UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

ANALYSE STATISTIQUE DES VALEURS DE RÉSISTIVITÉ ÉLECTRIQUE DU PERMAFROST DE MONTAGNE

◊◊◊

Travail de Recherche en Géographie

Un travail de Jeanbastien Kittel

Supervisé par Dr. Christian Hauck

Rendu le 28 juin 2018

Table des matières

0. Résumé/Abstract page 4
1. Introduction page 5
2. Problématique page 6
3. Considérations théoriques page 8
   - 3.1 Limites des données incluses dans le champ de recherche page 8
   - 3.2 Température, résistivité électrique et permafrost page 8
   - 3.3 Pertinence de la méthode statistique avec des données issues d'ERT page 8
   - 3.4 Algorithmes page 9
     - 3.4.1 Algorithme utilisé pour l'interpolation polynomial page 9
     - 3.4.2 Cluster k-médoïdes page 10
     - 3.4.3 Cluster en linkage page 10
4. Méthodologie page 11
   - 4.1 Base de données originale page 11
   - 4.2 Chronologie de la recherche page 13
     - 4.2.1 (a) Exploration page 13
     - 4.2.2 (b) Production page 14
     - 4.2.3 (c) Clustering page 17
5. Résultats page 18
   - 5.1 (1) Talus slope page 18
   - 5.2 (2) Rock glacier et (7) Relict rock glacier page 19
   - 5.3 (4) Weathered Bedrock page 20
   - 5.4 (5) Bedrock page 21
   - 5.5 (8) Ice-rich sediments et (11) Inconnus/autres page 22
   - 5.6 Variables de pourcentage page 25
   - 5.7 Variables de résistivité page 26
   - 5.8 Analyse en composantes principales page 27
6. Discussion page 29
   - 6.1 Poly1 pour corriger poly3 page 29
   - 6.2 Autres corrections page 31
   - 6.3 Commentaire sur les méthodes de clustering page 31
   - 6.4 Cluster et classes de rock glacier page 33
   - 6.5 Commentaire sur les unités produites page 33
   - 6.6 Validation

7. Conclusion page 35
7.1 Types/landforms page 35
7.2 Programme Iceclust2019 page 36

8. Bibliographie page 37

9. Annexes page 38

0. Résumé /abstract

[La tomographie, en tant que figure visuelle, est le résultat produit par l'investigation des propriétés de résistivité électrique du sol par la méthode ERT (Electrical resestivity tomography). En situation de permafrost, la création d'un profil permet une estimation à vue d'œil de la quantité de sol gelé. L'approche physique, basée sur le pourcentage de porosité, de glace et d'eau permet aussi une estimation de la quantité de sol gelé. Ce travail propose une approche différente, basée sur la statistique. L'objectif est d'identifier à partir de quelles grandeurs de résistivité, le sol est à considérer comme gelé. La méthode se base sur un tri des valeurs de résistivité issues d'une tomographie pour produire une courbe mathématique strictement croissante. Cette courbe est ensuite interpolée au degré 1 ou 3. À partir de l'interpolation, la valeur minimale de résistivité électrique pour le permafrost est connue. Cette valeur permet d'estimer un pourcentage. Quand la quantité de sol gelé est connu, la quantité de glace, riche ou pauvre, est estimée avec un cluster par k-médoïdes. Une grande quantité d'échantillons est ainsi analysée. Les tomographies ayant des formes géomorphologiques similaires sont regroupées avec un cluster de type linkage. Cinq catégories de formes géomorphologiques les regroupent : « Talus slope », « rock glacier », « weathered bedrock », « bedrock » et « ice-rich sediments ». Les estimations de sol gelé doivent encore être validées par l'approche physique, la géomorphologie et l'étude périglaciaire. Le regroupement en catégorie permet d'identifier des propriétés typiques pour un type de forme en situation de permafrost, ou atypique pour certains échantillons. En plus d'une estimation du sol gelé, l'approche statistique, qui n'en est qu'à ses débuts, propose aussi une estimation du sol saturé en eau.]

1. Introduction

Ces 20 dernières années, le département des géosciences de l'Université de Fribourg a collecté une importante quantité de données sur le permafrost. Ces données sont en partie constituées par des mesures ERT (Electrical resestivity tomography) en terrain de montagne ou polaire, là où le permafrost est présent. Le permafrost est suivi par les chercheurs car il tend à se réchauffer et fini par fondre à une échelle globale (Biskarborn et al., 2019). Dans ce travail, une importante quantité de mesures ERT, de différents pays, sont regroupées et analysées.

Le résultat d'une ERT est un profil. Une tomographie de résistivité électrique est une figure/une image sur laquelle on voit à quel(s) endroit(s) et à quelle(s) profondeur(s) la résistivité du sol est la plus élevée (généralement là où le sol est gelé), mais aussi où la résistivité est basse (généralement là où il y a de l'eau et de sels conducteurs dissous). Différentes tâches ont lieu en amont de l'interprétation d'une ERT. L'acquisition des données sur le terrain (data acquisition) puis le traitement par ordinateur (data processing), tel que l'inversion ou le filtrage, sont les étapes qui mènent à la représentation visuelle d'une tomographie et à son interprétation (Hauck & Kneisel, 2008). Évaluer la présence et la quantité de permafrost sur une ERT demeure impossible sans une bonne connaissance du terrain. Traditionnellement, l'interprétation d'une ERT concernant le permafrost implique la connaissance des techniques géophysiques mais aussi de l'environnement périglaciaire dans ses dimensions géomorphologiques, climatique et glaciologique. En effet, beaucoup de caractéristiques du sol sont à prendre en compte car elles modifient les résistivités mesurées. Elles concernent principalement la nature de la roche, la porosité, la saturation en eau, l'épaisseur de la couche active (la partie du sol qui dégèle et regèle chaque année), la contenance en argile, la conductivité du liquide dans les pores et dans une moindre mesure, la forme des particules et donc de la géométrie des pores (Supper et al., 2104), mais surtout la contenance en glace, car le sol peut être gelé sans contenir de glace. Toutes ces caractéristiques ne sont pas forcément répertoriées, du moins pas systématiquement pour chaque ERT. Elles sont souvent des éléments constitutifs d'une interprétation globale. De plus, l'interprétation d'une ERT intervient souvent dans une synthèse avec d'autres mesures de terrain, par exemple avec l'acquisition de mesures sismiques ou radar du sol, ou par un relevé morphologique ou géologique du terrain. Par exemple, l'étude de Sjöberg et al. (2015) utilise la méthode GPR (ground penetrating radar) parallèlement à l'ERT. Cette double méthode permet de mieux définir la limite en sol gelé et non-gelé.

Dans ce travail, l'interprétation de la résistivité du sol est basée sur une approche statistique. Un maximum de propriétés pertinentes seront explorées et répertoriées pour chaque jeu de données constitutives d'une ERT. Toutes ses données pourront ensuite être regroupées et analysées en cluster. Il s'agira d'identifier des fourchettes de valeurs de résistivités spécifiques au milieu où le permafrost apparaît ; glacier rocheux (« rock glacier »), dépôt sédimentaire (« sediments »), glissement de terrain (« talus slope ») ou roche mère (« bedrock »). Toutes ses situations reposent sur une approche par classification.

2. Problématique

Beaucoup de tableaux résument les résistivités électriques en fonction du type de matériel présent dans le sol (Hauck et Kneisel, 2008). On les appelle tableaux de résistivité spécifique (figure 1). Ces tableaux présentent une fourchette de valeur, et non pas une valeur unique, pour chaque type de matériel.

Material	Electrical resistivity (kΩ m)
Clay	0.001–0.1
Sand	0.1–5
Gravel	0.1–0.4
Granite	5–1000
Gneiss	0.1–1
Schist	0.1–10
Ground water	0.01–0.3
Frozen sediments/ground ice/mountain permafrost	1–1000
Glacier ice (temperate)	1000–100 000
Air	infinity

Figure 1 : Table de résistivité selon matériaux. L'unité des chiffres est en kΩ*m. on remarque que la glace peut avoir des résistivités nettement supérieures à des sédiments gelés. (Hauck et Kneisel, 2008)

Pourquoi n'y a-t-il pas une valeur unique pour chaque matériel ? Car le sol n'est pas une matrice homogène. Plusieurs matériaux peuvent être mélangés et se succéder en couche. La température peut varier à l'intérieur du sol. Une variation de température modifie la résistivité mesurée. Ces tableaux ne sont pas calculés uniquement pour des matériaux gelés. Ces tableaux ne sont pas calculés uniquement pour des situations où le permafrost est présent. À cela s'ajoute que la quantité d'eau et de glace change avec les saisons et les conditions météorologiques. Le contenu d'eau dans le sol fait aussi varier la conductivité. On parle de mesure en condition sèche ou en condition humide, selon la quantité d'eau non gelée en surface et sous la surface. Différents sels peuvent être présents dans l'eau du sol, ce qui fait aussi varier la conductivité. On ne peut

donc pas en regardant une simple mesure de la résistivité à une profondeur donnée savoir de quoi est composé le sous-sol. A contrario, on peut savoir à l'aide d'une ERT, à quelle profondeur et en quelle quantité un matériel est présent si on sait qu'il est probable de le rencontrer. Si on sait qu'il y a de la glace, on peut savoir à quelle profondeur. C'est le contraste de valeurs de résistivité entre différents points de la tomographie qui permet de signaler la présence ou non de tel ou tel élément connu et attendu.

Ce travail cherche aussi quelles résistivités apparaissent mais en s'appuyant sur une grande série de mesures de terrain qui regroupent des caractéristiques multiples. L'objectif est de comparer, non pas uniquement des éléments différents, mais de milieux différents et des situations différentes afin de répondre à la question de recherche suivante :

Quelles valeurs typiques de résistivité électrique sont mesurées en fonction de caractéristiques hydrologiques, morphologiques, géographique et physiques du terrain en situation de permafrost ?

Pour y arriver, ce travail cherchera pour chaque ERT à isoler statistiquement une partie des données qui seront représentatives d'une partie de l'état du terrain (gelé ou pas, saturé en eau ou pas) pour en connaître la résistivité locale. Les échantillons seront ensuite explorés et corrélés. Chaque ERT sera sujette à un cluster. L'objectif final sera de définir des résistivités moyennes typiques pour des milieux et des caractéristiques précises. Pour ce faire la méthodologie devra s'appuyer sur des critères statistiques connus et robustes mais aussi justifier de nouvelles manières de diviser les jeux de données. Les difficultés statistiques de ce travail seront d'être à la fois analytiquement et synthétiquement pertinentes. Il s'agit à la fois de trier et diviser chaque fichier, mais aussi de rassembler plusieurs jeux de données ensembles. Le domaine de recherche et celui des statistiques de l'environnement, des sciences de la terre et de la géophysique.

3. Considérations théoriques

3.1 Limites des données incluses dans le champ de recherche

Ce travail ne se focalisera pas sur la qualité de la représentation graphique d'une tomographie, ni sur une propriété unique lié au processus de formation du permafrost ou des environnements glaciaire et périglaciaire. Ce travail doit relever le défi de mettre en lumière des variables communes ou des valeurs de résistivité de référence en fonction de la quantité d'informations disponibles qui accompagnent un jeu de données. La température du sol ne fait, par exemple, pas parti des informations disponibles.

3.2 Température, résistivité électrique et permafrost

En théorie, la résistivité augmente linéairement quand la température décroit jusqu'au point de congélation. Comme le permafrost n'apparait que sous le 0°C, une croissance linéaire des résistivités indique (dans un milieu homogène) une baisse de la température du sol qui reste néanmoins positive. Si la température passe sous le point de congélation, la résistivité augmente exponentiellement jusqu'à la température la plus basse du permafrost (Hilbich et al., 2008). Cette loi s'exprime ainsi :

$$\rho = \rho_0 * e^{-b(T)} \quad (1)$$

Avec ρ résistivité mesurée en Ω*m, ρ_0 une constante, -b une constante en K^{-1} et T la température négative en Kelvin .

La partie de gelé et non-gelé du sol ont donc des ordres de grandeurs de résistivités différentes. La valeur de résistivité minimum qui définit où le sol n'est plus gelé, change dans tous les terrains et doit être déduite. Dans des conditions particulières, par exemple des palsas dans un milieu marécageux, des résistivités de 1000 à 1700 Ωm ont été utilisée comme seuil de référence (Sjöberg et al., 2015). En général dans les Alpes les résistivités attendues sont de l'ordre de 10 kΩm à 3 MΩm (Supper et al., 2014).

3.3 Pertinence des données issues d'ERT

L'ERT est une méthode d'investigation du sol non invasive par injection de courant électrique dans le sol. Elle n'est pas représentative de tous les composants et de toutes les couches du sol. Sa sensibilité est grandement déterminée par la quantité d'air d'eau et de glace qui compose le sol. En ce sens, faire des liens avec le type de matière sédimentaire ou structurale (qui influe autant sur la porosité que sur le contenu en eau) est

forcément complexe. En ce sens on peut se demander si les mesures ERT peuvent être statistiquement prises au sérieux. Les résistivités apparentes, c'est-à-dire les résistivités directement mesurées sur le terrain comportent une grande part d'erreur dues à l'interférence du ou des milieu(x) que traverse le courant électrique. Un set de données ERT doit donc avoir été inversé et évalué (data processing) pour être jugé fidèle à la réalité. L'inversion est un algorithme qui permet d'observer à partir d'un set de données de résistivités apparentes, la résistivité spécifique dans une projection en 2D (Hauck et Kneisel, 2008). En plus de transformer les résistivités, l'inversion inclue aussi dans la projection en 2D de la topographie, ce qui très important dans le contexte alpin et de montagne. Dans ce travail tous les sets de données ERT ont été inversés et sont ainsi à l'image de la réalité. La base de données, qui est constituée de toutes les ERTs est fiable.

3.4. Algorithmes

Cette partie explique uniquement comment fonctionne, en théorie, les fonctions et les algorithmes utilisés dans ce travail. Le but et la manière dont ils ont été utilisés par la suite, est décrit dans la méthodologie.

3.4.1 Algorithme utilisé pour l'interpolation polynomial

L'interpolation polynomiale de degré 3 est résolue en utilisant la fonction « *polyfit* » de Matlab®. Cette fonction évalue simplement le meilleur polynôme au sens des moindres carrés. Pour ceci, la fonction utilise un algorithme basé sur une matrice de Vandermonde (Mathworks, 2009a) :

$$\begin{pmatrix} x_1^n & x_1^{n-1} & \cdots & 1 \\ x_2^n & x_2^{n-1} & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ x_m^n & x_m^{n-1} & \cdots & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \qquad (2)$$

Avec *m*, le nombre de points interpolant *x*, *n* le degrés du polynôme, *P(1)…P(n+1)* les coefficients du polynôme, *y(0)…y(m)* les résistivités en chaque points x.

Deux problèmes ont été rencontré dans ce travail. Les données sont parfois mal conditionnées. Si l'écart entre les points de mesures est de h=1, ce qui arrive si on trie des points numérotés de 1 à *m* tel que : {x(1), x(2), … , x(*m*-1), x(*m*)} avec *m* le nombres de points, le programme Matlab® propose de mieux conditionner les données. Ce problème se résolve en utilisant des valeurs centrées réduites pour les points interpolants.

