



HAL
open science

Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies

Marc Dumont, Pierre-Alexandre Reninger, Alexandre Pryet, Guillaume Martelet, Bertrand Aunay, Jean-Lambert Join

► To cite this version:

Marc Dumont, Pierre-Alexandre Reninger, Alexandre Pryet, Guillaume Martelet, Bertrand Aunay, et al.. Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies. *Journal of Applied Geophysics*, 2018, 157, pp.1 - 9. 10.1016/j.jappgeo.2018.06.020 . hal-01855450

HAL Id: hal-01855450

<https://brgm.hal.science/hal-01855450v1>

Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Agglomerative hierarchical clustering of airborne electromagnetic data for multi-scale geological studies

M. Dumont ^{a,b,*}, P.A. Reninger ^c, A. Pryet ^d, G. Martelet ^c, B. Aunay ^b, J.L. Join ^a

^a Laboratoire Géosciences, Université de La Réunion, Institut de Physique du Globe de Paris, UMR 7154, Sorbonne Paris Cité, CNRS, 97744, Saint Denis, France ^b BRGM, Saint-Denis, La Réunion, France

^c Bureau de Recherches Géologiques et Minières (BRGM), UMR 7327, BP 36009, 45060 Orléans, France

^d EA 4592 Géoressources et Environnement, Bordeaux INP and Université Bordeaux Montaigne, Pessac, France

ABSTRACT

Airborne electromagnetic methods provide detailed subsurface resistivity imaging over extensive areas. The in-version of electromagnetic measurements can be conducted with a quasi-3D spatially constrained inversion scheme, which yields numerous vertical resistivity soundings. So as to conduct the interpretation, these soundings can be interpolated to obtain a 3D resistivity model. However, large surveys result in huge resistivity models, which can be challenging to interpret with 2D or even 3D views. We propose a complementary approach for the interpretation of airborne electromagnetic surveys based on agglomerative hierarchical clustering. With this statistical classification method, the numerous 1D vertical resistivity profiles distributed over the area of interest can be summarized in a 2D horizontal map. Inverted electromagnetic soundings are aggregated into clusters according to their vertical resistivity profiles. These clusters define areas with a consistent vertical resistivity response, i. e. to the first order, areas with comparable vertical (hydro)-geological structure. The clustering method is illustrated with an extensive airborne electromagnetic survey conducted over Reunion Island. After a description of the method, we show that the proposed classification facilitates the confrontation between airborne geophysical results and geological or hydrogeological observations/data. We highlight that hierarchical clustering is of high interest for multi-scale studies, from regional to local scales. This approach introduces a new way to use geophysical surveys to map areas with specific geological/hydrogeological behaviors.

1. Introduction

Airborne electromagnetics (AEM) aims at imaging subsurface resistivity contrasts. It can cover extensive areas with virtually no topographic and vegetation limitations within the same hydrological period (Fountain, 1998). Though developed originally for mining purposes, recent improvements of airborne time domain electromagnetic systems make it suitable for groundwater and environmental studies (Siemon et al., 2009; Sørensen and Auken, 2004). AEM surveys consist of thousands to hundreds of thousands of 1D electromagnetic measurements acquired along flight lines. The inversion of these latter can be conducted with a quasi-3D spatially-constrained inversion scheme (Viezzoli et al., 2008), which provides a consistent dataset at the regional scale. To be interpreted, these results are usually confronted with local or land-based observations

such as geological, geophysical or hydrogeological data. However, a straightforward confrontation of large 3D resistivity models with hundreds of wells (e.g. geological logs, electrical water conductivity), geological maps and climatic information tends to be challenging in complex geological settings.

After inversion, AEM surveys are generally visualized as 2D profiles of vertical resistivity soundings or 2D maps at various depths or elevations. Profiles and maps are useful for the interpretation but they cannot display succinctly the 3D information provided by the AEM resistivity model (Danielsen et al., 2003). Furthermore, the interpretation of such results has to be done with a fine tuned color scale, which necessarily highlights some contrasts at the expense of others. Alternatively, coupled with borehole data, 3D geological structures can be highlighted, located and characterized (Auken et al., 2008; Christensen et al., 2015) but this cannot be automated for large datasets. Alternatively, resistivity contrasts linked to specific geological contacts can be represented as iso-resistivity depth maps or interpolated resistivity slope maps (Reninger et al., 2016). This remains quite specific as both approaches require simple geological contexts to be reliable. Inverted vertical resistivity soundings can also be interpolated into a 3D resistivity model (Pryet et al., 2011) displaying the 3D nature of subsurface resistivity

* Corresponding author at: Laboratoire Géosciences, Université de La Réunion, Institut de Physique du Globe de Paris, UMR 7154, Sorbonne Paris Cité, CNRS, 97744, Saint Denis, France.

E-mail address: dumont.mrc@gmail.com (M. Dumont).

variations. However, it is challenging to confront 3D resistivity models to 1D or 2D (hydro)-geological data. As a consequence, 3D resistivity models are hardly usable for detailed spatial analysis. Thus, the interpretation of regional resistivity models derived from large AEM surveys often turns to be challenging.

We propose a clustering approach to summarize on 2D horizontal maps the 3D resistivity information contained in AEM datasets. Our methodology is based on a statistical algorithm, the Agglomerative Hierarchical Clustering (AHC) (Murtagh and Legendre, 2011; Ward, 1963). Soundings are sorted in function of their vertical resistivity distribution and gathered into clusters. These clusters define areas with consistent vertical resistivity structures. The obtained map directly features the repartition of 3D resistivity variations and allows regional preliminary interpretation in terms of geological or hydrogeological environment.

First, we detail the implementation of the statistical process on AEM data. Second we illustrate the AHC approach over an extensive AEM survey conducted over Reunion Island.

2. Materials and methods

2.1. Airborne electromagnetic data

Our methodology is applicable to electromagnetic surveys, either in frequency domain or in time domain (Siemon et al., 2009). This study focuses on helicopter-borne time domain electromagnetic for the illustration of the methodology. This method uses high-powered transmitter loops to induce electric currents in the ground and measure the resulting magnetic field. The measured signal, i.e. electromagnetic decay, allows bulk resistivity to be estimated. For a more detailed discussion on electromagnetic theory the reader is referred to Nabighian (1991). Airborne systems have been developed for surveying large areas in a short time lapse. They are not limited by surface water, topography or dense vegetation. Such surveys can be used to investigate lateral resistivity contrasts of large geological structures. They result in datasets consisting of thousands to hundreds of thousands of electromagnetic decays. Various processing stages and the process of an inversion have to be performed to transform these decays into 1D vertical resistivity soundings.

