Evaluation of Alternative Conceptual Models Using Interdisciplinary Information: An Application in Shallow Groundwater Recharge and Discharge

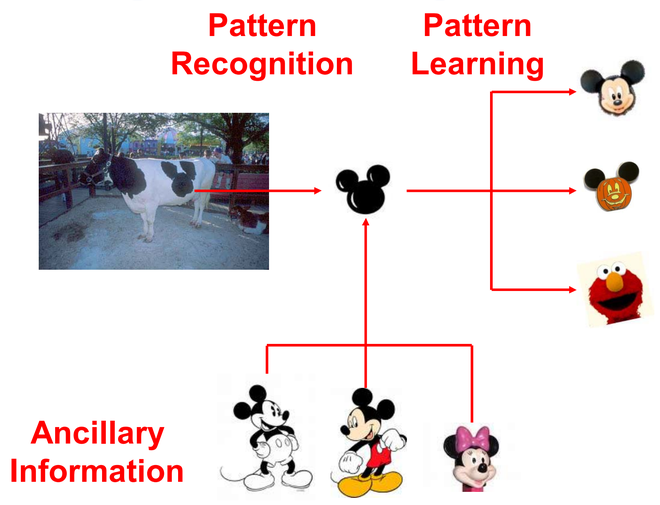
Paper # H31G-0738, Abstract 969

Yu-Feng Lin[1], Peter Bajcsy[2], Albert J. Valocchi[3], Chulyun Kim[2], and Jihua Wang[3], ({yflin, pbajcsy, ckim5, valocchi, jwang41}@uiuc.edu)
[1] Illinois State Water Survey, [2] National Center for Supercomputing Applications, [3] Department of Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign
Sponsors: National Center for Supercomputing Applications - Faculty Fellow Program , U.S. Geological Survey / National Institutes for Water Resources - National Competitive Grants

Alternative Conceptual Models

Natural systems are complex, thus extensive data are needed for their characterization. However, data acquisition is expensive; consequently we develop models using sparse, uncertain information. When all uncertainties in the system are considered, the number of alternative conceptual models is large. Traditionally, the development of a conceptual model has relied on subjective professional judgment. Good judgment is based on experience in coordinating and understanding auxiliary information that is correlated to the model but difficult to quantify in a mathematical model. For example, groundwater recharge and discharge (R&D) processes are known to relate to multiple information sources such as soil type, river and lake location, irrigation patterns and land use. Although hydrologists have been trying to understand and model the interactions between each of these information sources and R&D processes, it is extremely difficult to quantify their correlations using a universal approach due to the complexity of the processes, the spatiotemporal distribution and uncertainty. There is currently no single method capable of estimating R&D rates and patterns for all practical applications. Chamberlin (1890) recommended use of "multiple working hypotheses" (alternative conceptual models) for rapid advancement in understanding of applied and theoretical problems. Therefore, cross-analyzing R&D rates and patterns from various estimation methods and related field information will likely be superior to using only a single estimation method.

The Concept of Pattern Recognition and Learning for Evaluating Alternative Models



Cross-Analysis Using Ancillary Information

A GIS plug-in package, PRO-GRADE, was developed to help hydrogeologists estimate R&D in a more efficient way than conventional methods. The Pattern Recognition Organizer (PRO-GIS) in the PRO-GRADE package uses numerical methods and image processing algorithms to estimate and visualize shallow R&D patterns and rates with GIS. PRO-GRADE includes (but is not limited to) a Groundwater Recharge And Discharge Estimator (GRADE-GIS) using a finite difference mass balance approach in 2D and steady state. GRADE-GIS only requires data for water table, bedrock elevations and hydraulic conductivities. It can provide a fast initial estimate prior to planning labor-intensive and time-consuming field R&D measurements.