D'un point de vue théorique: *« Polynomials are unbounded, oscillatory functions by nature. Therefore, they are not well-suited to extrapolating bounded data or monotonic (increasing or decreasing) data »* (Mathworks, 2009a). Il est effectivement peu approprié de chercher un polynôme sur une courbe croissante uniquement, mais essentiellement en termes d'efficience et d'optimisation numérique, par exemple pour faire une prédiction du point x(*m*+1). Dans notre cas, cela n'a pas posé de problème.

3.4.2 Cluster k-médoïdes

La recherche de cluster par un algorithme k-médoïdes teste pour chaque classe, si un changement de points comme nouveau médoïde (un médoïde est le centre d'une classe), augmente ou diminue la distance entre les points à l'intérieur de la classe. L'algorithme utilisé : « *partionning around medoids* » (PAM) est lent pour un grand nombre de données mais plus précis. Il fonctionne en deux temps (Mathworks, 2009b).

- Étape 1 : Tous les points sont assignés au médoïde le plus proche (au début de l'algorithme le premier point est choisi au hasard, il est possible d'optimiser l'algorithme en choisissant ce point de départ).

- Étape 2 : Tous les points à l'intérieure de la classe sont testés en tant que potentiel médoïde, le point qui minimise au mieux les distances intra-classe est choisi comme nouveau médoïde.

En conclusion : à la fin de l'algorithme, chaque point est assigné au médoïde le plus proche, et que chaque médoïde et au centre de sa classe (figure 2). La séparation finale des classes a lieu quand la somme des variances des différentes classes est définitivement minimum.

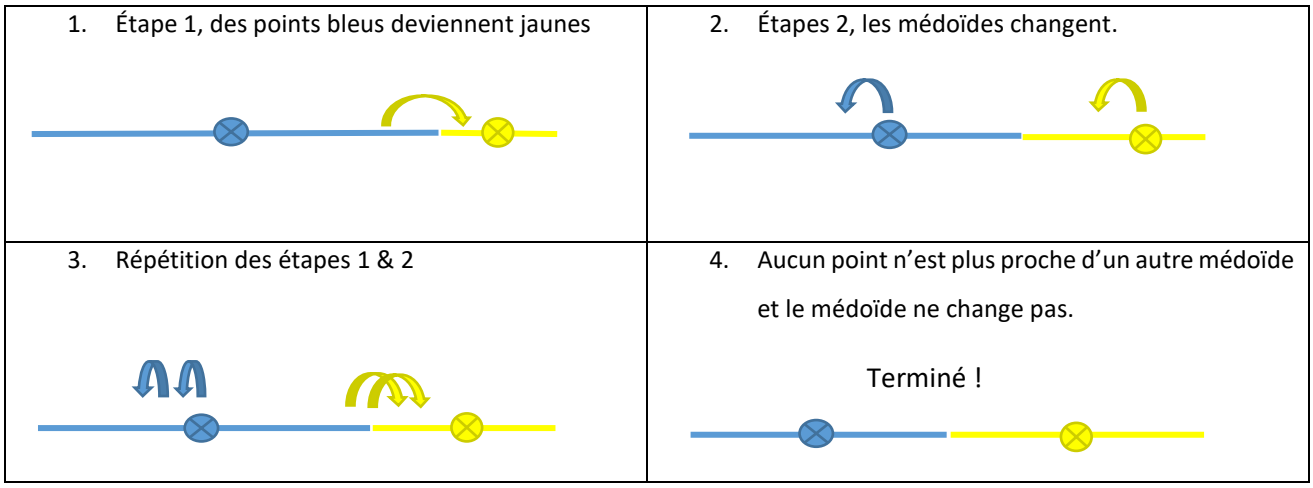


Figure 2 : schéma de l'algorithme K-médoïdes pour une seule variable avec une distribution linéaire.

3.4.3 Cluster en linkage

Un cluster en linkage est une classification hiérarchique des échantillons en fonction de leur ressemblance. Dans ce travail, c'est la distance euclidienne minimum qui est la méthode utilisée pour cette classification. Un dendrogramme est la représentation visuelle la distance entre les échantillons qui ressemble à un arbre (figure 3). Le nombre de classe décidées indique ou les branches de l'arbre sont coupées en commençant par le tronc.

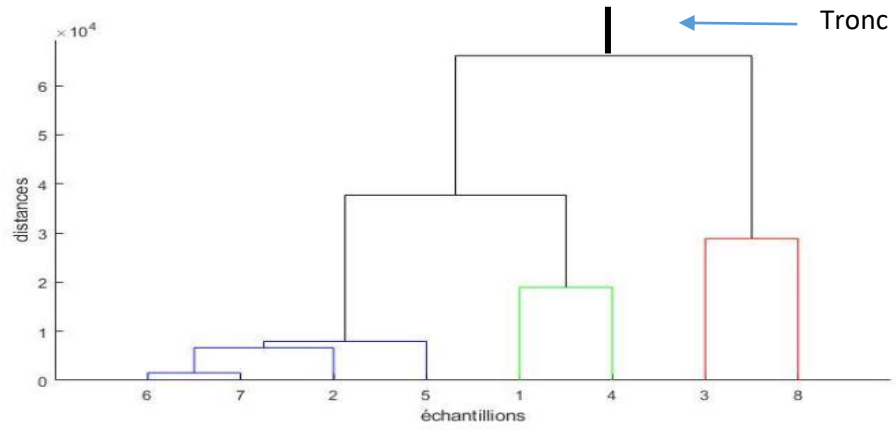


Figure 3 : Exemple de dendrogramme à huit échantillons. Le partitionnement est ici de 3 classes.

Le cluster de type linkage mesure la similarité des échantillons deux à deux (Trauth et al., 2006). La représentation graphique convient bien à ce type de cluster parfait pour de petits échantillons (n<30). Un nouveau cluster divise toujours un seul des clusters déjà identifié en deux. On peut ainsi rajouter une classe au cluster uniquement en séparant une classe préexistante. Par exemple, pour rajouter une classe au cluster (figure 3), et passer de 3 à 4 classes, il faudrait séparer la classe rouge en deux, puisque la distance entre les échantillons y est la plus élevée.

4. Méthodologie

4.1 Base de données originale

La base de données est constituée de 84 tomographies ERTs, que l'on appellera dans la suite du travail : échantillons. Trois variables de base accompagnent chaque échantillon : l'altitude moyenne, la localisation et la forme géomorphologique (landform, en anglais).

La forme géomorphologique est ici à comprendre comme la structure sur laquelle la tomographie a eu lieu. La dénomination landform remplacera le terme « forme géomorphologique » pour toute la suite du travail. Les principales landforms sont les suivantes: Talus slope, rock glacier, debris covered rock glacier, weathered bedrock, bedrock, relict rock glacier. On dénombre aussi des landforms mixtes tels que : sediments/protalus

rempart, talus slope/moraine, agricultural field/sediment et bien d'autres. Dans ce cas, le premier terme fait référence, sauf pour « block field/bedrock ».

Les localisations sont : Argentine, Chili, Espagne, Italie, Norvège et Suisse. Les échantillons ont été regroupés (tableaux 1 et 2) : Le type « block field/bedrock » a été assimilé au groupe 4 pour laisser le groupe 5 constitué de bedrock uniquement. L'Italie et la Suisse ont été regroupé dans le même groupe car il s'agit de pays alpin. L'Argentine et le Chili ont été regroupé car ils partagent la même chaîne de montagnes, et que le nombre d'échantillons argentins n'est que de deux. Six échantillons ont la localisation, l'altitude et/ou le landform qui manquent. Ils sont regroupés dans la catégorie de localisation 11 (tableau 2). Les premières colonnes du fichier Liste_84ERT_20variables.xls résument les variables attribuées à chaque échantillon.

Landforms	**Catégories incluses**	**nombre**
(1) Talus slope	Talus slope (talus slope/moraine - talus slope/sediment)	8
(2) Rock glacier	Rock glacier (debris covered rock glacier)	24
(4) Weathered b.	Weathered bedrock (weathered bedrock with debris cover – block field/bedrock)	11
(5) Bedrock	Bedrock (bedrock plateau)	11
(7) Relict rock g.	Relict rock glacier	3
(8) Ice-rich s.	Ice-rich sediments - iceWedge - ice-cored moraine - debris cov. glacier - dead ice	13
(11) Inconnus/autres	Catégorie inconnue - solifluction lobes - agricultural field/sediment - sediment - ground moraine	14

Tableau 1 : Regroupement des échantillons tel qu'on peut le voir dans la colonne « type » du fichier Liste_82ERT_20variables.xls

Localisation	**Catégories incluses**	**Nombre**
Groupe 1	Italie et Suisse (car ce sont des pays Alpin)	37
Groupe 2	Argentine et Chili (même chaine de montagne)	19
Groupe 3	Norvège	19
Groupe 11	Catégorie inconnue (6 échantillons) et Espagne (1 échantillon)	7

Tableau 2 : Regroupement des échantillons tel qu'on peut le voir dans la colonne « localisation » du fichier Liste_82ERT_20variables.xls

Chaque échantillon possède 3 variables mais est en soit une tomographie d'ERT (chapitre 1). Une tomographie compte entre 400 et plus de 10'000 points de mesures, qui comme dans une grille, voit chaque croisement donné par une coordonnée horizontale X et verticale Y, la mesure de résistivité par Z. Il est

important de précier que la valeur de résistivité n'est pas celle mesurée directement sur le terrain, mais celle issue du traitement par inversion (chapitre 3.3). Elle est considérée comme fiable. Pour chaque échantillon, le set de données est écrit sous la forme d'un fichier .xyz ou .txt.

4.2 Chronologie de la recherche

La méthodologie de ce travail se résume en 3 étapes clés. Trouver une manière efficace d'estimer la quantité de sol gelé, de sol non gelé, de sol saturé en eau sur une tomographie d'ERT, c'est-à-dire sur un échantillon ((a) Exploration). Développer un programme qui analyse tous les échantillons de la même manière ((b) Production). Regrouper les échantillons en catégorie de landforms et chercher des clusters avec les estimations produites ((c) Clustering).

4.2.1 (a) Exploration

La première partie du travail a été consacrée à développer un programme capable d'isoler la partie de sol gelé, du sol non gelé et de sol potentiellement saturé d'eau. Trois différentes manières de découper une tomographie ont été mises à l'essai. Ces découpages sont intitulés poly3, Z (k-médoïdes) et Zlog. Tous les échantillons ont été évalués avec ces trois découpages pour observer leurs efficacités. Deux exemples représentatifs, CER2018 et GUOL3, sont commentés (annexe 1 & 2). Il ressort de cette étape que :

- Le découpage poly3, qui fonctionne sur une interpolation de degré 3 (chapitre 3.4.1), a été retenu comme base du « programme Iceclust2019 » pour l'évaluation de tous les échantillons. Les raisons sont qu'il identifie le mieux la partie de sol non gelé, qu'il ne surestime pas la proportion dédiée aux basses résistivités, qu'il peut être amélioré en modifiant certains paramètres, que si le découpage poly3 est mauvais il peut être remplacer par un découpage poly1.
- Le second découpage, utilise une méthode à l'origine développée pour du clustering. Il s'agit de la fonction en k-médoïdes (chapitre 3.4.2). Dans notre cas le résultat est plus proche d'une séparation que d'un cluster, mais la méthode est inchangée. Ce découpage s'est révélé imprécis à distinguer la partie gelée de la partie non-gelée. Il a été gardé pour distinguer de manière automatique la partie du sol gelé riche en glace, de la partie pauvre en glace, car il a été efficace (figure 4a et 4b) à distinguer la différence de résistivité dans la partie gelée qui évolue exponentiellement (équation (1)).
- Le troisième découpage, intitulé cluster Zlog, est le même cluster que le second mais avec les logarithmes des résistivités comme éléments à partitionner. Ce cluster n'a pas apporté d'élément remarquable ou positif. Il n'a pas été gardé pour la suite.

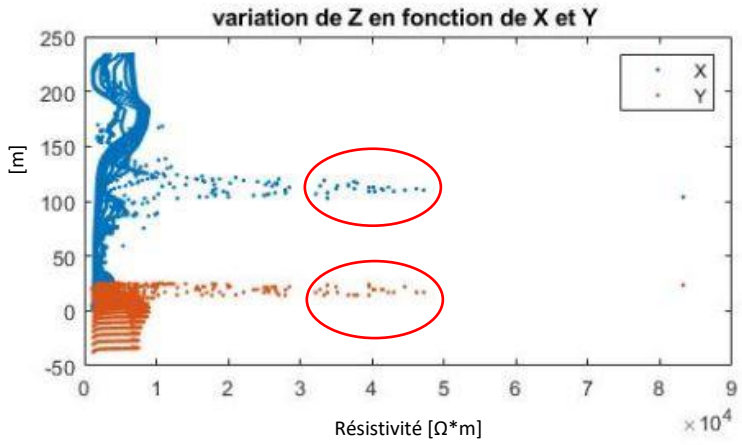


Figure 4a : variation des résistivités (Z) en fonction de la profondeur (Y) et de l'horizon (X) de la tomographie d'ERT F10_Barriales. On remarque qu'un groupe de point a des résistivités supérieures à 30'000 Ω*m. On peut les interpréter comme du sol gelé potentiellement riche en glace.

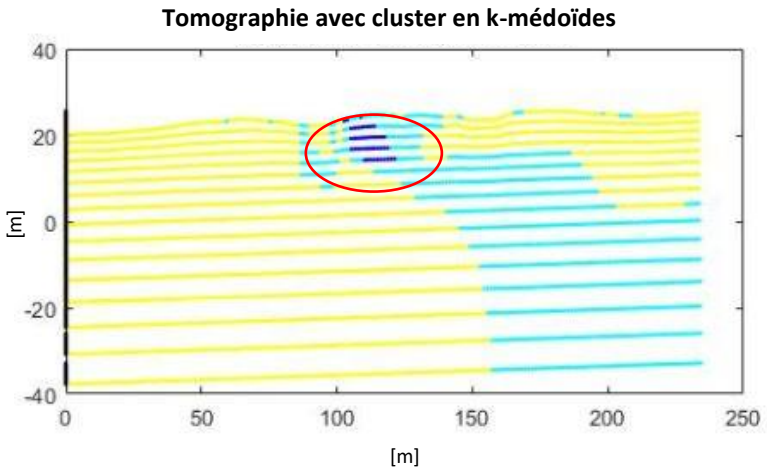


Figure 4b : tomographie de F10_Barriales avec une séparation Z k-médoïdes en 3 classes. Les points bleus foncés correspondent aux points à haute résistivité (cercle rouge).

4.2.2 (b) Production

Pour que les estimations soient comparables entre elles, mais aussi dans un souci de reproductibilité, ce travail c'est concentré dans un deuxième temps, sur un programme qui donnerait une estimation définitive des quantités et des moyennes : le programme Iceclust2019.m (annexe 3)

Le nombre de catégories estimées est passé de trois à quatre car la partie gelée est identifiée comme riche en glace ou comme pauvre en glace avec un cluster K-médoïdes. Des points de trop haute résistivité, intitulés « toohighpts » faussent l'estimation du cluster k-médoïdes de la partie du sol gelé riche en glace. Un point est défini « toohighpts » quand il est supérieur de 15% à la moyenne des trois points qui le précédent. Le programme Iceclust2019 réduit leur résistivité à 1% de plus que la valeur qui la précède. Cette correction ne s'applique que sur les 30 derniers points au maximum. Parfois, aucun point ne doit être corrigé. Si un point bleu « · » est remplacé par une astérisque « * » sur la courbe des résistivités triées croissantes (figure 5), c'est que celui-ci a été corrigé à la baisse.

