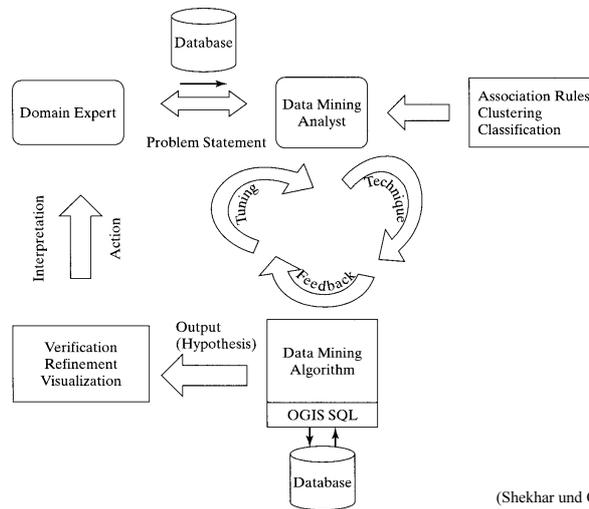


Räumliches Data Mining

Spatial Data Mining

- **Data Mining** = Suche nach "interessanten Mustern" in sehr großen Datensätzen => explorative Datenanalyse
- auch: **Knowledge Discovery in Databases (KDD)**
- verbreitete Anwendungen:
 - Analyse von Kundenprofilen ("Payback-Card" u.ä.)
 - Beurteilung der Kreditwürdigkeit
- Beispiel:
 - "Käufer, die nachmittags Windeln kaufen, kaufen oft auch gleichzeitig Bier" => genutzt für Platzierung der Waren

Data Mining



(Shekhar und Chawla 2003)

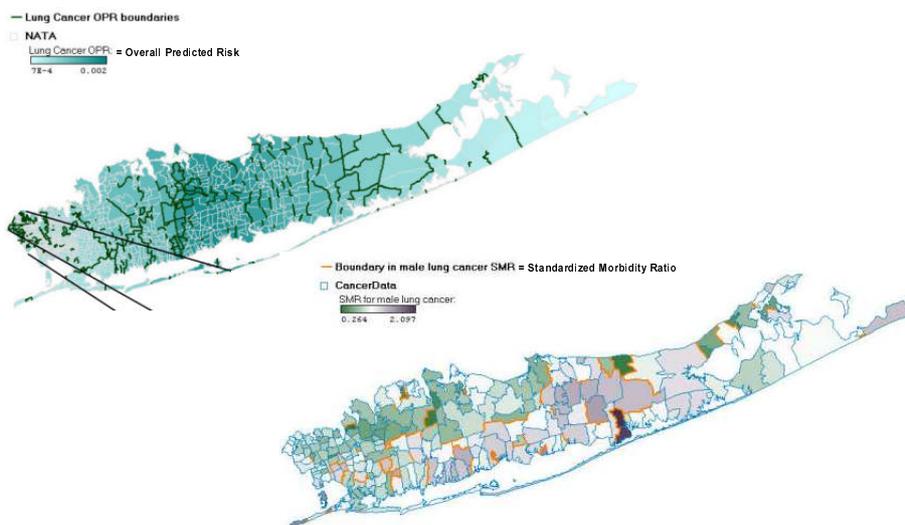
Räumliche Korrelation

Historische Beispiele für Spatial Data Mining

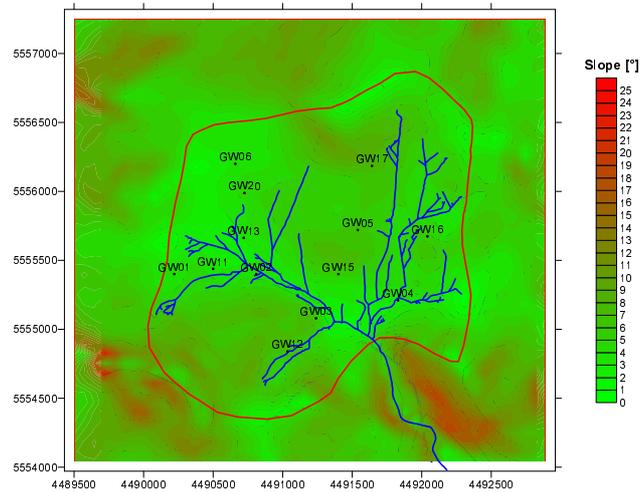
(Räumliche Korrelationen)

- Identifizierung eines kontaminierten Brunnens in London während einer Cholera-Epidemie (1855)
- Zusammenhang zwischen hohem natürlichem Fluorid-Gehalt und guter Zahngesundheit in Colorado Springs (1909)

Lungenkrebs auf Long Island (Jacquez and Greiling 2003)



