

# Internship TNO

---

<b>Project title:</b>	<b><i>Spatial constraint 3D clustering methods for electrical resistivity models of the subsurface as derived from airborne Electro-Magnetics</i></b>
<b>Supervisor TNO:</b>	Jan Gunnink
<b>E-mail:</b>	<a href="mailto:jan.gunnink@tno.nl">jan.gunnink@tno.nl</a>
<b>Telephone:</b>	0611700519
<b>Location:</b>	TNO – Geological Survey of the Netherlands, Princetonlaan 6, Utrecht
<b>Duration:</b>	6 months
<b>Apply before:</b>	24 February 2017
<b>Start date</b>	1 March 2017
<b>Type:</b>	Internship
<b>Compenstion:</b>	Euro 460 / month

## Introduction

Clustering is a technique to derive “similar groups” in often large datasets. Clustering techniques have been used in a wide range of applications, e.g. machine learning, pattern recognition, bioinformatics, data compression, computer graphics.

Recently, new clustering techniques have developed that take into account the spatial configuration of the data. This means that, besides a similarity based on the attributes of the data, the spatial position in relation to other data points is taken into consideration. This leads to more spatial coherent groups of clustered data. Especially in the medical profession, functional Magnetic Resonance Imaging (fMRI) are subject of elaborate, spatially constrained clustering techniques to

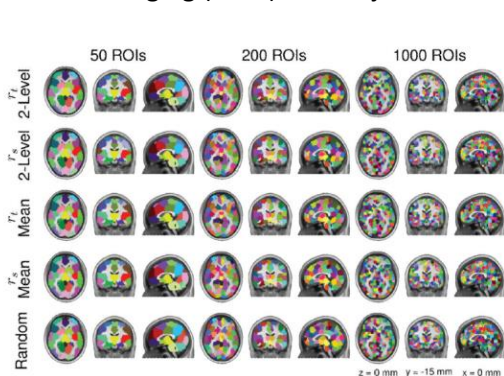


Figure 1: Region of Interest from fMRI, results from clustering

determine so-called “Regions of Interest”, see Figure 1. Since these regions are often clustered in 3D space, clustering techniques need to take the spatial layout into account. Normalized Cut Spectral clustering makes use of so-called “similarity graphs”, which are presented in a similarity matrix. This graph is then cut into a number of clusters, by minimizing a cost function. For more details see Craddock et al., 2012 and Luxburg, 2007.

## Spectral clustering for Earth Science data

Environmental data is inherently 3D because the earth is 3D. Geophysical instruments measure the response of the subsurface to an impulse (natural or man-made). The response to this impulse is dependent on the characteristics of the subsurface. For example, the response of the subsurface to earthquakes is depending on the density of the rocks and the ability to transfer the earthquake



Figure 2: Helicopter with measurement frame

waves through the rocks. Another example is the response of the subsurface on a man-imposed electromagnetic field, which is induced by equipment under a helicopter (Fig. 2). This results (after processing) in a semi-3D image of the electrical resistivity of the subsurface, sometimes up to 100's of meters. This 3D model often shows regions of similar resistivity, that can be detected by the human eye, see for an example Fig. 3. From a geological point of view, these clusters sometimes are meaningful and sometimes they are not. It would be useful to be able to influence the clustering algorithm with

geological *a-priori* information. This needs to be investigated more closely, but a Bayesian

framework, using conditional probabilities, might be an option to explore. Also the spatial continuity that is expected (from a geological point of view) needs to be considered. Spectral Clustering is not widely used in environmental research, while it offers (potentially) great advantages in terms of spatial continuous clusters that are coherent with geological knowledge.

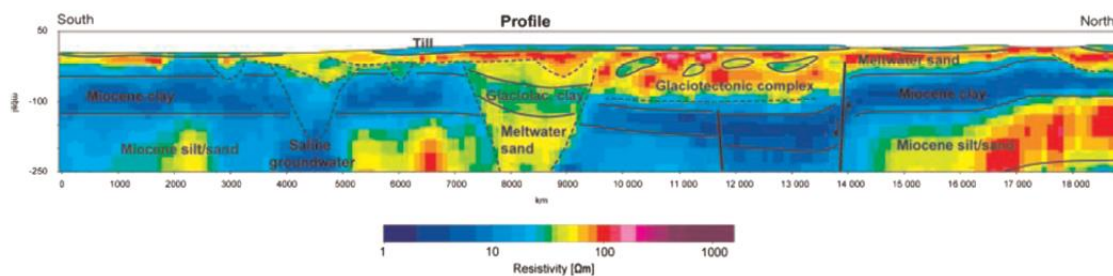


Figure 3: Manual interpretation of Airborne EM derived resistivity data

## Geological Survey of the Netherlands - TNO

The Geological Survey of the Netherlands (GSN) is part of TNO and is responsible for supplying professional users and the public with adequate and up-to-date information of the subsurface. This includes data, interpretations and models. International research activities include cooperation in EU-funded projects, with partners that have advanced geophysical techniques at their disposal, like

airborne EM. Results from this the airborne EM technique are very promising, since they provide dense data for large area's at reasonable costs. The amount of data is so large that manual interpretation is not feasible and, from a point of repeatability, not recommended. Techniques that were used to interpret the data up to now include statistical regression and Artificial Neural Networks. Clustering has not been used on this type of data, and the research on the ability to cluster the 3D data with spatial constraints is therefore a new, promising research area, and highly relevant, given the fact that airborne EM data is being increasingly collected in the Netherlands and abroad.

### *Traineeship*

The plan is to start with existing software code (e.g. used in fMRI analyses, but also other applications) to analyze its usability for AEM data. After selecting a suitable algorithm and computer code, additional coding and implementation of amended algorithms is next. We have a strong preference for the scripting language Python, and some of the existing code is already programmed in Python. Coding in other scripting languages (e.g. MatLab) can be done but needs to be wrapped into Python. Results will be documented in a TNO-report and / or in a scientific paper.

### *Competences of the applicant*

- Full-time available during the internship
- Studying Artificial Intelligence / Statistics / Computer Sciences or similar subject
- Interest in Machine Learning and open-minded towards the subject of the traineeship
- Experience with computer programming (Python or similar)
- Experience in Earth Sciences subjects (not a prerequisite) is a positive
- Enthusiastic and a team-player
- To be able to communicate with non-experts on Machine Learning (e.g. Geologists)
- Has a good command of the English language

For more information contact me, see details on top.

Do you want to apply? Send your letter of motivation and CV to: [jan.gunnink@tno.nl](mailto:jan.gunnink@tno.nl)

### *References*

Craddock, C. R, G. Andrew James, Paul E. Holtzheimer III, Xiaoping P. Hu, and Helen S. Mayberg, 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 2012, August; 33(8); doi: 10.1002/hbm.21333

von Luxburg, U, 2007. A tutorial on Spectral Clustering. *Stat Comput* (2007) 17: 395. doi:10.1007/s11222-007-9033-z