Fonctionnement du programme Icclust2019.m (à comparer avec les figures 5 et 6) **:**

1. Courbe des résistivités triées dans l'ordre croissant. (courbe bleu sur le graphique …)

2. Évaluation par un polynôme de degré 3. (courbe orange sur le graphique, « curvefit »)

(figure 5) Possibilité de bifurquer en degré 1 (figure 6) …étape 9

3. Identification du minimal et du maximal de la courbe polynomiale. (points rouges)

(Si pas de min ou de max alors recherche du point d'inflexion (forme de losange))

4. Création d'une ligne entre le minimal et le maximal.

(Si pas de min ou de max, création de la ligne avec des points situés avant et après le point d'inflexion. Exemple : 250 points avant et après, voir « sample », (annexe 3).

5. Toutes les résistivités 5% en dessus sont du sol gelé.

6. Toutes les résistivités 5% en dessous indique le sol saturé en eau. la partie saturée s'arrête à la « lowvalue » (rond jaune)

7. Le programme modifie les points « toohighpts »

8. Le cluster en K-médoïdes sépare la partie riche et pauvre en glace (rond violet et rond vert) … FIN.

9. Le programme fait une interpolation entre la plus basse résistivité et la plus haute. (courbe orange)

10. Soustraction de la valeur de la courbe bleue à la valeur de la courbe orange.

11. l'écart maximum indique le début du sol gelé

Retour à l'étape 7.

le minimum indique la « lowvalue » (Rond jaune)

12. Si le minimum est au point 1sur la courbe bleue, le programme effectue une deuxième recherche de maximum, sur la première moitié de points (du 1er point rouge au losange) avant le sol gelé.

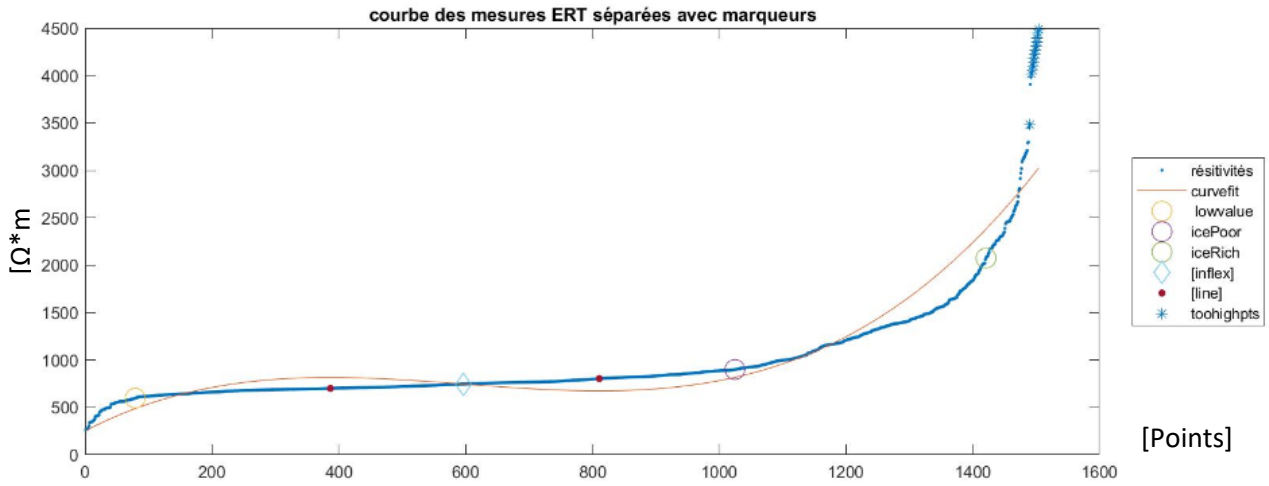


Figure 5 : Courbes des résistivités de l'échantillon ALBW2018, on peut y observer toutes les actions du programme Iceclssut2019.m en mode polynôme de degrés 3.

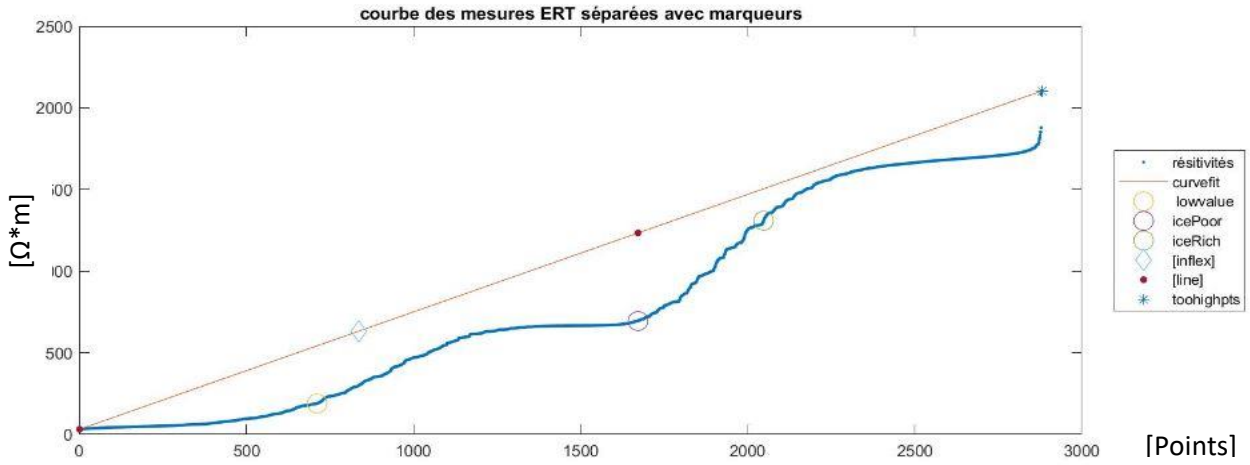


Figure 6 : Courbes des résistivités de l'échantillon SVAL_End. On peut y observer toutes les actions du programme Iceclssut2019.m en mode polynôme de degrés 1. Le programme a effectué une deuxième recherche de maximum entre le point 1 et le point situé sous le losange.

Au final, le programme Iceclust2019 fournit 17 variables pour chaque échantillon (tableau 3). Ces estimations sont rassemblées et ajoutées aux variables d'altitude, de landform et de localisation dans le fichier 84échantillons_20variablesparERT.xls. Pour chaque échantillon, une fiche récapitulative de 5 graphiques (annexe 4) est créé et rajouté au dossier DATA_Iceclust_plot.

Prct_low	% du sol saturé en eau
Prct_palier	% du sol non gelé
Prct_icePoor	% du sol gelé pauvre en glace
Prct_iceRich	% du sol gelé riche en glace

Prct_iceTotale	% du sol gelé total
Mean_low	moyenne des résistivités su sol saturé en eau [Ω*m]
Mean_palier	moyenne des résistivités du sol non gelé [Ω*m]
Mean_icePoor	moyenne des résistivité du sol gelé pauvre en glace [Ω*m]
Mean_iceRich	moyenne des résistivité du sol gelé riche en glace [Ω*m]
Mean_iceTotal	moyenne des résistivité du sol gelé total [Ω*m]
low_value	valeur maximum des résistivité basses
low_value_pt	n° du point ou se situe la low_value
icePoor_value	valeur minimum du sol gelé pauvre en glace
icePoor_value_pt	n° du point ou se situe la icePoor_value
iceRich_value	valeur minimum du sol gelé riche en glace
iceRich_value_pt	n° du point ou se situe la iceRich_value
Too_high_pts	nombre de point ou la résistivité a dû être corrigée à la baisse

Tableau 3 : Variables calculées par le programme Iceclust2019.

4.2.3 (C) Clustering

En langue française « cluster » signifie partitionnement ou regroupement. L'approche par « clustering » désigne plusieurs méthodes algorithmiques utilisées dans des domaines divers, tel que l'économie, la psychologie, les sciences naturelles, et qui ont pour but de diviser un groupe en un nombre connu ou inconnu de classes. C'est un processus de discrétisation qui peut aboutir à une classification en différents groupes ou à l'émergence d'hypothèse(s) nouvelle(s), par exemple en identifiant un idéal de partitionnement ou en identifiant des échantillons différents ou éloignés du groupe.

Pour chaque groupe de landforms, un cluster en linkage (chapitre 3.4.3) avec différentes combinaisons de variables de résistivités moyennes (tableau 3) sont envisagés. Les variables de pourcentage ont été écartées pour la production de clusters car on ne connait ni les dimensions, ni le positionnement du profil de la tomographie. Les variables de résistivités moyenne sont en ce sens plus fidèles à la réalité car elles se mesurent sur des sections déterminées.

Après avoir procédé à des clusters intra-landform, il est intéressant de comparer les landforms entre eux, ainsi que tous les échantillons entre eux. Une analyse en composante principale (PCA) entre tous les échantillons, en incluant leur appartenance à un landform, va mesurer de potentielles similarités. Cette analyse présente l'avantage d'être visuellement explicite malgré une constitution basée sur plusieurs variables. Il est nécessaire, au préalable d'identifier, des échantillons trop différents des autres qui pourraient masquer la variabilité de l'ensemble par leur poids relatif trop important (Caloz et Collet, 2011).

5. Résultats

L'essentiel des graphiques de ce chapitre se base sur des valeurs de résistivités. En référence au point 5.7 les échantillons MCO2018, SWHmoraine, ROTH et ROTV ne sont pas représentés.

5.1 (1) Talus slope

Deux classes distinctes se démarquent. La première, en bleu, a des résistivités moyennes de sol non-gelé qui sont 5 à 6 fois inférieures à la résistivité moyenne de sol gelé (figure 7). La seconde, en jaune, a des résistivités moyennes de sol gelé supérieures à 40'000 Ω*m. L'échantillons RBS2017 a une résistivité de sol non-gelé environ 30 fois inférieure à celle de son sol gelé. DRE2014 a une résistivité moyenne de sol non gelé 10 fois inférieure.

Les résistivités moyennes du sol non-gelé des échantillons sont en général comprises en 800 Ω*m et 6000 Ω*m pour talus slope. C'est la fourchette la plus petite de tous les landforms.

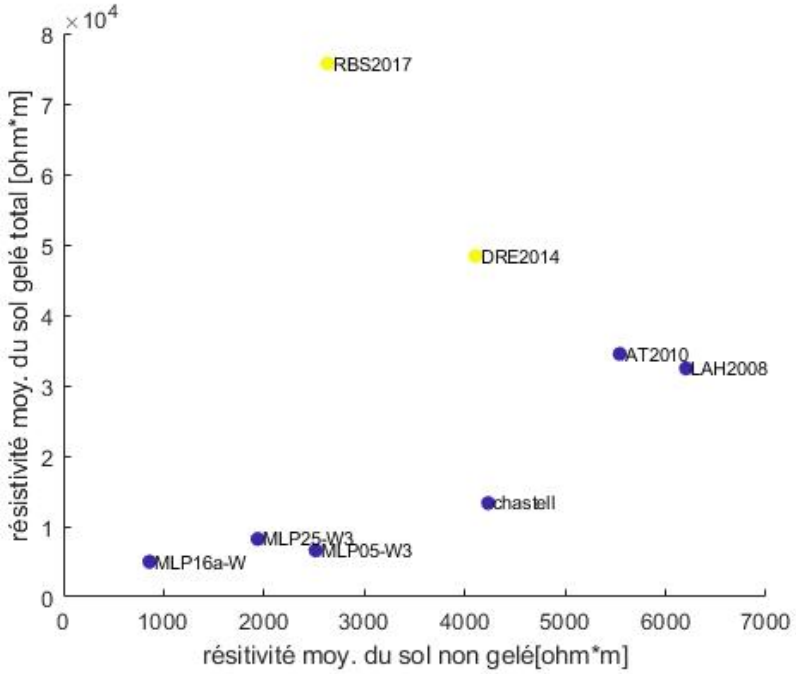


Figure 7 : Landform (1) Talus slope, représenté graphiquement avec les variables de sol gelé et non-gelé, réparti en deux classes à partir d'un cluster en linkage selon mean_low, mean_palier, mean_icePoor et mean_iceRich.

5.2 (2) Rock glacier et (7) Relict rock glacier

Trois classes sont identifiables sur la représentation graphique (figure 8) qui compare la quantité de sol gelé riche et pauvre en glace. Cette représentation s'est révélée plus intéressante que la comparaison entre résistivités moyennes du sol gelé et non-gelé. Les différentes classes sont bien séparées. La première classe (en jaune) est constituée des échantillons dont les résistivités moyennes du sol gelé pauvre en glace sont inférieures à 10'000 Ω*m. Tous les Relict rock glacier appartiennent à cette première classe. la seconde classe (en violet) a des résistivités moyenne de sol gelé riche en glace supérieure à 100'00 Ω*m. La troisième classe (en vert) est constituée de cinq échantillons dont les résistivités moyennes de la partie du sol gelé riche en glace est supérieure à 350'000 Ω*m. Sans compter MLP16bw, muraglj et RB3eRt, le rapport entre sol gelé riche en glace et pauvre en glace est de 2.67

Il n'y a pas de lien entre résistivités et une variable de pourcentage. Les pourcentages de sol gelé riche en glace sont très épars pour le landform Rock glacier et va de 1 à 15%. Le pourcentage de sol gelé pauvre en glace va de 5 à 45%. Cette variation est aussi due à l'emplacement du profil sur le glacier rocheux. En ce sens, l'analyse des résistivités moyennes est plus fiable que celle des pourcentages.

La différence entre la classe jaune et la classe violette pourrait être due à la profondeur de la partie gelée et riche en glace du sol. Cette réflexion est menée dans la discussion, chapitre 6. Les résistivités moyennes de sol non-gelé monte jusqu'à 30'000 Ω*m et de sol gelé total jusqu'à 150'000 Ω*m.

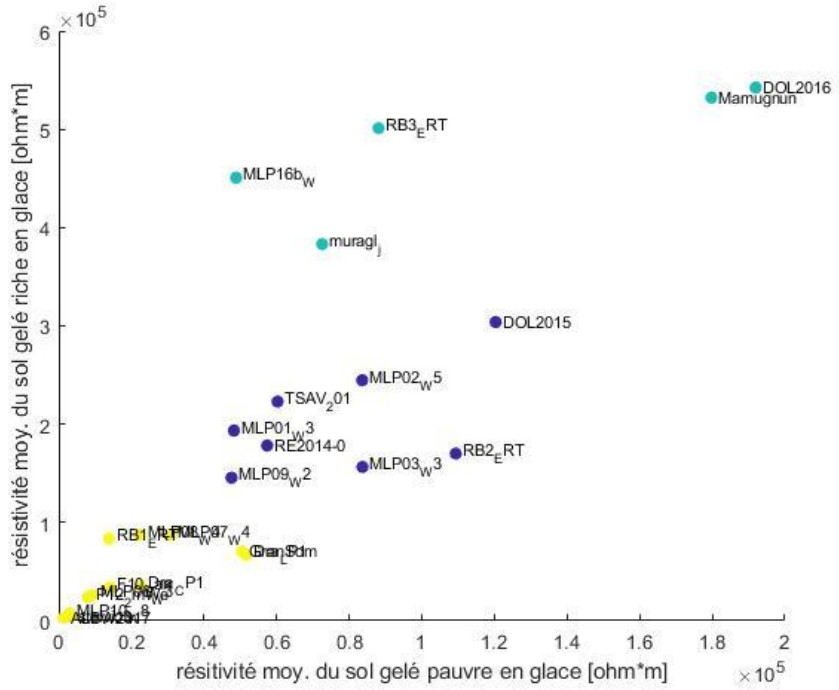


Figure 8 :

Landforms (2) et (7), représentés graphiquement avec les variables de sol gelé riche et pauvre en glace, répartis en trois classes à partir d'un cluster en linkage selon mean_low, mean_palier, mean_icePoor et mean_iceRich.

