Supplement to the manuscript 'Modeling atmospheric aging of small-scale wood combustion emissions: distinguishing causal effects from non-causal associations'

Ville Leinonen, Petri Tiitta, Olli Sippula, Hendryk Czech, Ari Leskinen, Sini Isokääntä, Juha Karvanen, and Santtu Mikkonen

Table of contents

Figures

Tables

S1 Motivation for time series filtering

The measurement $y_{k,t}$ at time t for time series y_k is the sum of the true value $\alpha_{k,t}$ and the measurement error $\varepsilon_{k,t}$. These measurement errors comprise sampling from the chamber ($\varepsilon_{k_{sampling},t}$), error related to the estimation of particles and gases losses to chamber walls ($\varepsilon_{k_{wall-loss},t}$), and error related to processing of measurement instrument data from raw data into more useful form ($\varepsilon_{k_{processing},t}$).

Measurements $y_{k,t}$ were presented as

$y_{k,t} = \alpha_{k,t} + \varepsilon_{k,t}$ (A.1)

$\varepsilon_{k,t} = \varepsilon_{k_{sampling},t} + \varepsilon_{k_{wall-loss},t} + \varepsilon_{k_{processing},t}$ $(A.2)$

where the error term $\varepsilon_{k,t}$ is independent in time and follows a specified distribution presenting all uncertainties. The state $\alpha_{k,t}$ describes the estimate of the real state of the variable in the chamber and the error term $\varepsilon_{k,t}$ represents the error related to the estimation of the state.

Understanding the evolution of the state $\alpha_{k,t}$ of the variable is the question of interest. We would like to understand the factors affecting the change of state during the aging of emissions. Therefore, we estimated the state $\alpha_{k,t}$ from measurements $y_{k,t}$ for each variable.

S2 Methods for simulation studies

Two simulated data sets (see Table 1 in the main text) were formed to study how model would perform in a situation where we know the correct evolution and structure. Both data sets were formed using R-package deSolve (Soetaert et al., 2010). Data sets describe the evolution of Ordinary Differential Equation (ODE) system which length is 100. The difference between data sets is the way how differential equations of variables are linked to each other. In smaller data sets, differential equations are following the Laws of Mass Action, applied in R-package

episode (Mikkelsen, 2017; Seinfeld and Pandis, 2016, see Table S1 for coefficients used in the system). In larger data set, equations have been formed independently for each variable, using pre-defined causal structure for variables (see Table S2 for coefficients used in the system. We called these data sets as simulated data sets throughout the text.

Table S1 Coefficients for smaller simulated dataset.

$\Delta(x)$	Predictor	Coefficient
V2	V1+V3	0.021
V1	V2	0.0225
V4	V2	0.0225
V1	V2+V4	0.015
V5	V2+V3	0.02
V2	V5	0.01
V4	V5	0.01
V6	V2+V5	0.02
V2	V6	0.03
V5	V6	0.03
V7	V6	0.015
V7	V2+V8	0.02
V2	V7	0.03
V8	V7	0.03
V3	V4	0.015

Table S2 Coefficients for larger simulated dataset.

Gases			Particle chemical composition		
$\Delta(x)$	Predictor	Coefficient	$\Delta(x)$	Predictor	Coefficient
NO2	NO2*NuclM	-2.30E-08	NO3	NO3*O3	8.15E-04
O3	PTR3	1.59E-03	NO3	NO3*AitM	-1.33E-07
O3	O3*PTR1	4.19E-04	NO3	NO3*NuclM	-4.67E-08
O3	O3*NuclM	-7.85E-07	NO3	NO3*CoarseM	-4.21E-08
O3	O3*CoarseM	-1.22E-07	NO3	AitM*O3	1.52E-08
O3	NuclM*PTR3	1.06E-07	NO3	AitM	1.37E-10
O3	PTR1*CoarseM	4.90E-08	NO3	AitM*CoarseM	1.19E-12

OH	PTR2*OH	-2.44E-02
PTR1	PTR1	-6.59E-03
PTR1	PTR3*PTR1	3.00E-05
PTR1	PTR1*NuclM	-1.08E-07
PTR1	NuclM	1.42E-08
PTR1	CoarseM*PTR1	-6.58E-09
PTR1	PTR3*NuclM	1.31E-09
PTR1	NuclM*CoarseM	2.45E-12
PTR2	PTR2	5.86E-03
PTR2	AitM*PTR2	-1.17E-08
PTR3	PTR3	-1.22E-03
PTR3	PTR3*PTR2	6.09E-05
PTR3	PTR3*CoarseM	2.05E-09

Particle size

$\Delta(x)$	Predictor	Coefficient
AccM	AccM*PTR3	-1.03E-05
AccM	NuclM*CoarseM	9.83E-07
AccM	NuclM*AccM	-1.26E-07
AccM	AccM*CoarseM	-1.03E-08
AitM	AitM	4.39E-02
AitM	PTR3*AitM	-2.23E-04
AitM	AitM*AccM	6.12E-09
CoarseM	AccM*NO2	8.82E-06
CoarseM	NuclM*CoarseM	-7.89E-08
CoarseM	AitM*CoarseM	-5.05E-08
CoarseM	AitM*AccM	1.36E-08
NuclM	NuclM	-2.02E-03
NuclM	NuclM*CoarseM	1.77E-07
NuclM	AccM*NuclM	-5.19E-08
POA1	AitM*POA1	-1.58E-08
POA2	POA2	9.55E-02
POA2	PTR1*POA2	-1.20E-02
POA2	POA2*O3	1.34E-03
POA2	AitM	5.99E-07
POA2	AitM*POA2	-5.06E-08
SOA1	O3*SOA1	-7.46E-05
SOA1	NuclM*SOA1	1.25E-07
SOA2	O3	2.81E-02
SOA2	POA2*SOA2	-1.57E-03
SOA2	SOA2*O3	-8.86E-04
SOA2	SOA2	3.40E-04
SOA2	NuclM*SOA2	-7.10E-08
SOA3	SOA2*POA2	6.42E-04
SOA3	SOA2*SOA1	4.16E-04
SOA3	SOA3	-1.49E-04
SOA3	SOA3*SOA2	1.26E-05
SOA3	NuclM*SOA3	-1.76E-07

Several questions of interest existed related to the properties of input data set and data pretreatment (Table S3). Firstly, we were interested to study how the precision of the measurements by analytical instrumentation is related to the model fit, which is assessed by

using different proportions of random noise mimicking the measurement error and the number of measurements made in time. To mimic the measurement error, normally distributed random noise was added to the variables. The standard deviation of the added random noise have been proportional to the standard deviation of the simulated variables due to evolution process. The proportion of the standard deviation of the random noise have been called fraction of added uncertainty. The number of measurements during the same evolution length (100) was altered to be between 26 and 401.

Secondly, we were interested to know whether the methods we applied to increase the quality of data are increasing the quality of the fit, i.e. accuracy of fit and obtained causal structure. Does filtering or smoothing of time series improve the fit and accuracy of prediction and is there an optimal time resolution to which data should be averaged?

Thirdly, the amount of necessary prior information was the question of interest. We were interested to study the importance of prior information given to a causal discovery algorithm to the modeled structure. Does addition of prior information improve the accuracy of modeled structure and how much prior information is necessary to get a reasonably good fit for the model.

