Supplement of Atmos. Chem. Phys., 21, 12595–12611, 2021
https://doi.org/10.5194/acp-21-12595-2021-supplement

Supplement of

Aerosol formation and growth rates from chamber experiments using Kalman smoothing

Matthew Ozon et al.

Correspondence to: Kari E. J. Lehtinen (kari.lehtinen@uef.fi)

S1 Used assumptions for the Extended Kalman Filter

	Quantity	simulated data	H_2SO_4-NH_3	α-pinene ozonolysis	HIO_3
time step	Δ_t [s]	40			1.82
size-	q [-]	16			32
discretization	d_0 [nm]	1.795			1.747
$d_i = d_0 r^{i-1}$	r [-]	1.117			1.057
	$g^{0\|0}$ [nm h^{-1}]	10^{-3}			
	$\Gamma_g^{0\|0}$ [nm^2 h^{-2}]	1			0.25
	σ_g [nm h^{-1}]	1			1.25
growth rate g	T_g [s]	300			60
	ζ_g [-]	0.95			
	δ_g [-]	$0.4l = 6.4$	6.4	6.4	12.8
	a_g [-]	2			
	$\lambda^{0\|0}$ [s^{-1}]	$\lambda_{wall}^{0\|0} = 1.63 \cdot 10^{-3}/d_p$[nm]			
		$\lambda_{dil}^{0\|0} = 1.72 \cdot 10^{-4}$		$\lambda_{dil}^{0\|0} = 1.58 \cdot 10^{-4}$	$\lambda_{dil}^{0\|0} = 1.72 \cdot 10^{-4}$
	$\Gamma_\lambda^{0\|0}$ [s^{-2}]	$\left(0.1\,\lambda^{0\|0}\right)^2$			
loss rate λ	σ_λ [s^{-1}]	10^{-7}			
	δ_λ [-]	$0.1l = 1.6$			0.1=3.2
	a_λ [-]	1			
	$J^{0\|0}$ [cm^{-3} s^{-1}]	0.07			1.39
	$\Gamma_J^{0\|0}$ [cm^{-6} s^{-2}]	0.25			100
nucleation rate J	σ_J [cm^{-3} s^{-1}]	0.5			10
	T_J [s]	300			60
	ζ_J [-]	0.95			
size-distribution	$N^{0\|0}$ [cm^{-3}]	0			
N	$\Gamma_N^{0\|0}$ [cm^{-6}]	$(10\Delta_i)^2$			
measurement		$\Delta H = 0.5 \cdot (\alpha H + \|H(d+\delta_d) - H(d)\|)$			
error model	α [-]	0.2	0.5		
	δ_d [nm]	$2.3 \cdot 10^{-3} d_i$			

Table S1: Used assumptions for the Extended Kalman Filter for all datasets.

S2 Source of errors in the discrete aerosol general dynamics equation (GDE)

The state equations (Eq. (3) and Eq. (4)) include the noise terms $w^k$and e^k, which are approximated as Gaussian errors. These terms include errors due to discretization, model and parameter uncertainties, as the problem is formulated in a discretized space with uncertainties on both the evolution and measurement model as well as uncertainties on the parameters governing the evolution model. The discretization in size and time is illustrated in Fig. S1. In the following, we derive expressions, which approximate these error terms, such that their relative magnitudes can be estimated.

