# Use of Random Forests in Small Area Estimation

Keven Bosa

Statistics Canada

---

When domain sample sizes are small, design-consistent direct estimators of population parameters can be unstable. To improve the precision of direct estimators, the Fay-Herriot (F-H) area-level model is often used. This model relies on the validity of a linear linking model which specifies the relationship between the parameter of interest and the auxiliary variables. The linearity assumption is not always reasonable, which requires a modification of the linking model, for example using a piecewise linear model. As the demand for small area estimates increases, it becomes more and more relevant to evaluate nonparametric linking models, such as random forests, to determine if they can bring some robustness against departure from the linearity assumption. We have chosen to investigate random forests mainly for two reasons: i) they can be applied to a mixture of categorical and continuous auxiliary variables, and ii) they produce predictions which remain in the range of observed values. The F-H model also required the variance of the linking model and the smooth design variance as input parameters. We also proposed to use random forests to estimate these parameters. During this talk, I will assess the properties of the use of random forests in the F-H model through a simulation study.