Graphical Models as Surrogates for Complex Ground Motion Models

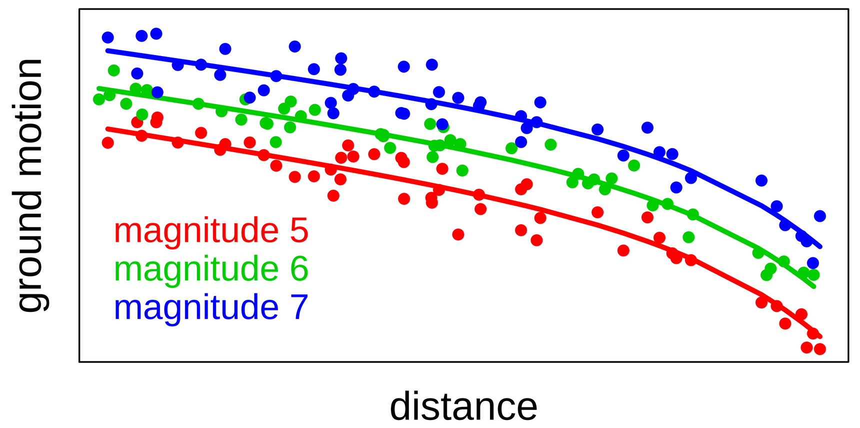
K. Vogel, C. Riggelsen, N. Kuehn, F. Scherbaum — Institute of Earth and Environmental Science, Potsdam University
contact: kvog@geo.uni-potsdam.de

1. Ground Motion Models

- describe ground motion caused by earthquakes given event and site related predictor variables
- a popular model, based on physical principles, is the so-called stochastic model [a]
- because of its complexity, a surrogate model is often used instead
- usually a regression function is fitted through data generated by the stochastic model; the functional form is derived from expert knowledge; e.g.:

$$f(M,R,SD,\kappa,V_S30) = a_0 + a_1 M + (a_2 + a_3 M)\ln\sqrt{a_4^2 + R^2} + a_5 M \cdot \ln SD + a_6 \kappa R + a_7 V_S30 + a_8 \ln SD$$

with $\kappa = \kappa_0 + t^*$, $t^* = \frac{R}{Q_0 V_{sq}}$ and $V_{sq} = 3.5\frac{km}{s}$.



- Directed Graphical Models are a viable alternative, having several advantages, e.g.
  - allow inference in all directions; all conditional distributions of interest can be calculated
  - prior knowledge can, but does not have to be included; no prior assumptions about functional form and physical relationships required
  - can learn the dependency structure of the variables from the data
  - can deal with missing values

[a] Boore, D.M.: Simulation of ground motion using the stochastic method. Pure and Applied Geophysics 160, 635-676 (2003)

2. Data set

Using a sample of the predictor variables

M - moment magnitude of the earthquake,
R - distance between source and site,
SD - stress released during the earthquake,
Q_0 - attenuation of seismic wave amplitudes in deep layers,
κ_0 - attenuation near the surface,
V_S30 - average shear-wave velocity in upper $30m$

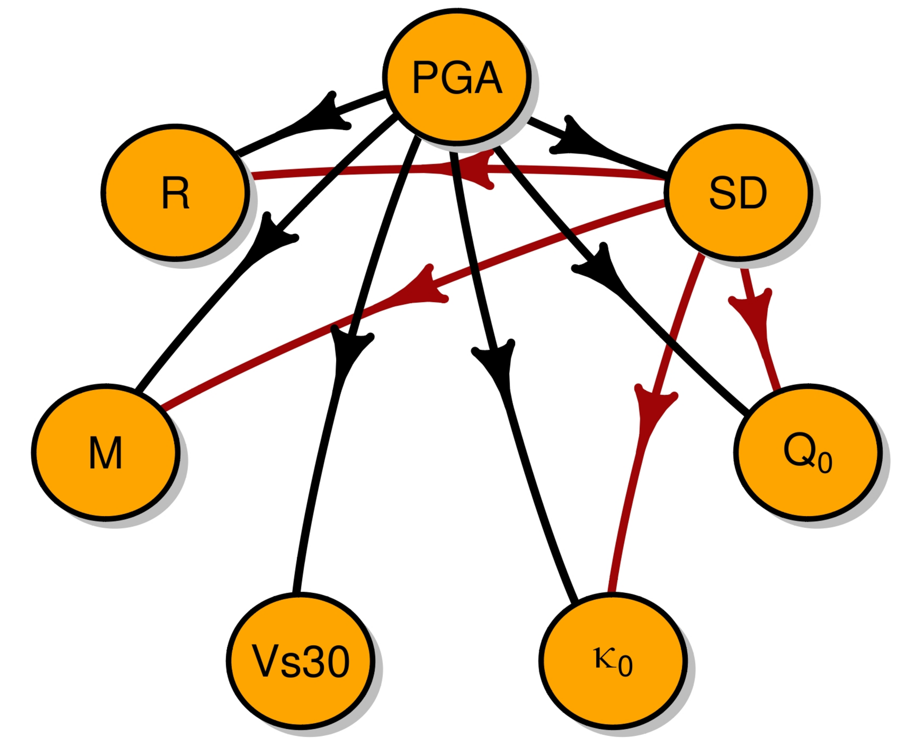
we apply the stochastic model to generate corresponding ground motion values, i.e.

