



Singular vector decomposition for sensitivity analyses of tropospheric chemical scenarios

N. Goris^{1,2,*} and H. Elbern^{1,2}

¹Institute of Energy and Climate Research, IEK-8, Forschungszentrum Jülich, Jülich, Germany

²Rhenish Institute for Environmental Research at the University of Cologne, Cologne, Germany

* now at: Geophysical Institute at the University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

Correspondence to: N. Goris (nadine.goris@gfi.uib.no)

Received: 13 May 2011 – Published in Atmos. Chem. Phys. Discuss.: 16 June 2011

Revised: 26 March 2013 – Accepted: 26 March 2013 – Published: 15 May 2013

Abstract. Observations of the chemical state of the atmosphere typically provide only sparse snapshots of the state of the system due to their insufficient temporal and spatial density. One possibility for optimisation of the state estimate is to target the observation of those parameters that have the largest potential of resulting in forecast improvements. In the present work, the technique of singular vector analysis is applied to atmospheric chemical modelling in order to identify the most sensitive chemical compounds during a given time period and prioritise them for measurement. Novel to the current work is the fact that, in the application of singular vector analysis, not only the initial values but also the emissions are considered as target variables for adaptive observation strategies. This specific application of singular vector analysis is studied in the context of a chemistry box model allowing for validation of its new features for two chemical regimes. The time and regime dependence of the ozone (O_3) and peroxyacetyl nitrate (PAN) formation potential of individual volatile organic compounds (VOCs) is investigated. Results show that the combined sensitivity of O_3 and PAN to individual VOCs is strongly dependent on the photochemical scenario and simulation interval used. Particularly the alkanes show increasing sensitivities with increasing simulation length. Classifying the VOCs as being of high, medium, little or negligible importance for the formation of O_3 and PAN allows for the identification of those VOCs that may be omitted from measurement. We find that it is possible to omit 6 out of 18 VOCs considered for initial value measurement and 4 out of 12 VOCs considered for emission measurement. The omission of these VOCs is independent of photochemical regime and simulation length. The VOCs selected for

measuring account for more than 96 % and 90 % of the O_3 and PAN sensitivity to VOCs, respectively.

1 Introduction

Typically measurements of the earth's environmental variables have sparse temporal and spatial density and hence provide only incomplete snapshots of the state of the system. Consequently, measurement configurations are sought to optimally garner information from the available monitoring capabilities.

The optimal targeting of observations is a topic that is frequently revisited in the field of numerical weather prediction. Here, adaptive observations of sensitive areas can reduce uncertainty and decrease forecast errors (Buizza et al., 2007). One approach to identifying sensitive system states is through the method of singular vector analysis by which the directions of fastest perturbation growth per finite time interval are determined. The application of singular vector analysis to numerical weather prediction was introduced by Lorenz (1965), who used the method to estimate the predictive capabilities of an idealised atmospheric model by computing the largest error growth. Due to its high computational cost, the method of singular vector analysis was not readily applied to realistic meteorological problems before the late 1980s. Since the singular vectors associated with the largest singular values contain the directions of fastest error growth (Buizza and Palmer, 1993), they are applied as reasonable tools to initialise ensemble forecasts. Their successful use in the ECMWF Ensemble Prediction System resulted in the

first application of targeted singular vectors in a field campaign (Montani et al., 1999). Several other field campaigns followed, including NORPEX (North Pacific Experiment), CALJET (California Land-falling Jets Experiment), the Winter Storm Reconnaissance Programs (WSR99/WSR00) and NATReC (North Atlantic THORPEX Regional Campaign). Buizza et al. (2007) investigated the results of the latter campaigns and stated that targeted observations lead to a larger reduction in forecast error than observations taken in random areas. However, the forecast improvements strongly depend on underlying regions, seasons, static observing systems, and prevailing weather regimes. In the present work, the successful use of singular vector analysis as an analytical tool in meteorology has prompted the extension of its use to chemical modelling.

In atmospheric chemistry, studies focusing on targeted observations are rare. The earliest stimulus for analysing uncertainties of the chemical composition was provided by Khatatov et al. (1999). By investigation of the linearised model these authors inferred that 9 linear combinations of the initial species' concentrations are sufficient to adequately forecast the concentrations of the complete set of 19 simulated species 4 days later. Since most instruments measure concentrations of individual species, the determination of linear combinations has only limited practical value. Yet, the work of Khatatov et al. (1999) inspired further examination of the sensitivity of the initial chemical composition. Liao et al. (2006) used singular vectors to estimate optimal adaptive measurements for chemical compounds. In this manner, application of the results to measurement strategies is feasible, as already demonstrated in the meteorological campaigns mentioned above. However, Liao et al. (2006) focused on the optimal placement of observation locations. While optimal monitoring locations of the standard parameter set wind, temperature, and humidity are of special interest in meteorology, singular vector analysis for atmospheric chemistry differs in at least two ways:

1. given the multitude of parameters impacting the chemical forecast, it is of special interest not only to identify optimal observation locations but also to identify a list of chemical compounds which have to be observed with preference; and
2. initial values are not the only important parameters in chemical modelling since emissions have a similarly large impact on forecast accuracy. While adaptive observations of initial values become less important with growing simulation length, the effect of emission rates on the final concentration increases.

Singular vector analysis accomplishes the identification of measurement priorities by detecting the most sensitive species (here equated to the fastest-growing uncertainties). Therefore, the first objective of the present work is the singular-vector-based sensitivity analyses of specific photo-

chemical scenarios, while the second objective is the generation of sensitivity-based measurement strategies.

Common sensitivity analysis methods for chemical box models are the "brute force" approach, adjoint sensitivity, and the direct decoupled method (DDM). While the brute force approach is simple to implement and computationally inexpensive, it becomes infeasible for systems with a large number of input parameters. The implementation of adjoint sensitivity and DDM is technically challenging, but both methods provide greater computational efficiency and accuracy (Dunker et al., 2002). The brute force approach, adjoint sensitivity and DDM all compute first-order local concentration sensitivity coefficients. In contrast, singular vector analysis provides the most sensitive variables for a given target function (here, the target function maximises the length of the perturbation vector at final time). While adjoint sensitivities are calculated for fixed final perturbations, the final perturbations of singular vector analysis are subject to optimisation. Hence, the singular vector sensitivities equal qualitatively the adjoint sensitivities if merely one final variable is taken into account. Considering multiple final variables, the results of the singular vector analysis are dominated by those final variables allowing for the largest perturbations. This is not necessarily the case for adjoint sensitivities. Since we aim to find the chemical species that cause the largest error growth in a chemical forecast (in order to reduce the error by conducting adaptive observations), singular vector analysis is our method of choice. Singular vector analysis is furthermore a well-proven approach that allows keeping the dimension of the problem flexible, so that the anticipated later application to a three-dimensional chemistry transport model is feasible in a computationally efficient way.

As a prequel to identifying measurement priorities by a fully coupled 3-dimensional chemical transport model (CTM), the present work considers the problem of measurement optimisation in a zero-dimensional model set-up. This box model context has the advantage of permitting an isolation of specific photochemical environments. Furthermore, the reduced numerical effort (CPU time) of the box model allows for a comprehensive investigation of special features of interest, like the singular vector dependence on varying time interval lengths (which is due to ageing of air over days) and varying initial times (i.e. the dependence on the placement of the start time in the diurnal cycle). For this purpose, two tropospheric scenarios, representing moderately polluted and polluted environments, are taken as base cases for studying the influence of different volatile organic compounds (VOCs) on the air pollution caused by the secondary pollutants ozone (O_3) and peroxyacetyl nitrate (PAN). Various studies have previously examined the formation potential of VOCs (e.g. Carter, 1994; Hakami et al., 2004; Kirchner et al., 2001). However, these studies focused mainly on the ozone formation potential, and the aspect of its time dependence is often disregarded. One of the few studies examining both the ozone and PAN formation potential of VOCs was

conducted by Derwent et al. (1998), who utilised the Master Chemical Mechanism (MCM) (Jenkin et al., 1997) to examine the photochemical ozone creation potential (POCP) and the photochemical PAN creation potential (PPCP) over several days. Despite performing model simulations for multiple days, the time dependence of POCP and PPCP was not explicitly investigated. Butler et al. (2011) calculated the tagged ozone production potential (TOPP) for a subset of the Master Chemical Mechanism and specifically focused on the TOPP progression over timescales of multiple days. Moreover, with the help of a tagging approach, they were able to backtrack the effect of VOC intermediates. In contrast, we apply the singular-vector-based sensitivities to determine a time-independent measurement priority of VOCs. Focus has been placed on polluted chemical regimes, since these environments are characterised by high population densities, and enhancements in air pollution forecasting (due to adaptive observations) are more likely to benefit human health. The critical VOCs identified thus can furthermore be of assistance in emission reduction policies. The experimental set-up has been restricted to only two final secondary pollutants in order to allow for validation of the singular vector technique.