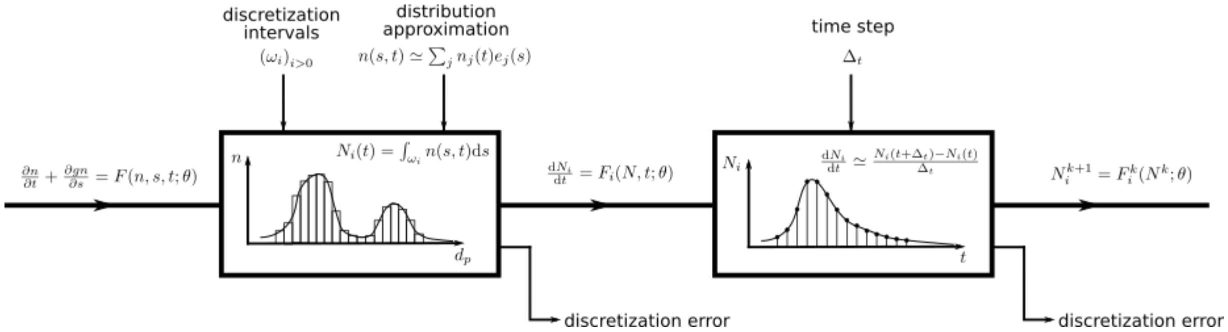

discretization intervals
$(\omega_i)_{i>0}$
distribution approximation
$n(s,t) \simeq \sum_j n_j(t)e_j(s)$
time step
Δ_t
$\frac{\partial n}{\partial t} + \frac{\partial gn}{\partial s} = F(n,s,t;\theta)$
$N_i(t) = \int_{\omega_i} n(s,t)\mathrm{d}s$
n
d_p
$\frac{\mathrm{d}N_i}{\mathrm{d}t} = F_i(N,t;\theta)$
N_i
$\frac{\mathrm{d}N_i}{\mathrm{d}t} \simeq \frac{N_i(t+\Delta_t)-N_i(t)}{\Delta_t}$
t
$N_i^{k+1} = F_i^k(N^k;\theta)$
discretization error
discretization error

Figure S1: Schematic view of the discretization process of the aerosol GDE in the FIKS framework. The discretization is a two-step process, 1) the size discretization by integration of the GDE over many size intervals (or bins) produces a collection of time ordinary differential equations describing the evolution of the concentrations in each size interval, and 2) the time discretization converting the time ordinary differential equations into difference equations (e.g. in the most simple case an Euler scheme, i.e. $\frac{\boldsymbol{dN_i}}{\boldsymbol{dt}} = \frac{\boldsymbol{N_i(t+\Delta_t)-N_i(t)}}{\boldsymbol{\Delta_t}}$).

S2.1 Size discretization

We start with the discretization in size, which could be a potential source of error as illustrated in Fig. 1. We used 16 discretization intervals from 1-10 nm for all presented results, except for the iodic acid experiment where we used 32 discretization intervals. The discretization in size affects the kernel functions (H^k; Fig. 1) but also the size distribution (N_i^k), the evolution model ($F(X^k)$), and the process parameters (g^k, λ^k, J^k).

S2.1.1 Size distribution approximation

The usual way to discretize the time-size population balance equations for aerosols includes as first step an integration over size intervals to get a system of time-ordinary-differential-equations. The discretization is introducing errors whose amplitude depends mostly on the coarseness of the discretization grid. An illustrative sketch of this discretization procedure is given in Fig. S2. The simplified practical relation between the approximated size-distribution $\hat{n}$ from the discretized concentrations $(N_i)_{i>0}$ is given by:

$$\hat{n}(d,t) \triangleq \sum_{i=1}^{M} \frac{N_i}{\Delta_i} I_{\omega_i}(d) \tag{S1}$$

where the rectangular function $I_{\omega_i}(d)$ is defined as 1 over the interval ω_i and 0 anywhere else. The higher order terms $O((d-d_i)^2)$ can be neglected for small size intervals where $d \sim d_i$. Therefore, the true size distribution can be locally approximated by a linear function shown by the blue solid line in Fig. S1. Now, we formulate the error $\varepsilon_i^N(d)$ between the true and the approximated size distribution for each interval ω_i:

$$\varepsilon_i^N(d) \ = n(d) - \hat{n}(d) = n(d_i) + \frac{\partial n}{\partial d}(d_i)(d-d_i) + O((d-d_i)^2) - \frac{N_i}{\Delta_i} \tag{S2}$$