PGA - the peak ground acceleration.

3. Graphical Models

A Graphical Model describes a joint probability distribution, e.g. $P(PGA, M, R, SD, Q_0, \kappa_0, V_S30)$, decomposing it into a product of (local) conditional probability distributions according to a directed acyclic graph, which encodes the conditional independences. We investigate graphical models admitting to different factorizations:

(Tree Augmented) Naive Bayes



NB: $P(M|PGA)\,P(R|PGA)\,P(SD|PGA)\,P(Q_0|PGA)\,P(\kappa_0|PGA)\,P(V_S30|PGA)\,P(PGA)$

Naive Bayes (black edges) and Tree Augmented Naive Bayes (black and red edges) learned from synthetic seismic data set

Naive Bayes (NB)

- all variables depend only on the target variable
  $\Rightarrow$ attributes are assumed to be independent
- perform well in classification tasks
- computationally simple

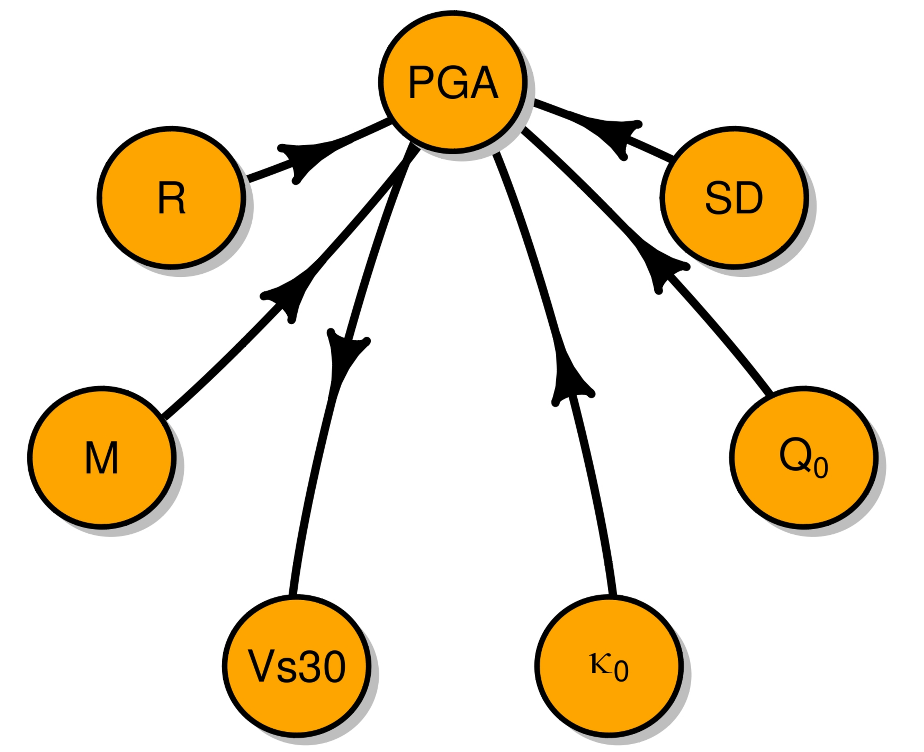
Tree Augmented Naive Bayes (TAN)

- each attribute can depend on one more variable in addition to the target variable
- relaxes independence assumption of Naive Bayes

Bayesian Networks (BN)

- structure learned from data
- no assumptions about dependencies required
- gives intuition about dependency structure of the underlying system

