

INVERSION MINÉRALOGIQUE PAR STRATE GÉOLOGIQUE

Vianney Bruned ¹ Alice Cleynen ² André Mas ³ Sylvain Wlodarczyk ⁴

¹ *Université de Montpellier, Schlumberger. VBruned@slb.com*

² *IMAG, Université de Montpellier. alice.cleynen@umontpellier.fr*

³ *IMAG, Université de Montpellier. andre.mas@umontpellier.fr*

⁴ *Schlumberger. SWlodarczyk@slb.com*

Résumé.

La détermination de la lithologie et de la proportion de fluides au sein d'une formation géologique est un objectif essentiel du pétrophysicien. Pour atteindre cet objectif ce dernier dispose d'un vecteur de mesures physiques réalisées tout au long d'un forage à intervalles de distance réguliers $L_1, L_2, \dots, L_n \in \mathbb{R}^d$. Reconstruire la lithologie revient à inférer pour chaque profondeur $i \in \{1, \dots, n\}$ les fractions volumiques $V_{i,m}$ des M types de minéraux contenus dans le sous-sol. Dès que $M > d$ on est confronté à un problème inverse mal posé (plus d'inconnues que d'équations). Les méthodes classiques actuelles sont basées sur des choix d'experts qui contraignent le problème. Nous proposons une approche bayésienne (ABC) et des méthodes d'optimisation globale afin de déterminer la lithologie par strate géologique.

Mots-clés. Problème Inverse, Approche Bayésienne, Optimisation globale.

Abstract.

The determination of the lithology and the proportion of fluids of a geological formation is a key objective for the petrophysicist. To achieve this goal, the latter have some physical measurements collected at regular intervals $L_1, L_2, \dots, L_n \in \mathbb{R}^d$ along the wellbore. Computing the lithology is equivalent to inferring, for each depth $i \in \{1, \dots, n\}$, the volumic fractions $V_{i,m}$ of the M mineral types contained in the subsurface. As soon as $M > d$, we are confronted to an ill-posed inverse problem (more unknowns than equations). The state of the art methods are based on the choice from expert which make the problem feasible. We propose a Bayesian method (ABC) and some global optimization methods to determine the mineralogy per geological stratum.

Keywords. Inverse Problem, Bayesian Approach, Global Optimization.

1 Contexte

En pétrophysique, déterminer la composition des roches et la porosité (mélange des différents fluides: eau, huile et gaz) d'une formation géologique est une étape cruciale. Pour atteindre cet objectif, on fore des puits d'explorations où l'on fait des mesures

physiques selon la profondeur du puits. Ces mesures se nomment diagraphies ou logs (en anglais) et ce sont les données dont dispose le pétrophysicien. Ces logs sont assimilables à des séries temporelles indexées par la profondeur du puits d'exploration. Ce sont des mesures des propriétés physiques de la roche: la radioactivité gamma, la densité volumique, la porosité par neutrino, la vitesse d'onde sismique, la résistivité... La composition minéralogique ou lithologie et la porosité se calculent pour chaque profondeur du puits. Par abus de langage, on parle de volume pour les fractions volumiques des éléments constituant la lithologie et les fluides. Il existe des modèles physiques qui relient la composition minéralogique et la porosité d'une formation avec les logs. Pour de nombreuses mesures, les modèles sont linéaires.

Formellement, le modèle statistique peut s'écrire de la façon suivante: si l'on dispose de d logs disponibles et que l'on fixe un modèle de minéraux et de fluides à M éléments, pour chaque profondeur n , on pose $L_n \in \mathbb{R}^d$, le vecteur des logs, $V_n \in \mathbb{R}^M$ le vecteur des volumes des minéraux et/ou des fluides, on modélise la relation logs-volumes par:

$$L_n = G(V_n) + X_n, \quad (1)$$

avec $X_n \sim N(0, \Sigma)$ où $\Sigma \in \mathcal{M}(\mathbb{R}^d)$ (matrice de covariance) et G un opérateur de \mathbb{R}^M dans \mathbb{R}^d qui est connu mais qui dépend des composants minéralogiques et/ou des fluides considérés. De plus, les $V_{i,n}$ sont contraints:

$$\sum_{i=1}^M V_{i,n} = 1, \\ V_{i,n} \geq 0 \text{ pour tout } i.$$

Dans la version linéaire de l'équation 1, l'opérateur G est matriciel et pour chaque log l :

$$L_n^l = \sum_{i=1}^M V_{i,n} \gamma_{l,i} + X_n^l. \quad (2)$$

où les $\gamma_{l,i}$ sont appelés endpoints et sont les valeurs physiques du minéral ou du fluide à l'état pur. Par exemple, le quartz ayant une radioactivité gamma de 30 *API*, son endpoint pour la radioactivité gamma est de 30. Les endpoints ne dépendent pas de la profondeur.

En pratique, l'équation 1 forme un problème inverse mal posé. En effet, il y a plus de 20 minéraux et/ou fluides possibles alors que l'on dispose seulement de 4-5 logs. On a alors $M \gg d$ d'où un système indéterminé ce qui oblige le pétrophysicien à faire un choix de minéraux qui dépend essentiellement d'informations externes (localisation géographiques du puits, études précédentes, expertise d'expert...). En pétrophysique, une pratique courante consiste à choisir un modèle lithologique qui respecte la contrainte $M \leq d + 1$ pour une zone géologique donnée provenant d'études sismiques précédentes.

Par un processus d’itération, le pétrophysicien raffine son modèle en regardant l’évolution de l’erreur de reconstruction des logs.

Une autre méthode consiste à regrouper les différents minéraux en famille. En effet, il existe 3 grandes familles de minéraux: les sables, les argiles et les carbonates. Les autres minéraux sont plus ”exotiques” et sont souvent aisément identifiables sur les logs. La condition $M \leq d + 1$ est alors aisément acquise mais se pose alors le problème des valeurs des endpoints. La composition des familles étant inconnue, les valeurs $\gamma_{l.}$ introduites au-dessus ne sont plus valides. Le pétrophysicien est là aussi contraint de fixer par zone les valeurs de ces endpoints synthétiques. Ces derniers dépendent alors de la strate géologique identifiée et sont à priori inconnus.

2 Contributions

L’objectif de ce travail est de mettre en œuvre, développer et/ou adapter des algorithmes mathématiques afin de guider les choix du pétrophysicien.

2.1 Mise en forme par strate du problème d’inversion

Tout d’abord, il nous faut trouver les strates géologiques où l’on peut faire l’hypothèse d’une lithologie stable. On se ramène ainsi à un problème de segmentation de séries temporelles multivariées bien développée dans la littérature statistique mais très peu dans le domaine de la pétrophysique. Pour chaque strate géologique, on suppose que la lithologie suit un modèle gaussien de moyenne inconnue et de variance fixé. Par exemple, les algorithmes Binary Segmentation [2] ou PELT [1] sont indiquées pour résoudre cette problématique. Il nous suffit juste de les adapter au cas multivarié. Par la suite, on suppose résolu le problème de segmentation et on considère chaque segment séparément.

Une formulation assez générale du problème consiste à regrouper les minéraux par famille, et ensuite à déterminer à la fois les endpoints par strate et les volumes des familles minéralogiques par profondeur de la strate. Pour la strate géologique k , on est amené à résoudre:

$$\min_{\gamma_{.,.k}} \sum_{n=\tau_k+1}^{\tau_{k+1}} \min_{V_{i,n}, i \in \Omega} \left\| L_n - \sum_{i \in \Omega} V_{i,n} \gamma_{.,i,k} \right\|^2$$

où Ω est l’ensemble des familles de minéraux, et τ_k et τ_{k+1} sont les indices de profondeur délimitant la strate k . On fait face à un problème d’optimisation qui n’est pas forcément convexe et la présence de minima locaux est possible. Des méthodes d’optimisation globales comme le recuit simulé [4] sont indiquées pour résoudre le problème. Cependant, nous explorons une piste différente basée sur l’inférence statistique.