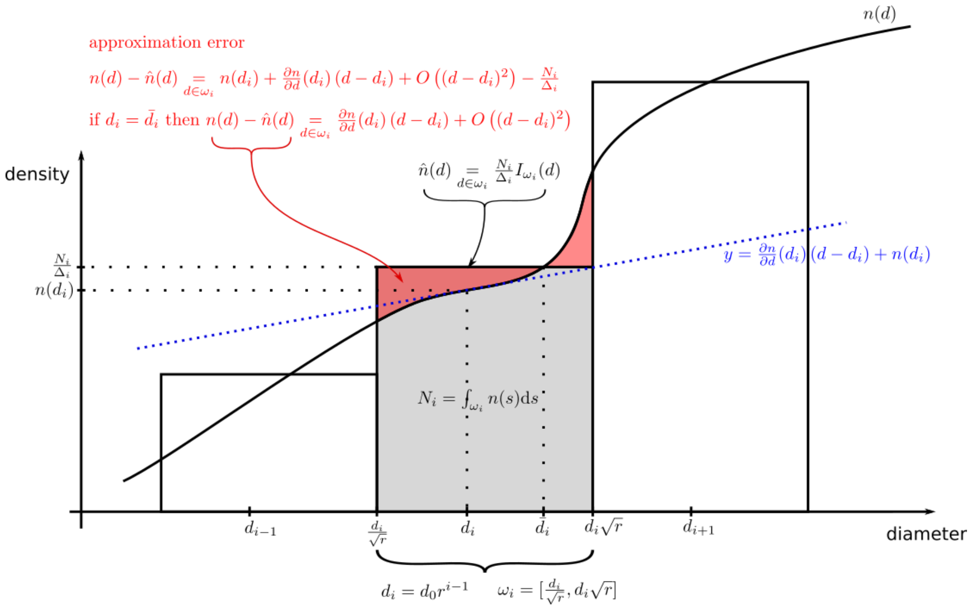
where O represents all terms of higher order and the time variable is dropped for clarity. If the size distribution at the centroid diameter $n(d_i)$ is assumed to be the mean value $\frac{N_i}{\Delta_i}$, then Eq. (3) can be simplified. As depicted on Fig. S2 the size distribution must take the mean value at least once over the size interval ω_i (at $d = \bar{d}_i$ in Fig. S1). We integrate the error term over the size-interval, which is by definition 0 (as the Taylor Series is an exact approximation if it includes all higher order terms), and obtain:

$$0 \stackrel{\text{def}}{=} \int_{\omega_i} \varepsilon_i^N(s)ds = n(d_i)\Delta_i + \frac{\partial n}{\partial d}(d_i)\Delta_i\left(\frac{d_i\sqrt{r}+\frac{d_i}{\sqrt{r}}}{2} - d_i\right) + O(\Delta_i^3) - N_i \tag{S3}$$

which allows for the definition of the error ε_i^N by expressing the approximation $n(d_i)\Delta_i$ through the Taylor series:

$$\varepsilon_i^N = N_i - n(d_i)\Delta_i = \frac{\partial n}{\partial d}(d_i)\frac{\left(\sqrt{r}-1\right)^2}{r-1}\frac{\Delta_i^2}{2} + O(\Delta_i^3) \tag{S4}$$

It can be seen that the error approaches 0, if r approaches 1, which means that the error is getting smaller as the size intervals shrinks.



$n(d)$
approximation error
$n(d) - \hat{n}(d) \underset{d\in\omega_i}{=} n(d_i) + \frac{\partial n}{\partial d}(d_i)\,(d-d_i) + O\left((d-d_i)^2\right) - \frac{N_i}{\Delta_i}$
if $d_i = \bar{d}_i$ then $n(d) - \hat{n}(d) \underset{d\in\omega_i}{=} \frac{\partial n}{\partial d}(d_i)\,(d-d_i) + O\left((d-d_i)^2\right)$
density
$\hat{n}(d) \underset{d\in\omega_i}{=} \frac{N_i}{\Delta_i} I_{\omega_i}(d)$
$y = \frac{\partial n}{\partial d}(d_i)(d-d_i) + n(d_i)$
$\frac{N_i}{\Delta_i}$
$n(d_i)$
$N_i = \int_{\omega_i} n(s)\mathrm{d}s$
d_{i-1}
$\frac{d_i}{\sqrt{r}}$
d_i
$\bar{d}_i$
$d_i\sqrt{r}$
d_{i+1}
diameter
$d_i = d_0 r^{i-1}$
$\omega_i = [\frac{d_i}{\sqrt{r}}, d_i\sqrt{r}]$

Figure S2: Representation of the sources of errors in the discretization of the size-distribution. The black curve is the true size-distribution $\boldsymbol{n}$ and the black rectangles represent the zeroth order approximation $\hat{\boldsymbol{n}}$. The red shaded area indicates the absolute error over the size interval $\boldsymbol{\omega_i}$, while the blue dashed line gives the first order Taylor approximation of $\boldsymbol{n}$ around the centroid diameter $\boldsymbol{d_i}$ of the $\boldsymbol{i^{th}}$ bin. $\bar{\boldsymbol{d}}_{\boldsymbol{i}}$ is the diameter for which the true size-distribution $\boldsymbol{n}$ intersects with its approximation $\hat{\boldsymbol{n}}$, which is not necessarily equal to the bin's centroid $\boldsymbol{d_i}$ and is potentially not unique.

