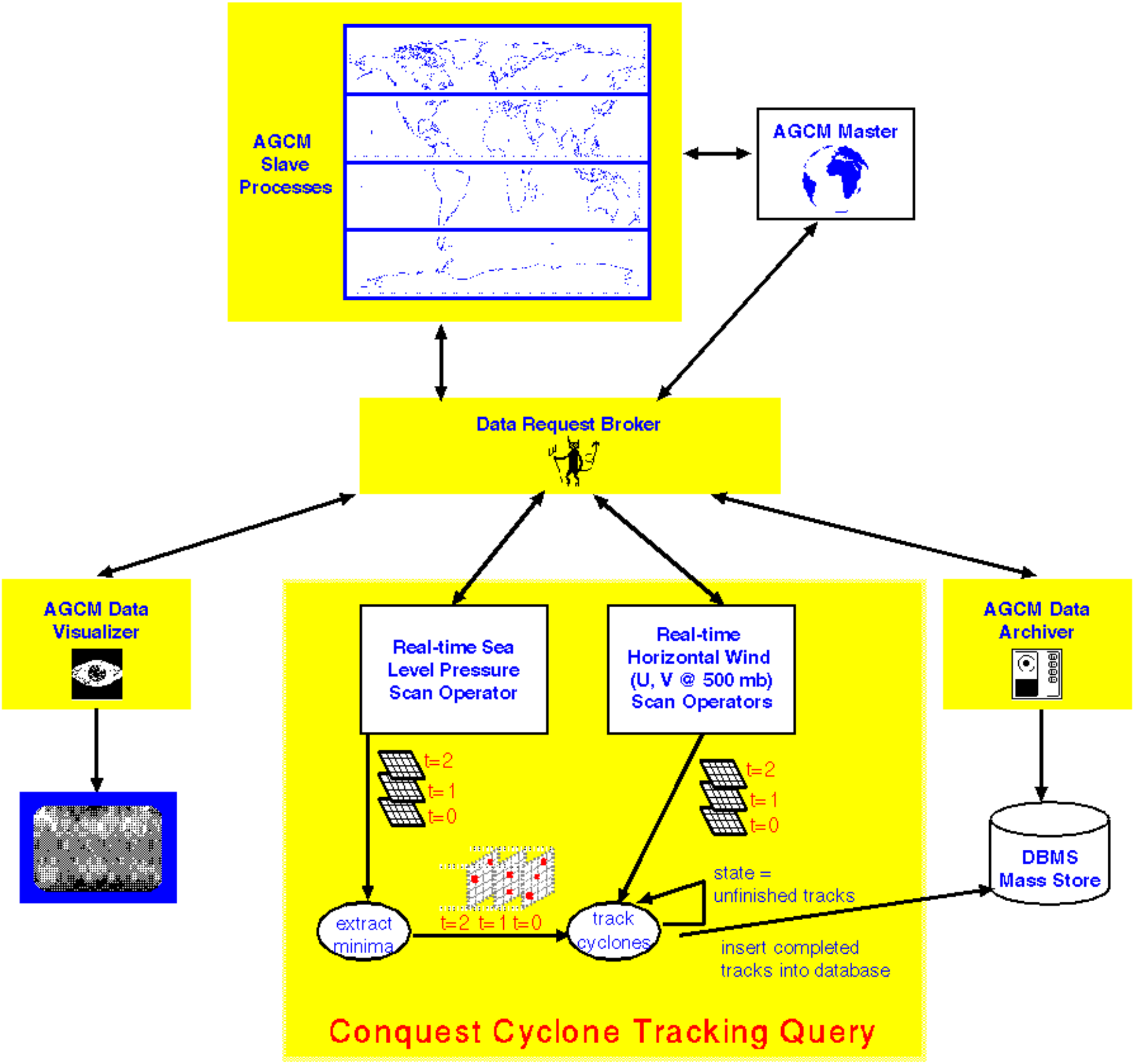
Real-Time

Everything on line – rapid access

Rapid systematic dissemination of results

Real Time Data Mining

- Super computing
- GRID



www.dml.cs.ucla.edu

ALTERRA WAGENINGEN UR

Knowledge Discovery in Disaster Management

Uncertainty

False alarm problems are a major issue for natural hazard warnings due to cost and trust

Our Approach

Real-Time

- Assembling the data
- Pre-processing the data
- Data Mining Task
- Interpreting the results

