
Analyses statistiques non-supervisées de spectres de masse haute résolution de chondrites carbonées.

Noé Le Becq^{*1}, Cédric Wolters¹, Laurène Flandinet¹, Lydie Bonal¹, Sylvain Doute¹,
Véronique Vuitton¹, and François-Régis Orthous-Daunay¹

¹IPAG – Université Grenoble Alpes, CNRS : UMR5274 – France

Résumé

L'analyse de la matière organique soluble par spectrométrie de masse très haute résolution Orbitrap permet de contraindre la composition moléculaire des chondrites carbonées et les processus auxquelles elles ont été soumises depuis leur formation. Bien que la très haute résolution soit un avantage certain pour mettre en évidence la diversité moléculaire des échantillons, l'interprétation des spectres est très complexe, rendant difficile le traitement et la comparaison systématiques d'un grand nombre d'échantillons. C'est pourquoi nous proposons de développer des méthodes d'analyses statistiques non supervisées dans le but d'obtenir un outil capable de traiter de façon systématique et objective de grands jeux de données. L'analyse en composantes principales (ACP) est l'un des outils les plus utilisés pour mettre en évidence des schémas de variation communs entre échantillons. Bien qu'ayant déjà démontré son efficacité sur des spectres de masse haute résolution de suies, l'ACP n'a jamais été utilisée sur des données très haute résolution de chondrites carbonées. Pour ce développement, nous utilisons un jeu de données constitué de 21 spectres de chondrites carbonées afin de voir s'il est possible de distinguer les similarités et les différences entre échantillons. Après avoir aligné les spectres selon les valeurs de m/z , nous avons effectué l'ACP sur la matrice nouvellement créée. Les échantillons et les observations sont ensuite tracés en fonction des composantes principales (PC). Leur position dans le nouvel espace formé des PC est utilisée pour faire du clustering. Les groupes devraient alors représenter des échantillons présentant des points communs dans leurs spectres. Cependant, la continuité apparente des valeurs propres associées aux PC complexifie le choix du nombre de composantes à prendre en considération. Plusieurs méthodes ont donc été testées pour estimer le nombre de dimensions mais les résultats furent très variés. A ce point, la structure, la résolution, et la dimensionnalité du jeu de données ont été explorées mais il reste encore à montrer que les résultats de l'ACP permettent de réellement remonter aux caractéristiques physico-chimiques des échantillons.

*Intervenant