Furthermore, the Spatial Pattern to Learn (SP2Learn) was developed to cross-analyze results from PRO-GRADE with ancillary field information, such as land coverage, soil type, topographic maps and previous estimates. The learning process of SP2Learn cross-examines each initially recognized R&D pattern with the ancillary spatial dataset, and then calculates a quantifiable reliability index for each R&D map using a supervised machine learning technique called Decision Tree. This JAVA-based software package is capable of generating alternative R&D maps if the user decides to apply certain conditions recognized by the learning process. The reliability indices from SP2Learn will improve the traditionally subjective approach to initiating conceptual models by providing objectively quantifiable conceptual bases for further probabilistic and uncertainty analyses.

Both SP2Learn and PRO-GRADE have been designed to be user-friendly and universal utilities for pattern recognition and learning to improve model predictions from sparse measurements by computer-assisted integration of spatially dense geospatial image data and machine learning of model dependencies.

Acknowledgements

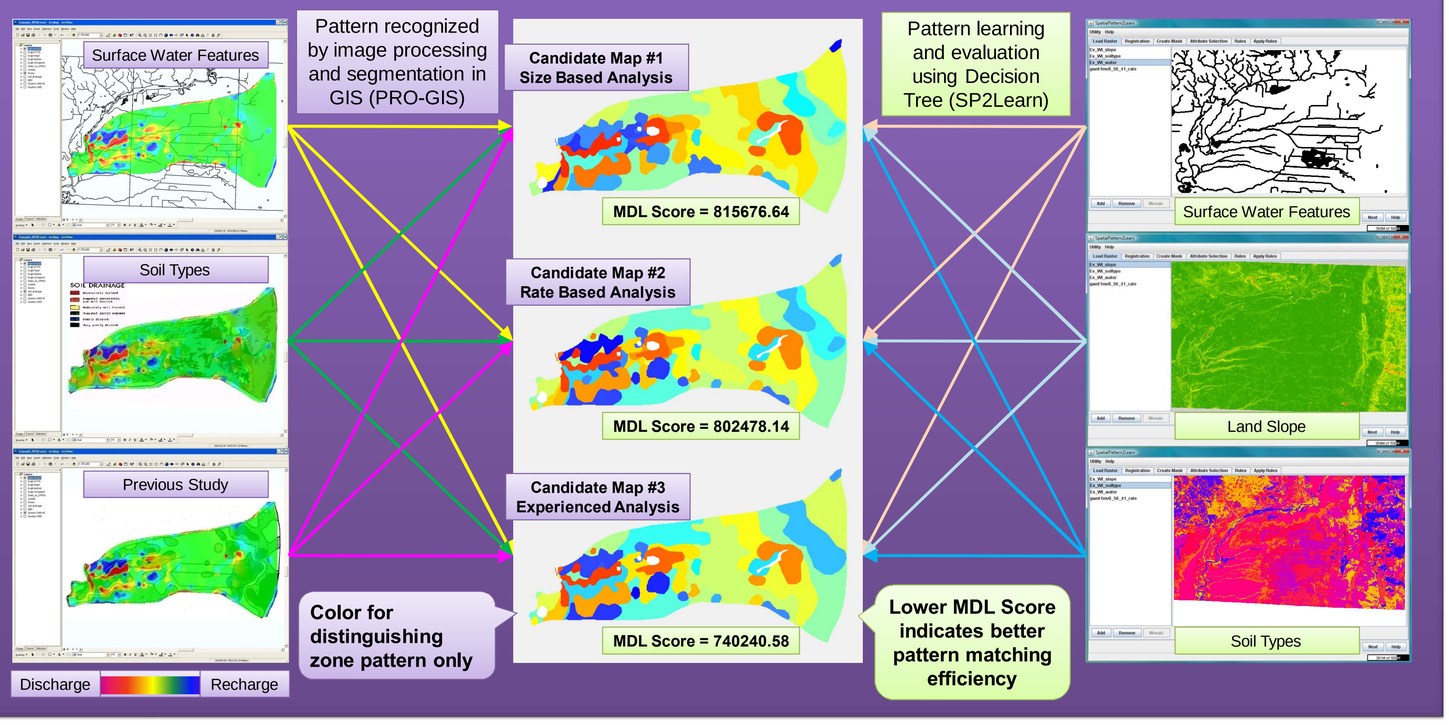
We appreciate Dr. Randall J. Hunt from U.S. Geological Survey (WRD, Middleton, Wisconsin), Dr. Ming Ye from Florida State University and the other beta testers for helping us to improve this software package. Data for the example case were provided by the US Geological Survey, Wisconsin Water Science Center, Middleton, Wisconsin; the Wisconsin Geological and Natural History Survey and the Central Wisconsin Groundwater Center of the University of Wisconsin – Extension.