5.3 (4) Weathered bedrock

La représentation graphique (figure 9) montre une première classe (en violet) qui contient la majorité des échantillons. Les échantillons de la classe jaune ont des résistivités moyennes de sol gelé supérieure à 50'000 Ω*m. Cette grandeur est comparable avec Talus Slope qui sépare les classes à 40'000 Ω*m.

Les échantillons qui ont les résistivités de sol gelé les plus hautes, c'est à dire KIST200 et ISK1+22, ont aussi les pourcentages de sol gelé les plus élevés ; 29% de sol gelé pour KIST200 et 30% de sol gelé pour ISK1+22. SCH2015 et GUOL1 ont des pourcentages inférieurs à 4% de sol gelé. Loma2tet a un pourcentage de sol gelé de 5%.

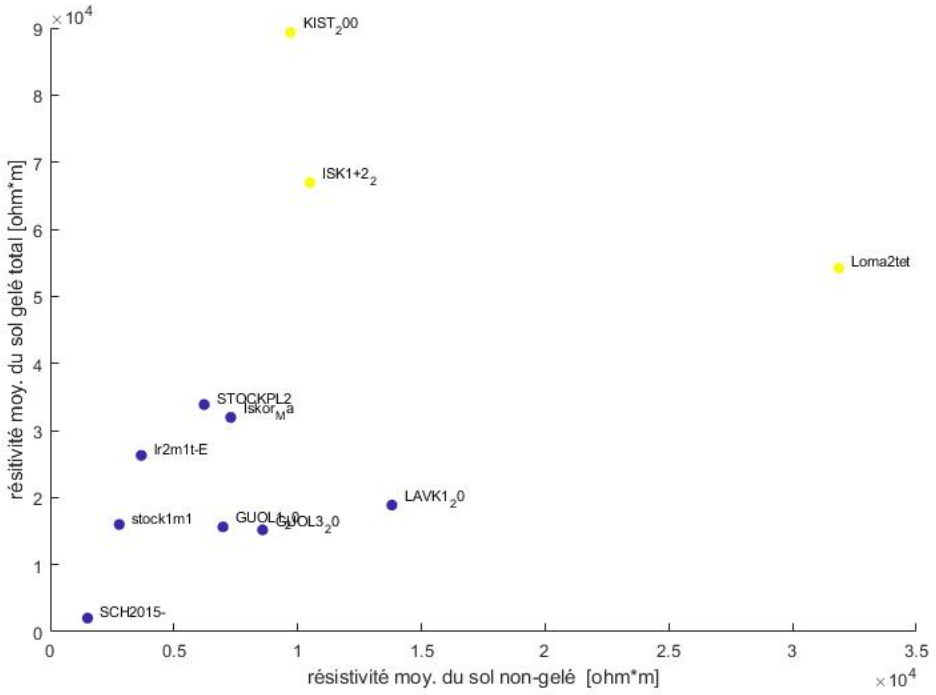


Figure 9 : Landform (4) weathered bedrock, représenté graphiquement avec les variables de sol gelé ou non gelé, réparti en deux classes à partir d'un cluster en linkage selon mean_low, mean_palier, mean_icePoor et mean_iceRich.

5.4 (5) Bedrock

Deux éléments se démarque sans avoir pour autant le même comportement (figure 10). Gamanjunn1 a une forte résistivité moyenne de sol gelé, mais une résistivité moyenne de sol non-gelé similaire à la majorité des échantillons (en violet). Une hypothèse serait que Gamanjunni1 et JUV_PACE, qui n'ont d'ailleurs pas le même pourcentage de sol gelé, soit 5.5% pour Gamanjunni1 et 30% pour JUV_PACE, contiennent dans les pores de la roche de la glace très pure, comparable à de la glace de glacier (figure 1) qui est à l'origine de cette haute résistivité. Leur substrat rocheux n'en demeure pas moins différent, pour preuve la différence de résistivité moyenne de sol non-gelé.

D'une manière générale, en comparant les figures 9 (Weathered Bedrock) et 10 (Bedrock), on constate que les résistivités moyennes des échantillons de landform Bedrock sont environ 60% plus élevées : Moyenne des résistivités moyennes du sol non-gelé pour Bedrock : 15'600 Ω*m, pour Weathered bedrock : 9'400 Ω*m. Moyenne des résistivités moyennes du sol gelé pour Bedrock : 51'800 Ω*m, pour Weathered bedrock : 33'700 Ω*m.

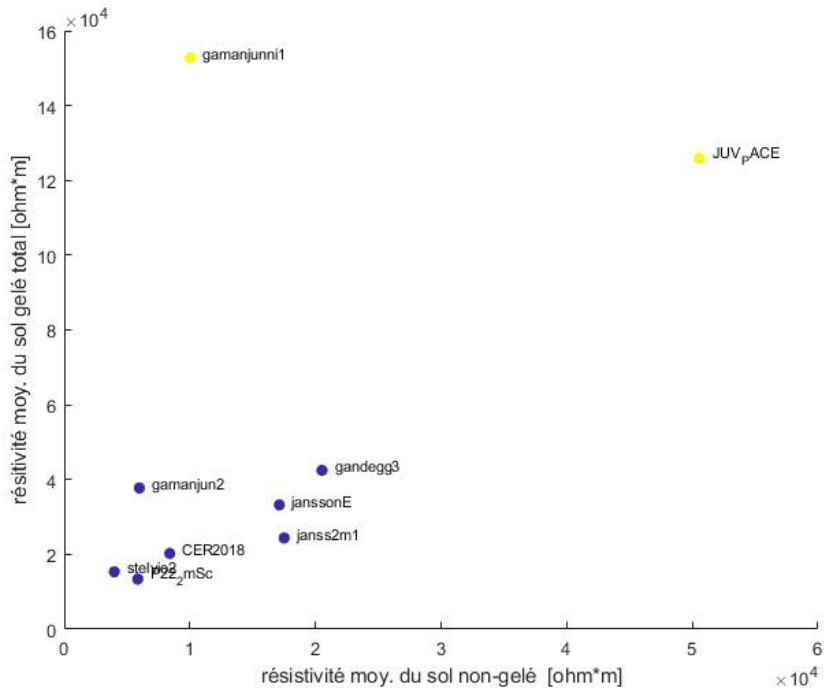


Figure 10 : Landform (5) Bedrock, représenté graphiquement avec les variables de sol gelé ou non gelé, réparti en deux classes à partir d'un cluster en linkage selon mean_low, mean_palier, mean_icePoor et mean_iceRich.

5.5 (8) Ice-rich sediments et (11) Inconnus/autres

Les échantillons de (11) Inconnus/autres sont traités avec (8) Ice-rich sediments (figure 11). Le but est de mieux cibler, visualiser et déterminer la possible appartenance des échantillons dont le landform reste inconnu (tableau 4), à une des catégories de landforms principales ; (1) Talus slope, (2) Rock glacier et Relict rock g., (4) Weathered bedrock, etc… y compris une auto-appartenance au groupe (11) inconnus /autres (tableau 5).

On voit que tous les échantillons « Ground moraine » (tableau 4), appartiennent tous à la classe jaune (figure 11). À l'exception de gandegg2, tous les échantillons du landform (8) Ice-rich sediments appartiennent à la classe jaune. Pour la classe jaune uniquement, le rapport entre sol gelé riche et pauvre en glace est de 1.47. Le même rapport pour les classes verte et violette réunies est de 3.5.

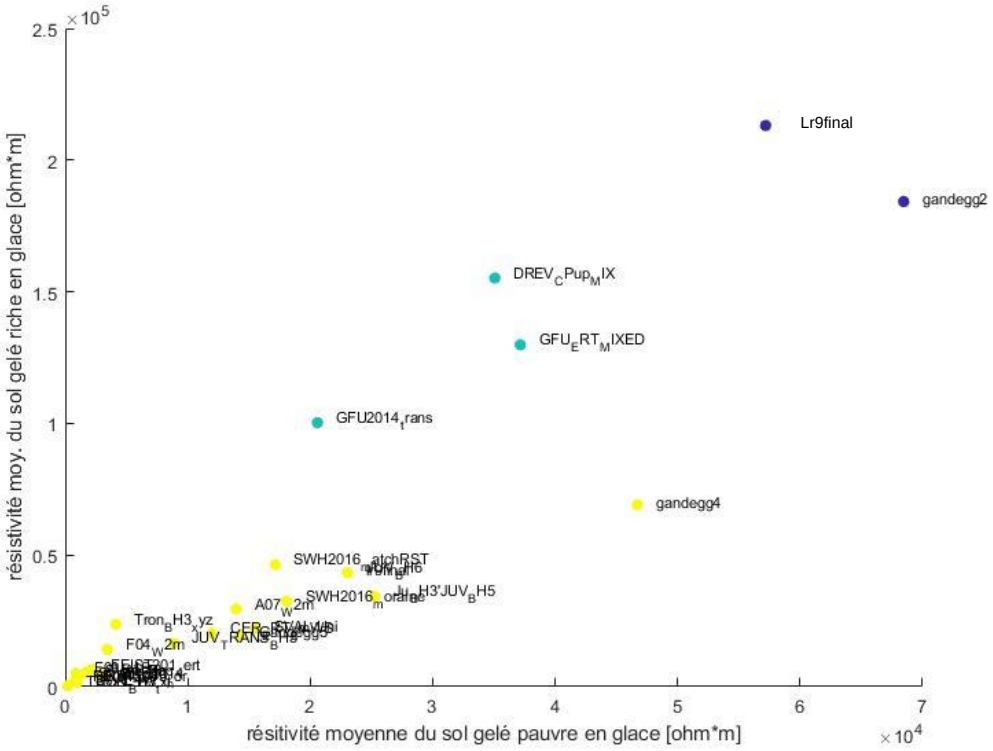


Figure 11 : Landform (5) Bedrock, représenté graphiquement avec les variables de sol gelé riche et pauvre en glace, réparti en tois classes à partir d'un cluster en linkage selon mean_low, mean_palier, mean_icePoor et mean_iceRich.

Le tableau 4 montre les échantillons qui composent les landforms (8) et (11). Pour connaitre le landform véritable des échantillons inconnus du groupe (11) le test statistique suivant a été effectué : Les variables qui composent chaque échantillon du groupe (11), à savoir les variables de pourcentage et de résistivité (tableau 3), ont toutes étés corrélés avec l'ensemble des échantillons. Un échantillon individuel a donc été comparé avec toute la liste des échantillons, par l'intermédiaire de leurs variables. Chaque corrélation supérieure à 0.99 est répertoriée (tableau 5). L'échantillon inconnus du groupe (11) est susceptible d'appartenir au groupe dans lequel le nombre de corrélation répertoriée est maximale.

(8) Ice-rich sediments	**Landform**	**(11) Inconnus**	**Landform**
A07_W2m	Ice-rich sediments	CER_ERT	inconnu
F04_W2m	Sediments	DREV_CPup_MIX	inconnu
F09_W15m	Sediments/solifluxion lobe	FRE2014	Agricultural field
FEIST201_vert	Sediments/rock glacier	GFU_ERT_MIXED	inconnu
FEIST2015_hor	Sediments/protalus rempart	GFU2014_trans	Solifluxion lobe
gandegg1	Ice-cored moraine	Ju_BH3	Ground moraine
gandegg2	Ice- cored moraine	JUV_BH5 / JUV_TRANS	Ground moraine
gandegg4	Sediments	JUV_BH6	Ground moraine
Gandegg5	Sediments	lr9final	inconnu
SWH2016_matchRST	Debris covered glacier	SWAL-end	solifluxion
SVAL-Uni	Sediments	SVAL_IW_LP (not SVAL_icewedge !)	inconnu
SVAl_icewedge	Coin de glace	TRON_BH3_txt	inconnu
		Tron_BH3_xyz	Ground moraine

Tableau 4 : Liste des échantillons des landforms (8) et (11).

Selon le tableau 5 :

- DREV_CPup_MIX, GFU_ERT_MIXED, GFU2014_trans sont corrélés préférentiellement à des échantillons du landform (2) Rock glacier er Relict rock glacier.
- Ju_BH3_0, Ju_BH5/JUV_Trans_BH5 et JUV_BH6 sont des Ground moraines qui n'ont pas de groupe de prédilection dans le tableau 5, mais qui s'assimile à la classe jaune (figure 11). Tron_BH3_xyz est une Ground moraine qui ressemble à des éléments du landform (8). Par déduction, on pourrait affirmer que les Ground moraines devraient être simplement assimilées au landform (8), ce qui peut sembler trivial. Pourtant JUV_BH5/JUV_Trans_BH5 ne semble avoir aucune affinité avec le landform (8), ce qui laisse suspecter pour cet échantillon, une faible quantité de glace dans le sol gelé.
- Lr9final pourrait être un élément du landform (4) Weathered bedrock.

- FRE2014, SWAL_end et SVAL_IW_LP ont des caractéristiques uniques (chapitre 5.6). Ils sont en somme des cas à part.

- TRON_BH3_txt est probablement une version modifiée de Tron_BH3_xyz, mais reste un mystère, du fait du peu de corrélations que l'on observe (tableau 5).

	(1)	(2)	(4)	(5)	(7)	(8)	(11)
DREV_CPup_MIX	4	7	5	-	1	2	3
GFU_ERT_MIXED	5	9	5	2	1	4	3
GFU2014_trans	4	10	3	-	1	2	3
Ju_BH3_0	-	-	2	2	2	2	2
JUV_BH5_ / JUV_TRANS_BH5	2	1	2	2	1	0	2
JUV_BH6	1	2	2	1	0	1	2
lr9final	-	1	2	-	-	1	1
TRON_BH3_txt	-	1	-	-	-	1	1
Tron_BH3_xyz	-	1	1	-	-	3	2

Tableau 5 : Liste de corrélations supérieures à 0.99 entre un échantillon individuel et tous les échantillons d'un landform. En vert, le landform dans lequel l'échantillon individuel est le plus susceptible d'appartenir.

5.6 Variable de pourcentage

Il est possible de comparer tous les échantillons en termes de pourcentage. Une variable de pourcentage produit par le programme Iceclust2019 indique quel part d'un profil, d'une tomographie, est considéré comme gelé (totale, riche ou pauvre en glace) ou saturée en eau ou encore simplement comme non gelée (tableau 3). Un graphique qui représente le pourcentage de sol gelé en fonction du sol non-gelé, indique implicitement quel est le pourcentage du sol est saturé en eau (figure 12), si les deux pourcentages réunis ne totalisent pas 100%. Tous les échantillons qui ne contiennent pas de sol saturé en eau, se retrouvent donc sur une ligne. Les résistivités moyennes typiques du sol saturé en eau oscillent entre généralement entre 200 et 10'000 Ω*m. Elles montent jusqu'à 20'000 Ω*m pour certains Bedrocks et Rockglaciers.