The question about necessity of the prior information is also related to the dependence of model fit and the correctness of the structure. Intuitively, one might think that the correct structure would produce the best fit for the evolution. As many of the variables are highly dependent, it is probable that we will fail to obtain the exactly correct structure between variables. In addition to the differences between obtained and real structure in the model, we are interested about the predictive value of the obtained model compared to the simulated evolution. If dependent variables which we use in the model to explain the evolution are correlated with real causes in the data set, the model might still be able to predict the evolution of emission in the chamber.

Table S3 Simulated experiments to study the model performance. Aim of simulation studies is to investigate how different parameters affect model capability to return correct structure and goodness of fit.

	How test was performed	Purpose of the test
Measurement frequency (Table S4 and S5)	Reducing or increasing number of simulated 'measurement' points in given time (100). Frequency of 1/4,1/2,1,2,4 time points was used. This means dataset has 401, 201, 101, 51 and 26 measurement points during time simulation time 0-100.	To see whether increasing or decreasing measurement frequency would help making model better. Increasing of points with same measurement noise would lead to lower signal/noise ratio?
Measurement uncertainty (Table S4 and S5)	Adding normally distributed random noise to the simulated evolution. This random noise was representing the possible measurement uncertainty.	Purpose was to see if the measurement uncertainty reduces both structural accuracy and fit accuracy.
Filtering and smoothing (Figure S1)	Applied filtering and smoothing to original variables of simulated dataset.	How filtering or smoothing would improve the fit and structure of the model? Is it reasonable to use filtering or smoothing of dataset before making a model?
Prior Information (Table S6 and S7)	Using prior information about the edges possible in the model. Using both correct and incorrect prior information. Fraction of information from all correct prior information was used as a measure of information.	Does addition of prior information help model to get correct structure and good fit?

Accuracy tests for causal discovery algorithms have been performed earlier (Scheines and Ramsey, 2016; Singh et al., 2017). However, those tests are dependent on the used data set. In our case, variables that explain the evolution of some variable do not originate directly from the discovery algorithm. Our situation differs from the tests performed earlier as the variables from the algorithm are used to form possible interaction variables, not directly to explain the evolution.

We measured the performance of the model in simulated data set by two ways. First is the accuracy of the model fit to the simulated data set: how well the model can capture simulated evolution and how well the model can predict the simulated evolution after fitted data set. Second can be called as structural accuracy: how well the underlying causal structure of variables can be returned by the model.

For measuring accuracy of the model fit, we compared the evolution obtained from the model to measurements. Evolution was then compared to true evolution, not including the error added to the simulations, using Root Mean Squared Error (RMSE) for each time series. To equally weight each time series when calculating RMSE, each time series were scaled by dividing those with its standard deviation before calculating RMSE. In further text, we refer to this scaled version as RMSE.

In addition to the accuracy of the model, we also evaluated the predictive accuracy of the model. We used the obtained coefficients from the model to predict further time steps of the evolution of the system. Then we compared the prediction to the same time steps from the real system and evaluate the accuracy of model prediction using RMSE. Prediction length was 30% of the simulated data set used to fit a model.

For measuring structural accuracy, we used adjacency precision (AP), adjacency recall (AR) (Scheines and Ramsey, 2016) and F-score (Singh et al., 2017). AP was defined as a fraction of correct edges in the model of all proposed edges. AR was defined as a fraction of correct edges in the model of all correct edges. F-score was defined based on AP and AR as

$F_{score} = 2 * \frac{AP*AR}{AP+AR}$.

In addition to F-score, we wanted to study whether incorrect predictors for variables were close to correct causes and whether the model could find a good replacement for each correct predictor that was not chosen for modeled structure. Correlation was used to measure closeness here. For each correct predictor we calculated correlation between it and each predictor in the model (for same$\Delta(x)$). The maximum of these correlations was taken as the value for that

predictor. This results that if the correct predictor was in the model, correlation was 1. In the results section we have calculated the mean value of correlations for all the predictors.

S3 Results of simulation studies

As expected, the model fit to the data was better when the error assigned on it was lower. The effect of measurement error is reported for simulated data sets (Tables S4 and S5). The effect of uncertainty was small for F-score and correlation between the correct edges and edges in the model. This was expected, as the error makes the data set noisier. Furthermore, as the simulations also studied the effect of time resolution in measurements, in both data sets the optimal number of measurements during one experiment was related to the amount of error applied (Tables S4 and S5). For lower error, higher time resolution leads to better results when the real change between subsequent time points was dominating the change observed and if many observations about the evolution process were available. In case of high error, fewer observations from the phenomenon were preferred, as the signal to noise ratio was low and increasing the number of observations would make it even lower. However, the dependence between sample size and fraction of error was weak. This is still opposite of what is expected and could be due to a small sample size.

We found that in some cases, there were unstabilities in computations. These can be seen as large numbers of RMSE in the tables S4 and S5. These are resulted by some variable that gets large values in the differential equation system, resulting the system to be unstable. We haven't been studied the exact reason for the unstability. This could be due to the inaccurate estimation of the structure, i.e. some incorrect edge in the model results a dependence that is having "large" coefficient, and that differential equation would result system to be unstable. Other option is that the estimate itself is overestimated, i.e., due to low number of points or inaccuracies in the measured differences that are modeled.

Table S4 Effect of measurement frequency and fraction of error to goodness of fit parameters of the model for larger simulated dataset. Variable 'Fraction of added uncertainty' is a standard deviation of random noise added to simulation data (sd of each time series is adjusted to 1). Number of edges is the number of edges in the model, RMSE is Root Mean Squared Error, F-score is a measure of structural accuracy, based on a number of correct and incorrect edges, see S1 for more information. Mean correlation is mean correlation between proposed and correct predictor variable. For each row, number of replications is 10. For number of observations = 401, and fraction of added uncertainty = 0.5, there is 2/10 cases, when some unstabilities result large values for RMSE.

Number of observations	Fraction of added uncertainty	Mean number of edges in the model	Mean of RMSE	Mean of RMSE for prediction	Mean of Fscore	Mean correlation of correct predictor and selected predictor
26	0.01	35.9	17.07	12.86	0.22	0.88
51	0.01	46.2	17.47	16.2	0.22	0.88
101	0.01	41.1	19.43	18.13	0.22	0.87
201	0.01	39.1	23	20.63	0.2	0.86
401	0.01	28	24.46	17.85	0.21	0.83
26	0.05	30	19.97	16.51	0.19	0.82
51	0.05	29.9	19.77	20.07	0.17	0.84
101	0.05	29.5	17.05	19.03	0.19	0.89
201	0.05	27.4	19.84	22.21	0.19	0.92
401	0.05	24.5	23.61	20.71	0.18	0.93
26	0.1	25.9	19.96	15.68	0.17	0.86
51	0.1	28.9	18.91	19.51	0.2	0.93
101	0.1	28	21.24	21.56	0.19	0.93
201	0.1	22.4	28.06	35.83	0.18	0.92
401	0.1	19.8	33.86	22.49	0.18	0.91
26	0.5	33.7	31.17	25.76	0.19	0.93
51	0.5	28.3	47.12	31.35	0.19	0.92
101	0.5	41.1	1757.27	38.03	0.22	0.94
201	0.5	44.3	344.31	42.27	0.23	0.94
401	0.5	46.1	2.19841E+18	71.84	0.23	0.93

Table S5 Effect of number of observations and fraction added uncertainty to goodness of fit parameters of the model in smaller simulated dataset. Variable 'Fraction of added uncertainty' is a standard deviation of random noise added to simulation data (sd of each time series is adjusted to 1). Number of edges is the number of edges in the model, RMSE is Root Mean Squared Error, F-score is a measure of structural accuracy, based on a number of correct and incorrect edges, see S1 for more information. Mean correlation is mean correlation between proposed and correct predictor variable. For each row, number of replications is 10. For number of observations = 401, and fraction of added uncertainty = 0.5, there is 1/10 cases, when some unstabilities result large values for RMSE.