The paper is organised as follows: the theory of singular vector analysis is presented in Sect. 2, in which the application to initial uncertainties and emission factors is also described. In Sect. 3, the model set-up is specified. The application of the zero-dimensional model to both tropospheric scenarios is described in Sect. 4. The results of this work are summarised and discussed in Sect. 5 and, finally, concluded in Sect. 6.

2 Singular vector analysis

By applying singular vector analysis to a forecast model, sensitive system states are identified. These sensitive system states are characterised by their ability to lead to significant forecast changes despite being only slightly perturbed. The leading singular vector reveals the direction of fastest perturbation growth during a finite time interval.

In the present work, singular vector analysis is applied to atmospheric chemical modelling to study the influence of initial concentrations and emissions on the temporal evolution of chemical compounds. These two parameters are chosen, since they both strongly determine the system's evolution. Meteorological fields, deposition velocities, and boundary conditions are other parameters to which the evolution of chemical species is sensitive, but these are beyond the scope of this study.

2.1 Uncertainties of initial values

Deterministic chemical forecasts propagate the concentrations of chemical species $\mathbf{c} \in \mathbb{R}^n$ (denoted in mass mixing ratios) forward in time. With \mathcal{M}_{t_1, t_F} denoting the model op-

erator starting at initial time, t_1 , and ending at final time, t_F , the model solution reads

$$\mathbf{c}(t_F) = \mathcal{M}_{t_1, t_F} \mathbf{c}(t_1). \quad (1)$$

Chemical models rely on initial values. These initial values have errors or uncertainties, $\delta\mathbf{c}(t_1)$. The problem of finding the most unstable initial uncertainty can be envisaged as a search of the phase space direction for the $\delta\mathbf{c}(t_1)$ that results in the maximum error growth.

Assuming the initial errors to be sufficiently small to evolve linearly within a given time interval, the evolution of initial uncertainties can be described with the tangent-linear model, \mathbf{L}_{t_1, t_F} :

$$\delta\mathbf{c}(t_F) = \mathbf{L}_{t_1, t_F} \delta\mathbf{c}(t_1). \quad (2)$$

The ratio between perturbation magnitudes at final time, t_F , and initial time, t_1 , can be used to define a measure of error growth, $g(\delta\mathbf{c}(t_1))$:

$$g(\delta\mathbf{c}(t_1)) := \frac{\|\delta\mathbf{c}(t_F)\|_2}{\|\delta\mathbf{c}(t_1)\|_2} \quad (3)$$

(see Liao et al. (2006) for a comprehensive discussion). Maximising this ratio with respect to the initial disturbance, $\delta\mathbf{c}(t_1)$, provides the direction of maximal error growth, $\delta\mathbf{c}_1(t_1)$. Since $g(\delta\mathbf{c}(t_1)) \geq 0$, the initial perturbation, $\delta\mathbf{c}_1(t_1)$, that maximises the squared error growth, $g^2(\delta\mathbf{c}(t_1))$, also maximises the error growth, $g(\delta\mathbf{c}(t_1))$. For convenience, the squared error growth is henceforth maximised:

$$\max_{\delta\mathbf{c}(t_1) \neq 0} g^2(\delta\mathbf{c}(t_1)) = \max_{\delta\mathbf{c}(t_1) \neq 0} \frac{\delta\mathbf{c}(t_F)^T \delta\mathbf{c}(t_F)}{\delta\mathbf{c}(t_1)^T \delta\mathbf{c}(t_1)}. \quad (4)$$

Using Eq. (2) the variable $\delta\mathbf{c}(t_F)$ may be eliminated from the squared error growth, to yield

$$g^2(\delta\mathbf{c}(t_1)) = \frac{\delta\mathbf{c}(t_1)^T \mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F} \delta\mathbf{c}(t_1)}{\delta\mathbf{c}(t_1)^T \delta\mathbf{c}(t_1)}. \quad (5)$$

Here, \mathbf{L}_{t_1, t_F}^T denotes the adjoint model of the tangent-linear operator, \mathbf{L}_{t_1, t_F} . The operator $\mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F}$ is also known as the Oseledec operator. Closer inspection shows that it is symmetric, and that the ratio given by Eq. (5) is a Rayleigh quotient. Applying Rayleigh's principle to problem (4) results in a search for the largest eigenvalue, λ_1 , and its associated eigenvector, $\mathbf{v}_1(t_1)$, of the following eigenvalue problem:

$$\mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F} \mathbf{v}(t_1) = \lambda \mathbf{v}(t_1). \quad (6)$$

Since the entire set of eigenvectors, $\mathbf{v}_i(t_1)$, of $\mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F}$ can be chosen to form an orthonormal basis in the n -dimensional tangent space of linear perturbations, the eigenvectors, $\mathbf{v}_i(t_1)$, $i=2, \dots, n$, define secondary directions of instability (Palmer, 1995). The influence of eigenvector $\mathbf{v}_i(t_1)$ is quantified by the magnitude of the square root of its associated eigenvalue,

λ_i . The term singular vector analysis refers to the fact that the square roots of the eigenvalues, λ_i , of $\mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F}$ are the singular values, σ_i , of the tangent-linear model \mathbf{L}_{t_1, t_F} . The associated left and right singular vectors, $\mathbf{u}_i(t_F) \in \mathbb{R}^n$ and $\mathbf{v}_i(t_1) \in \mathbb{R}^n$, of the operator \mathbf{L}_{t_1, t_F} are defined as satisfying the following conditions:

$$\mathbf{L}_{t_1, t_F} \mathbf{v}_i(t_1) = \sigma_i \mathbf{u}_i(t_F) \quad \text{and} \quad (7)$$

$$\mathbf{L}_{t_1, t_F}^T \mathbf{u}_i(t_F) = \sigma_i \mathbf{v}_i(t_1), \quad (8)$$

with $\|\mathbf{v}_i\|_2 = 1$ and $\|\mathbf{u}_i\|_2 = 1$. Combining these two equations,

$$\mathbf{L}_{t_1, t_F}^T \mathbf{L}_{t_1, t_F} \mathbf{v}_i(t_1) = \sigma_i \mathbf{L}_{t_1, t_F}^T \mathbf{u}_i(t_F) = \sigma_i^2 \mathbf{v}_i(t_1), \quad (9)$$

reveals that the eigenvectors, $\mathbf{v}_i(t_1)$, of the Oseledec operator are the right singular vectors of the tangent-linear operator \mathbf{L}_{t_1, t_F} . Hence, the right singular vector $\mathbf{v}_1(t_1)$ assigned to the largest singular value, σ_1 , of a chemistry-transport model characterises the direction of maximum error growth over a finite time interval, $[t_1, t_F]$. The singular value σ_1 is the maximum value of the original ratio as per Eq. (3) and defines the amount of error growth.

2.1.1 Application of special operators

In order to address specific questions of atmospheric chemistry, special operators are applied to the initial and final perturbations, namely the weight matrix, \mathbf{W}_t , and the projection operator, \mathbf{P}_t .

The weight matrix, $\mathbf{W}_t \in \mathbb{R}^{n \times n}$, contains the concentrations of all considered chemical species, s with $s = 1, \dots, n$, at time, t :

$$\mathbf{W}_t := \text{diag}(\mathbf{c}^s(t))_{s=1, \dots, n} \quad \forall t \in [t_1, t_F]. \quad (10)$$

Since the concentrations of the different species may vary by many orders of magnitude, perturbations of those species with larger concentrations or higher reactivity are expected to dominate the error growth. Application of the inverse weight matrix avoids this effect by scaling the absolute uncertainties, $\delta \mathbf{c}(t)$, with current concentrations, $\mathbf{c}(t)$.