En observant la figure 12, On constate que FRE2014 est à l'écart car il comporte la plus grande quantité de sol saturé en eau 43%. FRE2014 est aussi le seul échantillon classé comme « agricultural field » (tableau 4). SVALIcewedge, SVALRockglacier, CER_ERT et SWAL_IW_LP contiennent des pourcentages de sols gelés très hauts. 88% de sol gelé (dont 78 % riche en glace) pour SVALIcewedge. SVAL_unicalm possède à la fois une grande partie de sol gelé mais aussi une grande partie de sol saturé en eau, 20%, ce qui est rare.

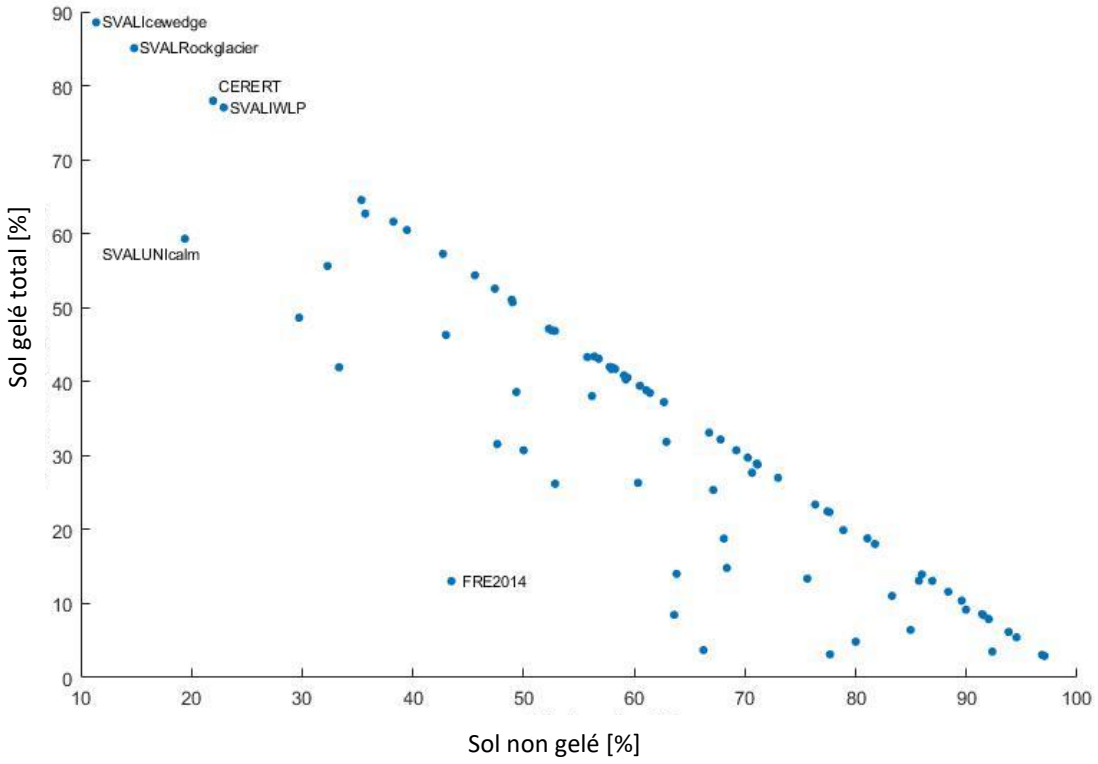


Figure 12 : graphique de pourcentage de sol gelé et non-gelé avec 84 échantillons. 6 échantillons se démarquent par des pourcentages particuliers. Leur nom est écrit.

5.7 Variable de résistivité

Certains échantillons ont des résistivités 2 à 5 fois plus élevées que la majorité (figure 13). Il s'agit de MCO2018, ROTH, ROTV, SWHmoraine, DOL2016 et Marmugnun. Ces six échantillons forment une classe à part ; la classe des grandes résistivités de sol gelé. Une hypothèse serait que tous ces échantillons possèdent une caractéristique commune qui expliquerait leurs hautes résistivités de sol gelé ; par exemple la nature du substrat rocheux. Cela reste à démontrer.

DOL2016 et Marmugnun ont été conservés pour la représentation graphique du landform (2) Rock glacier et Relict rock glacier (figure 8), ainsi que pour les PCAs (figure 14 et 15). Il a été jugé que ces deux échantillons ne masquent pas, ou ne déforment pas, ces représentations graphiques. Les quatre autres échantillons ont été retirés pour ne pas masquer la variabilité des autres échantillons (Caloz et Collet, 2011), c'est-à-dire occuper visuellement et en termes de grandeur de chiffres (les résistivités n'ont pas de limite maximale et sont parfois exprimées sur une échelle logarithmique pour corriger ce problème, alors que les pourcentages sont, dans ce travail, bornés à une échelle de 0 à 100%), l'essentiel du plan euclidien des représentations graphiques.

MCO2018 est un Rock glacier. La couleur jaune de son marqueur indique qu'un cluster en linkage à 6 classes sur 84 échantillons le positionne seul dans sa classe. Aucun autre échantillon ne lui ressemble. ROTH et ROTV sont des Bedrocks. Ils se ressemblent et ont des résistivités deux fois plus élevées que gamanjunni1 et JUVE_PACE (figure 10). SWHmoraine est répertorié comme « dead ice ».

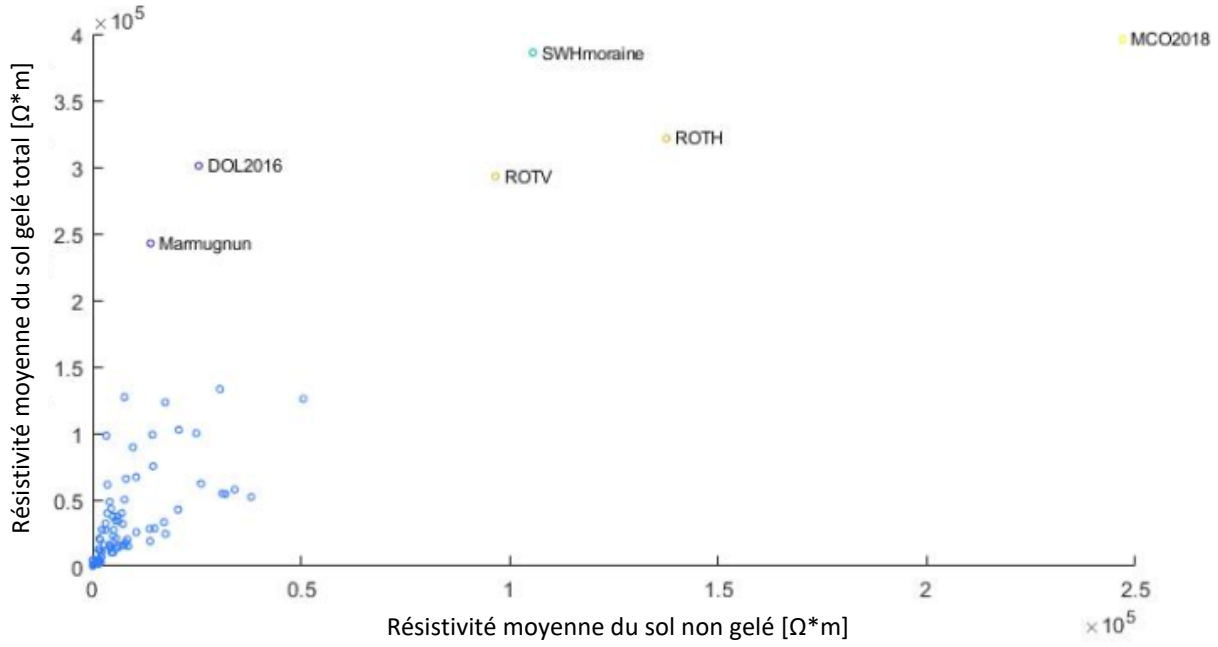


Figure 13 : graphique de résistivité du sol gelé et non-gelé avec 84 échantillons. 6 échantillons se démarquent par leur grande résistivité. Leur nom est écrit et la couleur de leur marqueur est différente.

5.8 Analyse en composantes principales

L'approche par PCA n'est pas concluante. Plusieurs combinaisons de variables ont été essayées. Aucune PCA ne révèle de groupes bien distincts (figure 14 et 15), ni d'orientations liés aux types/landforms (figure 14). Ceci signifie que les échantillons présentent un taux de variation élevée à l'intérieur des groupes qui leur sont attribués. La variable de Prct_low, qui n'est pas incluse dans la PCA (figure 6), présente aussi une forte diversité de type/landform parmi les échantillons concernés. Elle n'est pas incluse dans la PCA car de même que dans les figures 12 et 15, elle oriente les échantillons qui n'ont pas de sol saturé en eau sur une ligne et, pour une PCA basé sur des variables de pourcentage, masque les autres variables.

La grande hétérogénéité des échantillons, dont 95% de la variation est visible sur la figure 6, suggère que les conclusions doivent se basés sur les résistivités et sur l'identification de classes à l'intérieur des groupes de types/landforms.

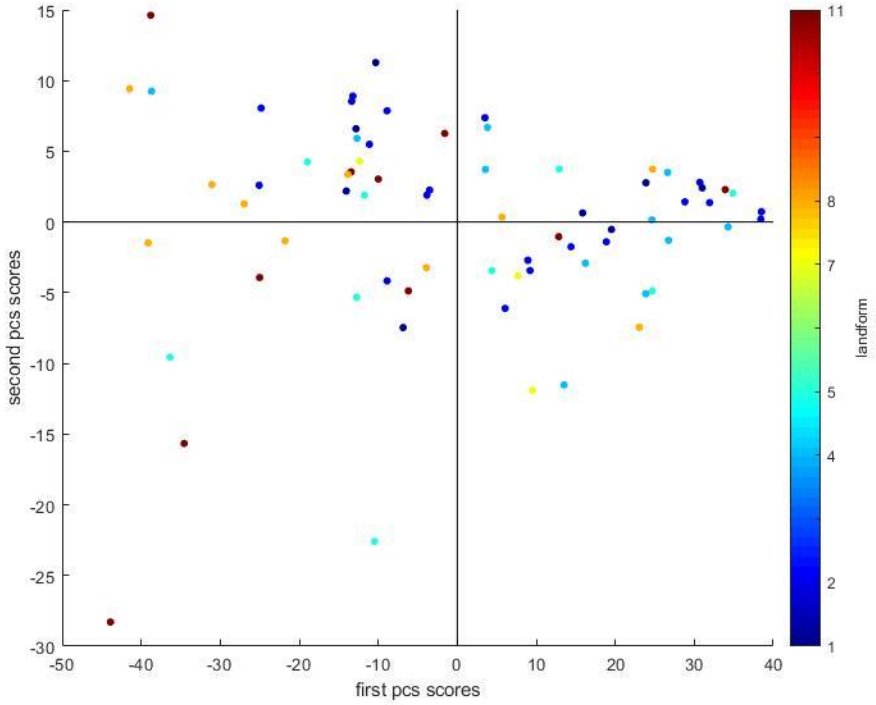


Figure 14 : PCA construite avec les variables Prct_palier, Prct_icePoor et Prct_iceRich. La première composante contient 83.95% de l'information, la deuxième composante contient 12.57% de l'information.

Les échantillons dont le Prct_low et de 0% ont forcément un Mean_low de 0 Ω*m. Cette caractéristique concerne 36 échantillons sur 84. Ces échantillons se retrouvent décalés à gauche et alignés (figure 15). Les échantillons qui ont des résistivités de sol gelé riche et pauvre en glace sont orientés vers la droite. Les échantillons qui ont des résistivités de sol non-gelé élevées sont orientés vers le haut. On remarque que les échantillons les plus à droite sont des Rock glaciers (bleu roi) (figure 15). Les landforms (1) Talus slope, (4) Weathered Bedrock et (11) Autres/inconnus ont tendance à être situé dans le coin inférieur gauche, ce qui signifie pour ces Landforms des résistivités moyennes de sol non-gelé plutôt basses. Les échantillons du landform (5) Bedrock sont au contraire situés plus haut, mais quand même à gauche.

Les landforms restent, dans la vision d'ensemble, confondus. On remarque à peine le landform (8) Ice-rich sediments, en orange. L'hétérogénéité des échantillons ne permet pas de procéder à un classement strict d'ordre de valeur pour les résistivités (par exemple avec une table de résistivité selon le landform).

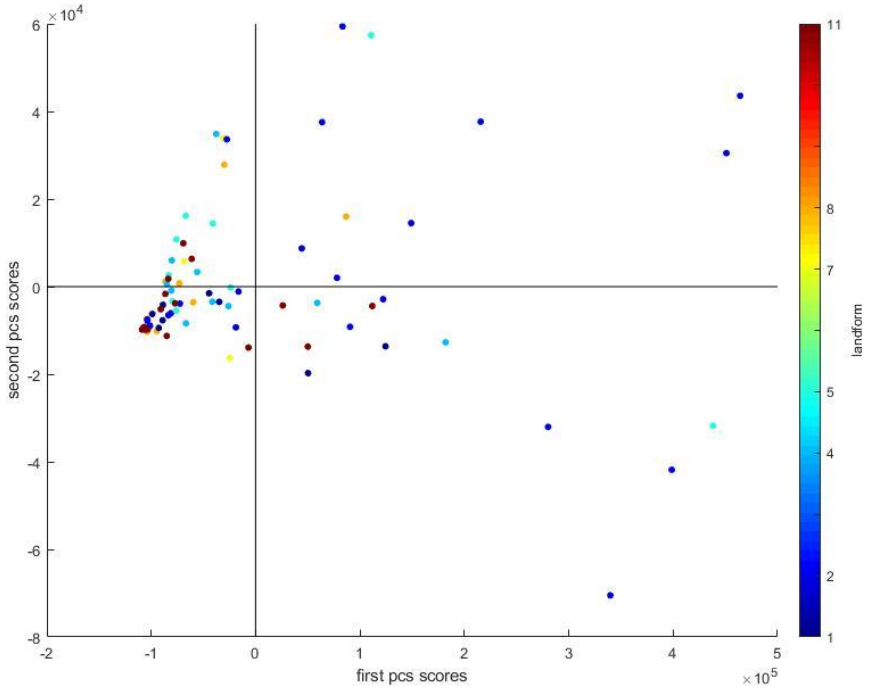


Figure 15 : PCA construite avec les variables Mean_low, Mean_palier, Mean_icePoor et Mean_iceRich. La première composante contient 61.12% de l'information, la deuxième composante contient 24.8% de l'information.

6. Discussion

6.1 Poly1 pour corriger poly3

Pour 13 échantillons, le découpage des parties respectivement gelée, saturé en eau ou non-gelée par polynôme de degré 3 a dû être remplacé par un découpage par un polynôme de degrés 1. Ces échantillons sont : A07_W2m, f10_Barr, LAH20018, RE2014, Dol2015, Gamanjun1, lr2m1-tE, lr9final, MLP02, MLP05, MLP10, MLP25 et RBS2017. Ces échantillons présentaient une mauvaise estimation des parties de sol saturés en eau ou de sol gelés. Il s'agit principalement de surestimation de ces parties. Ce problème a été récursif. La cause du problème a été à chaque fois la même. Si la croissance de la courbe des mesures de résistivités triées dans l'ordre croissant est trop faible (la courbe bleue est plate (figure 16a)), ce qui est amplifié par un grand nombre de mesures, alors (même si le polynôme trouve un point d'inflexion ou des extremum intéressants) la limite des 5% supérieurs ou inférieurs à ligne de palier est atteinte pour un changement de convexité ou de concavité non-signifiant vis-à-vis de l'ensemble des données.