Number of observations	Fraction of added uncertainty	Mean number of edges in the model	Mean of RMSE	Mean of RMSE for prediction	Mean of Fscore	Mean correlation of correct predictor and selected predictor
26	0.01	15.3	7.76	4.5	0.22	0.85
51	0.01	17.7	5.88	4.16	0.23	0.87
101	0.01	22.5	4.64	3.67	0.21	0.81
201	0.01	22.5	4.02	3.71	0.21	0.75
401	0.01	18.4	6.68	4.07	0.26	0.81
26	0.05	12.6	8.79	4.71	0.13	0.75
51	0.05	14.1	8.93	3.89	0.2	0.79
101	0.05	16.8	7.89	3.33	0.25	0.81
201	0.05	13.8	10.83	2.85	0.19	0.8
401	0.05	11.5	17.93	3.9	0.24	0.8
26	0.1	10.1	9.91	3.59	0.17	0.77
51	0.1	10.6	10.99	2.92	0.24	0.78
101	0.1	11.3	15.23	4.35	0.24	0.79
201	0.1	9.8	22.61	8.1	0.31	0.68
401	0.1	7.4	32.08	11.81	0.32	0.63
26	0.5	5.6	31.64	12.85	0.23	0.61
51	0.5	5	40.16	21.21	0.23	0.59
101	0.5	6	47.46	29.68	0.22	0.63
201	0.5	6.4	51.29	36.34	0.23	0.65
401	0.5	8.2	2.49773E+17	795977.48	0.26	0.76

Filtering or smoothing of the time series representing measurements reduces the error, especially when measurement uncertainty was high (Figure S1). These methods were also improving the accuracy of model prediction. To understand importance of applying smoothing or filtering method in a real data set, we need to understand whether the error in the real data set is large enough that applying filtering or smoothing method is reasonable. We assume the error related to the time series as the difference between these data sets. The standard deviation of the error was then compared to the standard deviation of the filtered data set to estimate the fraction of the error of the whole standard deviation in the time series.

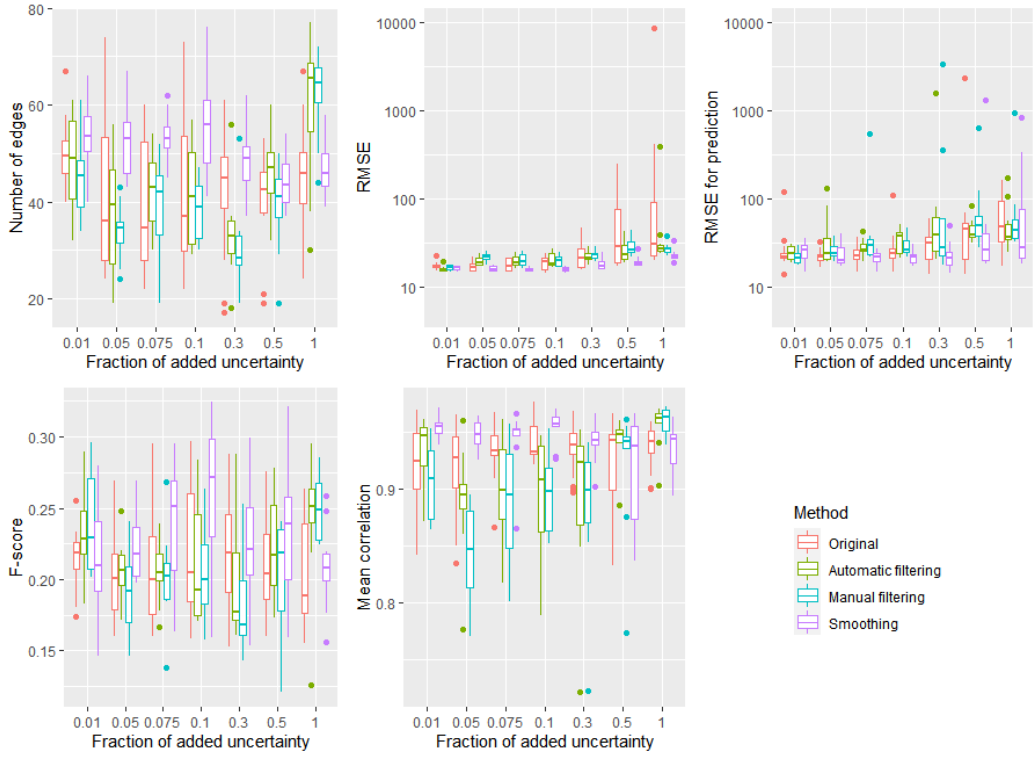

Figure S1 Effect of filtering and smoothing to the goodness of fit parameters. Method 'Original' means that time series is used without filtering or smoothing. In automatic and manual filtering, filter have been applied by selecting length of the filtering window by cross-validation (automatic) or by manual selection. Window length in smoothing have been chosen by cross-validation. Variable 'Fraction of added uncertainty' is a standard

deviation of random noise added to simulation data (standard deviation of each time series is adjusted to 1), see S1 for more information. Number of edges is the number of edges in the model, RMSE is Root Mean Squared Error, F-score is a measure of structural accuracy, based on number of correct and incorrect edges, see S1 for more information. Mean correlation is mean correlation between proposed and correct predictor variable. There is some additional points that are outside of the figure and are due to some unstabilities in the fit. For RMSE, there is one larger value for smoothing, two larger values for manual filtering and two for original time series, all with fraction of added uncertainty of 1. For predicted RMSE, there is one larger value for smoothed, uncertainty 0.1, three for manually filtered, with uncertainties 0.3, 0.3, and 1, and one for automatically filtered, with uncertainty 0.5.

In Aerosol Mass Spectrometer (AMS) and Proton-Transfer-Reactor Time-of-Flight Mass Spectrometer (PTR-ToF-MS) data sets and in coarse size bin (>300nm) of Scanning Mobility Particle Sizer (SMPS) measurements, the fraction of standard deviation of error was around 10-20% of the whole standard deviation. Variables measured by gas analyzers contained the least amount of error, approximately 1-5% of the standard deviation of filtered time series. However, because the fraction of error of some variables is high, applying of the filtering methods to real data set was reasonable.

Reasons why applying the filtering methods to the data set is not improving the fit before the error is relatively high could be assessed further but is not on the scope of this study. In addition to the fraction of error, the sample size is probably related to the necessity of the filtering method. These simulations for the larger data set were performed using the sample size of 100, which is the same sample size that was used in the shortest experiment.