Another objective is to examine the error growth of a limited set of chemical species. To meet this need a projection operator, $\mathbf{P}_t \in \mathbb{R}^{n \times n}$, is applied. This operator sets those entries of the perturbation vector that correspond to species or regions not targeted to zero (Barkmeijer et al., 1998). Thus, the projection operator is a diagonal matrix with binary entries. In case of a chemical box model, the projection operator is given by

$$\mathbf{P}_t := \text{diag}(p_i)_{i=1, \dots, n}, \quad p_i = \begin{cases} 1 & \forall i \in \mathcal{P}(t) \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

where $\mathcal{P}(t)$ denotes the set of selected chemical compounds.

Applying both the weight matrix and the projection operator allows one to investigate the relative influence of a set of

chemical compounds, at some initial time, t_1 , on another set of chemical compounds, at some later time, t . This gives the projected relative error, $\delta \mathbf{c}_{\text{pr}}(t) \in \mathbb{R}^n$:

$$\delta \mathbf{c}_{\text{pr}}(t) := \mathbf{P}_t \mathbf{W}_t^{-1} \mathbf{L}_{t_1, t} \mathbf{W}_{t_1} \mathbf{P}_{t_1} \mathbf{W}_{t_1}^{-1} \delta \mathbf{c}(t_1). \quad (12)$$

At initial time, t_1 , this formula reduces to

$$\delta \mathbf{c}_{\text{pr}}(t_1) = \mathbf{P}_{t_1} \mathbf{W}_{t_1}^{-1} \delta \mathbf{c}(t_1). \quad (13)$$

Hence, the projected relative error at final time, t_F , reads

$$\delta \mathbf{c}_{\text{pr}}(t_F) = \mathbf{P}_{t_F} \mathbf{W}_{t_F}^{-1} \mathbf{L}_{t_1, t_F} \mathbf{W}_{t_1} \delta \mathbf{c}_{\text{pr}}(t_1). \quad (14)$$

Inserting Eq. (14) into Eq. (3) yields the projected relative error growth

$$g_{\text{pr}}(\delta \mathbf{c}_{\text{pr}}(t_1)) := \frac{\|\delta \mathbf{c}_{\text{pr}}(t_F)\|_2}{\|\delta \mathbf{c}_{\text{pr}}(t_1)\|_2}. \quad (15)$$

The squared relative error growth, $g_{\text{pr}}^2(\delta \mathbf{c}_{\text{pr}}(t_1))$, is a Rayleigh quotient

$$g_{\text{pr}}^2(\delta \mathbf{c}_{\text{pr}}(t_1)) = \frac{\delta \mathbf{c}_{\text{pr}}(t_1)^T \mathbf{B}^T \mathbf{B} \delta \mathbf{c}_{\text{pr}}(t_1)}{\delta \mathbf{c}_{\text{pr}}(t_1)^T \delta \mathbf{c}_{\text{pr}}(t_1)}, \quad (16)$$

$$\text{with } \mathbf{B} := \mathbf{P}_{t_F} \mathbf{W}_{t_F}^{-1} \mathbf{L}_{t_1, t_F} \mathbf{W}_{t_1},$$

subject to the condition

$$[\delta \mathbf{c}_{\text{pr}}(t_1)](j) = \begin{cases} [\delta \mathbf{c}(t_1)](j) / [\mathbf{w}(t_1)](jj) & \forall j \in \mathcal{P}(t_1) \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Here, $[\mathbf{x}](j)$ denotes the j -th component of a vector \mathbf{x} , while $[\mathbf{a}](jj)$ represents the element in j -th row and j -th column of a matrix \mathbf{A} . According to Rayleigh's principle, the phase space direction that maximises the Rayleigh quotient, as given by Eq. (16), is the solution, $\mathbf{v}_{\text{pr}1}(t_1)$, of the symmetric eigenvalue problem

$$\mathbf{B}^T \mathbf{B} \mathbf{v}_{\text{pr}}(t_1) = \lambda_{\text{pr}} \mathbf{v}_{\text{pr}}(t_1) \quad (18)$$

and is assigned to the largest eigenvalue, $\lambda_{\text{pr}1}$. This solution does not necessarily satisfy condition (17). In order to comply with condition (17), the eigenvalue problem as given by Eq. (18) is rearranged. Making use of the idempotence of the projection operator, it can readily be seen that

$$\mathbf{P}_{t_1} \mathbf{v}_{\text{pr}}(t_1) = \mathbf{v}_{\text{pr}}(t_1). \quad (19)$$

Multiplying Eq. (18) with \mathbf{P}_{t_1} and substituting relation (19) gives

$$\mathbf{B}_{\text{pr}}^T \mathbf{B}_{\text{pr}} \mathbf{v}_{\text{pr}}(t_1) = \lambda_{\text{pr}} \mathbf{v}_{\text{pr}}(t_1) \quad (20)$$

$$\text{with } \mathbf{B}_{\text{pr}} := \mathbf{P}_{t_F} \mathbf{W}_{t_F}^{-1} \mathbf{L}_{t_1, t_F} \mathbf{W}_{t_1} \mathbf{P}_{t_1}.$$

The new symmetric eigenvalue problem ensures a feasible solution and complies with condition (17). The vector, $\mathbf{v}_{\text{pr}1}(t_1) \in \mathbb{R}^n$, assigned to the largest eigenvalue, $\lambda_{\text{pr}1}$, of

Eq. (20) is the right singular vector of the operator \mathbf{B}_{pr} , and it is henceforth denoted by the projected relative singular vector. Its projected relative singular value, σ_{pr1} , is the square root of the associated eigenvalue, λ_{pr1} .

Inserting the identity matrix, \mathbf{I}_n , instead of either the projection operator or the weight matrix into Eq. (12) provides the relative error, $\delta c_r(t)$, and the projected error, $\delta c_p(t)$, respectively. Following the steps above, the projected error growth, $g_p(\delta c_p(t_1))$, the relative error growth, $g_r(\delta c_r(t_1))$, and their associated symmetric eigenvalue problems can be established. The eigenvectors assigned to the largest eigenvalues are denoted by the projected singular vector, $\mathbf{v}_{p1}(t_1) \in \mathbb{R}^n$, and the relative singular vector, $\mathbf{v}_{r1}(t_1) \in \mathbb{R}^n$, respectively. Accordingly, the square roots of the largest eigenvalues are the associated projected singular value, σ_{p1} , and relative singular value, σ_{r1} , respectively.

2.2 Uncertainties of emissions

Emissions and initial values are of equal importance in the skillful prediction of the evolution of chemical species (Elbern et al., 2007). This section generalises the application of singular vector analysis to incorporate sensitivities with respect to emission rates. In contrast to initial concentrations, the total emitted amount of a species, s , varies in time. Due to these variations of emissions in time, the associated directions of largest error growth differ for each time step, $t \in [t_1, t_F]$. Consequently, their identification results in one application of singular vector analysis per time step, $t \in [t_1, t_F]$, yielding an ill-posed problem with high computational expenditure. By dealing with a time-invariant emission factor, e_{fi} , for species i , the set of singular vector analyses can be reduced to a single singular vector analysis per time interval $[t_1, t_F]$. Here, the diurnal profile shape of emission rates is taken as a strong constraint, such that only uncertainties in the amplitude are analysed. This choice is reasonable since the relative daily evolution of the emissions is far better known than the total emitted amount in a grid cell (Elbern et al., 2007). The results of the singular vector analyses quantify for which species further emission strength ascertainment is most useful. To the authors' knowledge this problem has not been addressed in the context of singular vector analysis previously.

