Clustering analysis of soil gas chemical ratios as a potential geochemical tool for surface geothermal exploration

Ángel M. González-Moro$^{1,2}$, Luca D’Auria$^{1,3}$, and Nemesio M. Pérez$^{1,3,4}$

$^{1}$Instituto Volcanológico de Canarias (INVolCAN), 38240 La Laguna, Tenerife, Canary Islands, Spain (anmorgon@gobiernocanarias.org)

$^{2}$Universidad Politécnica de Madrid (UPM), Madrid, Spain

$^{3}$Instituto Tecnológico y de Energías Renovables (ITER), 38611 Granadilla de Abona, Tenerife, Canary Islands, Spain

$^{4}$Agencia Insular de la Energía de Tenerife (AIET), 38611 Granadilla de Abona, Tenerife, Canary Islands, Spain

Geochemistry is a fundamental tool in surface geothermal exploration. In particular, the analysis of the composition of the soil atmosphere, the measurement of diffuse CO$_2$ flux and of the gas $^{222}$Rn activity are important parameters to detect and characterize the contribution of volcanic/hydrothermal sources in the diffuse soil degassing.

The analysis of the soil atmosphere usually consists of determining the chemical and isotopic composition of the gases, including concentrations and molar ratios of multiple chemical species (e.g. He, H$_2$, N$_2$, Ar, Ne, O$_2$, CH$_4$, CO$_2$), as well as the C isotopic ratios ($^{13}$C/$^{12}$C). In practice a single geochemical survey provides tens of different parameters for each sampling point. Taking into account that a typical survey is composed of hundreds of sampling points, the huge amount of collected data requires effective data mining tools to perform analyses going beyond the simple mapping of concentrations and/or ratios and to detect hidden patterns in the dataset.

Among the most effective multivariate statistical tools is clustering analysis. This technique allows determining the presence of groups of points showing a given degree of similarity. In this work we used and compared two different clustering techniques: the K-means and the DBSCAN algorithms, applying them to a geochemical dataset related to surveys realized in 2010 in the southern part of the island of Tenerife (Canary Islands Spain) with the aim of geothermal exploration. We show how the clustering analysis allows determining the presence of areas characterized by a similar chemical and isotopic composition. The use of standard geochemical tools allows interpreting the nature of these areal groups in terms of their relevance for the purposes of surface geothermal exploration.