

# Chutes and Ladders: metamorphic conditions of exhumed simulated rocks

Buchanan C. Kerswell <sup>1</sup>Matthew J. Kohn <sup>1</sup>Taras V. Gerya <sup>2</sup>

<sup>1</sup>Department of Geosciences, Boise State University, Boise, ID 83725

<sup>2</sup>Department of Earth Sciences, ETH-Zurich, Sonneggstrasse 5, Zurich 8092, Switzerland

## Key Points:

- 
- 
-

## Abstract

## 1 Introduction

## 2 Methods

### 2.1 Numerical model

### 2.2 Marker tracing

A set of markers,  $M = m_1, m_2, \dots, m_i$ , are initialized randomly in the model domain. Markers are traced by: 1) selecting all markers within a 760 *km* wide and 11 *km* deep section extending from  $x = 500$  *km* to approximately the trench, 2) save marker properties and positions, 3) if timestep  $\leq \alpha$ , 4) update markers. Where  $\alpha$  is the number of timesteps for marker tracing.

Slab rollback eventually leads to mechanical interference between the stiff convergence region centered at  $x = 500$  *km* and trench sediments. Sediments pile up against the barrier as the accretionary wedge deforms against the convergence region. Thickening trench sediments flatten the slab causing intense crustal deformation of the forearc and backarc regions. The abrupt change in dynamics makes marker P-T conditions meaningless after interference begins. The number of timesteps for marker tracing,  $\alpha$ , is chosen algorithmically by computing the topographic height of the sediment pile against the convergence region. Marker tracing stops when the sediment pile becomes the overall topographic high, usually within one or two timesteps after interference. Marker paths from timestep  $\leq \alpha$  are used for classification.

### 2.3 Marker classification

Tracing marker pressure, temperature, and x-z-position at each timestep is enough to compute characteristics of marker PTt paths. However, only markers recovered from the subducting slab are relevant for comparison to PT estimates of natural rocks. The main challenge, therefore, is to first classify markers as either *subducted* or *recovered* without an inherited class label.

At the heart of our marker classification algorithm is a finite Gaussian mixture model (GMM) fit by Expectation-Maximization (EM, Dempster et al., 1977). We derive GMM and EM below. Please note that GMM fit by EM is a general purpose clustering algo-

rithm broadly used in pattern recognition, anomaly detection, and estimating complex probability distribution functions (e.g., Banfield & Raftery, 1993; Celeux & Govaert, 1995; Figueiredo & Jain, 2002; Fraley & Raftery, 2002; Vermeesch, 2018). Before deriving the details of marker classification, we hypothesize that features computed from a PTt path may distinguish subducted from recovered markers. If true, clustering algorithms like GMM may reliably classify markers by their dissimilarity along any number of dimensions (e.g., Dy & Brodley, 2004), like maximum pressure.

### 2.3.1 Gaussian mixture model

Let the traced markers represent a  $d$ -dimensional array of  $n$  random independent variables  $x_i \in \mathbb{R}$ . Assume markers  $x_i$  were drawn from  $k$  discrete probability distributions with parameters  $\Phi$ . The probability distribution of markers  $x_i$  can be modeled with a mixture of  $k$  components:

$$p(x_i|\Phi) = \sum_{j=1}^k \pi_j p(x_i|\Theta_j) \quad (1)$$

where  $p(x_i|\Theta_j)$  is the probability of  $x_i$  under the  $j^{th}$  mixture component and  $\pi_j$  is the mixture proportion representing the probability that  $x_i$  belongs to the  $j^{th}$  component ( $\pi_j \geq 0$ ;  $\sum_{j=1}^k \pi_j = 1$ ).

Assuming  $\Theta_j$  describes a Gaussian probability distributions with mean  $\mu_j$  and covariance  $\Sigma_j$ , Equation 1 becomes:

$$p(x_i|\Phi) = \sum_{j=1}^k \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j) \quad (2)$$

where

$$\mathcal{N}(x_i|\mu_j, \Sigma_j) = \frac{\exp\{-\frac{1}{2}(x_i - \mu_j)(x_i - \mu_j)^T \Sigma_j^{-1}\}}{\sqrt{\det(2\pi \Sigma_j)}} \quad (3)$$

Estimates for parameters  $\mu_j$  and  $\Sigma_j$ , representing the center and shape of each cluster, are found by maximizing the log of the likelihood function,  $L(x_i|\Phi) = \prod_{i=1}^n p(x_i|\Phi)$ :

$$\log p(\Phi|x_i) = \log \prod_{i=1}^n p(x_i|\Phi) = \sum_{i=1}^n \log \left[ \sum_{j=1}^k \pi_j p(x_i|\Theta_j) \right] \quad (4)$$