Electromagnetic noise and anthropic couplings have to be rejected by automatic or manual filtering (Auken et al., 2009; Reninger et al., 2011). After the data pre-processing step, two methods can be used to obtain vertical resistivity profiles from electromagnetic measurements. The first option is to compute apparent resistivity curves in function of time for each decay (Fitterman and Stewart, 1986). The second option is to invert each decay curve to obtain a resistivity model. In this study, the second option was chosen, with the quasi-3D Spatially Constrained Inversion scheme. It provides more continuous and geologically consistent resistivity models (Viezzoli et al., 2008). This inversion yields a 1D vertical resistivity sounding for each AEM decay. A resistivity model is defined by a number of layers with a resistivity and a thickness. Two different types of inverted soundings are commonly used: (i) “smooth” models with a large number of layers (e.g. 20 to 30) with a fixed thickness, or (ii) “few-layer” models with a few number of layers (e.g. 4 or 5), for which both resistivity and thickness of the layers can vary during the inversion. These models are used for a smooth or a contrasted imaging, respectively. In this study, we focus in smooth inversion models generally used to inverse AEM datasets at regional scale. During the inversion and for each sounding, the depth of investigation (DOI) is estimated (Christiansen and Auken, 2012). Thus, after processing, the AEM dataset consists of 1D inverted soundings divided in layers, each defined by a resistivity and a thickness, down to the DOI.

The clustering approach presented in this paper could be achieved with different input data, such as the decay curves, the apparent resistivity curves or the resistivity soundings. The reliability of clustering approach depends on the consistency and the reliability of classified data. The classification of decay curves is hardly feasible because of the three following reasons. First, the number of usable time gates can vary significantly from one sounding to another. Second, they are not corrected for the exact location of emission and reception loops (i.e. flight altitude and effective loop areas). These variations create inconsistency between EM decays and would include bias in the classification. Third, the clustering results would not straightforwardly inform on vertical resistivity variations. Derived directly from decay curves, apparent resistivity also holds the uncertainty from EM noise and loops locations, which could bias the clustering approach too. So as to maintain the same number of parameters for each sounding, limit geophysical uncertainty and facilitate the interpretation, the AHC is therefore conducted on 1D inverted resistivity soundings, without consideration to the DOI. The parameters used for the AHC are the resistivity values of each layer of the smooth soundings.

2.2. Agglomerative hierarchical clustering

Statistical algorithms have long been used to classify observations in geosciences. They are based on the degree of similarity between individual observations. The most popular method is partitional clustering, as K-means (Bedrosian et al., 2007; Di Giuseppe et al., 2014; Paasche et al., 2006). It is based on the classification of the dataset in a specified number of clusters. Contrarily to the K-means method, Agglomerative Hierarchical Clustering (AHC) reveals the genetic relation between each observation. AHC classifies the dataset starting with as many clusters as observations. It iteratively groups “similar” observations to form a hierarchical tree, until all observations are gathered in a single cluster. The result can be visualized at different levels of the hierarchical tree, thus providing global to detailed synthetic views of the analyzed dataset, i.e. a multi-scale cartographic analysis (Martelet et al., 2006). Here, we aim at classifying a large number of 1D inverted resistivity soundings (the observations) into a series of clusters. The AHC process can be divided in three steps.

First, resistivity values are log-transformed to take into account the resolution of EM method. Thus, the standard score of logarithmic resistivity values, $\rho \sim_{i,k}$, is calculated in order to limit the impact of extreme values and increase the weight of vertical trends in the clustering process (Eq. 1).

$$\rho \sim_{i,k} = \frac{(\log \rho_{i,k}) - \mu_i}{\sigma_i} \quad (1)$$

where $\rho_{i,k}$ is the inverted resistivity values of layer i from sounding k . μ_i , σ_i are respectively the mean and the standard deviation of all the resistivity values of layer i .

Second, the AHC algorithm starts with as many clusters as AEM soundings in the dataset. The squared Euclidean distance (Eq. 2) is used to measure pairwise dissimilarities (i.e. distance):

$$\delta^2(c_1, c_2) = \sum_{i=1}^m (\rho \sim_{i,1} - \rho \sim_{i,2})^2 \quad (2)$$

where c_1 and c_2 are two AEM soundings now considered as two clusters of m layers. The Euclidean distance is the sum of m squared distances between $\rho \sim_{i,1}$ and $\rho \sim_{i,2}$ which are normalized resistivity values of the i -th layer of soundings c_1 and c_2 , respectively.

Third, after calculating the dissimilarity matrix, the entire dataset is iteratively merged until obtaining one final cluster. At each iteration, the two closest clusters are merged to minimize the total within-cluster

variance and maximize the distance between clusters. The Euclidean distance between the new cluster and the others is calculated with the Ward's minimum variance (Eq. (3) - (Ward, 1963)).

$$\delta(c_1 \cup c_2, c_3) = \left(\frac{|c_1||c_3|}{|c_1| + |c_2| + |c_3|} \delta^2(c_1, c_3) + \frac{|c_2||c_3|}{|c_1| + |c_2| + |c_3|} \delta^2(c_2, c_3) - \frac{|c_3|}{|c_1| + |c_2| + |c_3|} \delta^2(c_1, c_2) \right)^{1/2} \quad (3)$$

where $c_1 \cup c_2$ is the new cluster formed by clusters 1 and 2, and c_3 is another cluster, $||$ is the number of AEM soundings gathered in the cluster and δ^2 the Euclidean squared distance (Eq. (2)). For further details about the Ward method, the reader can refer to Murtagh and Legendre (2011).

The AHC process stops when the whole dataset is merged within a single cluster. Then, two graphics are usually used to analyze the results (Fig. 1). The dendrogram (Fig. 1A) presents the hierarchical evolution of the clustering process. Its vertical shape can be used to interpret cluster scattering and the overlap between them. Each node represents a pairwise aggregation. The ordinate length between two nodes represents the Euclidean distance between two merged clusters. Second, the evolution of the Euclidean distance as a function of the number of clusters can be displayed (Fig. 1B). It represents the distance between the last two clusters merged at each iteration.

The selection of the number of clusters has a direct impact on the interpretation and should be conducted with care. The number of clusters chosen for the interpretation should be consistent with the necessary level of detail. But this choice should also consider the minimization of the overall Euclidean distance (Fig. 1B). A high Euclidean distance results in more heterogeneity within clusters. Hence, a critical number of clusters should be defined. Under this limit (diagonal stripes, Fig. 1), dissimilarities within clusters may be too large for a relevant classification. However, few explained clusters (in terms of natural behavior) are preferable to many unexplained clusters. The selection of the number of clusters has a direct impact on their interpretation however the sub-clusters of any cluster of interest are accessible at all stages after clustering. In this paper, the selected classification degree is given in square brackets, e.g. cluster 8 [11d] is the cluster 8 for the 11-cluster classification. For a given number of clusters, each cluster is characterized by its median vertical resistivity profile and by the density of resistivity occurrence within each cluster at each depth. Higher densities indicate statistically more common resistivity at given depths and the density width informs on the resistivity dispersion.

The distribution of clusters can be plotted on a 2D map, but the information is only available at sounding locations. Interpolation is

performed in order to obtain a continuous cluster map. This can be conducted with the nearest neighbor interpolation, described in (Franke, 1982). The interpretation of such result is not dependent on a color scale and allows highlights in 2D most of the vertical and/or lateral (i. e. 3D) resistivity contrasts imaged by the AEM method. Illustrations will be provided hereafter as part of the case study.