Fourthly, prior information was closely related especially to the correctness of the obtained dependence structure between variables in the data set. In some situations, there might be large amounts of prior information from previous studies available, which might help to construct the structure based on prior knowledge. This means that the correctness of the structure is high, and the causal inference based on the structure and the observed data set has higher quality.

From wood combustion emission experiments, we do not have much prior knowledge about the phenomenon. Even though there has been and currently is a vast amount of research about the oxidative aging of combustion emissions, the number of possible reactions occurring during experiments is very high and the reaction coefficients are therefore hard to estimate. Thus, the construction of the complete causal structure between variables might be out of reach based on low amount of prior information.

One question of interest is that how the incorrect edges in the structure are related to the missing or present correct edges. Based on the nature of the causal discovery algorithm one might think that the algorithm is not able to separate between causal and non-causal associations. Therefore, it might be that in the model the real cause is replaced with an indicative one. This decreases the accuracy of the structure, but not necessarily affect the fit or accuracy of prediction of the model. Especially if the dependence between a real and an indicative cause is strong. Decreased accuracy in the structure may produce incorrect interpretations of the causal effects.

We found that there is a small difference in both goodness of fit and structure accuracy parameters when the number of correct and incorrect edges in prior information is varied in the smaller, artificial data set. Larger amount of correct prior information led to slightly lower RMSE for the model and prediction (Table S6). Addition of correct prior information led to lower RMSE and higher F-score. For prediction, the information had not much effect. Incorrect prior information didn't affect the prediction accuracy, nor the F-score as would have been expected. F-score was even higher when there was more incorrect information available (Table S7). This indicates that the prior information is not optimally used in the current version of the model and should be handled better in the future.

Table S6 Effect of correct prior information for the goodness of fit parameters of the model in smaller simulated dataset. Fractions of correct and incorrect information about dependencies are X -> Y dependencies given as a prior information for causal discovery algorithm. Variable 'Fraction of added uncertainty' is a standard deviation of random noise added to simulation data (standard deviation of each time series is adjusted to 1). Number of edges is the number of edges in the model, RMSE is Root Mean Squared Error, F-score is a measure of structural accuracy, based on a number of correct and incorrect edges, see S1 for more information. Mean correlation is mean correlation between proposed and correct predictor variable. For each row, number of replications is 10, number of observations is 101, and fraction of added uncertainty is 0.5.

Fraction of correct information about dependencies	Fraction of correct info about non-dependencies	Fraction of incorrect info about dependencies	Fraction of incorrect info about non-dependencies	Mean number of edges in the model	Mean of RMSE	Mean of RMSE for prediction	Mean of Fscore	Mean correlation of correct predictor and selected predictor
0.05	0.05	0	0	10.4	15.85	4.98	0.2	0.79
0.3	0.05	0	0	22.1	8.99	4.85	0.36	0.87
0.5	0.05	0	0	31.2	7.62	4.74	0.44	0.9
0.05	0.3	0	0	11	15.36	4.95	0.23	0.8
0.3	0.3	0	0	20.7	12.31	4.83	0.36	0.83
0.5	0.3	0	0	30.9	7.07	4.54	0.43	0.9
0.05	0.5	0	0	10.8	14.55	4.93	0.23	0.8
0.3	0.5	0	0	22	6.73	5.2	0.38	0.88
0.5	0.5	0	0	32.3	5.19	5.33	0.48	0.93

Table S7 Effect of incorrect prior information for the goodness of fit parameters of the model in smaller simulated dataset. Fractions of correct and incorrect information about dependencies are X -> Y dependencies given as a prior information for causal discovery algorithm. Variable 'Fraction of added uncertainty' is a standard deviation of random noise added to simulation data (standard deviation of each time series is adjusted to 1). Number of edges is the number of edges in the model, RMSE is Root Mean Squared Error, F-score is a measure of structural accuracy, based on a number of correct and incorrect edges, see S1 for more information. Mean correlation is mean correlation between proposed and correct predictor variable. For each row, number of replications is 10, number of observations is 101, and fraction of added uncertainty is 0.5.

Fraction of correct information about dependencies	Fraction of correct info about non-dependencies	Fraction of incorrect info about dependencies	Fraction of incorrect info about non-dependencies	Mean number of edges in the model	Mean of RMSE	Mean of RMSE for prediction	Mean of Fscore	Mean correlation of correct predictor and selected predictor
0.3	0.3	0	0	20.7	12.31	4.83	0.36	0.83
0.3	0.3	0.3	0	32.5	6.2	5.07	0.45	0.9
0.3	0.3	0.5	0	41.9	2.23	5.15	0.51	0.95
0.3	0.3	0	0.3	21.1	9.14	4.59	0.38	0.87
0.3	0.3	0.3	0.3	33	5.51	5.61	0.49	0.91
0.3	0.3	0.5	0.3	44.6	3.16	5.96	0.5	0.95
0.3	0.3	0	0.5	22	10.5	4.62	0.4	0.86
0.3	0.3	0.3	0.5	32.7	5.34	5.09	0.48	0.92
0.3	0.3	0.5	0.5	44.2	2.43	4.6	0.51	0.95

S4 Additional information about the models for wood combustion experiments

Related to wood combustion experiments, evolution of variables that are not represented in the main text are represented in Figures S2-S7. Figures S2-S4 show the evolution in dark aging experiments and Figures S5-S7 in photochemical aging experiments. Tables S8 and S9 provide estimated coefficients used to describe evolution in dark and photochemical aging experiments.

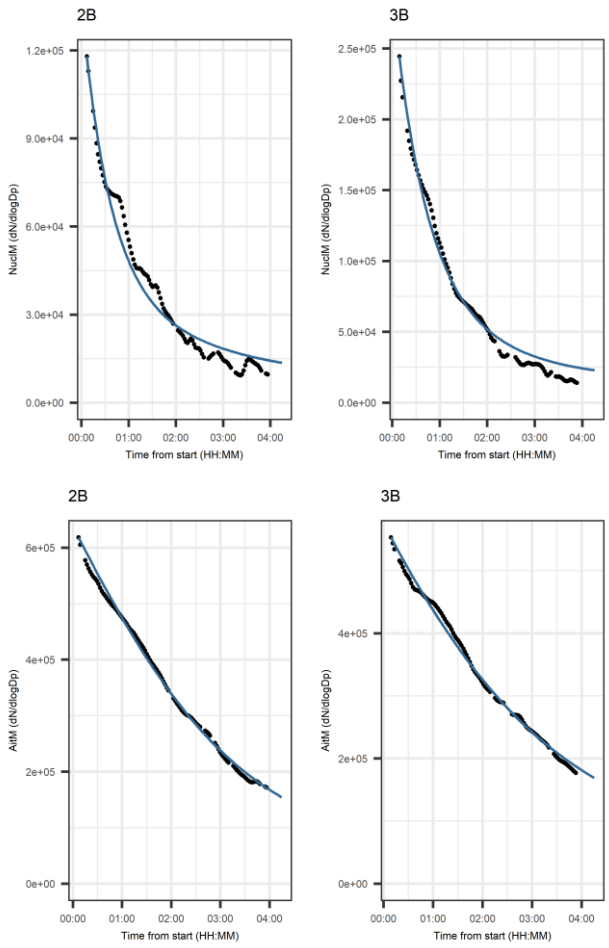

Figure S2 Evolution of size distribution variables in dark aging experiments. Size distribution (Nucleation mode (NuclM, < 25 nm) and Aitken mode (AitM, 25-100 nm)) are from

Scanning Mobility Particle Sizer (SMPS) measurements. Black points represent the filtered version of variable and the blue line shows the modeled evolution.