Hangneigung (Gradienten)



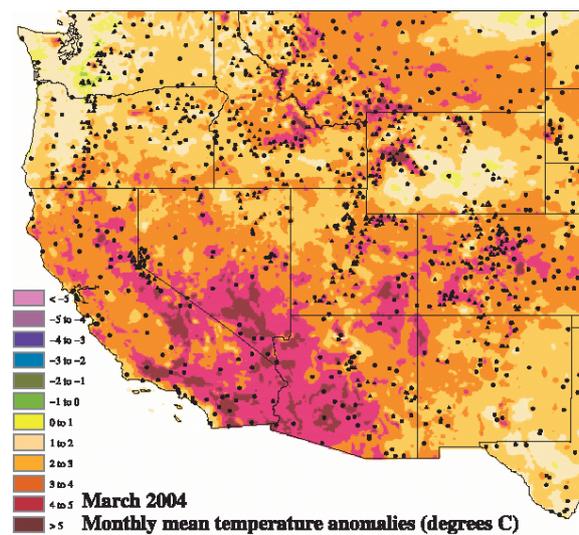
Verschneidung von Karten

Suche nach Zusammenhängen, wenn Ursache(n) und Wirkung

- nur lokal zusammen wirken,
- durch unterschiedliche Raumskalen gekennzeichnet sind,
- räumlich (und zeitlich) getrennt sind (Transportvorgänge),
- über anisotrope Beziehungen verknüpft sind,
- durch Fokussierung der Stoff- oder Energieflüsse verstärkt werden,
- ...

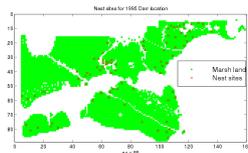
Räumlich beschränkter (lokaler) Zusammenhang

Beispiel: Lokale zeitliche Trends (Pagano et al. 2004)

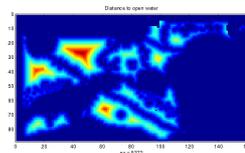


Unterschiedliche Raumskalen von Ursache und Wirkung

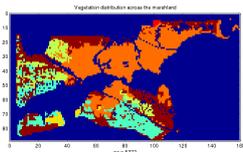
Brutplätze von Wasservögeln (S. Shekhar)



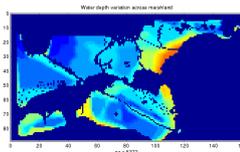
Nest locations



Distance to open water



Vegetation durability

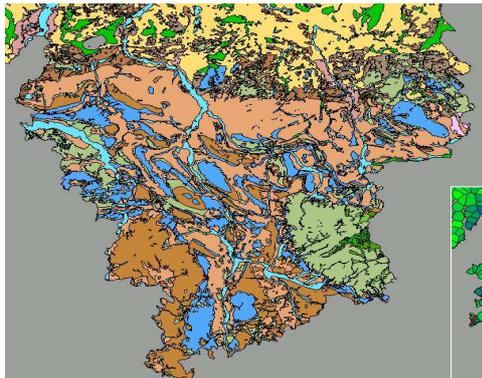


Water depth

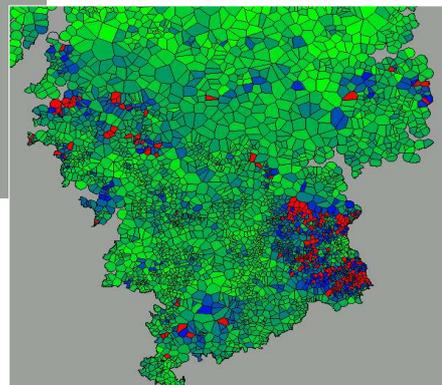
(<http://www.cs.umn.edu/research/shashi-group/Project/index.html>)

Räumliche Trennung von Ursache und Wirkung

Schwermetalle in Bachsedimenten



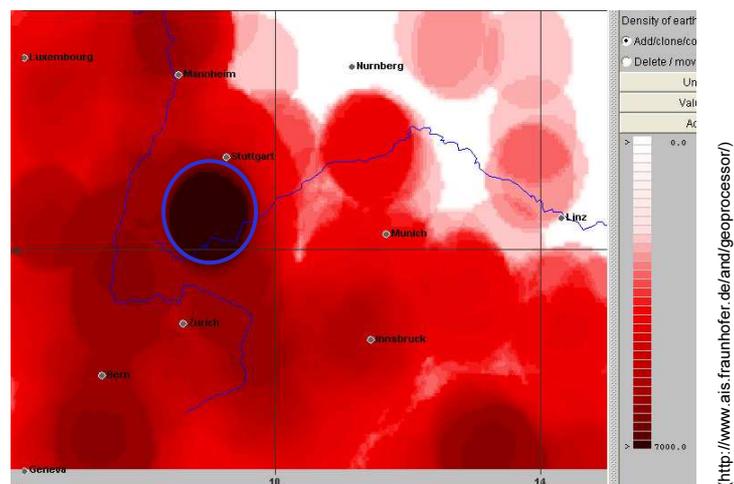
Südniedersachsen
(NLfB 2003)