3. Implementation of the clustering on Reunion Island

3.1. Regional survey on a volcanic island

We applied the proposed methodology to a large SkyTEM survey (Sørensen and Auken, 2004) conducted in 2014 over Reunion Island and an offshore ring (Martelet et al., 2014). About 350,000 Time Domain ElectroMagnetic (TDEM) decays were acquired during 3 months along 10,400 km of flight-lines (Fig. 2). The SkyTEM system operates two different moments for each EM decays: (i) a low moment (3100 A.m²) to ensure a good resolution at shallow depths and (ii) a high moment (160,000 A.m²) to increase the depth of investigation. Flight-lines were mainly oriented north-south with a 400-meter spacing, but additional lines with a different spacing and/or orientation were flown over areas of particular geological or hydrogeological interest. After standard data processing (Auken et al., 2009), 250,000 TDEM soundings have been conserved. A smooth inversion was then performed using a quasi-3D spatially constrained inversion (Viezzoli et al., 2008). The thickness of the 25 layers increases logarithmically from the surface to a depth of 350 m, starting with a 2 m thick first layer. The inversion quality is characterized with the residual (STD), which ranges between 0.5 and 2.5 with some peaks to 6 close to urbanized areas (Viezzoli et al., 2008).

Reunion Island is located in the southern part of the Mascarene Archipelago (southwestern Indian Ocean; Fig. 2). Two shield volcanoes have shaped this island by the succession of constructive (accumulation of lava flows and pyroclastic products) and destructive phases (erosive processes, landslides). The older *Piton des Neiges* volcano (North-West) was active before 2 Ma to 29 ka ago (Gillot and Nativel, 1982; Kluska, 1997). The younger *Piton de la Fournaise* (South-East) has been active since 450 ka (Lénat et al., 2012). The inter-tropical climate and cyclonic regime leads to significant erosion and rock weathering. Rainfall is highly contrasted between the windward side (up to 10 m/year) and the leeward side (lower than 2 m/year), influencing basalt weathering and erosion processes (Barcelo and Coudray, 1996). The geological heterogeneity leads to a wide variety of hydrogeological behaviors. In addition, the steep morphology of the recent volcano induces high variability between coastal and inland areas.

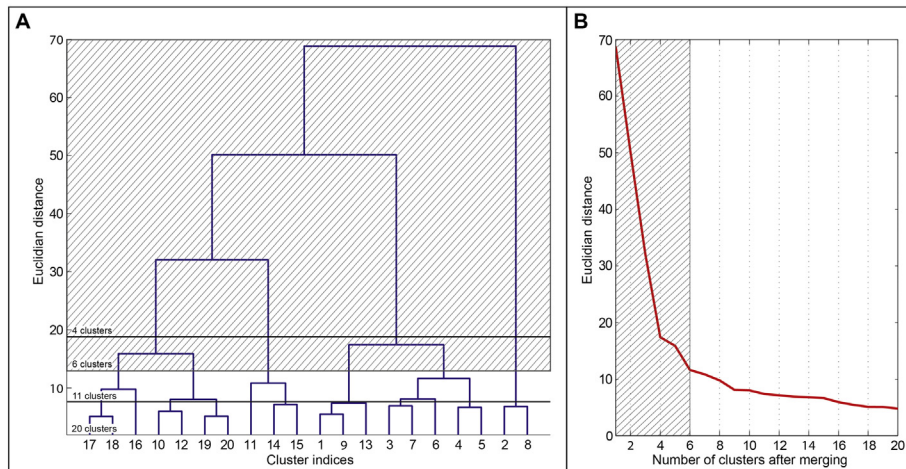


Fig. 1. AHC statistical graphics: (A) the dendrogram displays the hierarchical tree between 20 and 2 clusters. Black horizontal lines highlight 4, 6, 11 and 20 final clusters. (B) Euclidean distance between the two last clusters merged vs. the total number of clusters. Diagonal stripes correspond to levels of clustering with too large Euclidean distance, which are unreliable.

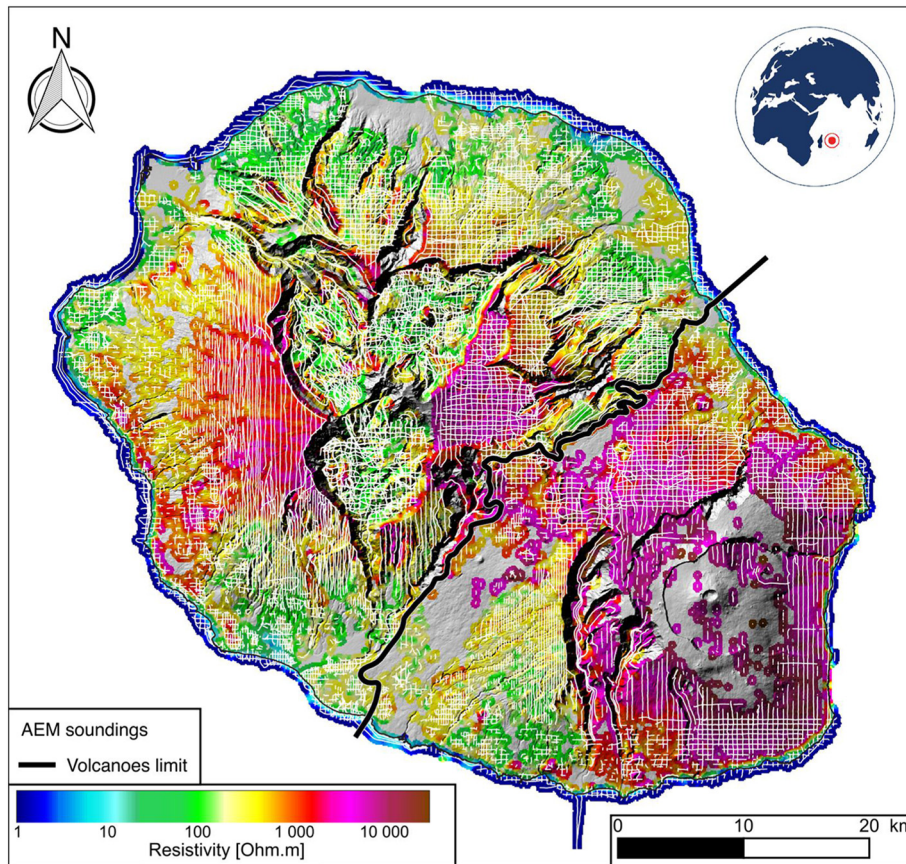


Fig. 2. Resistivity map between 20 and 30 m depth derived from SkyTEM survey of Reunion Island. AEM flight lines are displayed in white lines. The black line delineates the limit between the two volcanoes: Piton des Neiges and Piton de la Fournaise.

Geological, hydrogeological and geophysical studies have been performed over the last 30 years over Reunion Island. Thus, AHC results can be confronted with a wide variety of datasets.