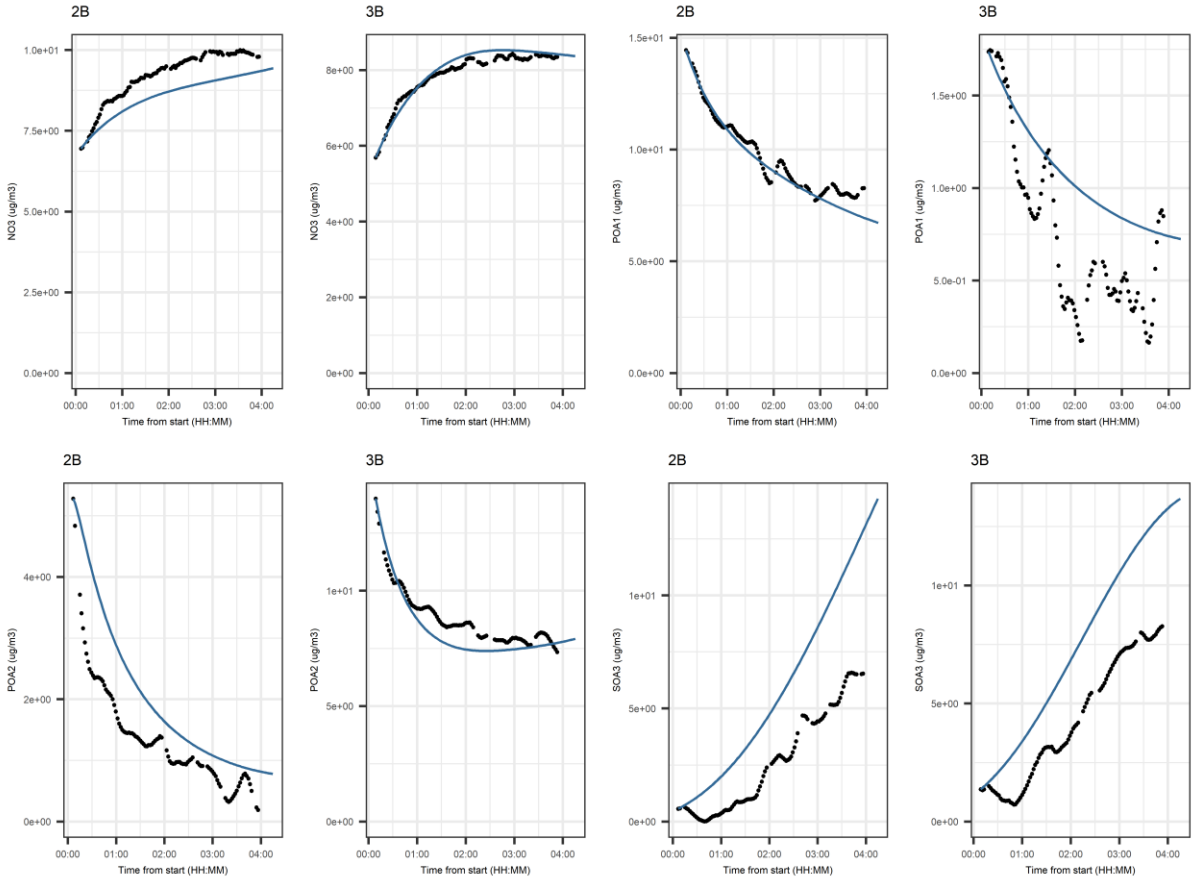

Figure S3 Evolution of nitrate (NO_3) signatured aerosol, primary organic aerosol (POA) factors 1-2, and secondary organic aerosol (SOA) factor 3 in dark aging experiments. All variables are from Aerosol Mass Spectrometer (AMS) measurements. POA factors are representing 1) biomass-burning OA and 2) hydrocarbon-like OA and SOA3 is representing formation by OH radicals. Black points represent the filtered version of variable

and the blue line shows the modeled evolution.