References

Stoertz, M.W., and K.R. Bradbury (1989), Mapping recharge areas using a groundwater flow model – a case study, Ground Water, vol. 27, no. 2, p.p. 220-228.

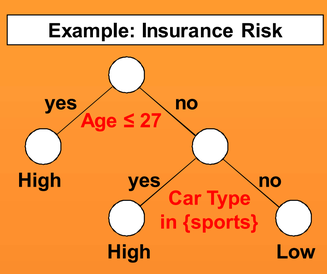
Lin, Y-F., and M.P. Anderson (2003) A Digital Procedure for Ground Water Recharge and Discharge Pattern Recognition and Rate Estimation, Ground Water, vol. 41, no. 3, p.p. 306-315.

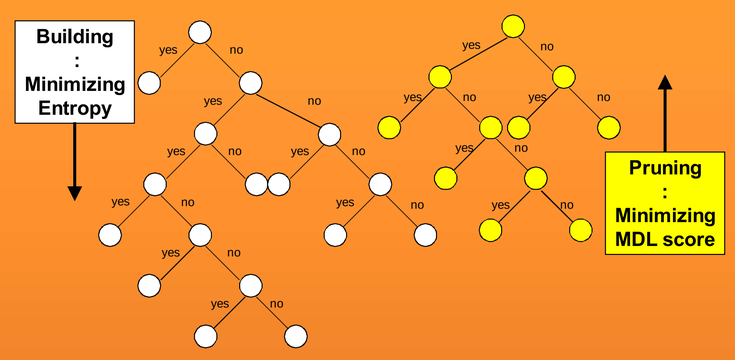
Spatial Pattern Recognition and Learning from Ancillary Information



Decision Tree: Machine Learning and Matching Efficiency

Age	Car Type	Risk
23	family	High
17	sports	High
43	sports	High
68	family	Low
32	truck	Low
20	family	High





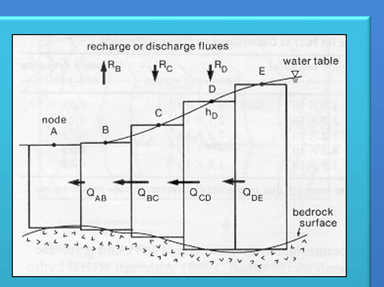
- Age ≤ 27 → Risk is high
- Age > 27 and Car Type is sports → Risk is high
- Age > 27 and Car Type is not sports → Risk is low

- Each internal node of the decision tree has a split point.
  - Split point : A test that determines how the data is partitioned
- The form of the split point
  - Continuous attribute A : name(A)<x where x is a value in the domain of A
  - Categorical attribute A : name(A) ∈ X where X ⊂ domain(A)

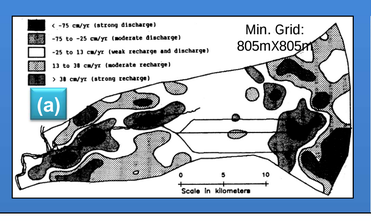
Make several decision trees and find common rules, then replace R&D criteria based on the rules. New R&D maps can be generated from the R&D criteria.

