Applying causal discovery algorithm to find predictors for transformation process of wood combustion emission

V. Leinonen[1], P. Tiitta[2], O. Sippula[2,3], H. Czech[2,4], A.Leskinen[1,5], J. Karvanen[6], S. Isokääntä[1], and S. Mikkonen[1,2]

[1]University of Eastern Finland, Department of Applied Physics, Finland.
[2]University of Eastern Finland, Department of Environmental and Biological Sciences, Finland.
[3]University of Eastern Finland, Department of Chemistry, Finland
[4]Helmholtz Zentrum München, Cooperation group "Comprehensive Molecular Analytics (CMA)", Germany
[5]Finnish Meteorological Institute, Kuopio, Finland
[6]University of Jyväskylä, Department of Mathematics and Statistics, Finland

Introduction

- Measurements of multiple measurement instruments (AMS, SMPS, PTR-MS and gas analyzers) have been combined to study the evolution of residential wood combustion emission. Statistical model called causal model have been applied to the combined dataset.
- Causal discovery algorithm have been applied for measured evolution of emission to study dependencies in and between gas- and particle-phase.
- Causal model have been formed based on the obtained structure to study the evolution of emission.

Chamber experiments

- Evolution of residential wood combustion emission have been studied in the chamber (Tiitta et al., 2016).
- In this study, four experiments, two dark-aging and two photochemical aging, were applied.
- The study combined statistical data-analysis and causal model (Pearl, 2009) to understand the evolution. The study attempts to understand the relationships between variables in the dataset and how those affect the evolution of the whole emission.
- We modeled the observed evolution of the measured variables by searching the structure for variables (Fig. 1). Evolution was investigated using causal discovery PC-algorithm (Wongchokprasitti, 2019). Algorithm was used for searching the dependencies between measured variables (X) and the change of the variable between time points ($\Delta(X)$).
- Prior information about known dependencies between variables were used to as input parameters of the algorithm.

Steps for creating the model

1. Use the causal discovery algorithm to search the potential dependencies between measured variables and measured changes ($\Delta(x)$:s) in dataset.

2. Form interaction variables from measured variables such that both variables in the interaction variable should be suggested by the algorithm. In addition, prior assumptions have been taken into account here. This means that if some variables are forbidden to affect $\Delta(x)$, those variables cannot be part of the interaction variable. Interaction variables are formed separately for each $\Delta(x)$.

3. Use LASSO (least absolute shrinkage and selection operator) to reduce the amount of selected predictor variables. Estimate the coefficients of each selected predictor using a linear model.

4. Calculate the modeled evolution using ODE system deSolve (Soetaert et al., 2010). The system uses estimated coefficients (step 3) as reaction coefficients and the first observation from the experiment as initial state.

Obtaining causal structure and reaction coefficients

- Causal structure represents researcher's assumption of dependencies between variables. The direction of an arrow represents the assumed direction of the dependence.
- Interactions (products) of two variables have been used as causing variables (Figure 3). Reaction coefficients for each arrow were estimated using linear regression model.

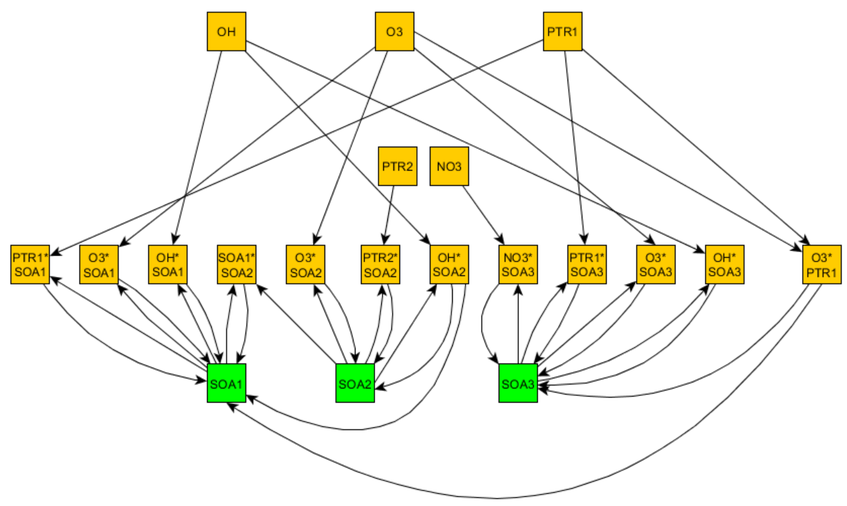

Figure 1: Causal graph obtained for photochemical aging experiments. Graph is only showing edges related to SOA variables

