
ORDINAL $PM_{2.5}$ DATA MINING WITH NON-PARAMETRIC CONTINUOUS BAYESIAN BELIEF NETS

Anca Hanea¹ and Winston Harrington²

¹ *Delft Institute of Applied Mathematics*

Delft University of Technology

The Netherlands

² *Resources for the Future*

Washington DC

United States

This paper introduces a Bayesian Belief Net (BBN) based approach for analysing the relationship between SO_2 emissions from power plants, and ambient concentrations of fine particulate matter of 2.5 micrometers or less in diameter, denoted $PM_{2.5}$. $PM_{2.5}$ exposure has adverse health effects for humans, hence we study this relationship with the goal of quantifying the health benefits/risks of emission reductions/increase. The data are gathered from electricity generating stations and from collection sites in the United States over the course of seven years.

A method for mining ordinal multivariate data using non-parametric BBNs is presented. The main advantage of this method is that it can handle a large number of continuous variables, without making any assumptions about their marginal distributions, in a very fast manner. Of course, rapid computations are of little value if the model itself cannot be validated.

We discuss the issue of validation and additionally answer questions of the following types: "Given emissions at a number of relevant plants, what is the effect on a given collector", or "Given measured ambient concentrations at one or more collectors, what can we infer about given emitters?"