Uncertainty

- Accuracy
- Efficiency
- Usability

Accuracy

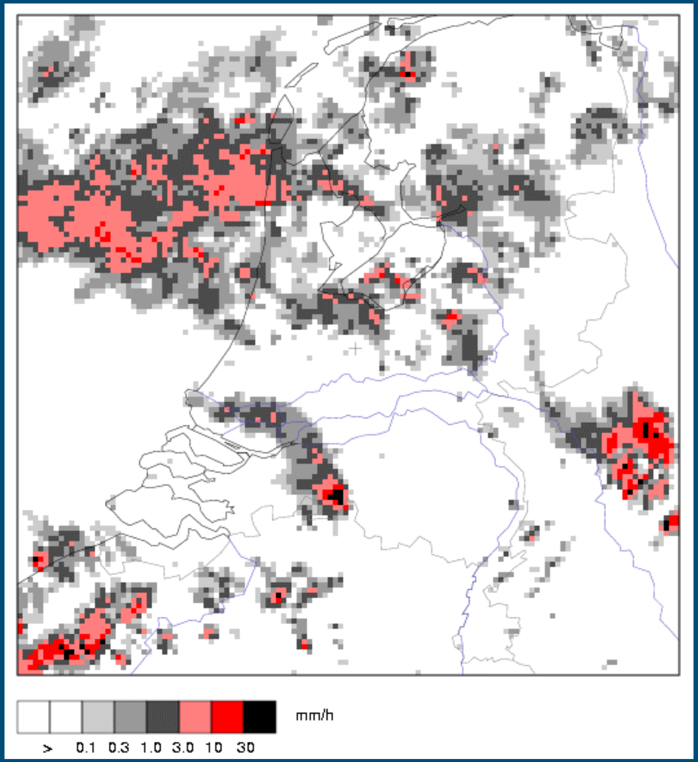
- Pre-Processing:

  The observations were clustered using the density based algorithm (Ester et al. 1996, Ester et al. 1998).

  - Clusters (patterns) → C,S (confidence, support)
  - Consistent patterns (high levels of support) separated from unique patterns (low levels of support - anomaly)

Target Data Set

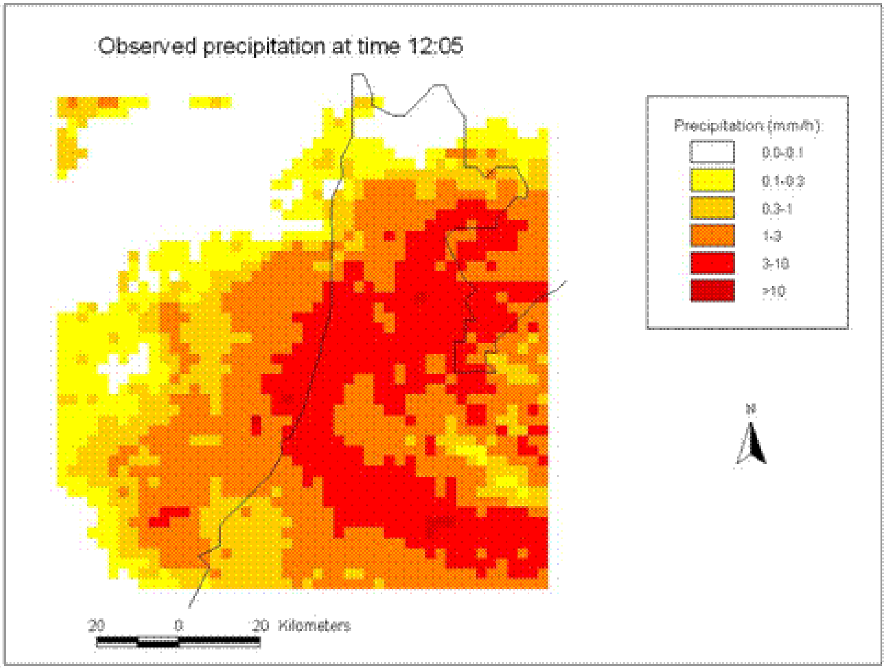
- Pseudo – Constant Altitude Plan Position Indicator
- every 5 minutes
- 19 September 2001
- Rainfall of 110 mm