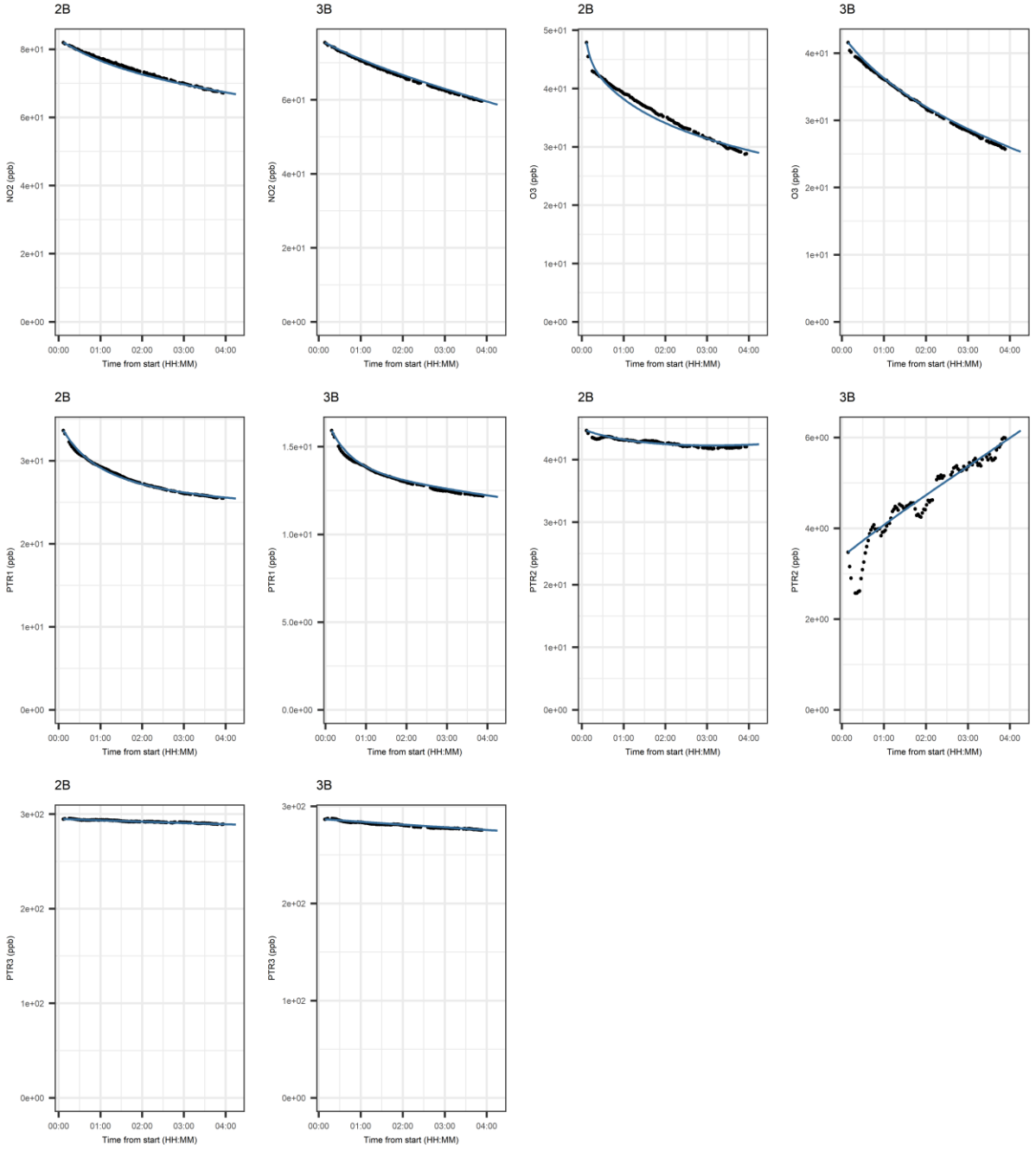

Figure S4 Evolution of gas variables (NO_2, O_3, and PTR factors 1-3) in dark aging experiments. NO_2 and O_3 are from gas analyzers and PTR factors are from Proton-Transfer-Reactor Time-of-Flight Mass Spectrometer (PTR-ToF-MS). PTR factors are representing 1) primary VOCs, 2) photochemical aging products, and 3) dark aging products. Black points represent the filtered version of variable and the blue line shows the modeled evolution.

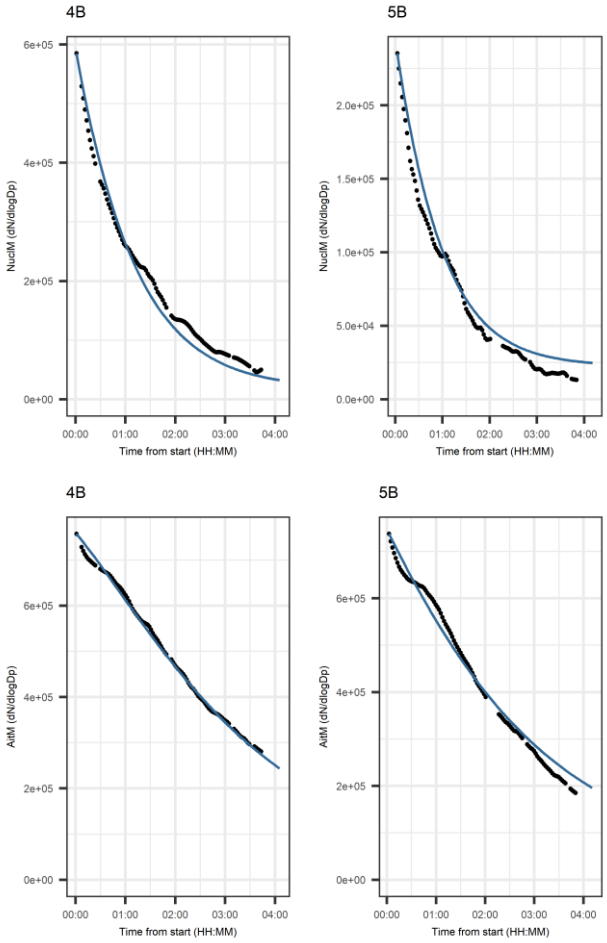

Figure S5 Evolution of size distribution variables in photochemical aging experiments. Size distribution (Nucleation mode (NuclM, < 25 nm) and Aitken mode (AitM, 25-100

nm)) are from Scanning Mobility Particle Sizer (SMPS) measurements. Black points represent the filtered version of variable and the blue line shows the modeled evolution.

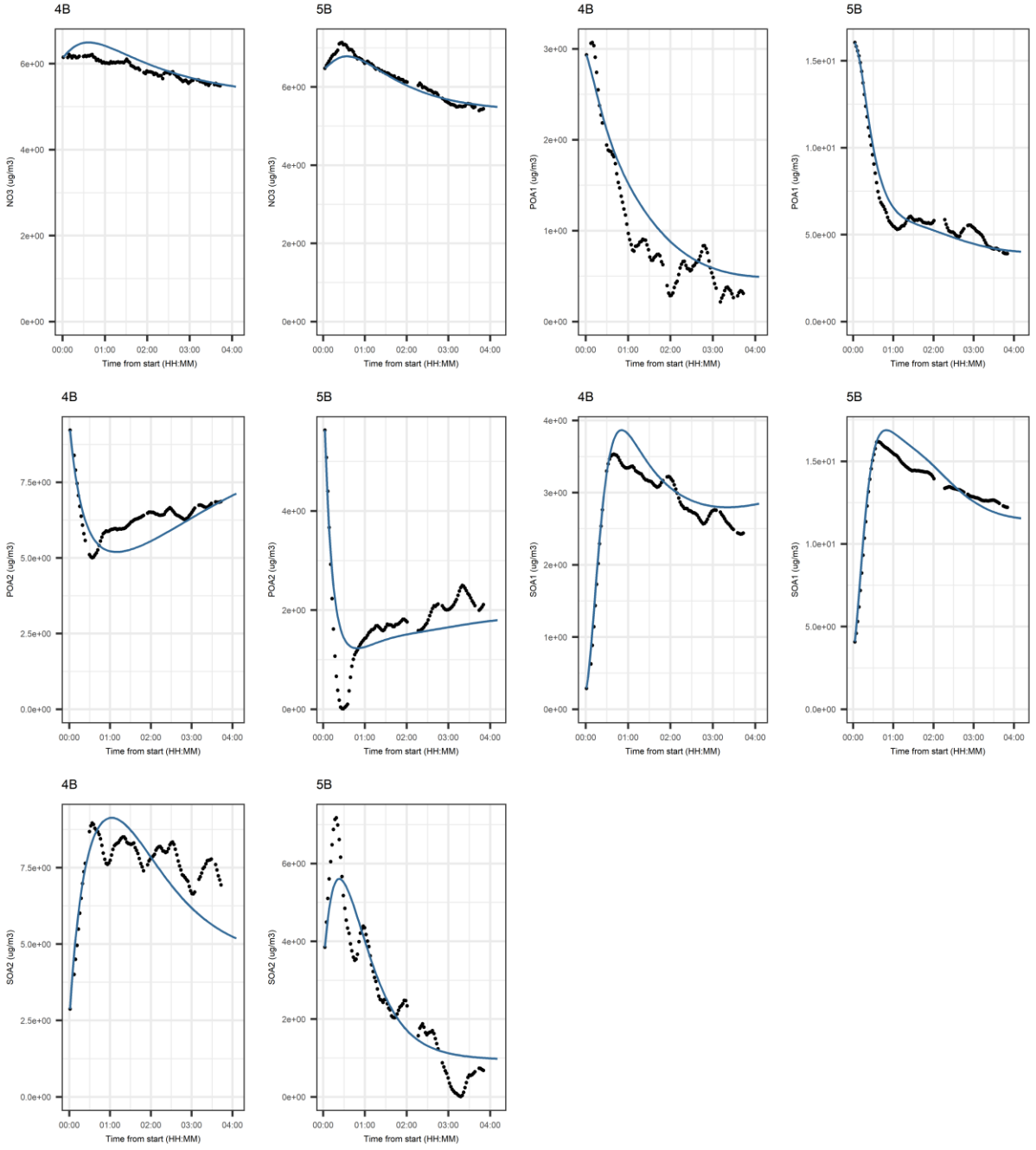

Figure S6 Evolution of nitrate (NO3) signatured aerosol, primary organic aerosol (POA) factors 1-2, and secondary organic aerosol (SOA) factors 1-2 in photochemical aging experiments. All variables are from Aerosol Mass Spectrometer (AMS) measurements. POA factors are representing 1) biomass-burning OA and 2) hydrocarbon-like OA and SOA factors are representing 1) formation by ozonolysis and 2) formation by nitrate/peroxy radicals. Black points represent the filtered version of variable and the blue line shows

the modeled evolution.