Modeling results for combustion experiments

- Based on the obtained structure and reaction coefficients, the evolution of the system was simulated using ODE-solver *deSolve* (Soetaert et al., 2010) in R environment. Same structure and coefficients were used for both experiments.

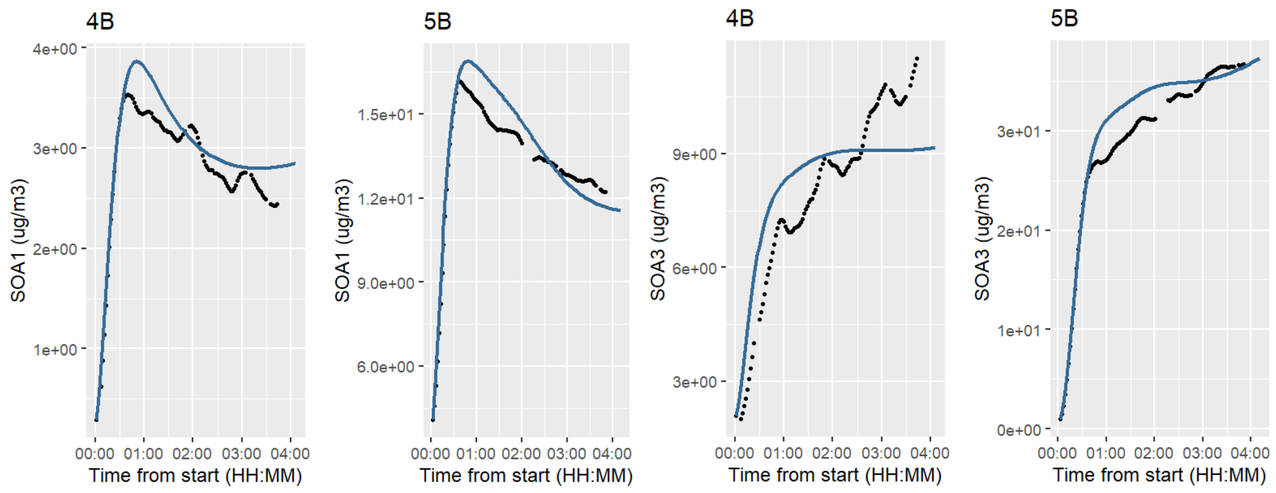

Figure 1: Evolution of SOA1 and SOA3 factors in photochemical aging experiments (4B and 5B). Black points represent the filtered version of variable, and blue line is the modelled evolution.

Conclusions

- Wood combustion emissions contain numerous reacting compounds with unknown reaction paths. Causal model suggests possible reaction paths for researcher to interpret.
- Causal discovery algorithm is based on the observed dependencies in the dataset. Without explicit prior information of the studied phenomenon, careful consideration of the truthfulness of the obtained dependencies is necessary.
- See more detailed information of this study from the paper which is currently in review (https://www.geosci-model-dev-discuss.net/gmd-2020-13/)

References

Pearl, Cambridge University Press, Cambridge, U.K, 2009
Soetaert et al. J. Stat. Softw. 33(9), 1-25, 2010
Tiitta et al., Atmos. Chem. Phys., 16, 13251-13269, 2016
Wongchokprasitti, rcausal: R-Causal Library, 2019

Contact

Ville Leinonen, MSc
ville.j.leinonen@uef.fi
+358503517579

UNIVERSITY OF EASTERN FINLAND