Highlight

* Geostatistical and machine learning methods show different strength and limitations.
* The model comparison concerns prediction intervals and model interpretation.
* Post-processing random forest with Lasso regression outperforms random forest.
* Geostatistical stacked learning outperforms stacked learning methods.
* It is important to model spatial structure in national NO2 mapping.
* Non-spatial machine learning methods may not fully capture the spatial process.
* Spatial and non-spatial CV strategies affect model comparison.