**Couplage des mesures géophysiques en utilisant une approche de « data mining »**

**Kawtar SABOR1, Damien JOUGNOT2, Barthélémy STECK3, Louis APFFEL4, Roger GUERIN5**

1 **Sorbonne Université, CNRS, EPHE, UMR 7619 METIS, 75252 Paris, France, EDF R&D Chatou, 6 quai Watier 78054, Chatou,** [kawtar.sabor@sorbonne-universite.fr](mailto:kawtar.sabor@sorbonne-universite.fr)

**2 Sorbonne Université, CNRS, EPHE, UMR 7619 METIS, 75252 Paris, France,** [damien.jougnot@sorbonne-universite.fr](mailto:damien.jougnot@sorbonne-universite.fr)

**3 EDF R&D Chatou, 6 quai Watier 78054, Chatou,** [barthelemy.steck@edf.fr](mailto:barthelemy.steck@edf.fr)

**4 EDF R&D Chatou, 6 quai Watier 78054, Chatou,** [louis.apffel@edf.fr](mailto:louis.apffel@edf.fr)

**5 Sorbonne Université, CNRS, EPHE, UMR 7619 METIS, 75252 Paris, France,** [roger.guerin@sorbonne-universite.fr](mailto:roger.guerin@sorbonne-universite.fr)

**Mots clefs**

**Couplage, tomographie de résistivité électrique, tomographie sismique de réfraction, data mining, DBSCAN (**Density-Based Spatial Clustering of Applications with Noise**)**

**Résumé**

**Un algorithme de traitement de données issu des approches de data mining est testé pour combiner les résistivités électriques et les vitesses sismiques issues de la sismique de réfraction. Une première validation numérique est réalisée dans l’attente d’une validation sur des mesures réelles issues d’une campagne de mesure en cours.**

**Abstract**

**A data mining algorithm was tested in order to jointly interpret electrical resistivity tomography and seismic refraction profiles. As a first step, the geophysical profiles used are numerical ones before applying this approach to real measurements data in further studies.**

**Introduction**

Les ouvrages hydrauliques en terre (digues de protection contre les inondations, canaux à charge permanente de navigation ou d’alimentation des usines hydroélectriques, barrages de tout usage) nécessitent d’être surveillés car ils peuvent présenter en cas de rupture un grave danger pour les usagers et pour les populations environnantes. Pour les ouvrages hydrauliques, une démarche préventive et périodique de surveillance est nécessaire et imposée depuis 2007 par la réglementation. Ce qui a poussé les exploitants dont EDF à développer et étudier les méthodes de reconnaissance et de surveillance des ouvrages en terre. En effet, pour satisfaire cette surveillance réglementaire, le diagnostic est basé sur des informations issues d’examens visuels et de mesures d’auscultation parmi lesquelles on trouve les mesures géophysiques non destructives. Chaque méthode étant sensible à une ou plusieurs propriétés particulières du sol, combiner les différentes méthodes permettrait de réduire l’incertitude des résultats obtenus en tenant compte de l’ensemble des sensibilités des différentes méthodes. Les approches de combinaison des mesures géophysiques sont continuellement en cours de développement (e.g. Gallardo and Meju, 2003, Looms et al., 2008, Doetsch et al., 2010). L’objectif de ce travail est d’explorer une nouvelle approche de couplage et de traitement des mesures géophysiques. Cette approche est inspirée des approches de traitement de données utilisées dans le cadre du « data mining ». Dans un premier temps une validation numérique sur des modèles synthétiques est réalisée afin de vérifier l’adéquation de cette approche à nos applications en mesures géophysiques et de développer une méthodologie d’application de cette approche dans le cadre du couplage des mesures géophysiques.