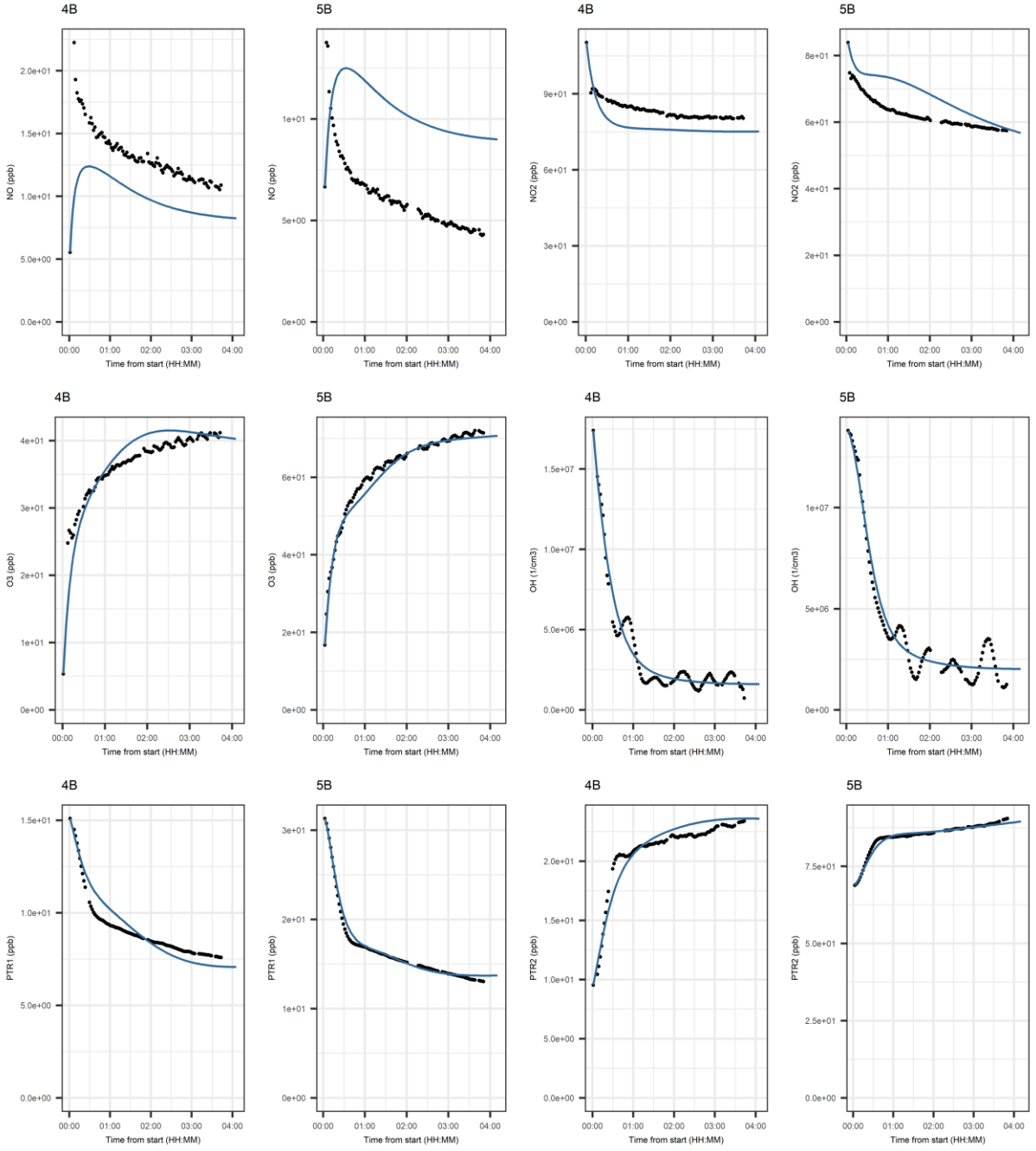

Figure S7 Evolution of gas variables (NO, NO_2, O_3, OH, PTR1, and PTR2) in photochemical aging experiments. NO, NO2 and O_3 are from gas analyzers. OH (derived from d9-butanol measurements), PTR1, and PTR2 are from Proton-Transfer-Reactor Time-of-Flight Mass Spectrometer (PTR-ToF-MS). PTR factors are representing 1) primary VOCs and 2) photochemical aging products. Black points represent the filtered version of variable and the blue line shows the modeled evolution.

Table S8 Coefficients for dark aging experiments. $\Delta(x)$ is the variable which its predictors are affecting. Estimate is the coefficient for the linear differential equation (1) in the main text. Gases (NO, NO2, O3) are measured by gas analyzers and PTR factors (PTR1-3) are from Proton-Transfer-Reactor Time-of-Flight Mass Spectrometer (PTR-ToF-MS) measurements. Particle chemical composition, nitrate (NO3) signatured aerosol and primary and secondary organic aerosol factors (POA1-2, SOA1-3) are derived from Aerosol Mass Spectrometer (AMS) measurements. Particle size distribution variables are representing typical aerosol modes, nucleation (NuclM, < 25 nm), Aitken (AitM, 25-100 nm), accumulation (AccM, 100-300 nm), and coarse (CoarseM, > 300 nm). Size distribution variables are derived from Scanning Mobility Particle Sizer (SMPS) measurements.

Gases

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate
NO2	NO2	-0.0515	PTR1	PTR1*PTR2	0.000213
NO2	O3*NO2	-0.00078	PTR1	PTR3*PTR1	-7.7E-05
NO2	PTR3*O3	0.00026	PTR1	PTR3*O3	7.22E-05
NO2	PTR3*NO2	0.00014	PTR1	PTR1	-6.7E-06
O3	O3	0.00664	PTR2	PTR2	0.00387
O3	NO2*O3	0.00541	PTR2	O3	0.000744
O3	O3*PTR3	-0.00249	PTR2	O3*PTR2	-0.00013
O3	O3*PTR2	-0.00066	PTR2	PTR1*PTR2	-1.7E-05
O3	NO2*PTR3	0.000569	PTR3	PTR2	0.0126
PTR1	PTR1*O3	-0.00111	PTR3	PTR3	-0.00037
PTR1	PTR2*O3	0.000436	PTR3	PTR3*PTR2	-3.8E-05

Particle chemical composition

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate
NO3	NO3	0.0656	POA2	NO2*POA2	-0.00515