Emissions impact the final state, $\mathbf{c}(t_F)$, according to the ordinary differential equations that describe the chemical evolution:

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(\mathbf{c}(t)) + \mathbf{e}(t). \quad (21)$$

Inserting the vector of emission factors, \mathbf{e}_f , leads to

$$\frac{d\mathbf{c}}{dt} = \mathbf{f}(\mathbf{c}(t)) + \mathbf{E}(t)\mathbf{e}_f, \quad (22)$$

where $\mathbf{E}(t)$ is a diagonal matrix with the vector of emissions $\mathbf{e}(t)$ on its diagonal. The tangent-linear model integration of

Eq. (22) describes the linear part of the evolution of perturbation $\delta\mathbf{c}(t_F)$ and is caused by uncertainties in the emissions, $\delta\mathbf{e}_f$, and in initial values, $\delta\mathbf{c}(t_1)$:

$$\delta\mathbf{c}(t_F) = \tilde{\mathbf{L}}_{t_1, t_F} \delta\mathbf{z}(t_1), \quad (23)$$

where

$$\tilde{\mathbf{L}}_{t_1, t_F} = (\mathbf{L}_{t_1, t_F}, \mathbf{L}_{t_1, t_F}^e) \text{ and} \quad (24)$$

$$\delta\mathbf{z}(t_1) = (\delta\mathbf{c}(t_1), \delta\mathbf{e}_f)^T.$$

For expository purposes, the perturbation is henceforth solely caused by emissions uncertainties ($\delta\mathbf{c}(t_1) = 0$). Since uncertainties in emission factors already denote a relative disturbance, only their relative impact is considered. The latter is expressed by the relative error at final time:

$$\delta\tilde{c}_r(t_F) = \mathbf{W}_{t_F}^{-1} \delta\tilde{\mathbf{c}}(t_F). \quad (25)$$

Here, the tilde denotes a perturbation caused by emissions uncertainties only. Since the final uncertainty contains only perturbations in concentrations of species, the original weight matrix, \mathbf{W}_t as per Eq. (10), is applied. From definition (25), the relative error growth is obtained:

$$\tilde{g}_r(\delta\mathbf{e}_f) := \frac{\|\delta\tilde{\mathbf{c}}_r(t_F)\|_2}{\|\delta\mathbf{e}_f\|_2}. \quad (26)$$

Considering the squared ratio and applying Eq. (23) yields

$$\tilde{g}_r^2(\delta\mathbf{e}_f) = \frac{\delta\mathbf{e}_f^T \tilde{\mathbf{B}}_r^T \tilde{\mathbf{B}}_r \delta\mathbf{e}_f}{\delta\mathbf{e}_f^T \delta\mathbf{e}_f}, \quad (27)$$

with $\tilde{\mathbf{B}}_r := \mathbf{W}_{t_F}^{-1} \mathbf{L}_{t_1, t_F}^e$.

The phase space direction that maximises the ratio given by Eq. (27) is the eigenvector, $\tilde{\mathbf{v}}_{r1}$, of the eigenvalue problem

$$\tilde{\mathbf{B}}_r^T \tilde{\mathbf{B}}_r \tilde{\mathbf{v}}_r = \tilde{\lambda}_r \tilde{\mathbf{v}}_r \quad (28)$$

associated with the largest eigenvalue, $\tilde{\lambda}_{r1}$. Since the solution, $\tilde{\mathbf{v}}_{r1}$, gives the right singular vector of operator $\tilde{\mathbf{B}}_r$, it is henceforth denoted as the relative singular vector with respect to emission uncertainties.

A special set of perturbation norms can be provided with the help of the projection operator, \mathbf{P}_t , as per Eq. (11). For the error growth of emission factor uncertainties, the projected relative singular vector can be calculated following Sect. 2.1.1.

2.3 Singular-vector-based sensitivity analyses

Based on different error norms, the singular vectors introduced in this section all identify the most rapidly growing perturbation (in the linear approximation) of a chemical scenario run over a given time period. The identified perturbations may be regarded as the most critical errors. The magnitude of the entry of a compound in the perturbation vector reveals how sensitive the considered forecast result is to

small errors in the initial concentrations or in the emission factors of the associated species. The order of magnitude of the singular vector entries therefore equals the order of sensitivity. According to Palmer (1995), dynamically sensitive model states (i.e. high singular vector entries) moreover identify valuable measurements sites. In a chemical box model context, where only one location is considered, dynamically sensitive model states identify chemical species that have to be measured with priority. Hence, singular vector analysis answers the question of measurement priority in terms of sensitivity ranking. It is thus critical that sensitivity ranking be done prior to the determination of optimal observation configurations.

3 Model design

In order to calculate singular vectors for a chemical box model, the software package Kinetic PreProcessor (KPP, Sandu and Sander, 2006) is applied. KPP offers a set of numerical integrators from which the second order Rosenbrock method is chosen for this study. Since Rosenbrock solvers have good stability properties and conserve the linear invariants of the system (Hairer and Wanner, 1991), they are well-suited for atmospheric chemistry applications. For Rosenbrock integrators, KPP includes tangent-linear and two different adjoint code generation schemes (discrete and continuous) with respect to initial values. Here, the discrete adjoint model is our adjoint model of choice, since it is more consistent with the forward and tangent-linear model than the continuous adjoint model. Both the tangent-linear and the adjoint model with respect to emissions have been coded by hand.

Tangent-linear routines have been checked for consistency with corresponding forward and adjoint modules. For consistency with the forward model, the gradient check ratio (Navon et al., 1992) is applied, defined as

$$d = \frac{\|\text{FWD}(\mathbf{x} + \alpha \delta \mathbf{x}) - \text{FWD}(\mathbf{x})\|_2}{\|\text{TLM}(\alpha \delta \mathbf{x})\|_2}. \quad (29)$$

The abbreviations FWD and TLM denote particular parts of the forward model and associated tangent-linear routines. These new abbreviations are introduced to allow for piecewise code checking. As the scalar parameter α approaches zero, ratio (29) should approach 1. Note that round-off errors dominate in cases where the magnitude of α is sufficiently small. Here, the limits of numerical precision are approached and results degrade. Within numerical limits, the new tangent-linear routines, as well as the complete tangent-linear model, demonstrate the required characteristics of Eq. (29) for considered test cases. Besides consistency with the forward model, the gradient check ratio also indicates the accuracy of the tangent-linear approximation. Application of the tangent-linear model is only justified if the considered perturbation, $\delta \mathbf{x}$, is small enough to ensure $d \approx 1$. For each simulation considered in this paper, the accuracy of

the tangent-linear assumption is tested prior to analysing its results. This is effected by inserting the first singular vector for $\delta \mathbf{x}$ into Eq. (29), whereas the associated forward and tangent-linear models are applied for FWD and TLM, respectively. Variations of scalar α reveal that setting $\alpha = 0.15$ is sufficient to gain $|1.0 - d| \leq 0.01$ (valid for every simulation). Within the bounds of numerical precision further reduction of α improves the test results. Therefore, the ratios are close enough to 1 to suggest that the tangent-linear assumption is sufficiently satisfied.

In order to verify the consistency of the tangent-linear and the adjoint model, both models are utilised to build the Jacobian matrix. For construction of the Jacobian matrices, the equations

$$\delta \mathbf{y}_i = \text{TLM}(\delta \mathbf{x}_i) \quad \text{and} \quad \delta \mathbf{z}_i = \text{ADJ}(\delta \mathbf{x}_i) \quad (30)$$

are executed for $i = 1, \dots, n$. Here, ADJ denotes adjoint routines, TLM denotes the associated tangent-linear routines, and $\delta \mathbf{x}_i$ is defined to be a unit vector where the entry of the i -th element is 1 and all other entries are zero. It follows that $\delta \mathbf{y}_i$ represents the i -th row of the Jacobian matrix constructed with the tangent-linear model, while $\delta \mathbf{z}_i$ represents the i -th column of the Jacobian matrix constructed with the adjoint model. Consistency is verified if, within the limits of machine precision, the elements of both Jacobian matrices are equal. Again, the new routines demonstrate correctness for single routines as well as for the complete model.