**Approche utilisée**

L’approche utilisée consiste en cinq étapes principales (Fayyad et al., 1996) :

* Sélection de données : Cette étape consiste à choisir les données adaptées à l’interprétation souhaitée. Dans le cadre de la prospection des ouvrages hydrauliques, la ou les méthodes géophysiques adaptées sont choisies selon l’objectif de la prospection (e.g. analyse de la structure géologique, recherche d’anomalie ou de fuite), la résolution et la profondeur d’investigation souhaitées et aussi le type de sol prospecté.
* Prétraitement des données : Le prétraitement des données consiste à éliminer les données aberrantes et extraire les paramètres d’intérêt qui seront utilisés à l’étape de l’exploration des données. Les paramètres d’intérêt sont regroupés sous forme d’un tableau ou d’une matrice contenant pour chaque point des paramètres ses coordonnées ou ses indicateurs ou son index et les valeurs des paramètres. Dans le cadre de notre application, cette étape consiste à nettoyer les mesures géophysiques en les filtrant par rapport au facteur géométrique ou par rapport au potentiel mesuré par exemple. Ensuite ces données sont inversées pour retrouver la grandeur géophysique qui est notre paramètre d’intérêt.
* Transformation des données : Cette étape consiste à appliquer des transformations mathématiques aux paramètres afin de mieux ressortir les résultats souhaités. Dans le cadre de nos applications, un exemple de transformation des paramètres est l’application du logarithme aux résistivités électriques inversées pour mieux ressortir les contrastes.
* Exploration de données « data mining » : Cette étape consiste à appliquer un algorithme d’analyse des paramètres. Il existe trois grandes classes d’algorithmes, la classification des données, la régression et le partitionnement des données (Pedregosa et al., 2011). Dans le cadre de notre application, le partitionnement est choisi. Le partitionnement consiste à extraire dans les paramètres des clusters. Le regroupement des paramètres en clusters se fait selon un critère de similitude analysé de manière statistique ou selon la densité des points représentant les paramètres. L’algorithme de partitionnement choisi dans ce travail est appelé DBSCAN (pour une description détaillée du fonctionnement de l’algorithme voir Ester et al., 1996).
* Interprétation des données : L’interprétation des données se base sur les résultats de l’étape précédente pour réaliser l’interprétation souhaitée qui peut être une prise de décision, une prévision ou une classification. Dans le cadre de notre cas d’application qui est l’analyse des mesures géophysiques, l’interprétation consiste à relier les clusters aux différentes classes de sol et de reconstruire ainsi la structure géologique de la zone prospectée.

**Exemple application**

Dans un premier temps, cette approche est testée sur des modèles synthétiques qui sont choisis en s’inspirant de cas d’intérêt pour le diagnostic et la reconnaissance des digues en terre. En effet, dans le cadre de la modélisation du comportement des ouvrages, un modèle géologique de l’ouvrage est nécessaire. Les méthodes géophysiques sont utilisées pour reconstruire la structure géologique le plus précisément possible. L’exemple présenté dans la Figure 1 représente les profils inversés de résistivités électriques et de vitesse sismique associés à un modèle bicouche avec une marche d’escalier qui peut représenter une variation du substratum et donc de l’interface digue substratum. Les inversions électriques et sismiques sont réalisées avec le même maillage.

**Résultats**

L’approche décrite précédemment est appliquée aux profils inversés de résistivités électriques et sismiques (Figure 1). A l’issue des étapes de prétraitement et de transformation de données suite auxquelles on se retrouve avec un tableau contenant les positions des points des profils et les valeurs de résistivités électriques et de vitesses sismiques associées. Ces paramètres peuvent être représentés sous forme d’un nuage de points dans le cas 2D (Figure 2a). L’application de l’algorithme de data mining DBSCAN permet de répartir le nuage en différents clusters, deux pour cet exemple. En repositionnant chaque point du nuage à sa position dans le profil, on retrouve bien le modèle synthétique (Figure 2b). Les deux clusters désignent les deux couches du modèle. L’interface en escalier n’est pas très explicite sur les profils électriques et sismiques seuls. L’interprétation des deux profils sismiques et électriques est aussi très biaisée par l’échelle de couleur utilisée. L’application de l’approche de data mining permet de reconstruire plus précisément la structure, la forme de l’interface et sa position exacte sont retrouvées.