3.2. Clustering process

Here, we describe how the AHC algorithm has been applied on the SkyTEM dataset. First, the statistical algorithm classifies the vertical resistivity models obtained by inversion and yields the associated dendrogram (Fig. 3). In order to choose the adequate level at which to cut the hierarchical clustering for a regional classification, we have compared different clustering levels with geological and hydrogeological maps (Bret et al., 2003; Descloitres et al., 1997; Join et al., 2005; Lénat et al., 2001; Salvany et al., 2012). At the island scale, the classification in 11 clusters (Fig. 3) highlights the major geological structures and is therefore considered appropriate for regional mapping. Second, the clustered 1D resistivity models have been interpolated to create continuous areas for cartographic interpretation (Fig. 4).

The first step to interpret the results is the analysis of the clustering process. Fig. 3 displays the dendrogram and the 11 cluster signatures. Cluster signatures are characterized by: (i) its vertical resistivity shape with the median curve and the median depth of investigation, (ii) the resistivity density scattering and (iii) the number of AEM soundings gathered into each cluster. According to the dendrogram (Fig. 3), five panels, named A to E, are characterized by a range of resistivity values and a vertical trend. First, highly conductive geophysical responses deviate from the rest of the dataset (group E – 1 Ω .m). This group agglomerates only offshore soundings affected by seawater. Second, highly resistive soundings (group D – above 500 Ω .m) are discriminated from resistive soundings (groups A and B – between 100 and 500 Ω .m) and conductive soundings (group C – between 10 and 100 Ω .m).

The vertical trend, characterized by the resistivity density, has to be analyzed from the surface to the DOI. Group A and C have decreasing trends in opposition with group B, D and E. Group B has a specific manner: resistivity decreases between 0 and 20 m and increases again down to the DOI. However, five clusters are not sufficient to explain the regional geological heterogeneity. Their signatures (median and densities) overlap too much to characterize any specific geological response. For each panel, a higher discretization level provides new clusters with comparable vertical resistivity trends but centered around different resistivity values (Fig. 3). The two characteristics (i.e. the vertical trend and the resistivity) are important for groundwater and geological interpretations. At the 11-cluster level the resistivity signatures are well individualized and the resistivity density is rather well focused around the median sounding (Fig. 3). For instance, some clusters such as number 3 or 7 [11d] are well defined and are mainly controlled by regional behaviors. Other clusters seem to be well defined only at some interval depths such as clusters 6 or 8 [11d] and could also be associated to a specific setting. Sometimes, high density of resistivity occurs under the DOI. This is directly related to the starting model and should not be considered.

When clusters are correctly characterized, the second step is to correlate cluster signatures and their spatial extensions with natural observations. The cluster map (Fig. 4B) differs from the resistivity map (Fig. 2). The latter simply describes the resistivity variations at a given depth interval, while the former integrates in 2D the 3D resistivity structure information. The cluster map summarizes 3D resistivity variations: all the main geological settings are characterized by different clusters. The younger geological domain, in green, is mainly characterized by resistive clusters 2, 8 and 10 [11d]. In contrast, the older part of the island is mainly represented by more conductive clusters such as clusters 3, 4, 7, 9 and 10 [11d]. The last major geological ensemble, detritic and alluvial formations (in yellow), are mainly represented by

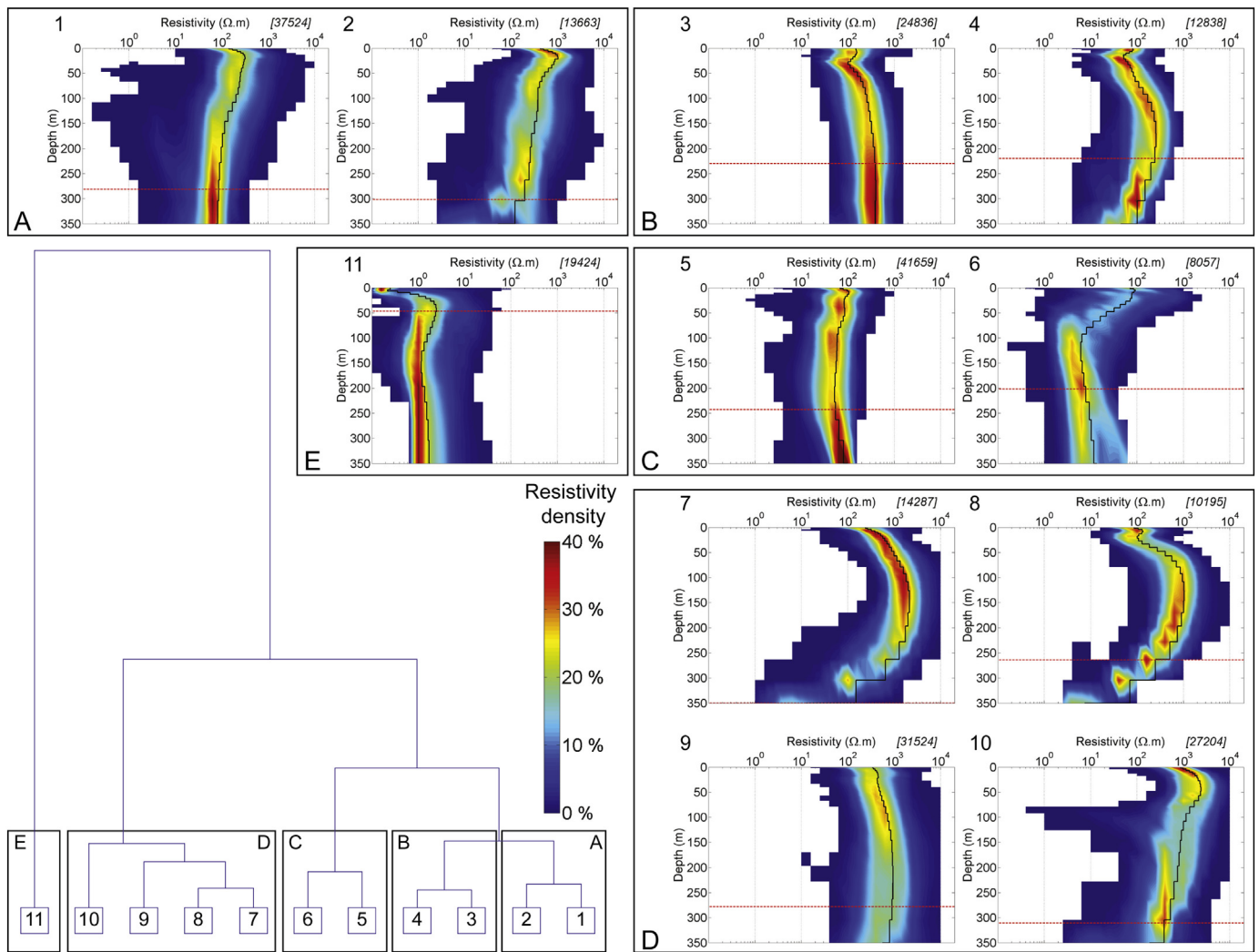


Fig. 3. The dendrogram with the 11 cluster signatures can be further interpreted in five sub-groups (A to E) according to the shape of the dendrogram. These sub-groups constitute families of contrasted major resistivity signatures; the number of AEM soundings within each cluster is specified in square brackets. Cluster signatures can be described with: (i) the median vertical profile (black line) and (ii) the density of resistivity occurrence (in color). The resistivity density informs in the resistivity occurrence within each cluster at each depth. The density is blanked for resistivity values without occurrence. For each cluster, the horizontal red dashed line represents the median DOI.

conductive clusters 1 and 5 [11d]. The two last clusters are correlated with hydrogeological behaviors: cluster 11 [11d] is located over the ocean, which explains low resistivity at the surface, and cluster 6 [11d] is restricted to the coastal area; its deep conductive layer might be controlled by saltwater intrusion. In conclusion, the cluster map highlights the main geological settings and two clusters are influenced by water mineralization. In the following, we present the interest of the AHC processing with two examples at the regional and local scales.