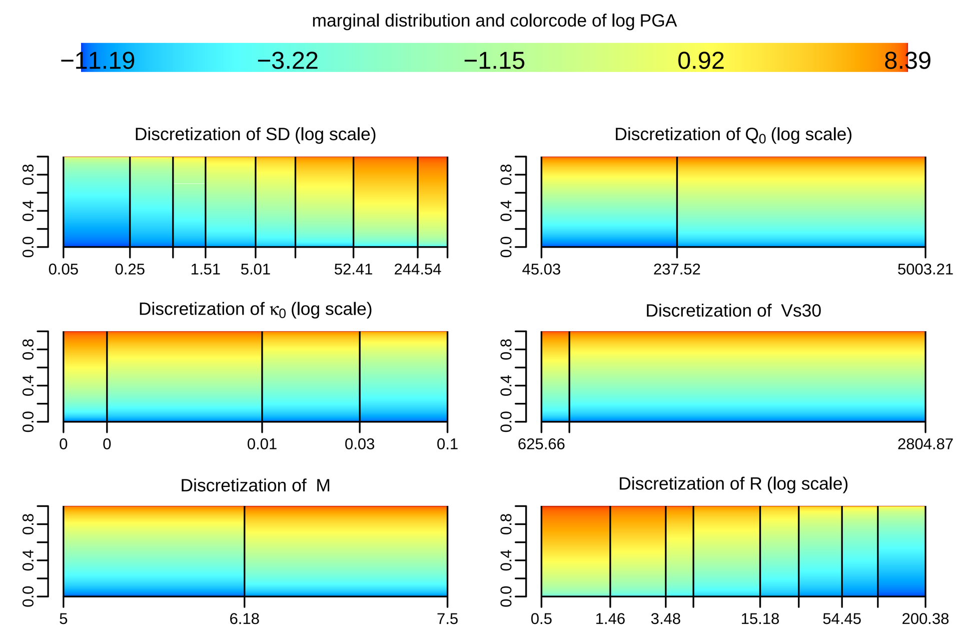
Bayesian Network



$P(M)\,P(R)\,P(SD)\,P(Q_0)\,P(\kappa_0)\,P(V_S30|PGA)\,P(PGA|M,R,SD,Q_0,\kappa_0)$

Bayesian Network learned from the synthetic seismic dataset of the stochastic model

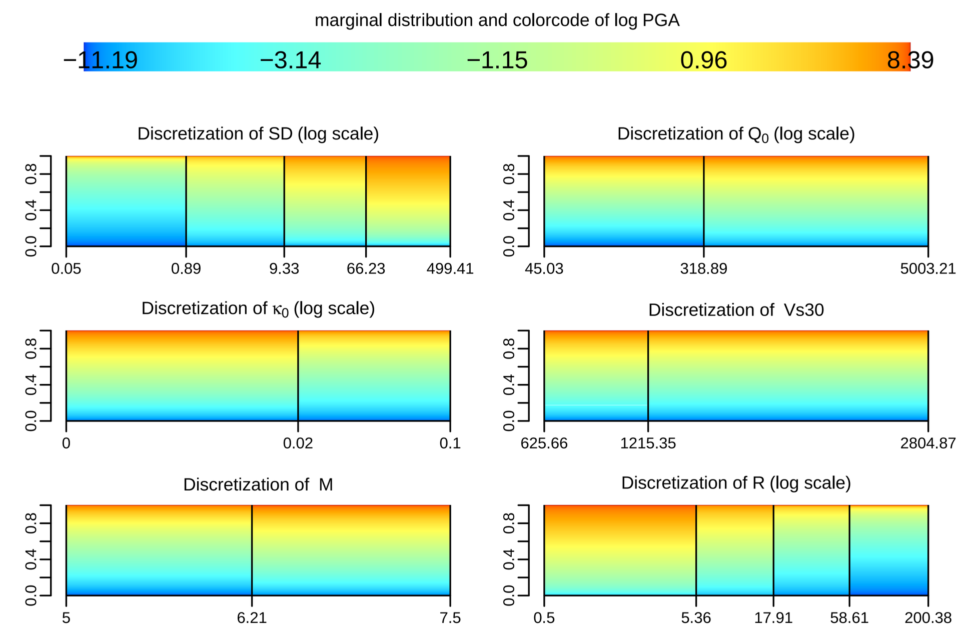
4. Automatic Discretization



Discretization of attributes and conditional distribution of PGA (color coded) for (Tree Augmented) Naive Bayes

- continuous variables are discretized to allow for distribution-free learning
- automatic discretization process, considering graph structure and distribution of the variables for discretization
- for (Tree Augmented) Naive Bayes *class entropy* is used[a]
- for Bayesian Networks, *Minimal Description Length* is applied[b]
- the methods are extended to allow for the usage of a continuous target variable

[a] Fayyad, U.M., Irani, K.B.: Multi-interval discretization of continuous-valued attributes for classification learning (1993)
[b] Friedman, N., Goldszmidt, M.: Discretizing Continuous Attributes While Learning Bayesian Networks. In Proc. ICML (1996)



Discretization of attributes and conditional distribution of PGA (color coded) found for Bayesian Network

5. Results

The three Graphical Models and a regression model (derived from expert knowledge) are compared in a 5-fold cross validation.

- on average all three Graphical Models perform better than the regression model
- without any prior knowledge about the model structure, the learned Bayesian Network enables for a correct interpretation of the (in)dependences between the variables
- the Bayesian Network detects a strong dependency of the peak ground acceleration on magnitude, distance, stress drop and κ_0

Mean squared errors of the prediction of $ln\ PGA$

	Regression	NB	TAN	BN
1.	0.666	0.488	0.509	0.569
2.	0.679	0.489	0.525	0.598
3.	0.688	0.566	0.566	0.583
4.	0.681	0.592	0.597	0.759
5.	0.650	0.473	0.494	0.579
Avg.	0.673	0.522	0.538	0.617

This poster participates in OSP

Outstanding Student Poster Contest