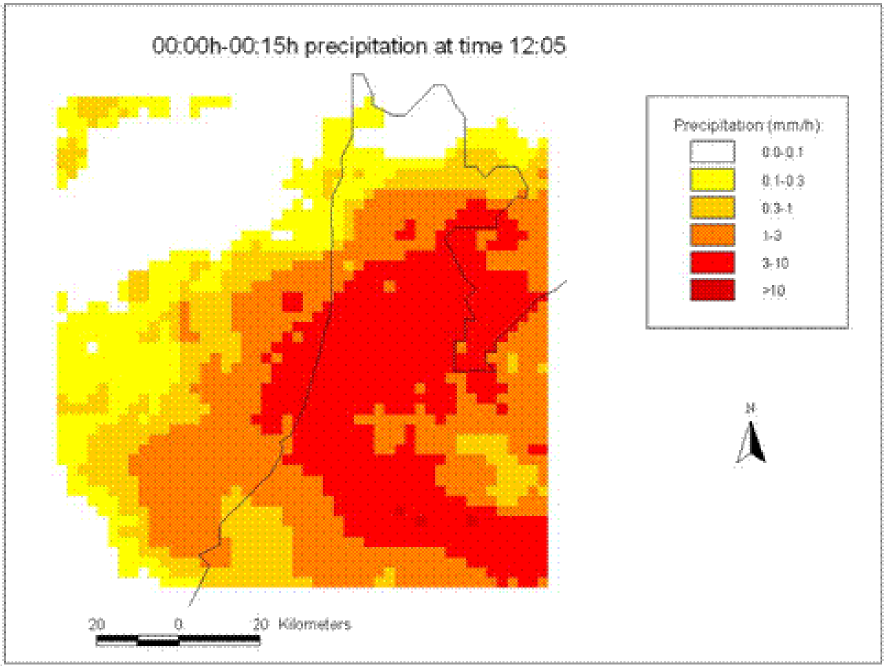


- pCAPPi image of the Netherlands and surrounding area (source: KNMI, 2004)

ALTERRA WAGENINGEN UR

Finding the clusters in time





Observed rainfall

Clustered

19 September 2001

ALTERRA
WAGENINGEN UR

Accuracy

- Data Mining Task:

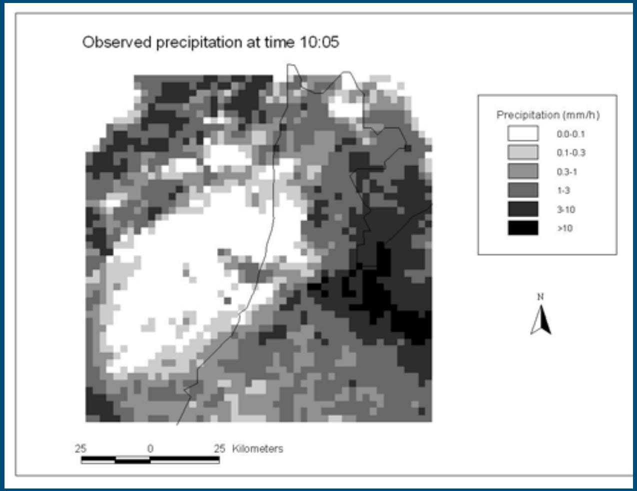
  Prediction of changes in the precipitation clusters using the supervised inductive learning method

  (Rao and Rao 1993).