3.3. Regional interpretation: basaltic weathering

Volcanic rock weathering is a major process in subtropical areas, strongly accentuated by trade winds and cyclonic phenomena in Reunion Island. Weathering has considerable impact on raw material prospecting, land-use planning, erosion process, risk and groundwater management. Weathering is controlled by two main factors, such as the age of the volcanic formations and the rainfall, which is correlated to wind exposure in oceanic islands. In terms of resistivity signature, un-weathered lava flows are highly resistive ($> 1000 \Omega.m$), whereas weathering increases clay content directly reducing the bulk resistivity (Descloitres et al., 1997). To understand their impacts, clusters on four basaltic areas have been compared (white dashed polygons, Fig. 4).

The first area, PN1, is an old basaltic slope (Fig. 4) with low rainfall rates (1 to 2 m y^{-1}). Its upper part is mainly represented by clusters 7 and 9 [11d] characterized by highly resistive cluster signatures (Fig. 3). The superficial layers of this cluster are less resistive, highlighting a weak weathering. The second slope, PN2, is made up of the same ancient basaltic formations. This area is wetter (3 to 4 m y^{-1}), increasing weathering process. Thus, the upper part is mainly characterized by more conductive clusters (3 and 9). While only the superficial layers are impacted on the dry leeward side, the wet windward side is characterized by deeper weathering. In contrast, the lower parts of these two slopes are characterized by less resistive clusters 1 and 5 [11d].

In opposition, the two other slopes are in the younger volcano. Slope PF1 is composed by old basaltic lavas in a dry area (1 to 2 m y^{-1} rainfall rates), whereas PF2 is composed by recent lava flows in a much wetter climate (4 to 10 m y^{-1} rainfall rates). PF1 is characterized by cluster 8 [11d], its first 50 m are less resistive than cluster 10 ($\approx 100 \Omega.m$). PF2 slope is characterized by the more resistive cluster 10 [11d] ($> 1000 \Omega.m$) related to unweathered lava flows. This difference can be associated to a significant variation in superficial weathering. Overall, the confrontation of the cluster map with geological and climatic information can be used to classify the four slopes according to their degree of weathering, in increasing order: (i) PF2, (ii) PN1, (iii) PF1 and (iv)

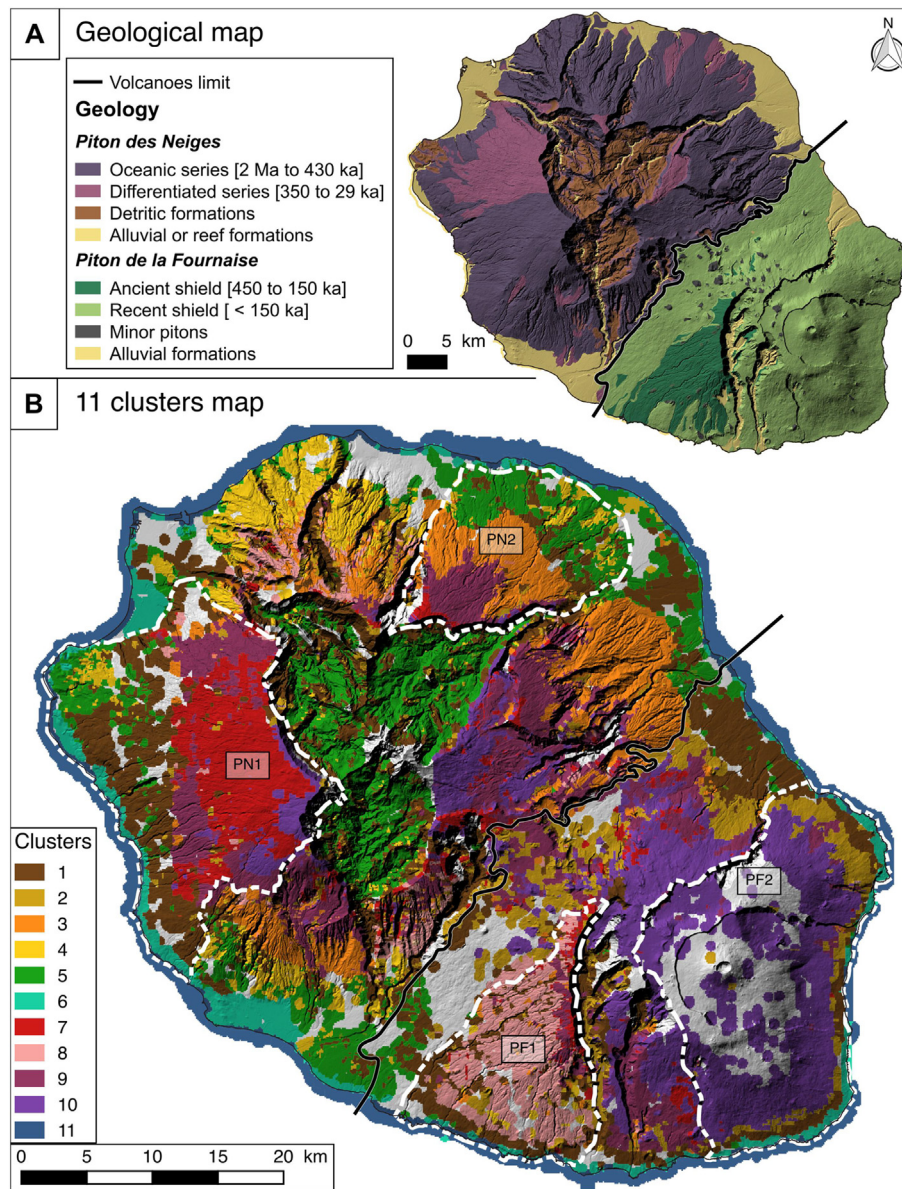


Fig. 4. The geological map [A] and [B] the 11 cluster map. The black line separates the two volcanoes: Piton des Neiges (north-west) and Piton de la Fournaise (south-east). Their activities are simplified in two periods with their age in square brackets. White dashed polygons highlight four specific areas with different weathering processes (see details in the text).