Fokussierung der Stoff- oder Energieflüsse

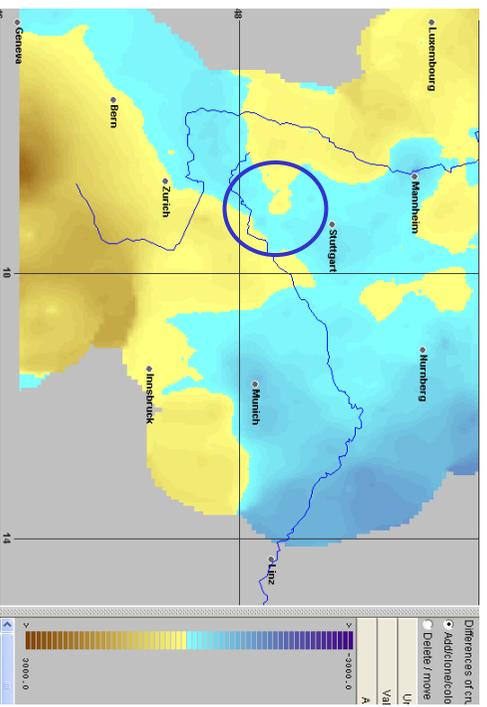
Erdbebenhäufigkeit

Intensität der Erdbeben ($M > 2$) in 50 km Umkreis von 365 bis 1979



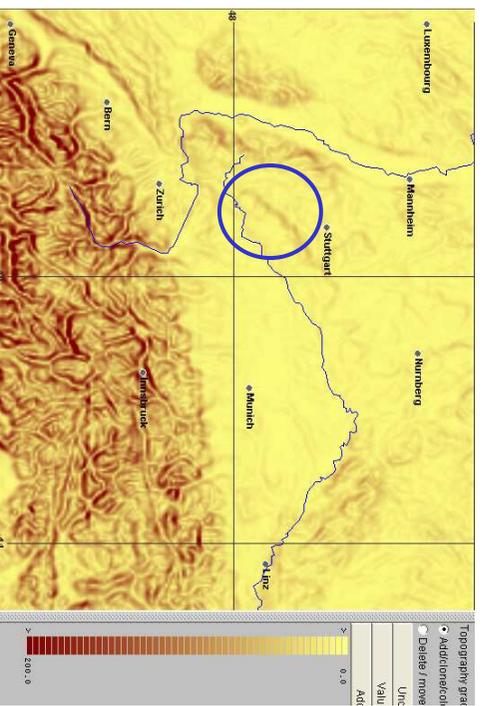
Bewegung der Erdkruste

Aktuelle vertikale Bewegung der Kruste



(<http://www.ais.fraunhofer.de/and/geoprocessor/>)

Topografischer Gradient



(<http://www.ais.fraunhofer.de/and/geoprocessor/>)

"Automatisierte" Verfahren

erschwert durch statistische Besonderheiten von Geodaten:

- räumliche Daten sind oft nicht normal- oder gleich-verteilt
- Einzeldaten sind i.d.R. räumlich autokorreliert (nicht unabhängig)
- oft liegen räumliche Trends vor
- Zusammenhänge sind oft nicht linear
- statistische Zusammenhänge sind oft lokal beschränkt

Spatial Data Mining

Prinzip: Identifizierung "Interessanter Muster" in großen Datensätzen durch

1. Assoziation / Korrelation
2. Gruppierung
3. Klassifizierung
4. Vorhersage

mittels unterschiedlicher Methoden

"Automatisierte" Verfahren: Beispiele

für (sehr) große Datensätze:

Genetische Algorithmen (GA)

- nicht-lineare Korrelation
- Vorhersage

Multi-layer Perceptrons (MLP) (Künstliche Neuronale Netze)

- nicht-lineare Korrelation
- Vorhersage

Classification and Regression Tree (CART)

- Regression
- Klassifizierung
- Vorhersage

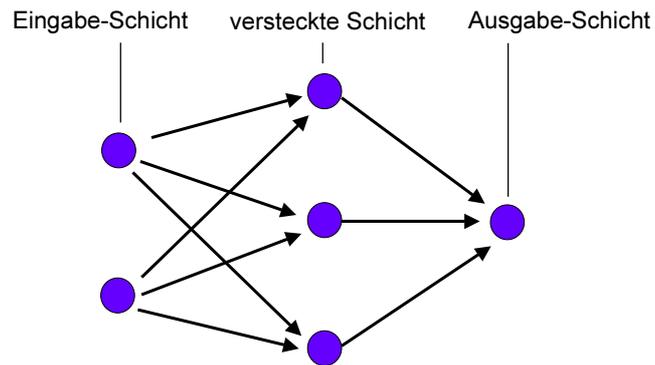
Genetische Algorithmen

Iteratives Verfahren:

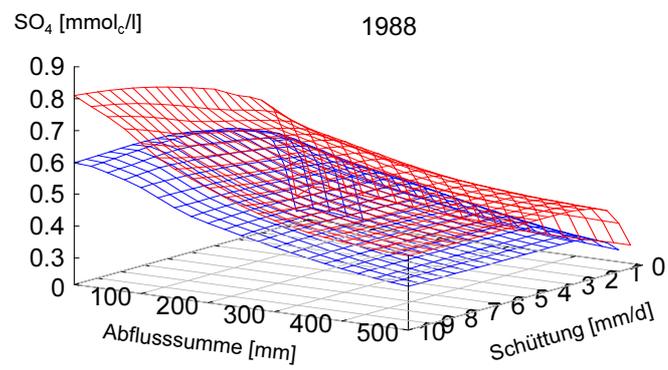
1. Start mit zufälliger Ausgangspopulation
2. Selektion der "fittesten" Algorithmen
3. Kreuzung (paarweise Kombination) von Algorithmen
4. Mutation: zufällige Variationen der Algorithmen
5. zurück zu 2.

Mehrschicht-Perzeptron

(2:3:1-Mehrschicht-Perzeptron)



Multivariate, nicht-lineare Regression

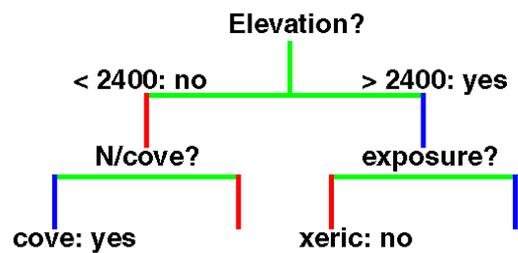


CART

= Classification and Regression Tree

Prinzip: Schrittweise binäre Aufteilung des Datensatzes, optimiert hinsichtlich der Vorhersage (numerische Eingrenzung) einer Variablen

Beispiel: Verbreitung einer montanen Pflanzenart



(www.env.duke.edu/lel/env214/le_strategy.html)

Data Warehouse

- "Daten-Kaufhaus": Daten einem breiten Anwenderkreis zugänglich machen

- oft Internet-basiert

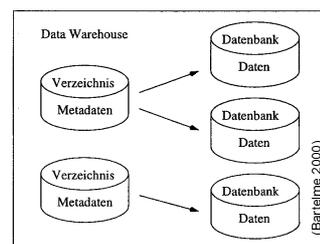
- Metadaten: **Data Directory**

- Anwendungen:

- Data Mining

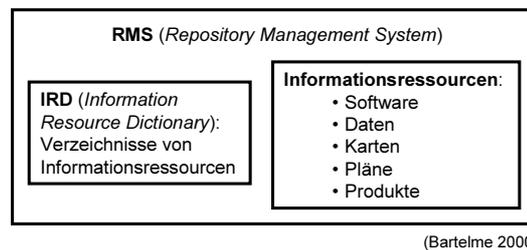
- Online Analytische Prozesse (**OLAP**): zur-Verfügung-stellen bestimmter Rechenoperationen auf der Anwenderseite

- Agenten-Technologie: Lösung von Recherche-Aufgaben anhand eines best. Benutzerprofils



Data Repository

- offen zugängliche Datenbank mit Informationen über Software, Daten, Dokumenten, etc., die ein Unternehmen produziert oder verwendet
(Mitschang 1998)
- Data Warehouse als Spezialfall eines Repository



Literatur

- Adam, N.R., Gangopadhyay, A. (1997): Database issues in geographic information systems. Kluwer Academic Publishers, Boston → *schnelle Übersicht*
- Bartelme, N. (2000): Geoinformatik. Modelle, Strukturen, Funktionen. 3. Auflage, Springer, Berlin → *gut verständliche Darstellung der theoretischen Aspekte*
- Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (1999): Geographic Information Systems. Vol. 1: Principles and technical issues. 2. Auflage, John Wiley & Sons, New York → *sehr umfangreich*
- Shekhar, S., Chawla, S. (2003): Spatial databases. A tour. Prentice Hall, New Jersey
- Zeiler, M. (1999): Modeling our world. The ESRI guide to geodatabase design. ESRI Press → *plastische Einführung, ArcInfo-orientiert*