  - Misuse rules
  - Anomaly detection rules
  - Normal behaviour rules

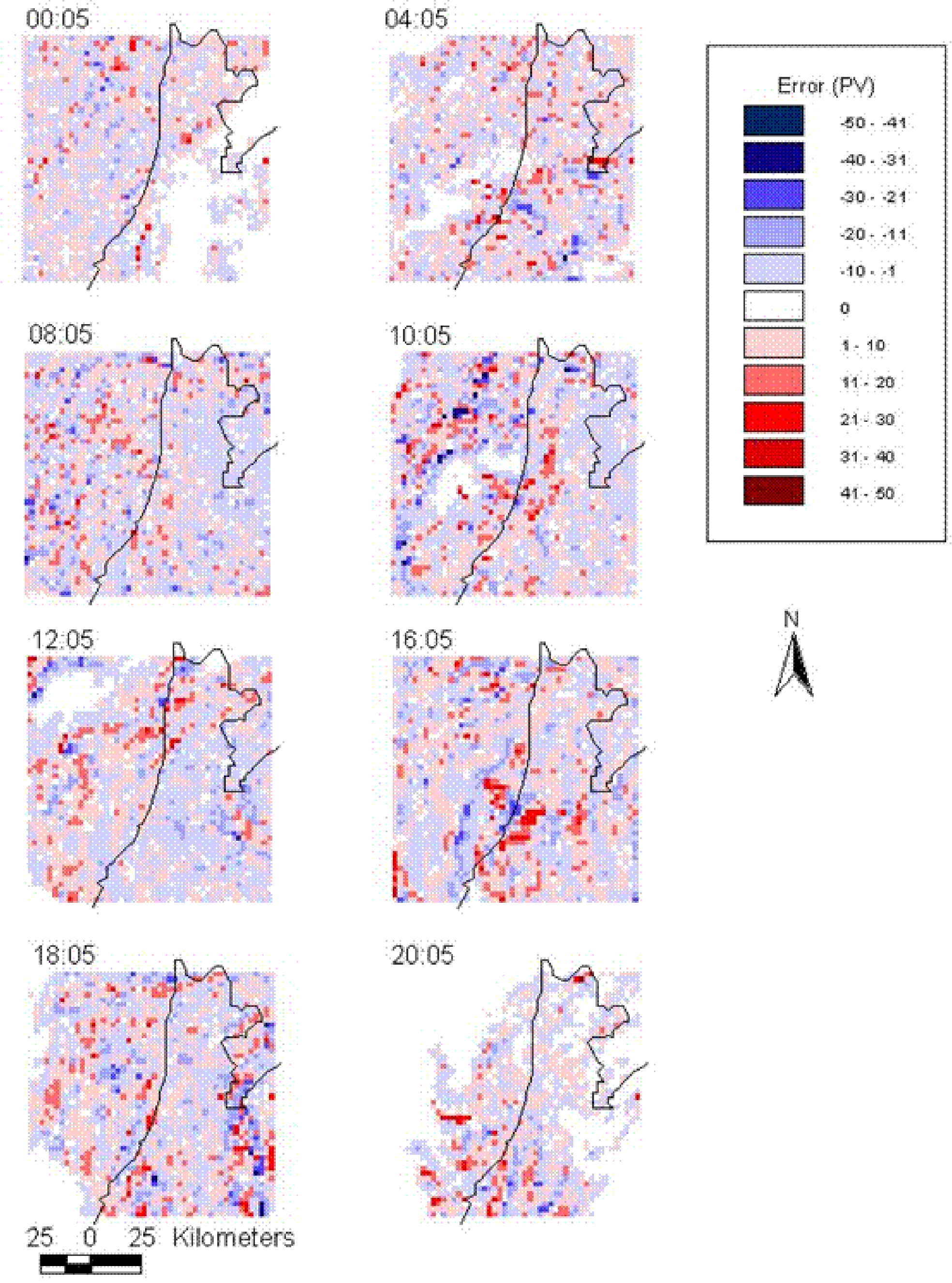
Data Mining

- Within a cluster:
  - Observed pixel values – 00:00, 00:15, 00:30 until 11:45
  - Predicted pixel values – 00:05, 00:10, 00:20 until 11:40
  - Validation – 12:00h until 23:45h



Results

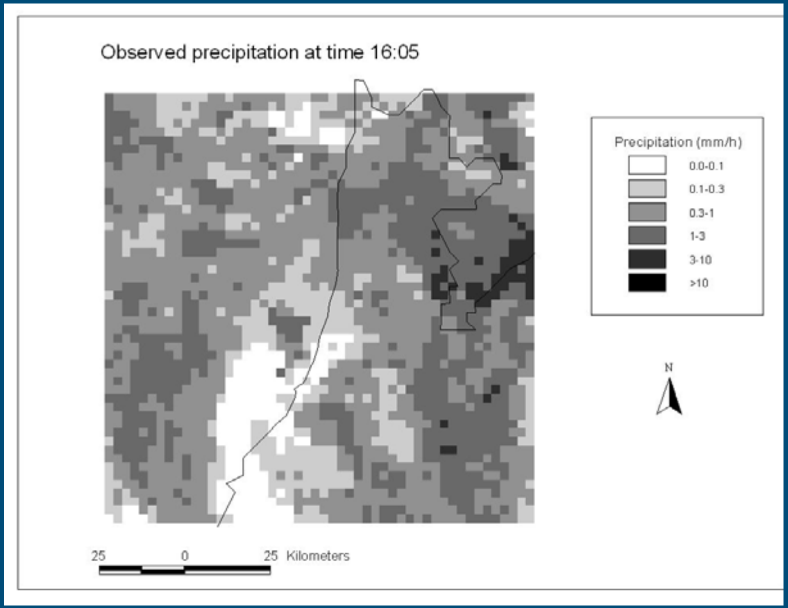
Error maps obtained from observations between 00:00h-00:15h



ALTERRA WAGENINGEN UR

Efficiency

- Computational Costs (time)
- How much does a rule cost?



Efficiency

- Rules can be computed:
  - From the first observation …………………. **Level 1**
  - During the capture of observations ………. **Level 2**
  - At the end of the observations ……………. **Level 3**
  - At the end, but requires access to data from potentially many other prior observations ... **Level 4**

Efficiency

- Cost Model : assigning weights to the levels
  - Level 1 : 1
  - Level 2 : 5
  - Level 3 : 10
  - Level 4 : 100

Efficiency

- Cost Model : defining multiple rules to generate the training sets.
  - Multiple training sets
  - Precision measures
  - Threshold values

Results

- This approach can reduce the computational cost by as much as 80% without compromising predictive accuracy.

Conclusions

- We were capable of correctly estimating the precipitation rates up to 12 hours ahead.
- We achieved an accuracy of 72% within the time interval 12:00h – 12:15h.
- The approach was suitable for mining the temporal dimension, but further studies are required for mining the spatial dimension.
- More research is needed for investigating usability issues.

ALTERRA
WAGENINGEN UR