Les échantillons concernés ont été corrigés manuellement, en forçant le programme à bifurquer en degré 1. L'avant et l'après de ces échantillons est dans le dossier ; mauvais poly3-vs-poly1. Un exemple de correction est RBS2017. Si Le découpage des parties sur la courbe bleue des résistivités triées croissantes est corrigé (figure 16b), la distribution des parties sur la tomographie est différente (figure 17a et 17b) et les statistiques changent. RB2017 passe de 37% sol gelé à 8.5% et de 12% de sol saturé en eau à 0%. Naturellement, la résistivité moyenne du sol gelé augmente et la résistivité moyenne du palier diminue.

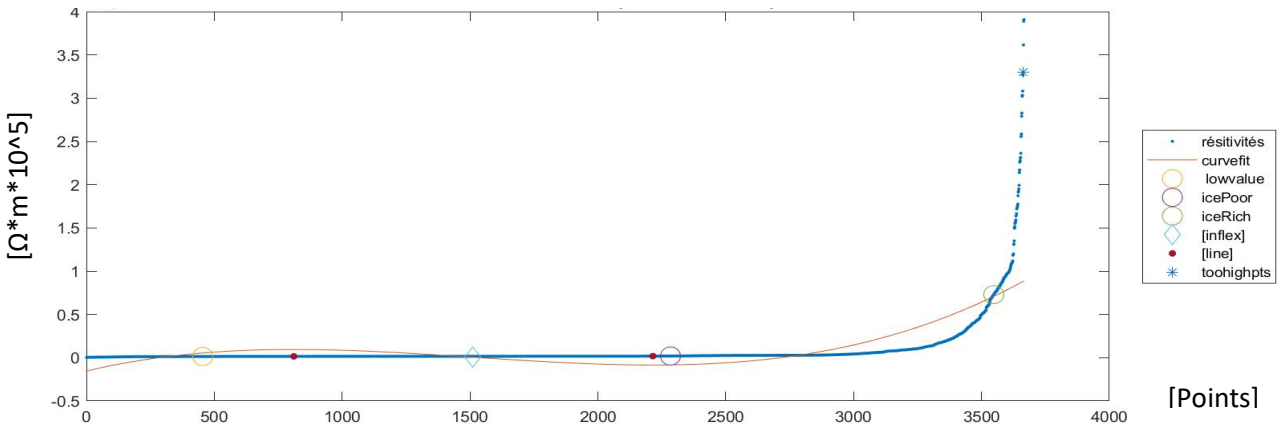


Figure 16a : Courbes des résistivités de l'échantillon RBS2017 et marqueurs selon le découpage poly3 (avant correction)

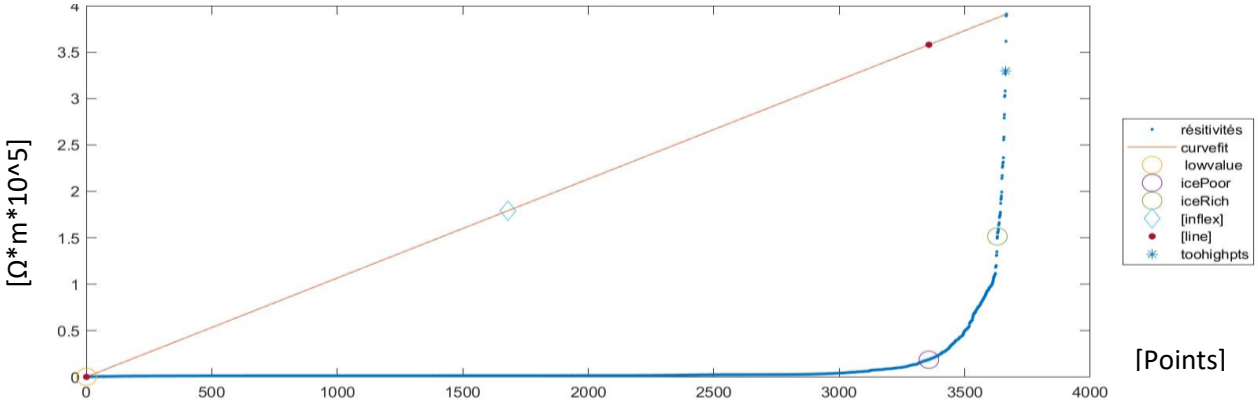


Figure 16b : Courbes des résistivités de l'échantillon RBS2017 et marqueurs selon poly1 (après correction)

La meilleure manière de détecter un mauvais découpage par le programme Iceclust2019, en plus d'observer le découpage sur la courbe des résistivités croissantes (figure 16a et 16b), est de procéder à une comparaison entre le découpage sur la tomographie (figure 17a et 17b, figure 4b) et la variation de de résistivité (Z) en fonction de la profondeur (Y) et de l'horizon (X) (figure 18, figure 4a). Le cercle rouge montre qu'il n'y a pas de hautes résistivités en profondeur (figure 18) et donc que le sol gelé pauvre en glace (bleu clair) n'as pas de raison d'être en dessous de -30m (figure 17a).

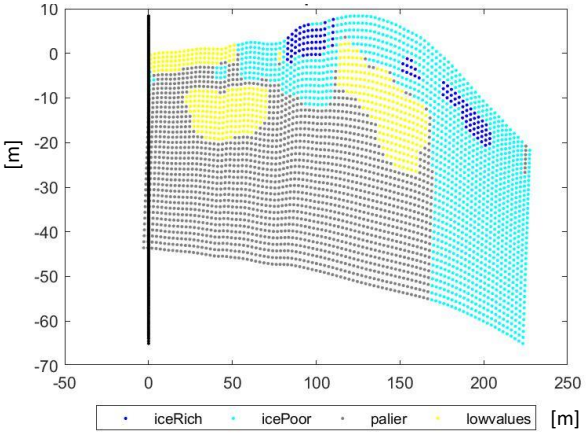


Figure 17a : Découpage de la tomographie de RBS2017 selon polynôme de degrés 3 (poly3).

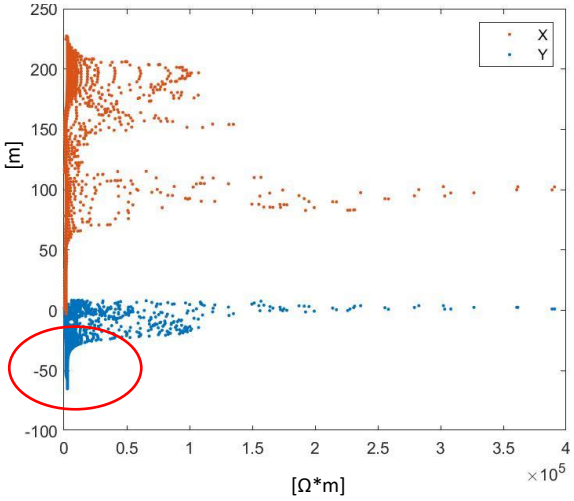


Figure 18 : Variation des résistivités en fonction de l'horizon (X) et la profondeur (Y).

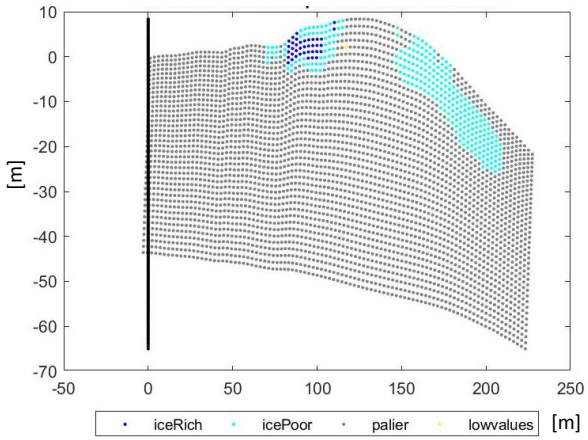


Figure 17a : Découpage de la tomographie de RBS2017 selon polynôme de degrés 1 (poly1).

Le programme Iceclust2019 pourrait donc encore être grandement amélioré. Est-ce que le programme ne devrait fonctionner uniquement avec un cluster de degré 1 ? Le risque majeur est que la partie de sol pauvre en glace soit grandement sous-évaluée selon la forme de la courbe des résistivités triées et la quantité de low value ignorée si la convexité en début de courbe est trop faible.

6.2 Autres corrections

Il est possible qu'un point jaune représentant une partie de sol saturé en eau est présent sur le graphique alors que les statistiques indiquent 0% de Prct_low. Il s'agit simplement d'un sacrifice pour le bon déroulement du programme.

La taille du « sample » (fonctionnement du programme Iceclust2019 : étape 4) n'a jamais été modifié de sa valeur de 25% à gauche et à droite du point d'inflexion.

Le minimum et le maximum ont été modifié manuellement trois fois pour les échantillons ; SVAL_Roc, SVAL_uni et CER_ERT. Ces trois échantillons présentent une structure en palier sur leur courbe de résistivités comme on peut le voir sur CER2018 (annexe 1, (g)) ou sur SVAL_end (figure6). Pour ces échantillons, le programme a fourni des résultats particulièrement mauvais avant l'inversion du minimum et du maximum ; aucune glace détectée. Après inversion, les pourcentages de partie englacée ont semblé élevés (respectivement 85%, 59% et 78%) mais n'ont plus été modifiés.

6.3 Commentaire sur les méthodes de clustering

Le clustering par k-médoïdes s'est révélé robuste quand le découpage des parties a dû être corrigé (chapitre 6.1). Une mauvaise estimation de sol gelé, n'affecte que peu l'estimation de sol gelé riche en glace. Par exemple ; après correction de LAH2018, la quantité de sol gelé à diminué de 61% à 20%, mais la quantité de sol gelé riche en glace seulement de 6% à 5%. Cela démontre qu'il existe une vraie différence entre ces deux parties de sol gelé dû à l'augmentation de la résistivité (équation (1)), et que l'estimation des distances intra-classe est une bonne manière de les évaluer (figure 2).

Le clustering par linkage produit des résultats un peu trop « arrangeants » pour être qualifié de robustes. De plus, la méthode basée sur les distances euclidiennes (chapitre 3.4.3) est relativement sensible à la grandeur des chiffres (à la valeur absolue des chiffres). Par exemple : La variable de sol gelé riche en glace (Mean_iceRich) prend trop de poids si on la compare au sol non-gelé (Mean_palier). C'est bonne chose de pouvoir manipuler rapidement le nombre de classe désiré, mais il est possible que le résultat visuel mérite

une deuxième confirmation. Il faut observer la forme du dendrogramme, qui peut se révéler mal construit (annexe 5).

6.4 Cluster et classes de rock glacier

Une cause de la division des glaciers rocheux en deux classes serait due à la profondeur de la glace sur le graphique de la tomographie (chapitre 5.2). Une différence de résistivité dans du sol gelé riche en glace est influencé par la nature de la glace et la quantité de particules mélangées à la glace. La quantité d'air dans le sol, qui augmente la résistivité (figure 1), peut aussi faire croire à de la glace.

L'échantillon MLP07 montre sur sa tomographie ; du sol riche en glace en profondeur et du sol riche en glace en surface (rond rouge ; figure 19a). La variation des résistivités en fonction de X et Y, montre que le sol gelé riche en glace est nettement plus haut en résistivité en profondeur. Ceci pour une grande quantité de points. En surface, les résistivités sont plus basses (rond rouge ; figure 19b). Néanmoins un point de mesures fait exceptions (rond vert) mais n'est pas plus élevé en résistivité que le maximum des points de mesures en profondeur.

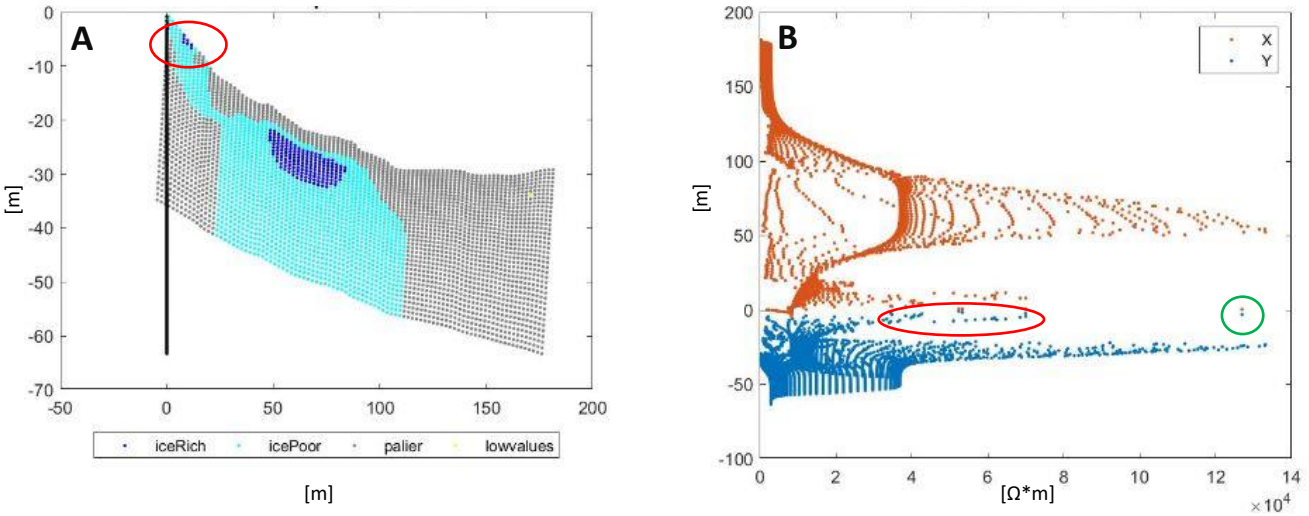


Figure 19a et 19b : comparaison graphique entre le découpage des sols de la tomographie (19a) et la variation de résistivités en fonction de l'horizon (X) et la profondeur (Y) (19b) de MLP07.

L'échantillon muraglj montre l'inverse. Les résistivités les plus hautes sont associées à la surface (rond rouge ; figure 20a et 20b). Leurs mesures semblent disproportionnées par rapport à celle des points qui les entourent. Selon Kneisel (2006), le glacier rocheux de Muragl présente des anormalités de permafrost peu profond, des parties froides aux marges du glacier et de la glace tempérée dans les parties plus centrale. Ces deux échantillons, MLP07 et muraglj ne peuvent fournir de réponse définitive notamment parce qu'ils n'ont pas été associés à la même classe, mais aussi parce que leurs mesures de résistivité n'appartiennent pas au même

ordre. MLP07 est estimé sur une échelle de 10^4 alors que muraglj est estimé sur une échelle de 10^5. Ceci témoigne d'une différence significative de milieu. MLP07 est chilien. Muraglj est suisse.

