Mineral Potential Modeling for Bauxite in the Northeast of São Paulo State, Brazil: Knowledge and Data-driven Approaches

This study presents the development and comparison of two predictive maps assessing bauxite favorability at a 1:200,000 scale, derived from distinct modeling approaches: one driven by expert knowledge and the other by data (knowledge and data-driven techniques, respectively). The study addresses the development of predictive maps using geophysical airborne survey data, geological maps, and X-ray fluorescence analyses to enhance understanding of regional mineral favorability.

Two mineral favorability maps for bauxite were generated, and different spatial modeling techniques were applied and compared, aiming to reduce the search area for new occurrences. The theoretical approach adopted was based on the concept of mineral systems. In the early stages of mineral exploration, knowledge-driven models should be executed before data-driven models to avoid unwanted biases since the judgment is commonly influenced by the geological characteristics of known deposits and the location of targets generated in previous data-driven modeling. Thus, in the first stage of this work, knowledge-driven modeling was performed, followed by data-driven modeling in the second stage. The main steps involved in producing the mineral favorability maps were: (i) definition of the dataset and spatial proxies based on the analysis of mineral system components in defining the spatial scale; (ii) selection of the most appropriate modeling techniques for the available data; (iii) processing and normalization of input data; (iv) generation of mineral favorability models; (v) model validation; (vi) selection of the most suitable and best-performing models; (vii) target generation.

The primary difference between the two models performed in this study, one based on *fuzzy logic* and the other on the *random forest* algorithm, lies in the permissiveness of classifying regions favorable for bauxite occurrence. The knowledge-driven model is more permissive, classifying more areas as favorable, whereas the data-driven model is more restrictive, identifying fewer areas as favorable. Although both models yield similar results, differences in permissiveness have implications for interpretation and application. The more permissive model may be more inclusive, identifying potentially favorable areas for exploration with greater confidence. However, it may also include areas less conducive to bauxite occurrence, increasing the risk of unsuccessful exploration. On the other hand, the *random forest* model, being more restrictive, can minimize false positives but may overlook areas of interest. Mineralization occurs preferentially in the rocks of the São José do Rio Pardo Suite, where a significant depletion of radioelements, particularly potassium, is evident. Notably, areas with low values of radioelements in higher altitudes exhibit substantial correspondence with areas considered favorable for bauxite.