Theme: Geospatial data analysis

Modelling discrepancies between predicted and measured Ammonia in Dutch Nature network using machine learning techniques

Problem description

Nitrogen (N2) is a colourless and odourless gas that forms about 78% of Earth's atmosphere. It is not harmful to humans or the environment by itself. However some of its compounds can be harmful to humans and environment. They include at least nitrogen oxides (NOx, a compound of nitrogen and oxygen) and ammonia (NH3, a compound of nitrogen and hydrogen). The amount of nitrogen oxides and ammonia in the air is called the concentration. Three factors can generally contribute to the so-called "nitrogen problem" in the Netherlands, 1) emission (how much gets into the air), 2) concentration (how much is in the air) and 3) deposition (how much gets on the earth surface). The Dutch national institute for public health and environment (RIVM) has developed an empirical model that predicts the ammonia concertation and deposition for the entire country. The model uses the input variables such as weather conditions, land use and traffic variables (e.g. number of cars, the emission of substances per car and data on the number of kilometres of asphalt in the Netherlands). There is also a national monitoring network that directly measure the ammonia and nitrogen oxides in a couple of stations (more than 80 stations) distributed around the Netherlands. There are however discrepancies between the results predicted from the model and those from the measurements. The observations from the monitoring networks are usually used to calibrate the model predictions.

Objectives of project

There are attempts to make the predicted and measured values compatible. For this purpose, data fusion methods using the Kriging method have been used to predict the discrepancies and therefore correct the model predictions. This method will mainly correct for the spatial variability of the discrepancies using the well-known Kriging method. Although the spatial variability plays a significant role, there are also other factors that can significantly affect the discrepancies. For example, temporal variability can play an important role when it is required to understand the temporal evolution of the collected data as time series. They can be used to implement short-term predictions (3 months than a year) for ammonia and nitrogen oxides. On top of these, the discrepancies can also be expressed as a function of model input parameters such as weather, land use and traffic variables. The underlying model for relating all the variables involved to the observed discrepancies requires modern fusion and assimilation method. The goal of this project is to use machine learning methods to make this possible. Having this model available allows to apply the corrections to the prediction model and therefore improve its performance.

Implementation aspects

The data processing algorithm is implemented in Matlab/ Python. I have developed a few Matlab/Python scripts to implement machine learning methods for geospatial data analytics. The scripts need to be adapted and applied to the data of LML (national air quality monitoring network) and MAN (monitoring network ammonia nature reserves) networks (data is available). The project requires other data sets such as meteorological data (from KNMI) and traffic data (CBS) as input in the training using the machine learning method. The discrepancies need to be expressed as a function of the input variables, which can be determined through the ML training. The performance of the method is then tested on the unseen data; those which have not been used in the network training step. This allows to calibrate/correct the predicted results from the available models, and hence downscale them spatially and temporally.

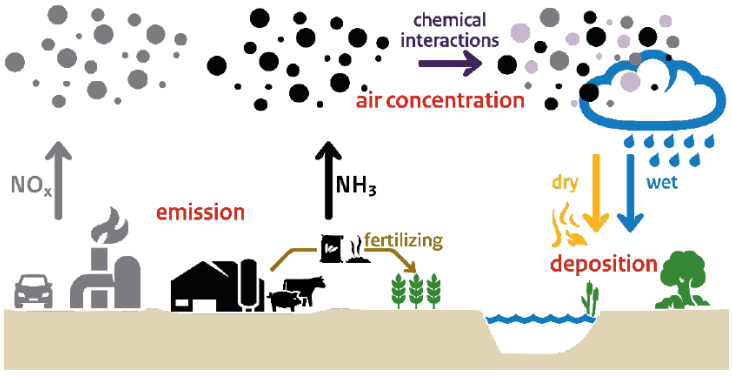

Image: RIVM report 2020-0076 of National Institute for Public Health and the Environment, the Netherlands

Information:
- Dr. Alireza Amiri-Simkooei
- Email: a.amirisimkooei@tudelft.nl
- Room 2.29