Two methods have been implemented for solving the eigenvalue problems, namely the power method and the implicitly restarted Arnoldi method. The power method (Mises and Pollaczek-Geiringer, 1929) is an iterative technique for computing the dominant eigenpair $(\lambda_1, \mathbf{v}_1)$ of a matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$. Here, only the case where \mathbf{A} is symmetric and $\mathbf{A} \in \mathbb{R}^{n \times n}$ is considered. If the dominant eigenvalue, λ_1 , is strictly greater in absolute value than all other eigenvalues, and if the start vector, $\mathbf{v}^{(0)} \in \mathbb{R}^n$, is not orthogonal to the eigenspace of λ_1 , then the sequences of vectors $\{\mathbf{v}^{(k)}\}_{k=1, \dots, s}$ and scalars $\{\lambda^{(k)}\}_{k=1, \dots, s}$ generated recursively by

$$\mathbf{v}^{(k)} = \mathbf{A} \mathbf{v}^{(k-1)} / \|\mathbf{A} \mathbf{v}^{(k-1)}\|_2 \quad (31)$$

$$\lambda^{(k)} = \mathbf{v}^{(k)T} \mathbf{A} \mathbf{v}^{(k)} \quad (32)$$

will converge to the dominant eigenpair $(\lambda_1, \mathbf{v}_1)$. The speed of convergence of the power method is proportional to $|\lambda_2/\lambda_1|$, and hence the rate of convergence is linear. The larger the absolute values of λ_2 with respect to λ_1 , the faster the power method converges.

An appropriate method to find the k largest eigenvalues and their associated eigenvectors is implemented in the open software ARnoldi PACKage (ARPACK, Lehoucq et al., 1998, Sorensen, 1996). This software supplies a package of FORTRAN77 subroutines for solving large-scale eigenvalue problems. For this purpose it requires a number of subroutines from the Linear Algebra PACKage and the Basic Linear Algebra Subprograms (LAPACK and BLAS, both available at <http://www.netlib.org/>). Depending on the properties

Table 1. Meteorological parameters for scenarios as defined by Poppe et al. (2001).

	PLUME	URBAN
Altitude (km)	0	0
Temperature (K)	288.15	298.00
Pressure (hPa)	1013.25	1013.25
Air (#/cm ³)	2.55×10^{19}	2.46×10^{19}
H ₂ O (%v/v)	1.00	1.00

of matrix **A**, ARPACK relies either on the Lanczos or the Arnoldi algorithm. Both methods have the ability to calculate the k largest eigenvalues and their associated eigenvectors in one iteration cycle. They are described in detail in Sorensen (1996).

4 Tropospheric chemistry scenarios

Poppe et al. (2001) introduced a set of six scenarios for modelling tropospheric chemistry covering conditions that are typical for the remote continental planetary boundary layer (scenario LAND, no emissions), the ocean (scenario MARINE, no emissions), the free troposphere (scenario FREE, no emissions), the moderately polluted planetary boundary layer (scenario PLUME, emissions), the polluted planetary boundary layer (scenario URBAN, emissions), and an urban plume with biogenic impact (scenario URBAN/BIO, emissions). Due to the fact that sensitivity to emissions is one of the main elements of the present work and since the experimental design focuses on forecasting of air pollution, only scenario PLUME and scenario URBAN have been selected for further investigation. Both scenarios include a varying burden of anthropogenic emissions. Scenario PLUME and scenario URBAN are defined via meteorological parameters (described in Table 1), chemical initial values (Table 2), emission strengths (Table 3) and photolysis frequencies. The chemistry reaction mechanism used is the second-generation Regional Acid Deposition Model (RADM2, Stockwell et al., 1990), which considers tropospheric gas-phase chemistry only. Poppe et al. (2001) simulated these scenarios with five different chemical solvers. All simulations started on 1 July, 12:00 LT, and ended five days (120 h) later on 6 July, 12:00 LT, and demonstrated excellent agreement between the participating numerical solvers. Hence, the scenarios are unambiguously described and may be used for testing numerical solvers. Simulations with the numerical solver employed in this work show nearly perfect agreement with those of Poppe et al. (2001). Consequently, the applied configuration may be regarded as suitable to model the chemical evolution of the considered scenarios.

Based on scenarios PLUME and URBAN, the present study focuses on analysing the sensitivity of the air pollution forecast to selected VOCs (in terms of initial values and

Table 2. Initial mixing ratios for the gas-phase constituents (ppb) of scenarios described by Poppe et al. (2001). NMHC denotes non-methane hydrocarbons.

	PLUME	URBAN
O ₃	50	30
NO	0.2	0.1
NO ₂	0.5	0.1
HNO ₃	0.1	1.5
CO	200	100
CH ₄	1700	1700
Isopr.	0	0
H ₂	500	500
H ₂ O ₂	2	2
HCHO	0	1
NMHC	0	0
SO ₂	0	0

Table 3. Emission strength as volume production rate in cm⁻³ s⁻¹ for scenario PLUME. For scenario URBAN, the emission strength of NO of scenario PLUME has to be multiplied by 5, while the emission strengths of all other species remains unchanged.

Species	Emission strength
NO	1.1×10^6
NO ₂	0.0
CO	2.4×10^6
CH ₄	0.0
SO ₂	2.2×10^5
VOC	3.0×10^6

emission factors). Here, we choose the RADM2 species O₃ and PAN as indicators of the air pollution caused by secondary pollutants. Since traceable results are of importance for the validation of the present algorithm, other secondary pollutants are not considered. VOCs belong to one of the major classes of directly emitted precursors of secondary pollutants. In the case of the RADM2 mechanism, VOCs comprise the model species ETH, HC3, HC5, HC8, OL2, OLT, OLI, ISO, TOL, CSL, XYL, HCHO, ALD, KET, GLY, MGLY, and DCB (detailed definitions of the RADM2 species abbreviations for the VOCs are given in Table 4). For these species, the total VOC emission strength of 3.0×10^6 cm³ s⁻¹ (see Table 3) was segregated into emission strengths per RADM2 compound according to Kuhn et al. (1998) and Middleton et al. (1990) (see Table 5). Both O₃ and PAN are built by photochemical reactions involving VOCs, oxides of nitrogen and sunlight. During their degradation, the VOCs have to compete for hydroxyl radicals (OH) in order to form peroxy radicals (RO₂) and the acetyl radical (CH₃CO), which later form O₃ and PAN, respectively. It is therefore to be expected that VOCs with higher OH reactivity have a larger impact on air pollution. Furthermore, VOCs with a higher carbon number

Table 4. VOCs of the RADM2 mechanism (according to Stockwell et al., 1990).

RADM2 species	Definition
<i>Alkanes</i>	
ETH	Ethane
HC3	Alkanes with HO rate constant (298 K, 1 atm) between 2.7×10^{-13} and $3.4 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$
HC5	Alkanes with HO rate constant (298 K, 1 atm) between 3.4×10^{-12} and $6.8 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$
HC8	Alkanes with HO rate constant (298 K, 1 atm) greater than $6.8 \times 10^{-12} \text{ cm}^3 \text{ s}^{-1}$
<i>Alkenes</i>	
OL2	Ethene
OLT	Terminal alkenes
OLI	Internal alkenes
ISO	Isoprene
<i>Aromatics</i>	
TOL	Toluene and less reactive aromatics
CSL	Cresol and other hydroxy substituted aromatics
XYL	Xylene and more reactive aromatics
<i>Carbonyls</i>	
HCHO	Formaldehyde
ALD	Acetaldehyde and higher aldehydes
KET	Ketones
3 GLY	Glyoxal
MGLY	Methylglyoxal
DCB	Unsaturated Dicarboxyl

Table 5. Emission strength per RADM2 VOC species for scenario PLUME and scenario URBAN. The total VOC emission of $3.0 \times 10^6 \text{ cm}^3 \text{ s}^{-1}$ was segregated according to Kuhn et al. (1998) and Middleton et al. (1990).

Species	Emission strength ($\text{cm}^3 \text{ s}^{-1}$)
ETH	1.02×10^5
HC3	1.09×10^6
HC5	3.23×10^5
HC8	1.93×10^5
OL2	1.94×10^5
OLT	0.93×10^5
OLI	0.80×10^5
ISO	0.0
TOL	2.43×10^5
CSL	0.0
XYL	2.20×10^5
HCHO	0.59×10^5
ALD	0.15×10^5
KET	1.36×10^5
GLY	0.0
MGLY	0.0
DCB	0.0

are likely to have a larger impact on air pollution since they have the potential to form more radicals in their degradation pathway and can thus form more O_3 or PAN, respectively. It has, however, to be kept in mind that not all VOCs have the ability to form the acetyl radical during their degradation path.