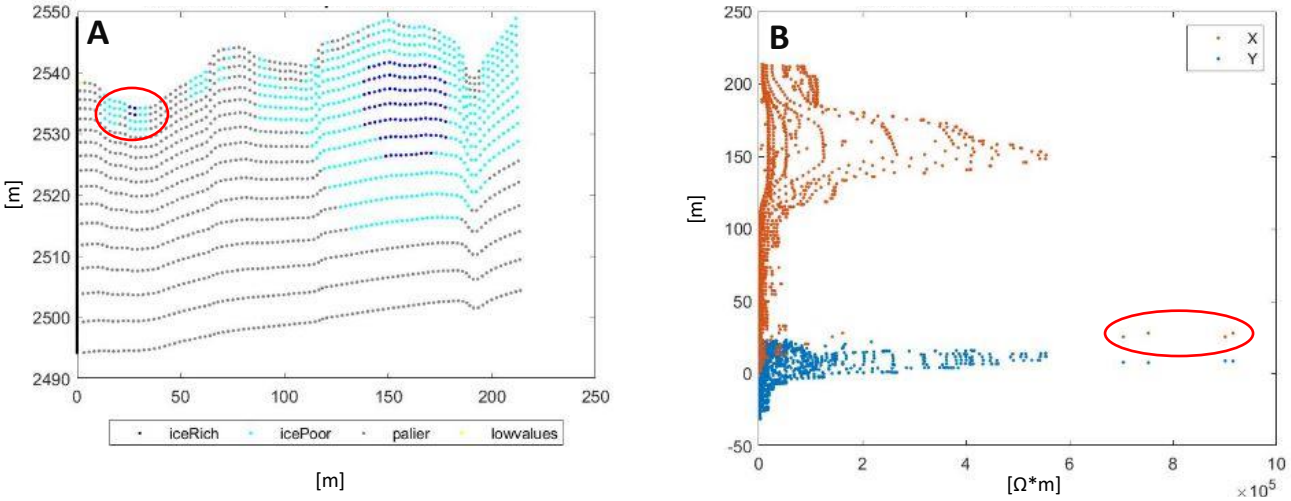


Figure 20a et 20b : Comparaison graphique entre le découpage de sol de la tomographie (20a) et la variation de résistivités en fonction de l'horizon (X) et la profondeur (Y) (20b) de muraglj.

6.5 Commentaire sur les unités produites

Dans ce travail les pourcentages estimés, c'est-à-dire les variables de pourcentage (tableau 3), sont relatives à la quantité de sol gelé ou non gelé, saturé d'eau, riche ou pauvre en glace. Traditionnellement les pourcentages produit dans l'analyse des ERTs est de nature physique et non statistique. Ils sont relatifs à la quantité de glace, d'air et d'eau (Hauck et al., 2011). Peut-on alors comparé ces deux mesures différentes ? Il est plausible que la quantité de glace soit proportionnelle à la quantité de sol gelé (Prct_iceTotal) et la quantité d'eau à la proportion de sol saturé en eau (Prct_low). Mais il est peu probable que la dépendance soit simple. Un sol peut-être grandement gelé et contenir relativement peu de glace. En ce sens, la comparaison entre les deux types d'estimation doit rester prudente en l'absence de certitude.

6.6 Validation

Il est impossible de valider tous les résultats produits par une méthode alternative. Néanmoins à titre d'exemple, 4 échantillons venant de Svalbard en Norvège, vont être comparé avec la méthode physique. Schwery (2016) indique que le volume de pore occupé par la glace pour SVAL_icewedge est d'environ 70%. L'estimation du sol gelé riche en glace pour cet échantillon est de 78%. Le rapport de terrain (Schwery, 2016) Indique que la partie superficielle du sol était humide et que le premier mètre pourrait contenir environ 50%

d'eau. L'estimation du sol saturé en eau est donc fausse pour cet échantillon puisqu'elle est de 0% (figure 21).

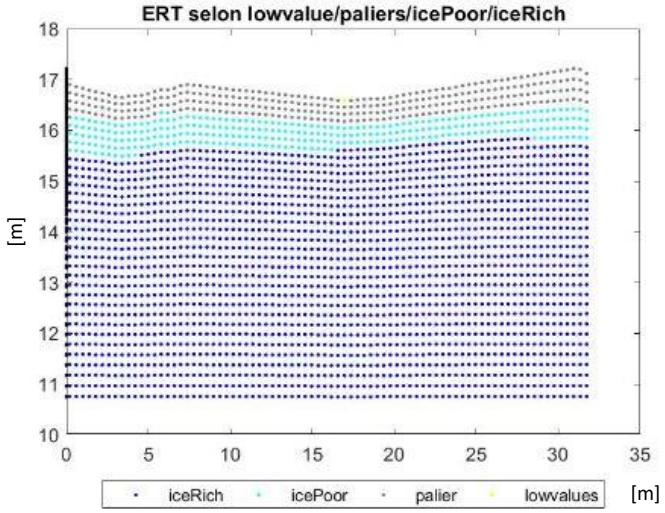


Figure 21 : Graphique ERT de SVAL_icewedge selon le programme Iceclust2019.

Les estimations suivantes comparent les résultats de Schwery (2016) avec celle du programme Iceclust2019 (tableau 6). Les estimations de Schwery sont basées sur l'approche physique du « 4 Phases Model ». Cette approche demande d'estimer la fraction de porosité dans la roche, c'est-à-dire l'air, pour être précise. Les pourcentages expriment la fraction de porosité occupée. La comparaison ne doit pas être considérée comme vrai mais uniquement comme expérimentale (chapitre 6.5).

	Water(%) (schwery, 2016)	ICE (%) (schwery, 2016)	Sol saturé en eau (Iceclsut2019)	Sol gelé total (Iceclsut2019)
SVAL-Uni	20 %	40 %	21 %	59 %
SVAL-end	40-60 %	40 %	25 %	41 %
SVAL-Roc	0 %	80 %	0.1 %	85 %

Tableau 6 : Comparaison de la méthode physique et statistique.

7. Conclusion

Beaucoup de relations statistiques restent mystérieuses. Notamment du point de vue explicatif. Pourquoi la classe jaune de (2) Rock glacier et différente de la classe violette ? Et qu'en est-il des classes séparées de (1) Talus slope et des autres landforms ? il faut savoir que chaque profil n'est pas connu. Ceci est aussi une bonne chose d'un point de vue scientifique. L'analyse gagne en objectivité si la recherche n'est pas orientée dans le but de démontrer des résultats déjà connus. L'objectif que fixait la question de recherche initiale est néanmoins atteint : Des valeurs typiques de résistivité sont collectées et déterminées pour des formes géomorphologiques typiques (landforms) mais aussi pour des caractéristiques physiques comme le sol saturé en eau ou du sol gelé pauvre en glace. Les caractéristiques géographiques, comme l'altitude ou la localité n'ont pas donné de résultats significatifs

7.3 Types/landforms

Le landform (1) Talus slope a la caractéristique d'être peu étendu en termes de résistivité moyenne de sol non-gelé ; jusqu'à 6000 Ω*m au maximum. Les résistivité moyennes de sol gelé montent jusqu'à 80'000 Ω*m. Si on se permet l'analogie, c'est une famille qui vit dans la même rue.

Le landform (2) Rock glacier et (7) Relict rock glacier est très éparse en termes de résistivité, même à l'intérieur des classes prises séparément. La classe verte a les résistivités les plus hautes. Les résistivités moyennes de sol gelé montent jusqu'à 300'000 Ω*m.

Le landform (4) Weathered bedrock est en partie similaire à (1) Talus slope. Les résistivités moyennes de sol gelé montent jusqu'à 90'000 Ω*m. Cependant, les résistivités moyennes de sol non-gelé sont jusqu'à 5fois plus hautes.

Le landform (5) Bedrock a des résistivités moyennes de sol non-gelé qui montent jusqu'à 50'000 Ω*m et des résistivités de sol gelé qui montent jusqu'à 160'000 Ω*m

Le landform (8) ice-rich sediments a effectivement les pourcentages de sol gelé les plus élevés soit environ 49% en moyenne en retirant gandegg2. On peut dire que c'est un groupe de personnes pas forcément de la même famille, mais qui vit dans la même rue.

7.4 Programme Iceclust2019

Plusieurs bonnes estimations pourtant loin d'être évidentes ont été réalisées. L'estimation de la partie su sol saturé en eau de FRE2014 a permis d'identifier dès le début cet échantillon à part. C'est en effet la seule ERT en terrain agricole. De plus, la tomographie a été réalisée en avril, ce qui pourrait laisser entendre que le terrain était effectivement gorgé d'eau à ce moment-là. L'estimation en sol gelé de SVALIcewedge est aussi plutôt rassurante car c'est logiquement le terrain qui a le plus de raison d'être englacé puisque c'est un coin de glace.

L'utilisation de la variable sol saturé en eau est généralement faible dans les représentations graphiques. Elle ne fournit pas d'explication applicable à une classe entière car cette variable marginalise souvent un échantillon par rapport au groupe.

La représentation visuelle sous la forme de grandes images à 5 ou 9 graphes (annexe 1, 2 et 4) qui a accompagné chaque estimation des variables des échantillons a réellement été d'un grand soutient et a guidé la recherche instinctivement, ce que les chiffres ont plus de mal à faire. En l'absence d'une autre source de comparaison pendant la recherche, ce fut un bon, simple, mais parfois peu robuste support de validation. Les graphiques et les courbes ont joué le rôle d'indices directeurs. Rien n'empêche maintenant de comparer ces données avec d'autres sources et d'en vérifier la validité. Rien n'empêche de ne garder qu'une partie du contenu produit et de rajouter des paramètres qui pourraient ouvrir de nouvelles voies.

8. Bibliographie

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V.E., Lewkowicz, A.G., Abramov, A. and Allard, M. (2019). Permafrost is warming at a global scale. *Nature communications*, *10*(1), 264.

Caloz, R., & Collet, C. (2011). *Analyse spatiale de l'information géographique*. PPUR Presses polytechniques.

Hauck, C., & Kneisel, C. (2008). *Applied geophysics in periglacial environments*. Cambridge University Press.

Hauck C., Böttcher M., Maurer H., (2011) A new model for estimating subsurface ice content based on combined electrical and seismic data sets. *The Cryosphere* **5**: 453-468.

Hauck, C. (2013). New concepts in geophysical surveying and data interpretation for permafrost terrain. *Permafrost and Periglacial Processes*, *24*(2), 131-137.

Hilbich, C., Hauck, C., Delaloye, R., & Hoelzle, M. (2008). A geoelectric monitoring network and resistivity-temperature relationships of different mountain permafrost sites in the Swiss Alps. In *Proceedings Ninth International Conference on Permafrost, June* (pp. 699-704).

Kneisel, C. (2006). Assessment of subsurface lithology in mountain environments using 2D resistivity imaging. *Geomorphology*, *80*(1-2), 32-44.

Mathworks (2019a) polyfit. Polynomial curve fitting: Limitations et Algorithms: https://www.mathworks.com/help/matlab/ref/polyfit.html [consulté le 9 avril 2019]

Mathworks (2019b). kmedoids. K-medoids clustering: More about: https://www.mathworks.com/help/stats/kmedoids.html [consulté le 9 avril 2019]

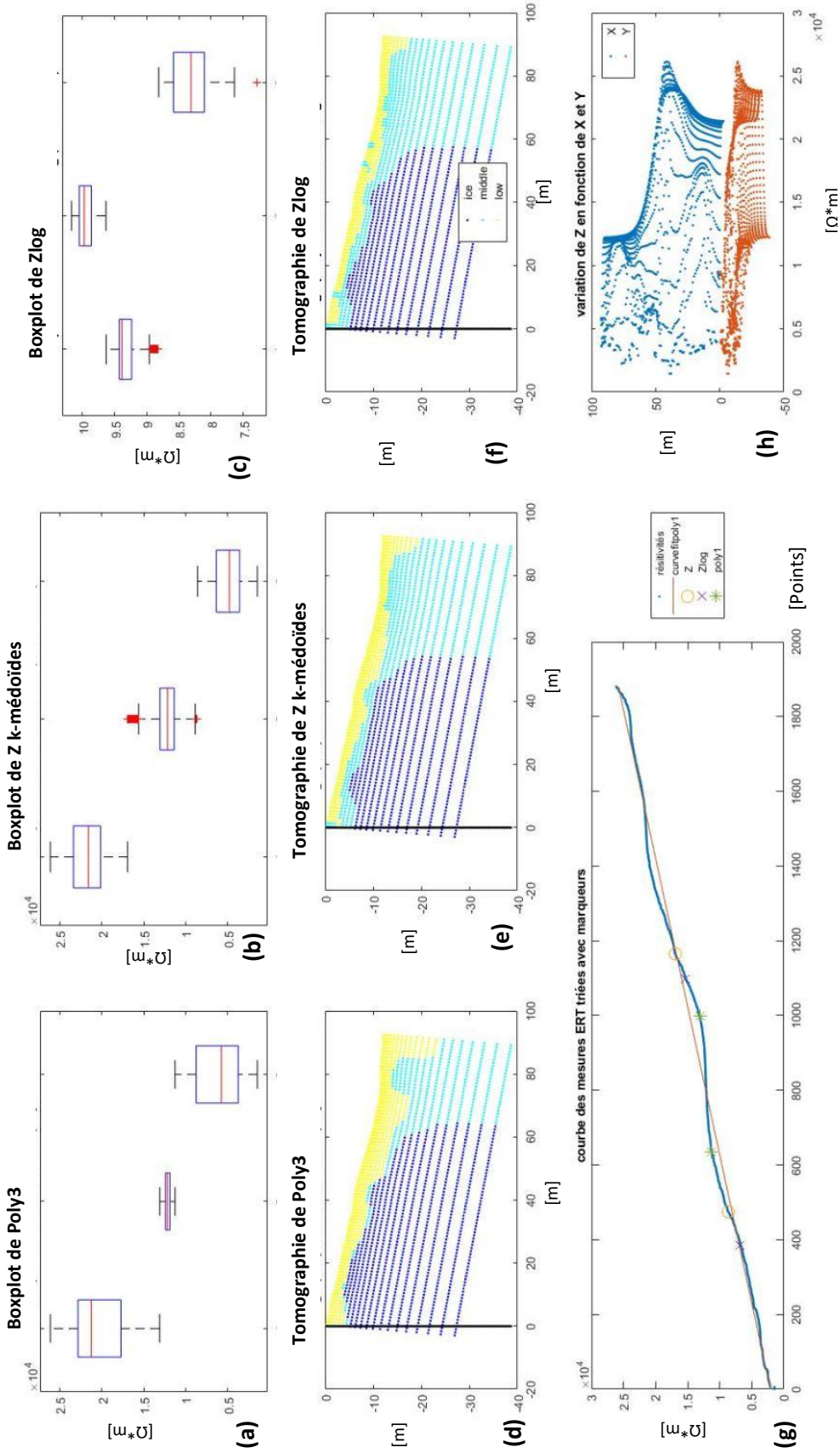
Sjöberg, Y., Marklund, P., Pettersson, R., & Lyon, S. W. (2015). Geophysical mapping of palsa peatland permafrost. *The Cryosphere*, *9*(2), 465-478.

Supper, R., Ottowitz, D., Jochum, B., Römer, A., Pfeiler, S., Kauer, S., Keuschnig, M. & Ita, A. (2014). Geoelectrical monitoring of frozen ground and permafrost in alpine areas: field studies and considerations towards an improved measuring technology. *Near Surface Geophysics*, *12*(1), 93-115.