PN2. Rainfall rates and the age of volcanic formations have a coupled impact on the weathering process. First, it appears that areas with extreme rainfall rates but recent lava flows are only slightly impacted by weathering process (i.e. PF2), somehow comparable to old dry volcanic slopes (PN1). Second, the older but drier slope PF1 is less weathered than PN2 (older and wetter). The AHC results bring new elements, which could be used to update the geological map and especially the regolith map of Reunion Island. Thus, the degree of weathering depends, first, on the rainfall rates, but the age of volcanic formations is also a major factor to take into account. The difference of weathering degree between windward and leeward sides appears to increase with time. Finally, on the older volcano, results would inform on the weathering process by significant temperature.

3.4. Local interpretation: saltwater intrusion

Saltwater intrusion is a major issue for groundwater management of oceanic islands, especially in the low coastal areas where the population concentrates and where water supply requires high pumping rates in

coastal aquifers. The significant decrease in bulk resistivity caused by saltwater allows its detection with AEM surveys (Viezzoli et al., 2012; Vittecoq et al., 2015). In Fig. 4, cluster 6 [11d] is only found in coastal areas. Its signature (Fig. 3) corresponds to saltwater intrusion with a resistive layer ($> 100 \Omega.m$) overlying a highly conductive layer ($< 10 \Omega.m$). In Fig. 5, we have selected a specific area located on the west coast to confront AHC results (i.e. cluster 6 [11d]) to groundwater measurements. The aim is to validate the interpretation of regional clusters at the local scale using the AHC multi-scale analysis.

At the regional scale, the selected area is mainly characterized by cluster 6 [11d] (Fig. 5). Using the hierarchical nature of the classification, we have increased the number of clusters from 11 to 18, and then the area is divided into two clusters: 9 and 10. Cluster 10 [18d] is characterized by one thick resistive layer ($1000 \Omega.m$). In contrast, cluster 9 [18d] presents a more complex shape with alternation of resistive ($100-200 \Omega.m$) and conductive layers ($5-20 \Omega.m$) in the first 50 m. The differences between them are the resistivity of the superficial layer and the depth of the conductive layer ($< 10 \Omega.m$). These differences are illustrated in the resistivity profile (Fig. 5): deep layers are highly conductive

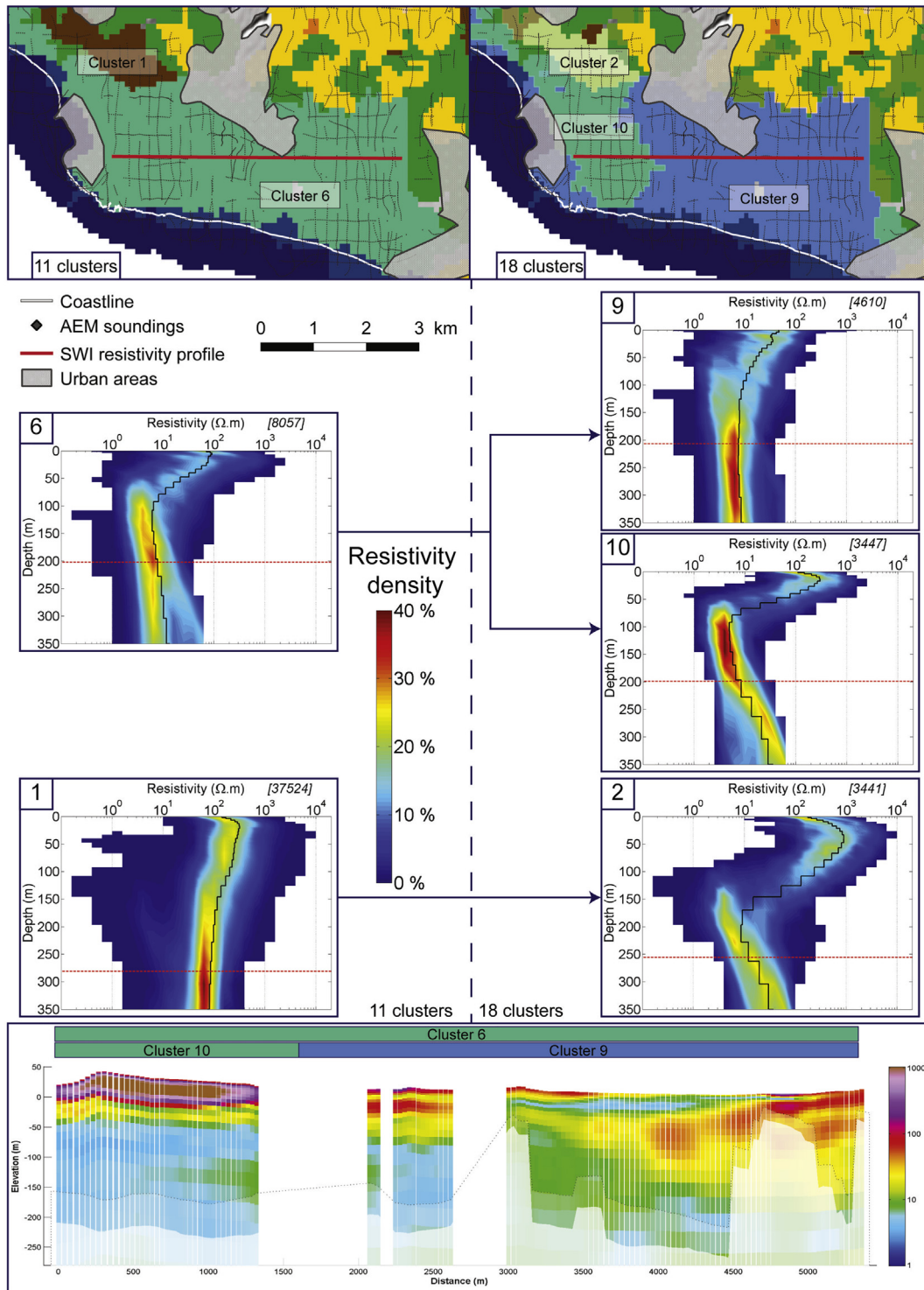


Fig. 5. Two cluster maps (11 and 18 clusters). For each map, clusters are characterized by their cluster signatures. Black line represents the median resistivity profile. The color scale represents the density of resistivity value for each cluster. The number of AEM soundings within each cluster is specified in square brackets. Below, the resistivity profile, located on the two maps by a red line, illustrates lateral resistivity variations in the area.

for all clusters, but superficial ones display sizeable resistivity and shape divergence from west to east.

These geophysical results have been confronted to geological and groundwater electrical conductivity logs (Bourhane et al., 2016). The deep conductive layer is related to saltwater intrusion, which plunges towards the east. Superficial layers of cluster 10 [18d] are related to dry basaltic sands. Conversely, superficial layers of cluster 9 [18d] are controlled by a salty lake and its deposits. The regional cluster 1 [11d]

is characterized by a superficial resistive layer (200–500 $\Omega \cdot m$) and a slow decrease with depth (Fig. 5). This cluster is located in both coastal areas and highlands. Thus, all the members of this cluster cannot be related to saltwater intrusion. In the selected area, cluster 1 [11d] becomes cluster 2 [18d] which is more contrasted and consistent with cluster 10 [18d]. Thus, cluster 1 [11d] can be related with deeper saltwater intrusion. The extension from 11 to 18 clusters allows the characterization of a local cluster in this specific area and cannot be extrapolated to the

whole cluster 1 [11d] area. Nevertheless, at the island scale the 11 cluster map permits a fast detection of deep saltwater intrusion areas.