2.2 Sélection minéralogique par ABC

Pour résoudre le problème de la composition minéralogique lorsque $M \gg d$, on propose une méthode de type Approximate Bayesian Method ou ABC [5]. Ici, l'objectif est de proposer différents modèles lithologiques (déterminer la présence des minéraux dans la strate) et non d'avoir une estimation précise des volumes. On pose le modèle d'inférence avec comme probabilité à priori une loi de Dirichlet pour les volumes (v) et un bruit gaussien pour le bruit des logs (L) et G connu.

$$\begin{cases} v \sim \mathcal{D}(\alpha_1, \dots, \alpha_M), \\ L \sim \mathcal{N}(G(v), \Sigma). \end{cases} \quad (3)$$

Pour ABC, on suit l'algorithme suivant:

- (i) On simule v à partir des probabilités à priori. On ajoute un bruit $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
- (ii) On calcule L^* correspondant.
- (iii) Si $\|L^* - L_n\|$ est plus petit qu'un certain δ , on accepte v .
- (iv) On revient à (i).

Une fois que l'on a effectué ABC pour chaque profondeur de la strate, on agrège toutes les lithologies obtenues par ABC en ne tenant pas compte de leur profondeur d'origine. Pour distinguer les modèles lithologiques les plus probables au sein de la strate, on applique sur les lithologies un algorithme de clustering basée sur la densité (Hierarchical Density-Based Spatial Clustering of Applications ou HDBSCAN) [3]. Cet algorithme nous permet d'éliminer certaines lithologies aberrantes. En prenant les principaux volumes des différents clusters, on a alors les principales hypothèses lithologiques sur le segment. On peut classer ces hypothèses selon la taille du cluster et/ou lorsque c'est possible (problème bien posé) son erreur d'inversion.

3 Résultats numériques

Nous avons testé la méthode de sélection minéralogique par ABC sur des données synthétiques et sur des données réelles. Sur la figure 1, on peut voir des boîtes à moustaches des résultats bruts d'ABC sur une strate de données simulées. Il s'agit de données synthétiques où l'on a trois minéraux: du quartz (élément 4) à 21%, de la smectite (élément 9) à 66% et de l'eau (élément 10) à 13%. On dispose de 4 logs et dans notre modèle on met 10 volumes: les éléments 1 à 3 sont des carbonates, les éléments 4 et 5 du sable, les éléments de 6 à 9 de l'argile et l'élément 10 l'eau. On voit sur les résultats bruts de ABC que l'on a du mal à identifier les éléments 6 et 9 (figure 1). En appliquant l'algorithme HDBSCAN sur les

résultats d'ABC de toute la strate, on trouve 2 clusters: un mélange des éléments 4-6-10 et un mélange des éléments 4-9-10. Sur la figure 3, on voit que les deux clusters sont caractérisés par leur composition en élément 6 et en élément 9. Après l'étape de HDBSCAN, on observe sur la figure 2 que l'on a réussi à bien identifier sur le cluster principal les éléments réellement présents. Toutefois, lorsqu'il y a un mélange de nombreux minéraux (supérieur à 5) en faible proportion il devient difficile de les identifier.

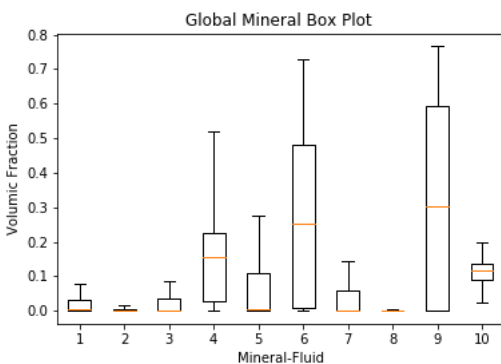


Figure 1: Boîte à Moustache des différents volumes issus de ABC de toute la strate.

