**Global, high-resolution statistical modelling of NO2**

Meng Lu1, Oliver Schmitz1, Kees de Hoogh2,3, Perry Hystad4, Luke Knibbs5, Derek Karssenberg1

1 Department of Physical Geography, Utrecht University, the Netherlands

2 Swiss Tropical and Public Health Institute, Basel, Switzerland

3 University of Basel, Basel, Switzerland

4 Oregon state University, USA

5 The University of Queensland, Australia

High spatial resolution (<100m) mapping of NO2 at various temporal scales (e.g., hours of the week, month, or year) provides opportunities to study the relationship between personal air pollution exposure and health over large populations. Statistical modelling of NO2 at the global scale provides high-resolution estimations for countries with deficient ground station measurements and provides air pollution maps and human exposures with consistent uncertainties for global health studies. Our objective is to develop spatiotemporally-resolved statistical learning models, understand the temporal dynamics of NO2 and the contributing sources, and open-source our global NO2 prediction maps at 100 m resolution. The global maps are provided at various temporal aggregations (e.g. separating between weekdays and weekends, day and night) and spatial aggregations (e.g. multiple gridded resolutions, administrative units) to facilitate global exposure assessment. To create these maps, we compiled from multiple sources a dataset of hourly NO2 measurements from more than 7000 ground stations over the globe, considerably larger in size and spatio-temporal coverage than used in recent high-resolution NO2 mapping studies. For statistical modelling, geospatial predictors include Sentinel-5 satellite (Tropomi instrument) measurements, variables relating to the emission sources (e.g., road network), dispersion processes (e.g., meteorological variables), elevation and Earth nightlights (from VIIRS nightlight data). We evaluate various statistical models including linear models, ensemble tree-based models, deep convolution models, stacked models with regularisation, and hierarchical modelling strategies and select the optimal model for mapping. Evaluation of models included uncertainty assessment as well as spatial validation methods.