This example shows, first, the ability of the AHC to provide quick views of a given 3D structure. Second, the hierarchical character of the method allows the integration of local studies in regional interpretation.

4. Discussion

The inversion of AEM datasets results in a large number of 1D vertical resistivity models. The visualization and interpretation of such 3D resistivity model are complex. The suggested AHC approach consists in classifying AEM vertical soundings into clusters. It is an original approach to delineate the spatial extension of 3D geological features on 2D maps. The case study in Reunion Island highlights the interest of the AHC algorithm for the interpretation of a large resistivity model in terms of geological and hydrogeological processes.

In some contexts, the interpretation of clusters can be challenging. The selection of the most efficient number of clusters should be conducted with care. The clustering process should be based on both statistical and geological/hydrogeological considerations. While the interpretation of some clusters is relatively easy, others with weak resistivity contrasts are more questionable at the regional scale. Such clusters may require a local focus to clarify their signatures at depth. This highlights the need to consider this approach at different scales and levels of details (hierarchical level and associated number of clusters). A classification with few clusters will provide a fast interpretation tool to delineate main resistivity behaviors at a regional scale. A classification with more clusters helps with more specific cluster signatures at local scale. The information collected at various levels of detail can eventually be integrated in a regional map assembling interpretations at different scales. For a proper interpretation, cluster signatures and maps should be confronted with resistivity profiles and available independent knowledge.

The diffusive nature of the electromagnetic method implies that information gradually fades away with depth. This loss of sensitivity is generally represented by the depth of investigation (DOI), which is different for each AEM sounding. Layers below the DOI should theoretically not be considered for the classification. However, the AHC algorithm requires the same number of parameters (here, number of layers) for each AEM sounding. A relevant option could be to weight the calculation of the Euclidean distance with the standardized sensitivity associated to each layer. On one hand this could limit the impact of unreliable deep resistivity values. On the other hand, the clustering might be too influenced by the DOI, at the expense of the resistivity. As the aim of our approach is to classify AEM soundings according to their resistivity profile, we preferred to disregard the DOI rather than introducing a bias in the resistivity classification. In our case, we expect that layers below the DOI have little impact on the classification. First, the median DOI is deep, implying that for the majority of soundings, only the last four layers are impacted. Second, the resistivity density (Fig. 3) shows that cluster signatures are relatively consistent under the DOI. During the inversion, layers with low sensitivity tend towards the starting model, which was spatially consistent and the same for most of the island; a more conductive starting model was chosen for the offshore ring and a more resistive for the volcanoes. Hence, we assume that the influence of the geophysical "artefacts" under the DOI on the classification is limited.

5. Conclusion

The AHC methodology is a reliable statistical tool clustering regional AEM datasets in 2D resistivity clusters. Compared to inverted resistivity models, it summarizes the 3D information from an AEM survey in two complementary results. First, the 2D map of interpolated clusters delineates areas where EM soundings display comparable vertical resistivity signatures. Second, each cluster signatures provide information on their

resistivity characteristics and their statistical reliability (i.e. heterogeneity of soundings within the cluster). Given the hierarchical nature of the clustering, AHC allows multi-scale analyses of the dataset, including local understanding in a regional approach. In heterogeneous areas, where local behaviors are hardly featured at the regional scale, the AHC is useful to individualize and focus geological and hydrogeological studies. Accordingly, this approach provides support in a wide variety of AEM applications/contextes, such as water management, mining prospecting, geotechnical studies and risk management. Finally, this methodology is a reliable tool for cross-studying multi-disciplinary results and provides 2D operational maps at regional and local scales; their interpretation is enhanced with the aid of resistivity profiles and confronted with available independent knowledge.

Acknowledgements

This work and SkyTEM survey in Reunion Island were funded by Région La Réunion, the European Union (through FEDER funding), the Office de l'Eau de La Réunion and the Bureau de Recherches Géologiques et Minières (BRGM). The authors thank Bruno Tourliere for his helpful advices on the AHC statistical method and Bernard Bourguin for cluster interpolation. We are also grateful for the language help provided by Hélène Magalon and Celia Northman.

References

- Auken, E., Christiansen, A.V., Jacobsen, L.H., Sørensen, K.I., 2008. A resolution study of buried valleys using laterally constrained inversion of TEM data. *J. Appl. Geophys.* 65: 10–20. <https://doi.org/10.1016/j.jappgeo.2008.03.003>.
- Auken, E., Christiansen, A.V., Westergaard, J.H., Kirkegaard, C., Foged, N., Viezzoli, A., 2009. An integrated processing scheme for high-resolution airborne electromagnetic surveys, the SkyTEM system. *Explor. Geophys.* 40:184–192. <https://doi.org/10.1071/EG08128>.
- Barcelo, A., Coudray, J., 1996. New annual isohyet and rainfall maximum map for the piton de la Fournaise massif (Reunion island). *J. Water Sci.* 9:457–484. <https://doi.org/10.7202/705262ar>.
- Bedrosian, P.A., Maercklin, N., Weckmann, U., Bartov, Y., Ryberg, T., Ritter, O., 2007. Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models. *Geophys. J. Int.* 170:737–748. <https://doi.org/10.1111/j.1365-246X.2007.03440.x>.
- Bourhane, A., Dumont, M., Bonnier, J., 2016. Mise en place d'un réseau de suivi des intrusions salines dans les aquifères côtiers de La Réunion (Office de l'Eau de La Réunion). Office de l'Eau http://www.eaureunion.fr/fileadmin/user_upload/Etudes/ETUDE_01046.PDF.
- Bret, L., Fevre, Y., Join, J.-L., Robineau, B., Bachelery, P., 2003. Deposits related to degradation processes on piton des Neiges volcano (Reunion Island): overview and geological hazard. *J. Volcanol. Geotherm. Res. Volcanic Hazards* 123:25–41. [https://doi.org/10.1016/S0377-0273\(03\)00026-X](https://doi.org/10.1016/S0377-0273(03)00026-X).
- Christensen, C.W., Pfaffhuber, A.A., Anschütz, H., Smaavik, T.F., 2015. Combining airborne electromagnetic and geotechnical data for automated depth to bedrock tracking. *J. Appl. Geophys.* 119:178–191. <https://doi.org/10.1016/j.jappgeo.2015.05.008>.
- Christiansen, A.V., Auken, E., 2012. A global measure for depth of investigation. *Geophysics* 77:WB171–WB177. <https://doi.org/10.1190/geo2011-0393.1>.
- Danielsen, J.E., Auken, E., Jørgensen, F., Søndergaard, V., Sørensen, K.I., 2003. The application of the transient electromagnetic method in hydrogeophysical surveys. *J. Appl. Geophys.* 53:181–198. <https://doi.org/10.1016/j.jappgeo.2003.08.004>.
- Desclotres, M., Ritz, M., Robineau, B., Courteaud, M., 1997. Electrical structure beneath the eastern collapsed flank of Piton de la Fournaise volcano. Reunion Island: Implications for the quest for groundwater. *Water Resour. Res.* 33:13–19. <https://doi.org/10.1029/96WR02673>.
- Di Giuseppe, M.G., Troiano, A., Troise, C., De Natale, G., 2014. K-means clustering as tool for multivariate geophysical data analysis. An application to shallow fault zone imaging. *J. Appl. Geophys.* 101:108–115. <https://doi.org/10.1016/j.jappgeo.2013.12.004>.
- Fitterman, D., Stewart, M., 1986. Transient electromagnetic sounding for groundwater. *Geophysics* 51:995–1005. <https://doi.org/10.1190/1.1442158>.
- Fountain, D., 1998. Airborne electromagnetic systems - 50 years of development. *Explor. Geophys.* 29:1–11. <https://doi.org/10.1071/EG98001>.
- Franke, R., 1982. Scattered data interpolation: tests of some methods. *Math. Comput.* 38: 181–200. <https://doi.org/10.2307/2007474>.
- Gillot, P.-Y., Nativel, P., 1982. K Ar chronology of the ultimate activity of Piton des Neiges volcano, Réunion Island, Indian Ocean. *J. Volcanol. Geotherm. Res.* 13:131–146. [https://doi.org/10.1016/0377-0273\(82\)90024-5](https://doi.org/10.1016/0377-0273(82)90024-5).
- Join, J.L., Folio, J.L., Robineau, B., 2005. Aquifers and groundwater within active shield volcanoes. Evolution of conceptual models in the Piton de la Fournaise volcano. *J. Volcanol. Geotherm. Res.* 147:187–201. <https://doi.org/10.1016/j.jvolgeores.2005.03.013>.
- Kluska, J.M., 1997. Evolution magmatique et morpho-structurale du piton des neiges au cours des derniers 500 000 ans. Université d'Orsay - Paris XI.

