Geophysical Research Abstracts
Vol. 17, EGU2015-3503-1, 2015
EGU General Assembly 2015

Making sense of crowdsourced observations: Data fusion techniques for real-time mapping of urban air quality

Philipp Schneider, Nuria Castell, Matthias Vogt, William Lahoz, and Alena Bartonova
NILU - Norwegian Institute for Air Research, Kjeller, Norway (ps@nilu.no)

With an ever-increasing amount of environmental observations available through crowdsourcing, one of the major emerging challenges is how to best make sense of the vast amount of collected observations and how to provide citizens and other end-users with a relevant value-added product. Regarding air quality, a high-density network of low-cost sensors provided by crowdsourcing has significant potential for improving spatial mapping in general and in urban areas in particular. However, most datasets of observations made within a crowdsourcing framework contain substantial data gaps and the observations are generally point measurements, which are only representative of a relatively small area. This poses a significant challenge for mapping applications. One way to overcome these issues is to combine the crowdsourced data with spatially continuous data from a model.

We present a novel data fusion-based technique for combining real-time crowdsourced observations with model output that allows to provide highly detailed, up-to-date maps of urban air quality. The EU-funded CITI-SENSE project is deploying a dense network of low-cost sensors measuring air quality in eight cities around Europe. These crowdsourced observations are used for mapping urban quality in real-time by fusing them with data obtained from statistical or deterministic air quality models. Data fusion techniques allow for combining observations with model data in a mathematically objective way and therefore provide a means of adding value to both the observations and the model. The observations are improved by filling spatio-temporal gaps in the data and the model is improved by constraining it with observations. The model further provides detailed spatial patterns in areas where no observations are available. As such, data fusion of observations from high-density low-cost sensor networks together with models can contribute to significantly improving urban-scale air quality mapping.

The data fusion approach used here is based on geostatistics and has been implemented as a fully automated system running in real-time. Initial testing of the methodology has been performed in the city of Oslo, Norway. First results indicate that the system is capable of producing detailed urban air quality maps that are based on adjusting the model output with new information from the crowdsourced observations. The system limits the spatial impact of the observations to their immediate surroundings as specified by the theoretical semivariogram. Evaluation of the methodology is being carried out using the leave-one-out cross validation technique and simulated datasets.

As the resulting value-added product, the detailed urban air quality maps can then further be used for providing personalized information about air quality. This near real-time information will help users to find the currently least polluted route through a city or to track their individual exposure to air pollution while moving through the city.