SS 2014-2015, SV-SVS, No. DISAL-SP74 Project Summary

Generating Lausanne Pollution Maps using an Artificial Neural Network approach on the OpenSense Dataset

Fanny Jeanneret-Grosjean

Professor : Alcherio Martinoli
Assistant(s) : Ali Marjovi/Adrian Arfire

This project's objectives were to analyse various parameters related to air pollution, and to use data obtained by sensors attached on buses in Lausanne (Switzerland) to develop a model that generates complete maps from explanatory and land-use variables such as meteorological or air pollution ones. The acquisition system, depending on the public transportation schedule and trajectories, has an irregular temporal and spatial sampling rate. The particular measure investigated here is the lung-deposited surface area (LDSA), which varies proportionally to the ground concentration of fine particles and has the interest to be a quite straightforward value to measure.

Neural networks were used to combine explanatory variables optimally and model LDSA concentrations considering different time resolutions. The models implemented only focused on the center of Lausanne, and extrapolating the results to the whole dataset should most likely lead to similar results as what was already obtained. 10-fold cross-validation and systematic search were used to determine the optimal size of the hidden layer and decay rate. Regarding the different modelling options that were considered, time variables were observed to have a positive impact on the results. It was limited due to the quality of the results, but it could help improve significantly a model that is already good. Street characteristics were also increasing the quality of the results, and highlighted the fact that the variability of the data among the streets is a challenge to model. Unfortunately, handling this variability by combining the inputs in a particular way for each street does not seem possible in a single neural network, so the model structure might need to be improved. However, a more complex structure will also represent a higher computational cost, so a trade-off has to be done.

As non-linearities were observed in the dataset, a solution able to handle them properly should be considered. Several artificial neural networks or simpler models combined together might be considered, for example by separating the data in several datasets on which the linearity assumption would hold. Another solution would be to consider a different discretization than the street segments, which would allow regrouping data points by similar characteristics and reducing the variability between the groups. Trying to arrange the data in clusters with self-organizing maps might be promising, as it would then allow the use of a different model on each cluster without falling in the excess of having a particular one for each street segment. The point that should be kept in mind is that having one different model for each street has a limited interest due to the lack of generalization that it means.