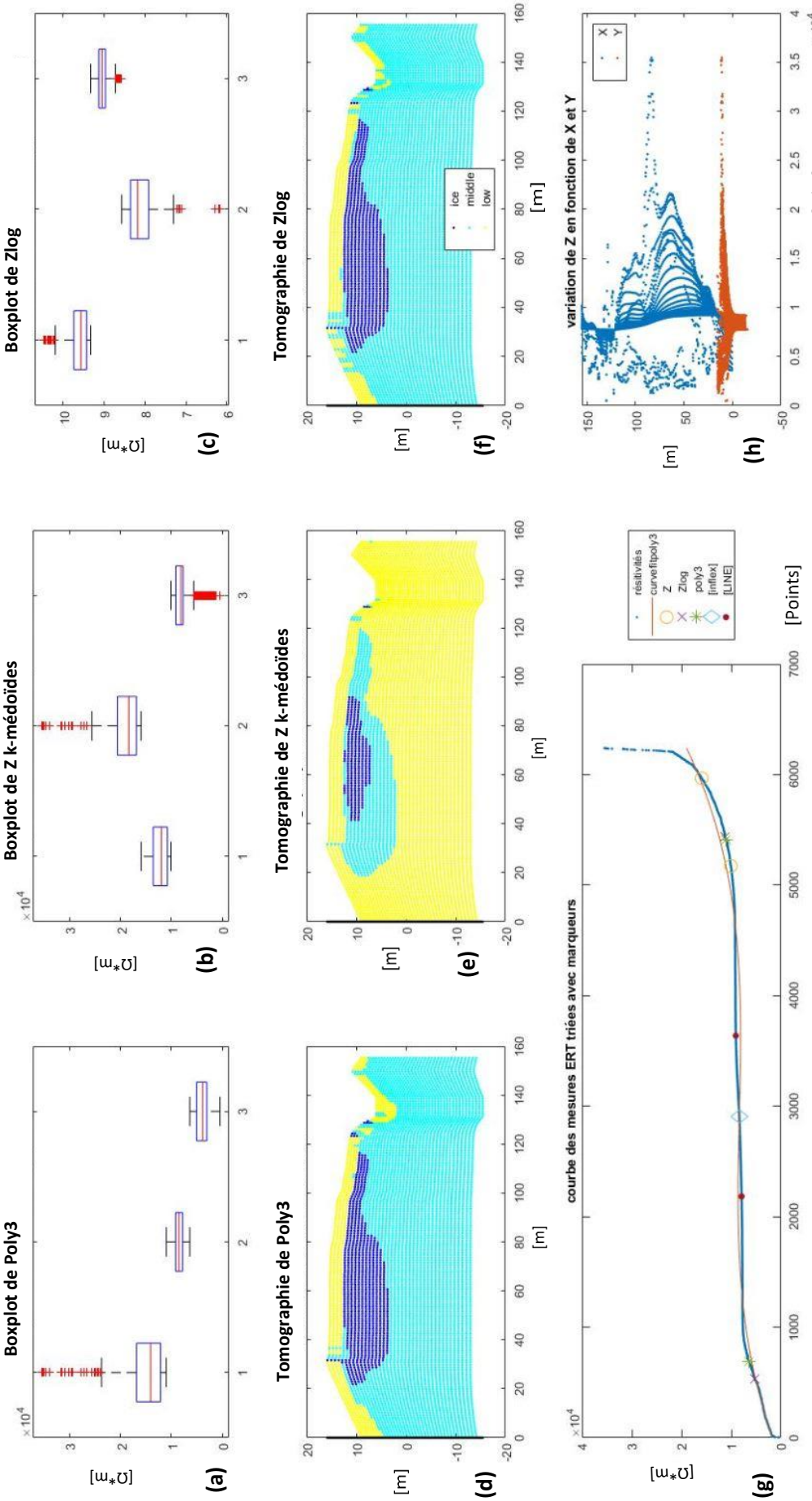
Schwery D. (2016) Permafrost in Svalbard: Quantification of ice and water contents by geophysical measurements and validation by ground truth observations from drill cores. Travail de Master, Fribourg, Faculté des Géosciences, Université de Fribourg, Suisse.

Trauth, M. H., Gebbers, R., Marwan, N., & Sillmann, E. (2006). *MATLAB recipes for earth sciences 2nd Edition*. Springer Berlin Heidelberg New York.

9. Annexes



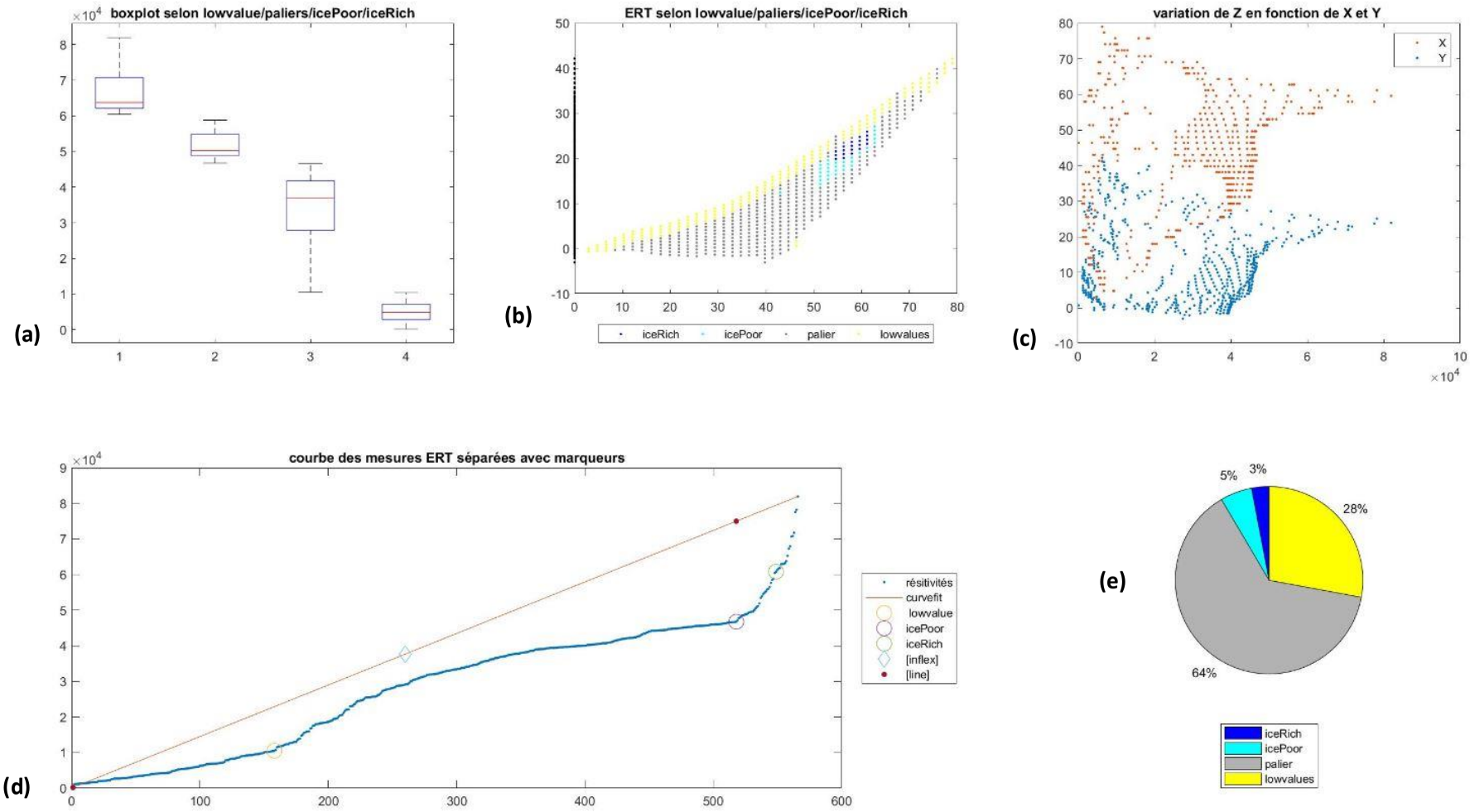
Annexe 1 : évaluation des découpages sur CER2018. On remarque que les découpages (d), (e) et (f) fonctionnent bien et donnent des résultats relativement similaires. (g) montre que le sol saturé en eau (en jaune) semble surestimé. (g) montre que le découpage poly3 a bifurqué en poly 1 pour cet échantillon. Cela est typique des échantillons qui ont une courbe de mesures en double ou triple palier. Ici le premier palier et au point 800 et le deuxième palier est environ au point 1500.



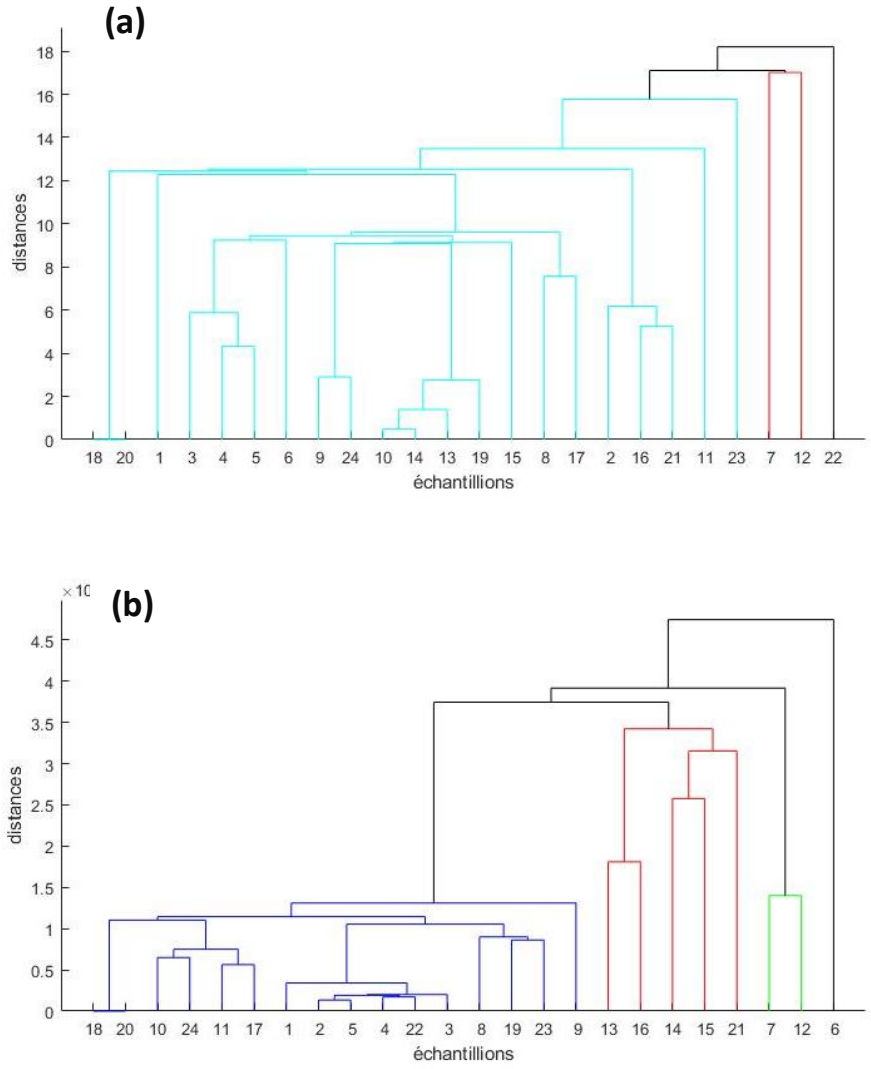
Annexe 2 : Recherche de découpage sur GUOL3. (b) montre que le cluster Z (k-médoïdes) surévalue la quantité de low. (a) montre qu'une quantité non négligeable de points (marqués par des croix rouges) ont des valeurs de résistivité supérieurs à l'estimation de la moyenne et de la déviation standard de la partie de sol gelé. Une estimation en 4 classes serait plus adaptée. (h) montre quel(s) endroit(s) des tomographies sont susceptibles d'être du sol gelé et aide à faire une comparaison visuelle.

```
1    % extraction de la matrice ERT format xyz opu txt%
2 -  clearvars
3 -  M=load('C:\Users\kittelj\Desktop\DATA2(A)\gamanjunni-2-ERT-data-modified-coordinates-m3-trim.xyz');
4 -  filename='gamanjun2';
5 -  Zor=M(:,3);
6 -  Z=sort(M(:,3)); zz=size(Z); z=zz(1); clear zz;
7 -  X=M(:,1);
8 -  Y=M(:,2);
9 -  sample=round(z/4);% toujours 25% à la base, 250pts pour un échantillon de 1000pts
10
11   %analyse par un polynome de degrès 3 ou 1%
12 -  x=(1:z)';
13 -  [p,S] = polyfit(x,Z,3);
14 -  P= (p(:,1)*x.^3) + (p(:,2)*x.^2) + (p(:,3)*x) + p(:,4);
15 -  TF = islocalmin(P); TF2=islocalmax(P); f1= diff(P); f2= diff(f1);
16 -  g= (find(f2<0&f2+1>0, 1, 'last' ));
17 -  if isempty(g)==1
18 -      g=0;
19 -  end
20   %recherche de point à valeur extrêmes hautes
21 -  too_high=0;
22 -  for k= z-30:z
23 -      K=[Z(k-1),Z(k-2),Z(k-3)];
24 -      if  Z(k)/mean(K)>1.15          % ici valeur à 15%, possible de modifié
25 -          Z(k)=Z(k-1)*1.01; too_high=too_high+1;
26
27 -      end
28 -  end
29
30 -  if g+sample(1)>z
31 -      y=[min(Z);max(Z)]; x=[find(y(1)==Z);find(y(2)==Z)];
32 -      [p,S] = polyfit(x,y,1); xk=(1:x(2));
33 -      P= (p(:,1)*xk+ p(:,2))';
34 -      ecart = P-Z;
35 -      g1=max(ecart);%!!!!!parfois écrire min à la place de max
36 -      i=find(ecart==g1);
37 -          ecart2=(1:i)';
38 -          for k=1:i
39 -               ecart2(k)=ecart(k);
40 -          end
```

Annexe 3 : Première lignes de code du programme Iceclust2019.m qui compte environ 200 lignes au total. On remarque le « sample » à la ligne 9 et la recherche de points à valeur haute à partir de la ligne 20. À la ligne 30 le programme envisage peut-être de bifurqué en mode interpolation de degré 1.



Annexe 4 : Résultat visuel du programme Iceclust2019 pour Dra_LP1. (d) montre que le programme a passé en mode poly1 car la ligne orange intitulée « curvefit » est linéaire. Toujours sur (d), on remarque que le programme a effectué une deuxième recherche de maximum car le marqueur lowvalue et aussi situé au deuxième plus grand écart entre la courbe des résistivités triées et curvefit. Le marqueur intitulé [inflex] n'est pas un point d'inflexion quand l'interpolation de la courbe est de degré 1. C'est la moitié de chemin entre le point 1 et le marqueur icePoor qui marque le début de la partie du sol gelé pauvre en glace. C'est sur cet intervalle que la recherche d'un deuxième maximum est effectuée par le programme.



Annexe 5 : l'image (a) est la représentation en branches (ou dendrogramme) d'un cluster en linkage de 4 classes. Le linkage n'est pas concluant car les branches de l'arbre se séparent en des distances toujours trop faibles pour y trouver des groupes. L'image (b) est un dendrogramme réussi, mais au cluster imparfait. Le nombre de classes est de 4. Pourtant le bon nombre de classe devrait être de 3. L'échantillon 6 est une « fausse classe ». Soit il s'agit d'un échantillon marginal, qui peut être écarté des données, soit il peut être assimilé à la classe verte si similitude il y a.