S2.1.2 GDE size-integration using the size distribution approximation

The size discretization of the size-distribution will also influence the evaluation of the general dynamic equation (GDE), i.e. the evolution model $F(X^k)$. Our exemplary derivation of the error term resulting from the discretization of the GDE only includes condensation/evaporation growth (g) and linear losses (λ) terms, but neglects the coagulation terms. The integral form of the GDE for each size interval ω_i is given by:

$$\int_{\omega_i}\left(\frac{\partial n}{\partial t}(s,t) + \frac{\partial gn}{\partial s}(s,t)\right)ds = -\int_{\omega_i}\lambda(s,t)n(s,t)ds \tag{S5}$$

Again dropping the time variable for clarity and by assuming well-behaved functions for the left-hand side term, namely the differentiation under the integral sign, we obtain:

$$\frac{dN_i}{dt} + g(d_i^+)n(d_i^+) - g(d_i^-)n(d_i^-) = -\int_{\omega_i}\lambda(s)n(s)ds \tag{S6}$$

where $d_i^+ = d_i\sqrt{r}$ and $d_i^- = \frac{d_i}{\sqrt{r}}$ are the lower and upper end of the size interval. Using the approximation Eq. (S4):

$$\frac{dN_i}{dt} + \frac{g(d_i^+)N_i}{\Delta_i} - \frac{g(d_i^+)\varepsilon_i^N}{\Delta_i} - \frac{g(d_i^-)N_{i-1}}{\Delta_{i-1}} + \frac{g(d_i^-)\varepsilon_{i-1}^N}{\Delta_{i-1}} = -\int_{\omega_i}\lambda(s)n(s)ds \tag{S7}$$

The right-hand side is approximated by a first-order expansion and by introducing the mean value of the linear losses in the size interval $\bar{\lambda}_i = \frac{1}{\Delta_i}\int_{\omega_i}\lambda(s)ds$, we obtain:

$$\int_{\omega_i}\lambda(s)n(s)ds = n(d_i)\int_{\omega_i}\lambda(s)ds + \int_{\omega_i}\lambda(s)\left(\frac{\partial n}{\partial s}(d_i)(s-d_i) + O((s-d_i)^2)\right)ds$$

$$= n(d_i)\int_{\omega_i}\lambda(s)ds + \lambda(d_i)\frac{\partial n}{\partial d}(d_i)\int_{\omega_i}(s-d_i)ds + \int_{\omega_i}O\left((s-d_i)^2\right)ds$$

$$= n(d_i) \int_{\omega_i} \lambda(s)ds + \lambda(d_i) \frac{\partial n}{\partial d}(d_i) \frac{\left(\sqrt{r}-1\right)^2}{r-1} \frac{\Delta_i^2}{2} + O(\Delta_i^3)$$

$$= n(d_i)\bar{\lambda}_i \mathit{\Delta}_i + \lambda(d_i) \frac{\partial n}{\partial d}(d_i) \frac{(\sqrt{r}-1)^2}{r-1} \frac{\mathit{\Delta}_i^2}{2} + O(\mathit{\Delta}_i^3) \quad \text{(S8)}$$

by substituting $n(d_i)$ through the approximation of Eq. (S4):

$$\int_{\omega_i} \lambda(s)n(s)ds = N_i\bar{\lambda}_i + \frac{\partial n}{\partial d}(d_i) \frac{(\sqrt{r}-1)^2}{r-1} \frac{\Delta_i^2}{2} (\lambda(d_i) - \bar{\lambda}_i) + O(\Delta_i^3). \quad \text{(S9)}$$

