Project Leader: Hugh Smithies

Project title: Identifying geologically plausible populations in whole-rock geochemical datasets

Project objective: Given a comprehensive geochemical dataset of igneous rocks of broadly similar type, from a geographically well-constrained region, and guidelines outlining the 'characteristics' expected of geologically plausible geochemical populations, apply machine learning and other analytical and visualisation techniques to group geochemical analyses (i.e. rocks) into potentially related populations.

Data Description: A dataset of ~600 whole-rock geochemical analyses of basaltic rocks from the regions between and around Kalgoorlie and Kambalda.

Project background: When basaltic magma flows solidify they form flat, continuous layers. Successive flows form stacked layers called stratigraphy. Particularly in old regions (our samples come from the Archean Yilgarn craton and formed 2700 million years ago) numerous deformational processes fold and slice stratigraphy into highly complicated geometric arrangements. An additional complication in places like the Yilgarn is that much of the rock is also covered by sand and so we have a very fragmented view of even the complexly deformed current geology.

Predictive mineral exploration depends on being able to reconstruct original stratigraphy. In rare cases where there is a clear visual distinction between the rock types forming that original stratigraphy, reconstructing original stratigraphy might be less difficult. In other cases, like the one presented here, all of the rocks are basaltic (dark, fine-grained and basically featureless) and so it is extremely difficult to visually relate samples to each other.

However, even in cases where basaltic units cannot be separated visually, different basaltic units should be identifiable based on persistent, albeit often only subtle, differences in their geochemistry.

The challenge here is, given a large number of geochemical analyses of basalts scattered throughout a reasonably small, but geologically complex area, we need to identify geologically plausible geochemical populations. A geologically plausible geochemical population is one that makes sense in terms of age (all components need to be the same age – we can assume this is satisfied in this exercise), makes sense in terms of geochemical relationships (form well-constrained clusters OR trends expected of evolving magmas) and makes sense spatially. Numerous geochemical groups ARE expected in the presented dataset, but the number is unknown. Some of these groups may originally have covered the entire area, others might have been more local – but geographical outliers need very careful consideration.

The current way of dealing with such problems would be to painstakingly plot all the data on a range of geochemical variation diagrams (these typically involve an x-axis that is an chemical component known to vary systematically in concentration as a magma evolves or crystallises) and simply look for geochemically valid patterns and relate these back to a map to assess the spatial viability of that population. It is an iterative process that could take weeks for a dataset like the one presented here. **If successful**, a more integrated IT approach would represent a significant time saving step in making geological reconstructions and in generally interpreting large geochemical datasets.