

---

# Exact Conditioning of Random Forest for Spatial Prediction

Francky Fouedjio\*<sup>1</sup>

<sup>1</sup>AngloGoldAshanti – Australie

## Résumé

Machine learning methods such as random forest are increasingly used for spatially predict a target variable in the presence of auxiliary information available everywhere within the study region. Even though random forest shows competitive prediction performance, it does not exactly honor the target variable's measured values at sampling locations, like other similar machine learning techniques. However, competitor techniques such as regression-kriging perfectly match the target variable's observed values at sampling locations by construction. Exactly matching the target categorical variable's measured values at sampling locations is often desirable in many applications in geosciences. This talk will present a random forest-based approach in which the target variable is exactly conditioned to the data. The proposed method's effectiveness is illustrated on the one hand using a synthetic dataset where the ground-truth is available everywhere within the study region, and on the other hand, using a real-world dataset comprising geochemical concentration data. It is compared with the regression-kriging and the classical random forest. It appears that the proposed approach can perfectly match the target variable's measured values at sampling locations while maintaining competitive out-of-sample predictive performance comparatively to regression-kriging and classical random forest.

---

\*Intervenant