Taking the derivative of Equation 4 with respect to each parameter,  $\pi$ ,  $\mu$ ,  $\Sigma$ , setting the equation to zero, and solving gives the Maximum Likelihood Estimators (MLE):

$$\begin{aligned}\pi_j &= \frac{N_j}{n} \\ \mu_j &= \frac{1}{N_j} \sum_{i=1}^n \omega_{ij} x_i \\ \Sigma_j &= \frac{1}{N_j} \sum_{i=1}^n \omega_{ij} (x_i - \mu_j)(x_i - \mu_j)^T\end{aligned}\tag{5}$$

where  $\omega_{ij}$  ( $\omega_{ij} \geq 0$ ;  $\sum_{j=1}^k \omega_{ij} = 1$ ) are membership weights representing the probability of an observation  $x_i$  belonging to the  $j^{th}$  Gaussian, and  $N_j = \sum_{i=1}^n \omega_{ij}$  represents the number of observations belonging to the  $j^{th}$  Gaussian. Please note that  $\omega_{ij}$  is unknown for unlabelled datasets, so MLE cannot be computed with Equation 5.

### 2.3.2 Expectation-Maximization fitting of Gaussian Mixtures

The EM algorithm estimates GMM parameters by initializing  $k$  Gaussians with parameters  $(\pi_j, \mu_j, \Sigma_j)$ , then by iteratively computing membership weights with Equation 6 (E-step), and updating Gaussian parameters with Equation 5 (M-step) until convergence (Dempster et al., 1977).

Let  $z_{ij} \in \{1, 2, \dots, k\}$  be a multinomial latent variable with joint distribution  $p(x_i, z_i) = p(x_i|z_{ij})p(z_j)$ .  $z_{ij}$  represents the unknown (unlabelled) classifications of  $x_i$  and takes values of  $j = 1, 2, \dots, k$ . Membership weights  $\omega_{ij}$  are equivalent to the conditional probability  $p(z_{ij}|x_i)$ , which represents the probability of observation  $x_i$  belonging to the  $j^{th}$  Gaussian. Using Bayes Theorem, the posterior probability  $p(z_{ij}|x_i)$  is:

$$p(z_{ij}|x_i) = \frac{p(x_i|z_{ij})p(z_j)}{p(x_i)} = \frac{\pi_j \mathcal{N}(\mu_j, \Sigma_j)}{\sum_{j=1}^k \pi_j \mathcal{N}(\mu_j, \Sigma_j)} = \omega_{ij}\tag{6}$$

which can be computed given initial estimates for  $k$  sets of Gaussian parameters  $\pi_j, \mu_j, \Sigma_j$  (E-step). With  $\omega_{ij}$ , new Gaussian estimates can be computed with Equation 5 (M-step).

EM is sensitive to local optima and initialization (Figueiredo & Jain, 2002), so a number of features were computed and tested in combination. Redundant or useless features (Dy & Brodley, 2004) were filtered out. We found a two-component mixture of  $x_i =$

max $_{1 \leq t \leq t_{max} P}$  and  $y_i = \sum_{t=1}^{t_{max}} \Delta P_t$  reliably classified markers among numerical models.  $x_i$  and  $y_i$  represent the maximum pressure attained along a marker PTt path and the sum total of all pressure changes along a marker PTt path, respectively.

We use general purpose functions in the R package **Mclust** (Scrucca et al., 2016) to fit Gaussian mixture models by EM.

### 2.3.3 Chutes or ladders?

GMM clustering typically classifies markers into 6-20 groups, so a final decision to classify a group as *subducted* or *recovered* is made by comparing the maximum marker pressure within each group to the maximum pressures of all markers. Groups that have maximum marker pressures well above the median overall marker pressure classify as *subducted*. If no groups have maximum marker pressures above 4 *GPa*, then all groups classify as *recovered*.

## 3 Results

(Table ??)

93	<b>4 Discussion</b>
94	<b>5 Conclusion</b>
95	<b>6 Open Research</b>

96    **Acknowledgments**

97            This work was supported by the National Science Foundation grant OIA1545903  
98    to M. Kohn, S. Penniston-Dorland, and M. Feineman.

99    **7 References**

## Appendix

- Banfield, J. D., & Raftery, A. E. (1993). Model-based gaussian and non-gaussian clustering. *Biometrics*, 803–821.
- Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, 28(5), 781–793.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1–22.
- Dy, J. G., & Brodley, C. E. (2004). Feature selection for unsupervised learning. *Journal of Machine Learning Research*, 5(Aug), 845–889.
- Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3), 381–396.
- Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), 611–631.
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *The R Journal*, 8(1), 289.
- Vermeesch, P. (2018). IsoplotR: A free and open toolbox for geochronology. *Geoscience Frontiers*, 9(5), 1479–1493.