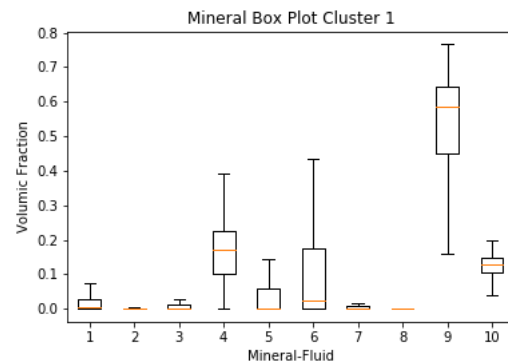


Figure 2: Boîte à Moustache des différents volumes issues du cluster principal du post-processing des résultats ABC par l'algorithme de clustering HDBSCAN.

Bibliographie

- [1] Eckley, I.A, Fearnhead, P. et Killick, R. (2012), Optimal detection of changepoints with a linear computational cost, *Journal of the American Statistical Association*, 107(500), 1590-1598.
- [2] Knott, M. et Scott, A.J. (1974), A cluster analysis method for grouping means in the analysis of variance, *Biometrics*, 30(3):507–512.
- [3] Campello R., Moulavi D. et Sander J. (2013), Density-Based Clustering Based on Hierarchical Density Estimates, *Advances in Knowledge Discovery and Data Mining* (pp. 160-172). Springer Berlin Heidelberg.
- [4] Gelatt C. D., Kirkpatrick S. et Vecchi M. P. (1983), Optimization by Simulated Annealing, *Science*, vol. 220, pp. 671-680.
- [5] Marin, J. M., Pudlo, P., Robert, C. P. et Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6), 1167-1180.

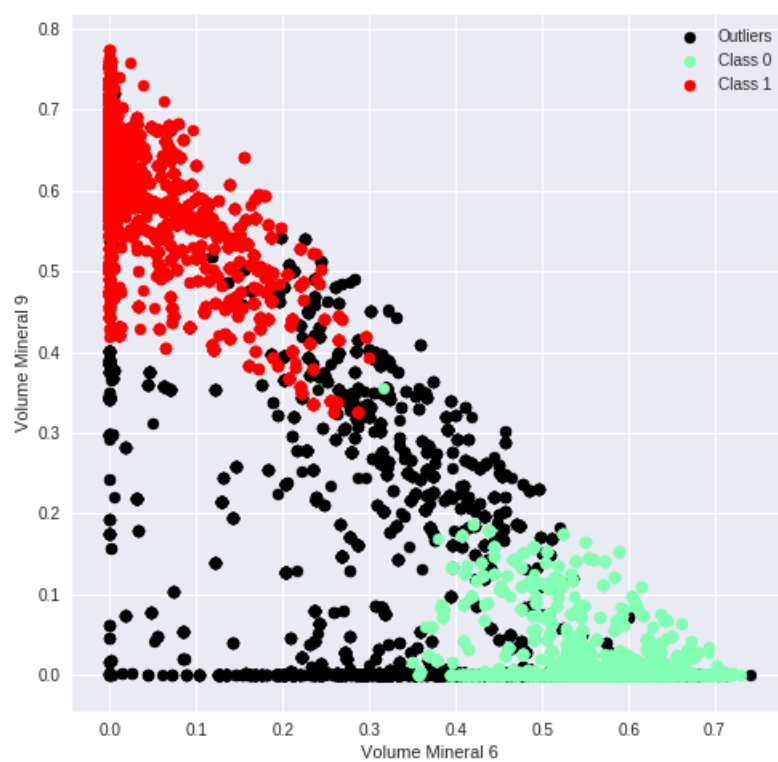


Figure 3: Projection des résultats de clustering (par HDBSCAN) des lithologies obtenues par ABC dans les éléments 6 et 9. L'algorithme a trouvé 2 clusters représentés en rouge (cluster principal) et en vert (cluster minoritaire). Les lithologies aberrantes sont représentées en noir.