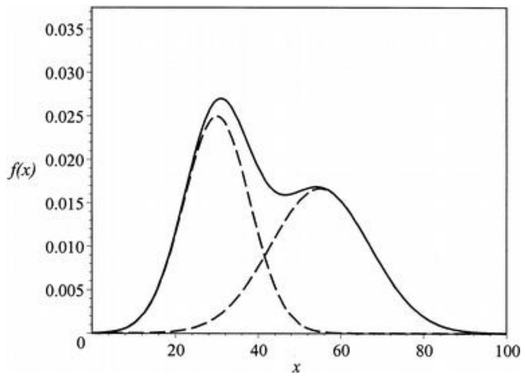
Semi-parameter clustering of well logs in the Fort Worth Basin, Texas

Bradley C. Wallet and Roderick Perez, ConocoPhillips School of Geology and Geophysics, The University of Oklahoma.

Summary

Crossplots have proven a valuable tool for analyzing well logs and seismic attributes. However, typical analysis involves the interpreter imposing his view upon the data as to what natural divisions and clusters are present in the data. Our work is focused upon data driven methods of probabilistic clustering to discover latent classes embedded in data sets. To this end, we use and entirely driven approach where clusters are defined using an iterative algorithm to find a maximum likelihood solution. Furthermore, we use a penalized likelihood criteria that allows us to compare models with different numbers of clusters. We apply these methods to Lame parameters calculated from well logs in the Lower Barnet Shale.

Mixture models



$$f(x) = \sum_{k=1}^{G} \omega_k \phi_k(x)$$

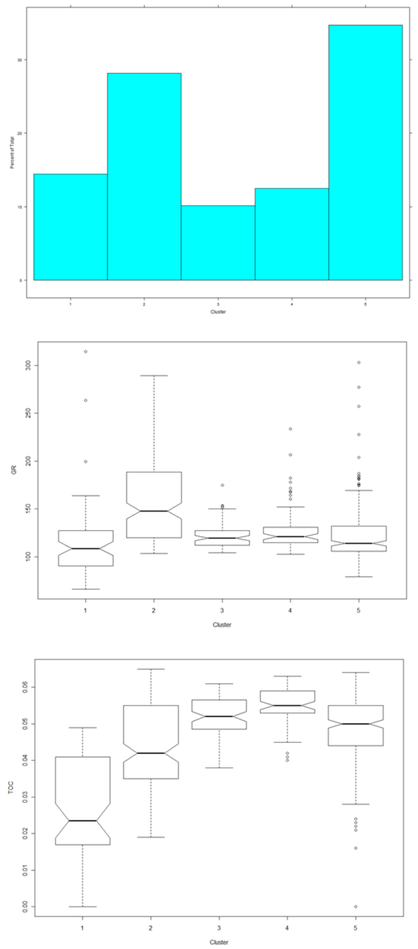
BIC

Adding terms (clusters) cannot decrease likelihood since associated weight of new term could be set to zero. Hence, we apply a complexity penalty which must be overcome by increased model fit.

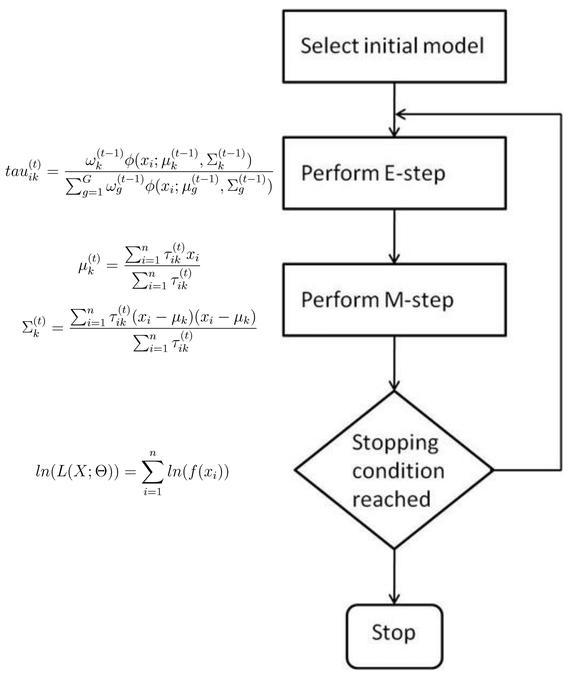
$$\mathrm{BIC} = -2ln(L) + ln(n)\kappa$$

We thus construct the best model we can find for different numbers of clusters. The different models are then compared based upon BIC, and the best, most parsimonious model is selected.