Observing that $\lambda(d_i) - \bar{\lambda}_i = \frac{\partial \lambda}{\partial d}(d_i)(\bar{d}_i - d_i) + O((\bar{d}_i - d_i)^2)$ and that $\bar{d}_i \in \omega_i$ ($|\bar{d}_\iota - d_i| \leq \Delta_i$), then:

$$\int_{\omega_i} \lambda(s)n(s)ds = N_i\lambda(d_i) - N_i \frac{\partial \lambda}{\partial d}(d_i)(\bar{d}_i - d_i) + O(\Delta_i^2) \quad \text{(S10)}$$

Now, putting back together all the terms:

$$\frac{dN_i}{dt} + \frac{g(d_i^+)N_i}{\mathit{\Delta}_i} - \frac{g(d_i^-)N_{i-1}}{\mathit{\Delta}_{i-1}} = -N_i\lambda(d_i) + W_i^{GDE} \quad \text{(S11)}$$

with

$$W_i^{\mathrm{GDE}} = N_i \frac{\partial \lambda}{\partial d}(d_i)(\bar{d}_i - d_i) + \frac{g(d_i^+)\varepsilon_i^N}{\Delta_i} - \frac{g(d_i^-)\varepsilon_{i-1}^N}{\Delta_{i-1}} + O(\Delta_i^2) \quad \text{(S12)}$$

The discretization error of the size-distribution evolution model (i.e. the GDE; W_i^{GDE}) can be written using the explicit form of ε_i^N, N_i (Eq. (S4)) and summarizing all terms of order Δ_i^2 leads to:

$$W_i^{\mathrm{GDE}} = \left(g(d_i^+) \frac{\partial n}{\partial d}(d_i) - \frac{1}{r} g(d_i^-) \frac{\partial n}{\partial d}(d_{i-1})\right) \frac{(\sqrt{r}-1)^2}{2(r-1)} \Delta_i + O(\Delta_i^2) \quad \text{(S13)}$$

From Eq. (S13) we see that the discretization error is of first order error whose linear coefficient depends on the size distribution, its first derivative and the growth rate. Note, that the linear loss terms are of second order in Δ_i and hence not considered further. If the order of magnitude of the error needs to be estimated, these values need to be roughly estimated. Altogether, it is obvious that if $r \to 1$, $\Delta_i \to 0$ and the error collapses.

S2.2 Time discretization and estimate of Γ_N^k

The time discretization of the GDE corresponds to the second panel of the block diagram in Fig. S1. The GDE is described as a system of ordinary differential equations:

$$\frac{dN_i}{dt} + \frac{g(d_i^+)}{\Delta_i} N_i - \frac{g(d_i^-)}{\Delta_{i-1}} N_{i-1} = -N_i\lambda(d_i) + W_i$$

This set of equations is discretized into difference equations for the time series $(t_k)_{k\geq 0}, t_k = t_{k-1} + \delta^k$. We use the notations $N_i^k = N_i(t_k)$, $g_i^k = g(d_i^+, t_k)$, $\lambda_i^k = \lambda(d_i, t_k)$, $W_i^{\mathrm{k}} = W_i(t_k)$ for all i and k, with $g_0^k = g(d_1^-, t_k)$. Using an explicit Euler scheme, we obtain the size-and-time discretized GDE:

$$N_i^{k+1} = N_i^k + \delta^k \left(\frac{g_{i-1}^k}{\Delta_{i-1}} N_{i-1}^k - \frac{g_i^k}{\Delta_i} N_j^k - \lambda_i^k N_i^k\right) + (w_N^k)_i \quad \text{(S14)}$$

with $(w_N^k)_i = \delta^k W_i^k + O((\delta^k)^2)$ the total error of the GDE part of the evolution model in Eq. (3), which is used as the estimate for Γ_N^k in the prediction step of the Extended Kalman Filter (Table 1, Algorithm 1).