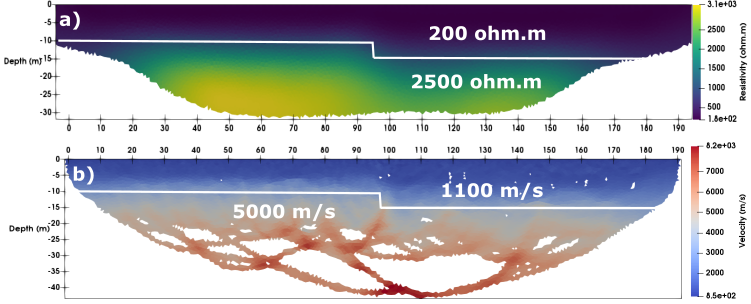


Figure 1 Exemple de cas utilisés pour valider l'approche d'analyse et de combinaison des données. a) Profil inversé de résistivité électrique. b) Profil inversé de vitesse sismique obtenue par sismique de réfraction. Le trait blanc désigne le modèle synthétique.

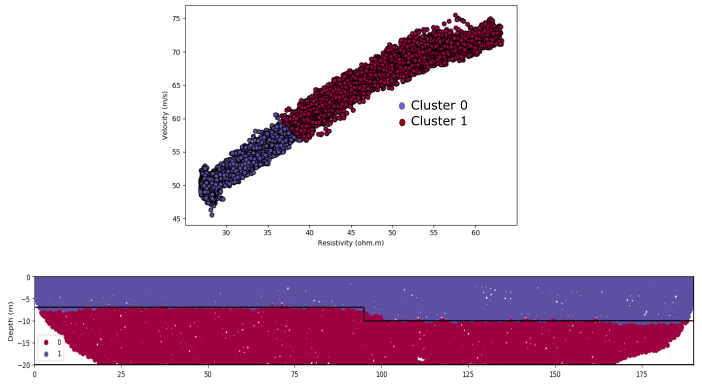


Figure 2 Résultat du traitement joint des résistivités électriques inversées et des vitesses sismiques inversées par l’approche de data mining. a) Résultat de l'application de DBSCAN sur les paramètres résistivités électriques et sismiques présentées sous forme d’un nuage de points b) Résultat de l'étape d'interprétation des paramètres suite à l’application de DBSCAN, qui permet de repositionner chaque point du nuage sur le profil, connaissant les coordonnées de chaque point. Les deux clusters désignent donc les deux couches qui se superposent bien avec le modèle synthétique en ligne noire.

**Conclusions et perspectives**

L’utilisation des algorithmes de data mining et particulièrement DBSCAN est prometteuse. Elle permet de mieux exploiter et d’automatiser l’analyse combinée des mesures géophysiques. L’utilisation de DBSCAN permet de retrouver précisément la structure du sol prospecté, ce qui permettra de fournir un modèle géologique plus précis des ouvrages et des digues en terre pour réaliser leur modèle numérique. Une campagne géophysique multi méthodes est en cours sur une portion d’une digue en terre afin d’acquérir les données qui serviront à la prochaine étape de validation sur des données terrain de cette nouvelle approche.

**Références bibliographiques**

Doetsch, J., Linde, N., Coscia, I., Greenhalgh, S. A., and Green, A. G. [2010] Zonation for 3D aquifer characterization based on joint inversions of multimethod crosshole geophysical data. *Geophysics*, **75**(6), 53–64.

Ester, M., Kriegel, H., Xu, X., and Sander, J. [1996] A density-based algorithm for discovering clusters in large spatial databases with noise. *Association for the Advancement of Artificial Intelligence*.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. [1996] Knowledge discovery and data mining: towards. *American Association for Artificial Intelligence*, **17**(3), 82–88.

Gallardo, L. A., and Meju, M. A. [2003] Characterization of heterogeneous near-surface materials by joint 2D inversion of DC resistivity and seismic data. *Geophysical Research Letters*, **30**(13), 2–5.

Looms, M. C., Binley, A., Jensen, K. H., Nielsen, L., and Hansen, T. M. [2008] Identifying Unsaturated Hydraulic Parameters Using an Integrated Data Fusion Approach on Cross-Borehole Geophysical Data. *Vadose Zone Journal*, **7**(1), 238.

Pedregosa, F., Weiss, R., and Brucher, M. [2011] Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.