Chutes and Ladders: metamorphic conditions of exhumed simulated rocks

Buchanan C. Kerswell [1] Matthew J. Kohn [1] Taras V. Gerya [2]

[1]Department of Geosicences, Boise State University, Boise, ID 83725

[2]Department of Earth Sciences, ETH-Zurich, Sonneggstrasse 5, Zurich 8092, Switzerland

Key Points:

-
-
-

Corresponding author: Buchanan C. Kerswell, buchanankerswell@u.boisestate.edu

Abstract

1 Introduction

2 Methods

2.1 Numerical model

A set of markers, $x_i = x_1, x_2, \ldots, x_i$, are initialized randomly in the model domain.

2.2 Marker tracing

Markers are traced for within a 760 km wide and 11 km deep section extending from the trench to 500 km from the left boundary (Figure **??**). Slab rollback eventually leads to mechanical interference between trench sediments and the stationary convergence region centered at 500 km from the left boundary. The fixed, high-viscosity, convergence region acts as a barrier to the incoming sediments, deforming the accretionary wedge into a rapidly thickening pile. The abrupt change in accretionary wedge geometry flattens the slab causing intense crustal deformation of the forearc and backarc regions. We consider the dynamics after interference begins unrepresentative of natural bouyancy-driven slab motion. Marker PTt paths are, therefore, increasingly meaningless after mechanical interference begins.

The number of timesteps for marker tracing t is chosen automatically for each model by computing the topographic surface profile for each timestep. Markers PTt paths are cut when the sediment pile deforming against the barrier becomes the overall topographic high, usually within one or two timesteps after interference.

2.3 Maker classification

Tracing marker pressure, temperature, and x-z-position at each timestep is enough to compute characteristics of marker PTt paths, like maximum pressure (Figure **??**). 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). Please note that GMM fit by EM is a general purpose clustering algorithm 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). We derive GMM in sec. 2.3.1 and EM in sec. 2.3.2.

Before deriving the details of marker classification, we hypothesize that features computed from a PTt path, like maximum pressure, may distinguish subducted markers from recovered markers. If true, clustering algorithms like GMM may reliably classify markers by their dissimilarity along any number of dimensions computed from marker PTt paths (e.g., Dy & Brodley, 2004).

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 Φ. The probability distribution of markers x_i can be modeled with a mixture of k components:

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

where $p(x_i|\Theta_j)$ is the probability of x_i under the j^{th} mixture component and π_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 Θ_j describes a Gaussian probability distributions with mean μ_j and covariance Σ_j, Equation 1 becomes:

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

where

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

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

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

Taking the derivative of Equation 4 with respect to each parameter, π, μ, Σ, setting the equation to zero, and solving for each parameter 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 ω_{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 ω_{ij} is unknown for unlabelled datasets, so MLE cannot be computed with Equation 5. We derive the solution to this problem in sec. 2.3.2.

We use general purpose functions in the `R` package `Mclust` (Scrucca et al., 2016) to fit Gaussian mixutre models. After Banfield & Raftery (1993), covariance Σ_j matrices are parameterized to be flexible in their shape, volume, and orientation (Scrucca et al., 2016):

$$\Sigma_j = \lambda_j D_j A_j D_j^T \tag{6}$$

where D_j is the orthogonal eigenvector matrix, A_j and λ_j are diagonal matrices of values proportional to the eigenvalues. This implementation allows fixing one, two, or three geometric elements of the covariance matrix among groups. That is, the volume λ_j, shape A_j, and orientation D_j of Gaussian clusters can change or be fixed among all k clusters (e.g., Celeux & Govaert, 1995; Fraley & Raftery, 2002). The parameterization of Equation 6 is chosen by measuring Bayesian Information Criterion (BIC, Schwarz & others, 1978) for all general parameterizations of Σ_j.

2.3.2 Expectation-Maximization fitting of Gaussian Mixtures

The EM algorithm estimates GMM parameters by initializing k Gaussians with parameters (π_j, μ_j, Σ_j), then by iteratively computing membership weights with Equation 7 (E-step), and updating Gaussian parameters with Equation 5 (M-step) until convergence (Dempster et al., 1977).

Let $z_{ij} \in \{1, 2, \ldots, k\}$ 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 $j = 1, 2, \ldots, k$. Membership weights ω_{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:

$$p(z_{ij}|x_i) = \frac{p(x_i|z_{ij})p(z_{ij})}{p(x_i)} = \frac{\pi_j \mathcal{N}(\mu_j, \Sigma_j)}{\sum_{j=1}^{k} \pi_j \mathcal{N}(\mu_j, \Sigma_j)} = \omega_{ij} \tag{7}$$

can be computed given initial estimates for k sets of Gaussian parameters π_j, μ_j, Σ_j (E-step). With ω_{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 (e.g., Dy & Brodley, 2004) were filtered out. We settled on a two-component mixture of:

$$\begin{aligned} x_i^{mP} &= \max_{1 \le t \le t} P \\ x_i^{sdP} &= \sum_{1}^{t} dP \end{aligned} \tag{8}$$

where x_i^{mP} and x_i^{sdP} represent the maximum pressure attained along a marker PTt path and the sum total of all pressure changes along a marker PTt path, respectively. The bivariate mixture model represented by Equation 2 and Equation 8 is fit with $k = 6$ Gaussian clusters using EM Equations 7, 5. The results of this step are arbitrary class labels $y_i \in 1, \ldots, k$ representing group assignment of markers x_i to one of k clusters. The next step is to determine which clusters and markers represent recovered or subducted markers.

2.3.3 Chutes or ladders?

GMM clustering classifies markers into 6 groups, so a final decision to classify a group as *subducted* or *recovered* is made by comparing each group's centroid (μ_j, 5) to the overall distribution of markers along x_i. Groups with centroids μ_j well above the median of x_i classify as *subducted* (Figure **??**). The exact threshold is defined as three-quarters of inter quartile range below the median.

3 Results

(Table **??**)

4 Discussion

5 Conclusion

6 Open Research

Acknowledgments

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

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.

Schwarz, G., & others. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*(2), 461–464.

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.