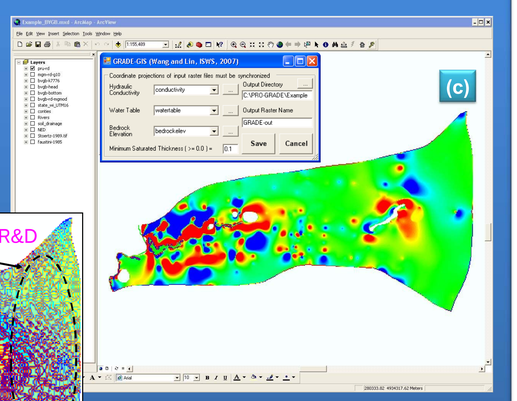
The Minimum Description Length (MDL) score for each candidate R&D map will be its quantifiable quality index.

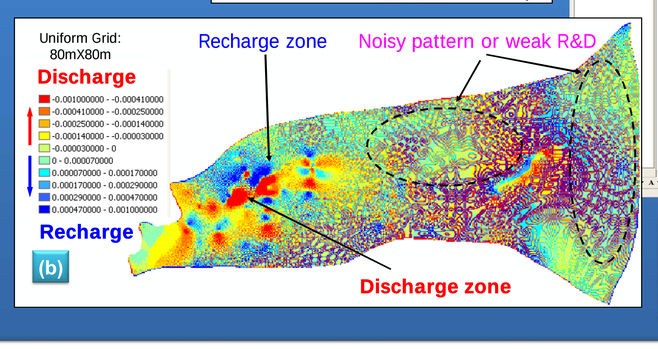
Groundwater Recharge and Discharge Estimation



Stoertz and Bradbury (1989) developed a 2D steady state R&D mapping method which modifies MODFLOW-88 to compute a water balance whereby lateral flow in and out of a cell is balanced by water that enters or leaves via groundwater recharge or discharge.



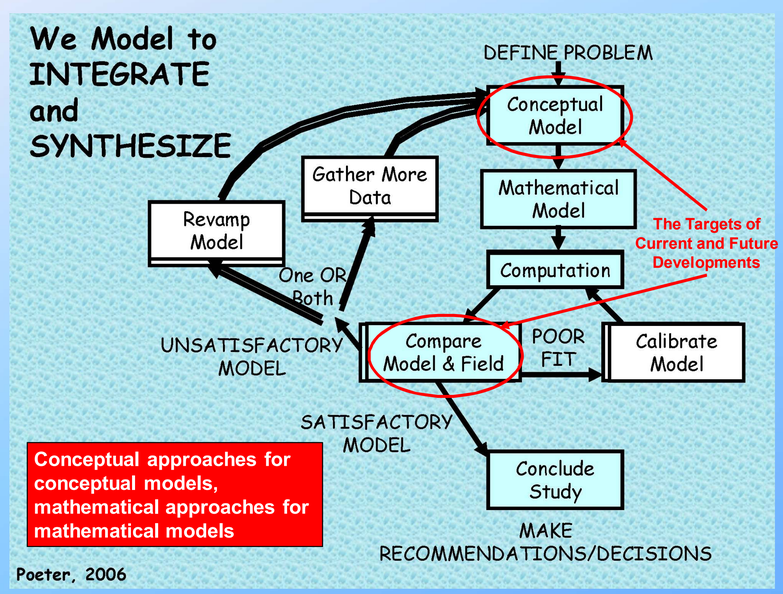




The resultant R&D maps from an modified MODFLOW code in (a) 1989, (b) 2000 and (c) a Visual Basic code as a GIS plug-in tool (GRADE-GIS)

Future Work

- Coupling machine learning algorithms with measurement uncertainty analysis in order to include uncertainties in the Decision Tree
- Extending quantitative measures to predictions and uncertainties that will bridge the gap between traditional subjective approaches for initiating conceptual models and advanced stochastic and uncertainty analysis
- Enhancing R&D estimation in 3D, pumping and transient conditions
- Exploring more applications in addition to R&D estimation, such as remote sensing images and topography patterns



Modified from National Ground Water Association Darcy Lecture 2006 by Eileen Poeter