Overall, we can conclude that the error due to the discretization of the GDE (time and size) is proportional to both, the size step and the time step, with a proportionality coefficient depending on the first derivative of the size distribution and the growth rate. From this set of difference equations, we can also easily infer a stability (non-divergent) and non-oscillatory criteria for fine size-and-time discretization (Gottlieb and Shu, 1998):

$$\forall i, k, \delta^k (\frac{g_i^k}{\Delta_i} + \lambda_i^k) \leq \frac{1}{2}. \quad \text{(S15)}$$

Hence, the discretized GDE in Eq. (S14) can only lead to physically meaningful solutions if the overall growth is slow enough for the chosen discretization grid. Consequently, if numerical diffusion is minimized by choosing a fine size discretization grid, also a proper time discretization grid satisfying the criteria in Eq. (S15) needs to be chosen. However, for a fine enough time discretization there might be not enough available measurements l. We therefore use a zero-padding technique to emulate a better time resolution, although the measurement operator has to change in time accordingly. The operator H^k should is set to the null operator (a matrix full of zeros) for the instants when the data have been set to zero. This way, the measurement update is non-informative and the estimates only rely on the model. In other words, the evolution model is computed several times for each actual instant of the dataset.

S2.3 Measurement operator discretization

Last, also the observation model includes an error term v^k, which apart from the measurement noise (first dropped here for simplicity), includes errors of the measurement operator. In the continuous case, the measured counts for channel i are given by the equation:

$$C_i^\star = \int_0^\infty H_i^\star(s;\theta)n(s)ds \tag{S16}$$

for the continuous size distribution n and channel efficiency $H_i^\star$ with parameters θ. However, the modeled counts C_i suffer from errors in the measurement model, which originate from either a discretization error $\varepsilon_i^{\mathrm{H}}$ or wrong assumptions in the measurement model e_i^{H}. Formally, the discrepancy between the true and the modeled counts can be written as:

$$C_i^\star - C_i = \varepsilon_i^{\mathrm{H}} + e_i^{\mathrm{H}} = \int_0^\infty H_i^\star(s;\theta)n(s) - H_i(s;\theta)\hat{n}(s)ds \tag{S17}$$

Using Eq. (S1) and $H_i^\star = H_i + \Delta H_i$, we obtain for the discretization error:

$$\varepsilon_i^{\mathrm{H}} = \sum_{j=1}^{M} \int_{\omega_j} (\frac{\partial n}{\partial s}(d_j)(s-d_j) + O((s-d_j)^2))\, H_i^\star(s;\theta)ds \tag{S18}$$

This can further be expanded, using the Taylor expansion, $H_i^\star(s;\theta) = H_{i,j}^\star + O(d-\bar{d}_j)$ around $\bar{d}_j$, which is the value of the diameter for which $H_i^\star(\bar{d}_j;\theta) = H_{i,j}^\star$:

$$\varepsilon_i^{\mathrm{H}} = \sum_{j=1}^{M} \frac{\partial n}{\partial d}(d_j)\frac{(\sqrt{r}-1)^2}{r-1}\frac{\Delta_j^2}{2} H_{i,j}^\star + O(\Delta_j^3) \tag{S19}$$

Similarly, the modeling error can be expressed as:

$$e_i^{\mathrm{H}} = \sum_{j=1}^{M} n(d_j) \int_{\omega_j} \Delta H_i(s;\theta)ds \tag{S20}$$

Rearranging the terms to form the total error $v_i = \varepsilon_i^{\mathrm{H}} + e_i^{\mathrm{H}} + \Delta y_i$ (including now the measurement noise Δy_i), and using the notation $\Delta\Psi_{i,j}$ for the averaged value of the operator modeling error, one gets:

$$v_i = \Delta y_i + \sum_{j=1}^{M} n(d_j)\Delta H_{i,j}\Delta_j + \frac{\partial n}{\partial d}(d_j)\frac{\left(\sqrt{r}-1\right)^2}{r-1}\frac{\Delta_j^2}{2} H_{i,j}^\star + O\left(\Delta_j^3\right) \tag{S21}$$

It can be seen that the numerical value v_i is either of order 1 or 2 in Δ_j. If the measurement model is perfectly known, the error is of order 2, but if the measurement model is uncertain, then the error becomes a first order — making the discretization error negligible and hence justifying the choice of neglecting the measurement operator discretization error in the used analysis. This estimate of v_i is then used to compute the covariance Γ_v^k in the Kalman gain matrix (see Eq. (10) and Eq. (11)) of the Extended Kalman Filter (Table 1, Algorithm 1).