NO3	POA2	0.0231	SOA1	SOA1	0.0352
NO3	POA2*NO3	-0.00673	SOA1	NO3*SOA1	0.00939
NO3	SOA2	0.00465	SOA1	SOA2*SOA1	-0.00405
NO3	SOA2*NO3	-0.00226	SOA1	O3*SOA1	-0.00164
NO3	NO3*POA1	-0.00212	SOA1	PTR2*SOA1	0.000556
NO3	POA1	0.000713	SOA1	O3*PTR2	5.62E-05
NO3	POA1*SOA2	0.000185	SOA2	SOA2	-0.0618
NO3	POA2*SOA2	-4.2E-07	SOA2	SOA2*O3	-0.0035
POA1	POA1	0.0327	SOA2	NO2*SOA2	0.00215
POA1	SOA1*POA1	0.00114	SOA2	NO2*O3	0.000308
POA1	NO2*POA1	-0.00062	SOA3	SOA3	-0.0497
POA2	POA2	0.0577	SOA3	SOA3*O3	0.00216
POA2	NO3*POA2	0.0128	SOA3	NO3	0.00145
POA2	O3*POA2	0.00553	SOA3	NO3*SOA3	-0.00029

Particle size

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate
AccM	AccM	-0.0027	CoarseM	AitM*AccM	2.99E-09
AccM	AccM*AitM	1.26E-08	CoarseM	AccM*CoarseM	2.94E-09
AccM	AccM*CoarseM	-1.2E-08	CoarseM	NuclM*CoarseM	-2E-09
AitM	AitM	-0.0124	CoarseM	AitM*CoarseM	3.71E-10
AitM	AccM	0.00142	NuclM	NuclM	0.0197
AitM	CoarseM*AitM	-4.1E-08	NuclM	AitM	0.00214
AitM	AitM*AccM	5.8E-09	NuclM	NuclM*CoarseM	-9.6E-08
CoarseM	CoarseM	-0.00573	NuclM	AitM*NuclM	-7.6E-08

Table S9 Coefficients for photochemical aging experiments. is the variable which its predictors are affecting. Estimate is the coefficient for the linear differen-tial equation (1) in the main text. Gases (NO, NO2, O3) are measured by gas analyzers and PTR factors (PTR1-3) are from Proton-Transfer-Reactor Time-of-Flight Mass Spectrometer (PTR-ToF-MS) measurements. Particle chemical composition, nitrate (NO3) signatured aerosol and primary and secondary organic aerosol factors (POA1-2, SOA1-3) are derived from Aerosol Mass Spectrometer (AMS) measurements. Particle size distribution variables are representing typical aerosol modes, nucleation (NuclM, < 25 nm), Aitken (AitM, 25-100 nm), accumulation (AccM, 100-300 nm), and coarse (CoarseM, > 300 nm). Size distribution variables are derived from Scanning Mobility Particle Sizer (SMPS) measurements.

Gases

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate
NO	NO	-0.0199	OH	PTR3*OH	-0.00055
NO	OH	2.26E-07	PTR1	PTR1	0.0533
NO	NO*OH	-1.6E-08	PTR1	PTR1*O3	-0.00189
NO2	NO2	0.0156	PTR1	PTR2*PTR1	0.000357
NO2	O3*NO2	-0.00036	PTR1	PTR2*O3	0.000121
NO2	OH*O3	6.25E-09	PTR1	PTR2	-8E-05
NO2	OH*NO2	-4E-09	PTR1	OH	3.43E-08
O3	PTR1	0.0969	PTR1	PTR1*OH	-6E-09
O3	O3	-0.0149	PTR2	PTR2	-0.0185
O3	PTR1*O3	0.000421	PTR2	O3	0.00312
O3	OH	1.08E-07	PTR2	O3*PTR2	0.000178
O3	OH*O3	-7.5E-09	PTR2	OH	2.64E-08
O3	OH*PTR1	6.92E-09	PTR2	OH*PTR2	9.9E-10
OH	PTR3	349	PTR2	OH*O3	9.25E-10
OH	PTR3*PTR1	28.9	PTR3	O3	0.0869
OH	PTR3*PTR2	4.11	PTR3	PTR3	0.0107
OH	OH	0.0303	PTR3	O3*PTR3	-0.00058
OH	PTR1*OH	0.00388	PTR3	PTR3*OH	-7.3E-10
OH	PTR2*OH	-0.00094			

Particle chemical composition

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate

NO3	O3	0.00231	SOA1	PTR1*SOA1	0.0034
NO3	NO3	-0.00155	SOA1	O3*SOA1	-0.00149
NO3	O3*NO3	-0.00044	SOA1	O3*PTR1	0.000707
NO3	OH*NO3	4.23E-10	SOA1	SOA1*OH	2.9E-09
NO3	OH*O3	5.68E-11	SOA1	SOA2*OH	7.59E-10
POA1	POA1*NO3	-0.0148	SOA2	SOA2	0.054
POA1	POA1*PTR1	0.00705	SOA2	O3*SOA2	-0.00138
POA1	POA1	-0.00475	SOA2	PTR2*SOA2	-0.00042
POA1	OH*POA1	-1.4E-08	SOA2	OH	4.18E-08
POA1	OH*PTR1	2.47E-09	SOA2	SOA2*OH	-3.3E-09
POA2	NO3	0.0137	SOA3	NO3*SOA3	-0.0358
POA2	POA2	-0.0091	SOA3	O3	0.0275
POA2	POA2*POA1	-0.0091	SOA3	PTR1*SOA3	0.0118
POA2	NO3*POA2	0.00171	SOA3	O3*SOA3	-0.00026
POA2	POA2*OH	-2E-09	SOA3	PTR1*O3	-3.2E-05
SOA1	SOA2*SOA1	-0.0078	SOA3	SOA3*OH	2.31E-09

Particle size

$\Delta(x)$	Predictor	Estimate	$\Delta(x)$	Predictor	Estimate
AccM	CoarseM	0.216	AitM	AitM*NuclM	-5.3E-09
AccM	AccM	-0.0477	CoarseM	CoarseM	-0.00544
AccM	AitM	5.48E-05	CoarseM	AitM*CoarseM	2.06E-08
AccM	NuclM*AitM	2.29E-07	CoarseM	NuclM*CoarseM	-1.8E-08
AccM	NuclM*AccM	-2.1E-07	CoarseM	NuclM*AitM	1.87E-09
AccM	CoarseM*AccM	-9.4E-08	NuclM	NuclM	-0.0194
AccM	AitM*AccM	9.21E-08	NuclM	CoarseM	0.00284
AitM	AitM	-0.0124	NuclM	CoarseM*NuclM	-6.2E-08
AitM	NuclM	0.0123			

References for Supplement

Mikkelsen, F. V.: episode: Estimation with Penalisation in Systems of Ordinary Differential Equations, [online] Available from: https://cran.r-project.org/package=episode, 2017.

Scheines, R. and Ramsey, J.: Measurement Error and Causal Discovery., CEUR Workshop Proc., 1792, 1–7 [online] Available from: http://www.ncbi.nlm.nih.gov/pubmed/28280453 (Accessed 20 February 2019), 2016.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics, edited by J. H. Seinfeld and S. N. Pandis, Wiley., 2016.

Singh, K., Gupta, G., Tewari, V. and Shroff, G.: Comparative Benchmarking of Causal Discovery Techniques, [online] Available from: http://arxiv.org/abs/1708.06246 (Accessed 28 May 2019), 2017.

Soetaert, K., Petzoldt, T. and Setzer, R. W.: Solving Differential Equations in R: Package deSolve, J. Stat. Softw., 33(9), 1–25, doi:10.18637/jss.v033.i09, 2010.