- Lénat, J.F., Gibert-Malengreau, B., Galdéano, A., 2001. A new model for the evolution of the volcanic island of Réunion (Indian Ocean). *J. Geophys. Res. Solid Earth* 106: 8645–8663. <https://doi.org/10.1029/2000JB900448>.
- Lénat, J.F., Bachèlery, P., Merle, O., 2012. Anatomy of Piton de la Fournaise volcano (La Réunion, Indian Ocean). *Bull. Volcanol.* 74:1945–1961. <https://doi.org/10.1007/s00445-012-0640-y>.
- Martelet, G., Truffert, C., Tourlière, B., Ledru, P., Perrin, J., 2006. Classifying airborne radiometry data with agglomerative hierarchical clustering: a tool for geological mapping in context of rainforest (French Guiana). *Int. J. Appl. Earth Obs. Geoinformation* 8: 208–223. <https://doi.org/10.1016/j.jag.2005.09.003>.
- Martelet, G., Reninger, P.A., Perrin, J., Deparis, J., 2014. Acquisition géophysique hélicoptère de l'île de La Réunion (Report No. RP-63818-FR). BRGM <http://infoterre.brgm.fr/rapports/RP-63818-FR.pdf>.
- Murtagh, F., Legendre, P., 2011. Ward's hierarchical clustering method: clustering criterion and agglomerative algorithm. *J. Classif.* 31:274–295. <https://doi.org/10.1007/s00357-014-9161-z>.
- Nabighian, M., 1991. Electromagnetic methods in applied geophysics. *Investigations in Geophysics. Society of Exploration Geophysicists* <https://doi.org/10.1190/1.9781560802686>.
- Paasche, H., Tronicke, J., Holliger, K., Green, A.G., Maurer, H., 2006. Integration of diverse physical-property models: subsurface zonation and petrophysical parameter estimation based on fuzzy c-means cluster analyses. *Geophysics* 71:H33–H44. <https://doi.org/10.1190/1.2192927>.
- Pryet, A., Ramm, J., Chilès, J.P., Auken, E., Deffontaines, B., Violette, S., 2011. 3D resistivity gridding of large AEM datasets: a step toward enhanced geological interpretation. *J. Appl. Geophys.* 75:277–283. <https://doi.org/10.1016/j.jappgeo.2011.07.006>.
- Reninger, P.-A., Martelet, G., Deparis, J., Perrin, J., Chen, Y., 2011. Singular value decomposition as a denoising tool for airborne time domain electromagnetic data. *J. Appl. Geophys.* 75:264–276. <https://doi.org/10.1016/j.jappgeo.2011.06.034>.
- Reninger, P.-A., Martelet, G., Perrin, J., Deparis, J., Chen, Y., 2016. Slopes of an airborne electromagnetic resistivity model interpolated jointly with borehole data for 3D geological modelling. *Geophys. Prospect.* <https://doi.org/10.1111/1365-2478.12465>.
- Salvany, T., Lahitte, P., Nativel, P., Gillot, P.Y., 2012. Geomorphic evolution of the Piton des Neiges volcano (Réunion Island, Indian Ocean): competition between volcanic construction and erosion since 1.4 ma. *Geomorphology* 136:132–147. <https://doi.org/10.1016/j.geomorph.2011.06.009>.
- Simon, B., Christiansen, A.V., Auken, E., 2009. A review of helicopter-borne electromagnetic methods for groundwater exploration. *Surf. Geophys.* 7. <https://doi.org/10.3997/1873-0604.2009043>.
- Sørensen, K.I., Auken, E., 2004. SkyTEM ? A new high-resolution helicopter transient electromagnetic system. *Explor. Geophys.* 35:194–202. <https://doi.org/10.1071/EG04194>.
- Viezzoli, A., Christiansen, A.V., Auken, E., Sørensen, K., 2008. Quasi-3D modeling of airborne TEM data by spatially constrained inversion. *Geophysics* 73:F105–F113. <https://doi.org/10.1190/1.2895521>.
- Viezzoli, A., Munday, T., Cooper, Y.L., 2012. Airborne electromagnetics for groundwater salinity mapping: case studies of coastal and inland salinisation from around the world. *Boll. Geofis. Teor. Ed Appl.* 53:581–599. <https://doi.org/10.4430/bgta0067>.
- Vittecoq, B., Reninger, P.A., Violette, S., Martelet, G., Dewandel, B., Audru, J.C., 2015. Heterogeneity of hydrodynamic properties and groundwater circulation of a coastal andesitic volcanic aquifer controlled by tectonic induced faults and rock fracturing – Martinique island (Jesser Antilles – FWI). *J. Hydrol.* 529 (Part 3):1041–1059. <https://doi.org/10.1016/j.jhydrol.2015.09.022>.
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58:236–244. <https://doi.org/10.1080/01621459.1963.10500845>.