Analysis of clusters



EM Algorithm



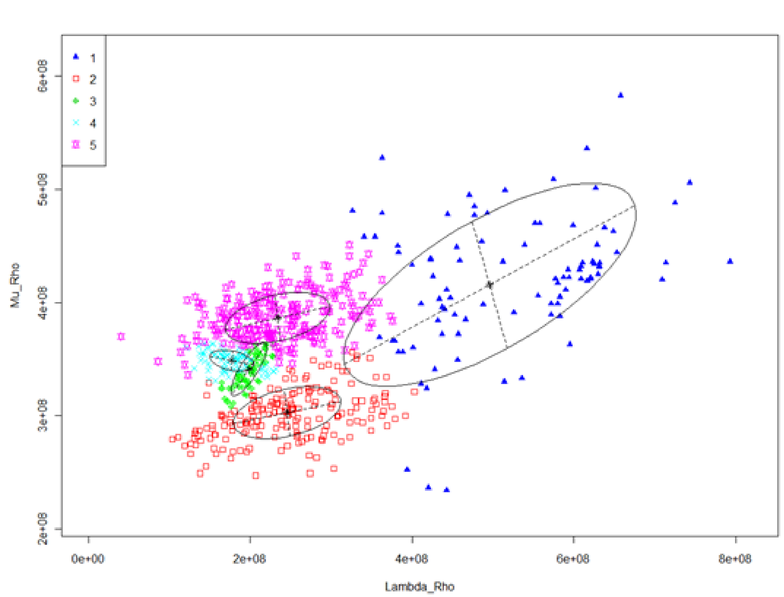
$$tau_{ik}^{(t)} = \frac{\omega_k^{(t-1)} \phi(x_i; \mu_k^{(t-1)}, \Sigma_k^{(t-1)})}{\sum_{g=1}^{G} \omega_g^{(t-1)} \phi(x_i; \mu_g^{(t-1)}, \Sigma_g^{(t-1)})}$$

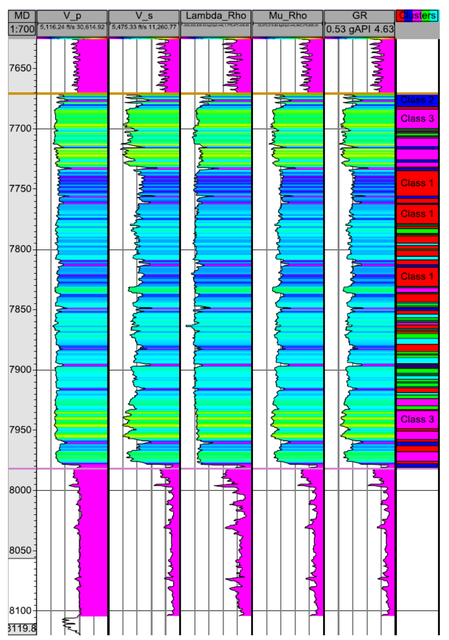
$$\mu_k^{(t)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)} x_i}{\sum_{i=1}^{n} \tau_{ik}^{(t)}}$$

$$\Sigma_k^{(t)} = \frac{\sum_{i=1}^{n} \tau_{ik}^{(t)} (x_i - \mu_k)(x_i - \mu_k)}{\sum_{i=1}^{n} \tau_{ik}^{(t)}}$$

$$ln(L(X;\Theta)) = \sum_{i=1}^{n} ln(f(x_i))$$

Clusters





Conclusions

- Mixture models give us a quantitative, object method for defining latent clusters in attribute space.
- Bayesian Information Criteria (BIC) allows us to compare models of differing complexities to find a parsimonious model that well fits our data.
- This approach fits well with a number of quantitative workflows where we are trying to do unsupervised classification of data based upon seismic and well attributes.