The sensitivity results associated with our experimental design identify which VOCs are to be measured to reduce VOC-caused uncertainties in air pollution forecasts. The error growth scheme applied here considers air pollution caused by O_3 and PAN as a whole, allowing the largest error growth to be dominated by only one of the considered final species.

The chemical evolution of air pollution is not only dependent on the chemical regime but also on the underlying simulation interval. It is therefore to be expected that significantly different singular vectors are to be found across a suite of simulations with different initial and integration times. For a comprehensive investigation of these effects a temporal singular vector diagram (TSVD) is implemented. Each TSVD consists of a complete set of singular vectors comprising different simulation lengths as well as several starting times within a chosen time interval. This allows for the study of different degrees of ageing. Spin-up considerations require the left boundary of the chosen time interval to be set to 2 July, 12:00, while the right boundary is set to be 6 July, 12:00. The hourly time points within this bound temporal interval are the starting points for a series of singular vector calculations. For the sake of clarity, the interval boundary of 2 July, 12:00, is henceforth called starting time, t_0 (corresponding to the first chosen starting point of the singular vector calculations), while the starting point of each individual singular vector calculation is called the initial time, t_1 . Further, the interval boundary of 6 July, 12:00, is denoted as the end time, t_n (corresponding to the last chosen starting point of the singular vector calculations), whereas the end point of each individual singular vector calculation is denoted as the final time, t_F . For each hourly ($\Delta t = 1 \text{ h}$) initial time, t_1 , $I=0, \dots, n-1$, within the interval $[t_0, t_n]$, a time row, $\text{TR}(I)$, of singular vector calculations is carried out, leading to $n = \frac{t_n - t_0}{\Delta t} = 96$ time rows per TSVD. Each time row, $\text{TR}(I)$, consists of n individual singular vector simulations, all starting at time $t_1 = t_0 + I \Delta t$, but differing in terms of simulation length, which equals $(m+1) \cdot \Delta t$ for calculation m , $m=0, \dots, n-1$. Figure 1 illustrates that time row $\text{TR}(0)$ consists of 96 simulations with fixed initial time, $t_1 = t_0$, and varying final times, t_F (in Fig. 1 each of these simulation is denoted as $[t_0, t_F]$, $F = 1, \dots, 96$). For the first simulation of $\text{TR}(0)$ the simulation length is $(m+1) \cdot \Delta t = 1 \text{ h}$ ($m = 0$). Therefore, the first simulation begins on July 2, 12:00, and ends one hour later on 2 July, 13:00 (in Fig. 1 the final time of 2 July, 13:00, is denoted as t_1 since the final time is one hour after starting time, t_0). Accordingly, the last simulation of $\text{TR}(0)$ has a simulation length of $(m+1) \cdot \Delta t = 96 \text{ h}$ ($m = 95$) and begins on July 2, 12:00, and ends on July 6, 12:00 (denoted as t_{96}). With

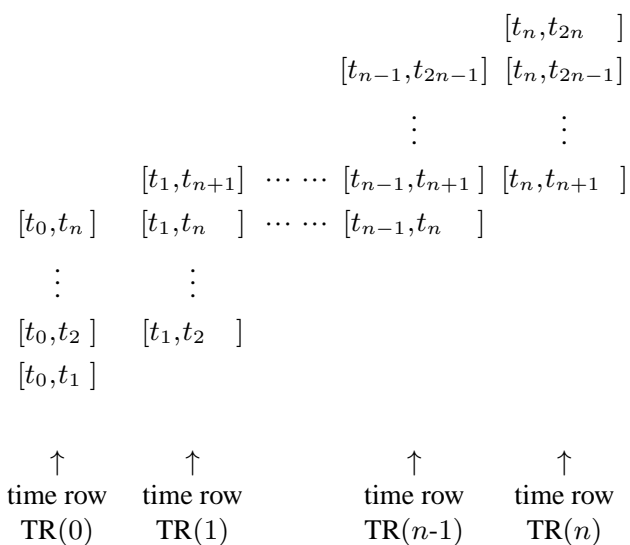


Fig. 1. Schematic presentation of the temporal singular vector diagram (TSVD). Here, each interval $[t_I, t_F]$ represents one singular vector calculation starting at initial time $t_I = t_0 + I\Delta t$, $I = 0, \dots, n-1$ and ending at final time $t_F = t_0 + F\Delta t$, $F = 1, \dots, 2n$, $F > I$.

96 different initial times and 96 different simulation lengths, each TSVD comprises $n^2 = 9216$ singular vector analyses.

In order to aid the interpretation of the TSVD results in the following subsections, the time dependence of the O_3 , PAN, OH, and VOC mixing ratios for both scenarios is depicted in Fig. 2.

4.1 Error growth of initial uncertainties

In the first stage of the present sensitivity study, focus is placed on changes in the final O_3 and PAN concentrations due to uncertain initial VOC concentrations. The associated singular vector analyses seek to answer the following question:

- To what extent does the uncertain initial concentration of an individual VOC contribute to uncertainties of the O_3 and PAN evolution?

In order to answer this question, the projected error growth (see Sect. 2.1.1) is determined with final projection on both O_3 and PAN and initial projection on all VOCs. Other species are not taken into account. Since the answer to the above question may differ depending on the scenario and the initial and final simulation times, it is of special interest to address the more in-depth issue:

- What is the degree of dependence of these results on the scenario selected and the time interval chosen, respectively?

Both questions may be answered by applying the TSVD of projected singular vectors to scenario PLUME and sce-

nario URBAN. The resulting sensitivities provide insight into the underlying chemical kinetics and are therefore meaningful. Yet their applicability for identifying optimal measurement strategies is limited, since the maximal projected error growth is dominated by the final variable, which allows for the largest absolute changes. Here, final variables are O_3 and PAN, and the sensitivity results for scenarios PLUME and URBAN are dominated by O_3 . The fact that small absolute changes in the final PAN concentration can be as harmful as large absolute changes in the final O_3 concentration is hereby ignored. In addition, some of the resulting sensitivities assign the largest absolute uncertainties to species with very small concentrations. Since these perturbations are not realistic, their associated optimal measurement strategies are not discussed in this section. Still, the analysis of the projected singular vectors is crucial to the understanding of the results of the relative error growth (see Sect. 4.2).

In order to gain insight into the results of a complete TSVD, the singular vectors are visualised by considering the vector components corresponding to a specific chemical compound separately. Since all singular vectors are set to unit length, a resulting vector component of 1 indicates that the O_3 and PAN concentration at final time is solely influenced by this particular compound. Figure 3 presents the resulting CSL section of a complete TSVD for both scenarios PLUME and URBAN. CSL has been chosen for representation since it shows the most prominent differences between sensitivities associated with simulations initialised during the daytime and sensitivities associated with simulations initialised during the nighttime. A significant difference between nighttime and daytime sensitivities is also present for the other, non-displayed VOCs. The main qualitative features of the O_3 and PAN sensitivity to VOCs are the following:

1. Initial time
The sensitivity to VOCs changes significantly depending on whether the associated simulation is initialised during the daytime or during the night. Furthermore, the behavioural pattern for sensitivities associated with simulations initialised during the daytime/nighttime recurs every day/night. Results with distinct initial times differ less for initial times within one regime (daytime or nighttime) than for initial times in different regimes; and
2. Simulation length
With growing simulation length, the influence of the individual VOCs changes significantly. Hence, simulation length is another influential feature.

For a direct comparison of the importance of different VOCs, the results of the singular vector analyses are categorised following the findings listed above. In order to present unambiguous results, a selected number of simulations are not considered for categorisation. First of all, simulations initialised during the nighttime and finalised before the first sunrise are not taken into account. In these special cases, the whole

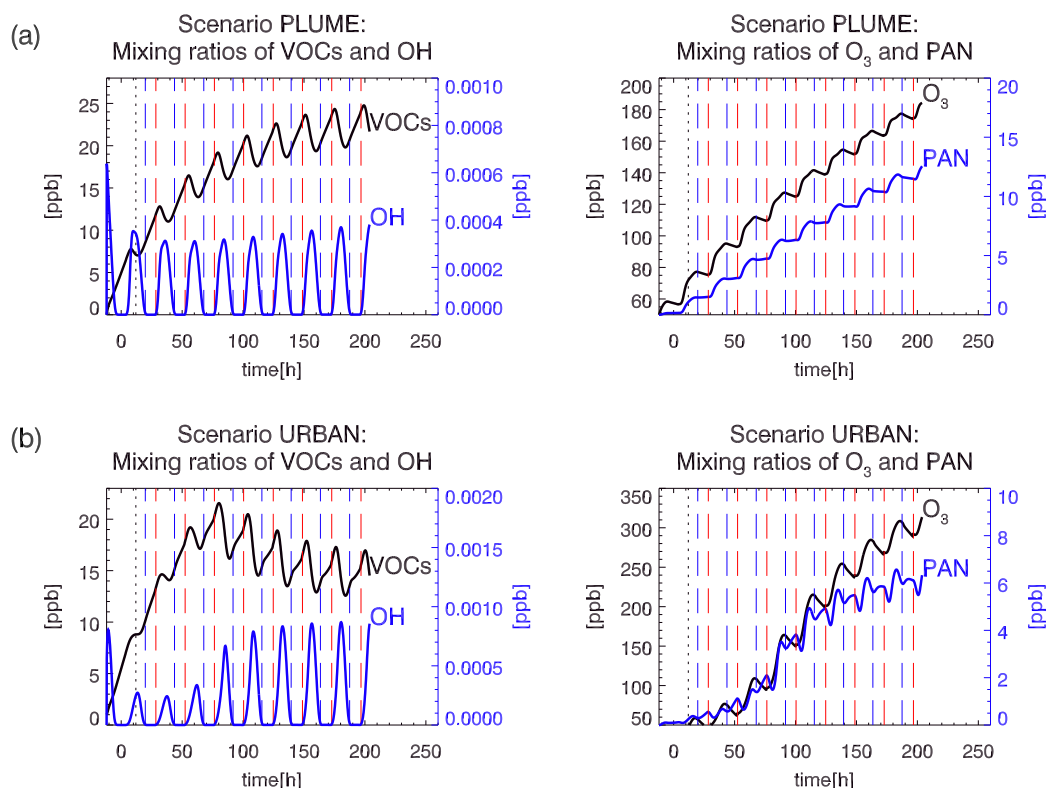


Fig. 2. Time dependence of the OH, VOC (left panel column), O₃ and PAN (right panel column) mixing ratios for scenarios (a) PLUME and (b) URBAN. Here, 12:00 corresponds to the starting time, t_0 , of the first singular vector calculation (2 July, 12:00), and 204h corresponds to the final time, t_F , of the last singular vector calculation (10 July, 12:00). Hence mixing ratios for all considered initial times, t_1 , and final times, t_F , of the TSVD are depicted. To aid interpretation, sunrises are marked with dashed red lines and sunsets with dashed blue lines, respectively.

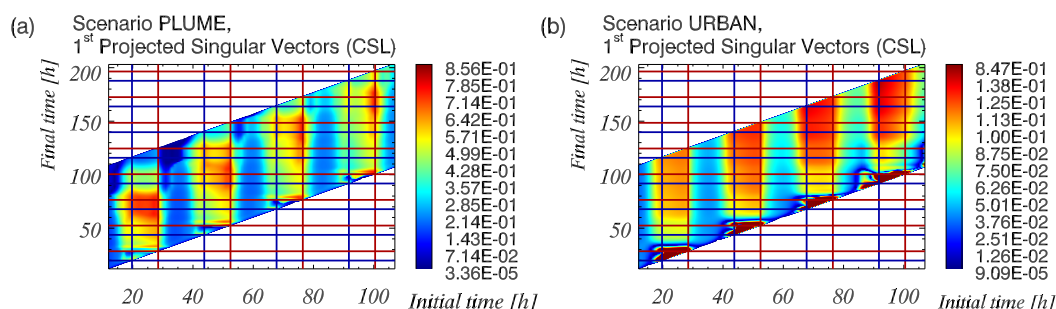


Fig. 3. CSL section of the TSVD of optimal projected singular vectors with respect to initial uncertainties for scenarios (a) PLUME and (b) URBAN. The organisation of the results follows the schematic diagram displayed in Fig. 1. Each colour pixel indicates an individual singular vector calculation with initial time, t_1 , and final time, t_F . To aid interpretation, singular vectors were set to unit length. Furthermore, sunrises are marked with red lines and sunsets with blue lines, respectively.

simulation takes place in the nighttime when there is no photochemistry. Secondly, simulations initialised during hours with rapidly decreasing or increasing insolation are disregarded for categorisation. More precisely, hours with rapidly increasing insolation are defined to be between sunrise and 3 h after sunrise, while hours with rapidly decreasing insola-

tion are defined to be between 4 h before sunset and sunset. During these hours, the behaviour of the calculated sensitivities can differ from the otherwise clearly defined behaviour of sensitivities associated with simulations initialised during the daytime. This is due to the fact that the transition from day- to nighttime behaviour does not occur abruptly for each

scenario but sometimes proceeds more steadily (see Fig. 3). Omitting these transition hours from categorisation enables a clear distinction between sensitivities associated with simulations initialised during the daytime and sensitivities associated with simulations initialised during the nighttime.

Subsequently, the results are categorised according to the “initial time” and “simulation length” criteria listed above. Attention is first paid to the “initial time” criterion. Thus the model output is categorised into results of simulations initialised during the daytime (category C_a) and results of simulations initialised during the nighttime (category C_b). In order to obtain general statements about the O_3 and PAN sensitivity to different VOCs, some statistical values are considered. For each category C_i , a mean O_3 and PAN sensitivity, m_i , is calculated according to

$$m_i(j) = \frac{1}{n_i} \sum_{i \in C_i} |v_i^*(j)|, \quad (33)$$

where n_i denotes the number of calculations belonging to category C_i , and $v_i^*(j)$ the normalised singular vectors components of species j belonging to category C_i . Furthermore, the absolute minimum and maximum values, c_i^+ and c_i^- ,

$$c_i^+(j) = \max_i |v_i^*(j)|, \quad (34)$$

$$c_i^-(j) = \min_i |v_i^*(j)|, \quad (35)$$

as well as the standard deviation,

$$st_i(j) = \sqrt{\frac{1}{n_i} \sum_{i \in C_i} (|v_i^*(j)| - m_i(j))^2}, \quad (36)$$

are calculated for each category, C_i , to examine the significance of the mean O_3 and PAN sensitivity, m_i .

Due to changing mixing ratios, sensitivity results of categories C_a and C_b may differ when the initial time of their associated simulations is during different days/nights. In order to capture this feature, the introduced statistical values are furthermore calculated for each day and night separately. In the case of these separately calculated statistics, it is henceforth always stated for which day/night the statistics are calculated (e.g. “statistics for category C_a , initial time at day 1”). Here, day i is defined as containing all results of simulations initialised between sunrise $S_{i-1}^+(t_0)$ and sunset $S_i^-(t_0)$, while night i comprises all results of simulations initialised between sunset $S_i^-(t_0)$ and sunrise $S_i^+(t_0)$. Sunrise $S_i^+(t_0)$ and sunset $S_i^-(t_0)$, $i=1, 2, 3, 4$, specify the i -th sunrise and i -th sunset after starting time, t_0 . Sunrise $S_0^+(t_0)$ equals starting time, t_0 .

By the second criterion (simulation length), simulation length needs also to be considered. Therefore categories C_a and C_b are further subdivided into 4 subcategories, depending on simulation length, leading to 8 categories in total. These categories are denoted as C_{a_k} and C_{b_k} , $k=1, 2, 3, 4$. For the subdivision into simulation length, sunrises are chosen as

separation criteria. In detail, each category, C_{a_k} or C_{b_k} , contains results of calculations ending between sunrise $S_{k-1}^+(t_1)$ and sunrise $S_k^+(t_1)$. In contrast to the previous definition of sunrise $S_i^+(t_0)$ (which specifies the i -th sunrise after starting time, t_0), sunrise $S_k^+(t_1)$, $k=1, 2, 3$ specifies the k^{th} sunrise after the initial time, t_1 , of each particular simulation. In addition, sunrise $S_0^+(t_1)$ is defined as equaling the initial time, t_1 , of each simulation, and sunrise $S_4^+(t_1)$ is defined as equaling the final time, t_F , of each simulation. As an example, for simulations initialised on 2 July, 12:00, the first sunrise is on 3 July, 04:45, while the first sunrise for simulations initialised on 3 July, 16:00, is on July 4, 04:45. Accordingly, category C_{a_1} represents simulations initialised during the daytime that have a rather short simulation length (between 1 h and 14 h), while category C_{b_4} represents simulations initialised during the nighttime that have a rather long simulation length (between 81 h and 96 h). Again, categories C_{a_k} and C_{b_k} are furthermore applied to each day and night separately following the conventions established above.

Table 6 displays the mean sensitivities and the associated standard deviations for categories C_a and C_b . In order to get an impression of the time dependency of the sensitivities, Fig. 4 depicts the mean sensitivities, their associated standard deviations as well as the absolute minimum and maximum values for categories C_{a_k} and C_{b_k} , $k=1, 2, 3, 4$ for day 1, night 1, day 4, and night 4. Here, only the ten most important VOCs are depicted. Figure 4 indicates that the sensitivity ranking of the VOCs is dependent on the chosen scenario, the initial time, t_1 , of simulation (day or night), and the length of simulation. In the following subsections (Sect. 4.1.1 and Sect. 4.1.2), the most important findings per scenario are described briefly. Here, each of the VOCs is grouped as being of *high importance* (mean sensitivity above 0.1), *medium importance* (mean sensitivity between 0.05 and 0.1), *little importance* (mean sensitivity between 0.001 and 0.05) or *negligible importance* (mean sensitivity below 0.001) for the formation of O_3 and PAN.

4.1.1 Scenario PLUME with final projection on O_3 and PAN

For simulations initialised during the daytime, the mean sensitivities of category C_a (see Table 6) suggest that O_3 and PAN are most sensitive to species CSL, DCB, MGLY (high importance), followed by species TOL, XYL, ALD, and OLI (medium importance). Species HC8, ISO, OLT, KET, HC3, HC5, and OL2 are of little importance, and the importance of GLY, HCHO, and ETH is negligible for the formation of O_3 and PAN. This sensitivity ranking is relatively constant for different initial days, with only the sensitivity values of CSL, TOL, OL2, and ISO changing significantly (see Fig. 4). The mean sensitivities of CSL and TOL increase between day 1 and day 4, and the mean sensitivities of OL2 and ISO decrease simultaneously. The impor-

Scenarios PLUME and URBAN, Projected Error Growth

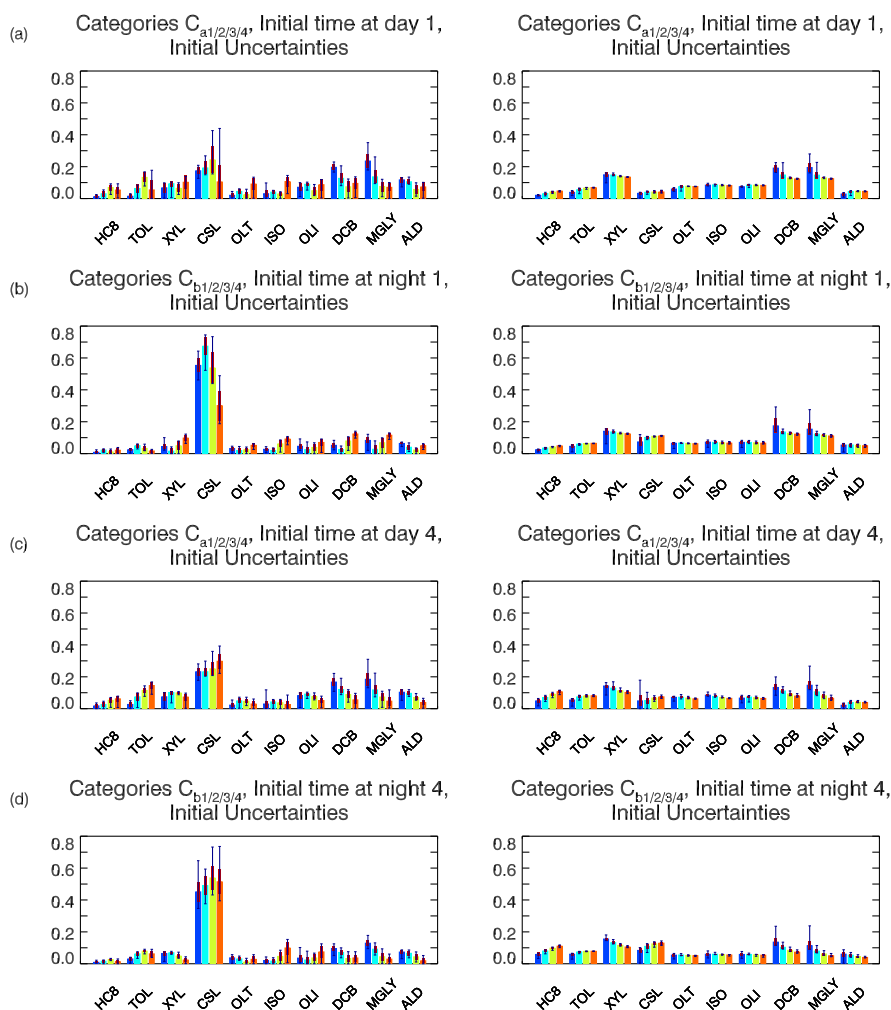


Fig. 4. Statistics of optimal projected singular vectors with respect to initial uncertainties for scenario PLUME (left panel column) and scenario URBAN (right panel column) for (a) day 1, (b) night 1, (c) day 4, and (d) night 4. Displayed are results for categories $C_{a1/2/3/4}$ and $C_{b1/2/3/4}$ for the VOCs with the most influence. All panels depict mean impact ($C_{a/b}$: blue bars; $C_{a1/b1}$: blue bars; $C_{a2/b2}$: turquoise bars; $C_{a3/b3}$: green bars; $C_{a4/b4}$: orange bars), minimum/maximum value (dark blue lines) and standard deviation (red bars).

tance of ISO for the formation of O₃ and PAN hereby decreases from “medium” to “little”, while the importance of the other VOCs remains unaltered. However, results show that the importance of each individual VOC is highly dependent upon simulation length (see Fig. 4). With growing simulation length the importance of MGLY, DCB, and ALD decreases from “high” to “medium”. Contrarily, TOL, HC8, KET, HC3, GLY, and ETH show increasing importance with increasing simulation length (from “negligible” to “little” for KET, HC3, GLY, and ETH, from “little” to “medium” for HC8, and, most prominent, from “little” to “high” for TOL).

For simulations initialised during the nighttime, the mean sensitivity to CSL is significantly higher than for simula-

tions initialised during the daytime (see Table 6), yielding an exceptionally high CSL sensitivity (high importance). According to the mean sensitivities of category C_b, MGLY and DCB are of medium importance, followed by XYL, TOL, ALD, ISO, OLI, OLT, GLY, OL2, HC8, KET, and HC3, which are of little importance for the formation of O₃ and PAN. Species of negligible importance are HC5, HCHO, and ETH. When compared to the sensitivity ranking of simulations initialised during the daytime, the most noticeable difference is the higher ranking of species ISO and GLY. While the sensitivity ranking stays approximately the same (compared to simulations initialised during the daytime), the sensitivity values themselves are generally smaller (except for

Table 6. Mean impact and standard deviation of VOCs for categories C_a and C_b . Considered are optimal projected singular vectors with respect to initial uncertainties and final projection on O_3 and PAN.

Species	Scenario PLUME, Category C_a	Scenario PLUME, Category C_b	Scenario URBAN, Category C_a	Scenario URBAN, Category C_b
ETH	0.006 ± 0.007	0.003 ± 0.002	0.007 ± 0.006	0.006 ± 0.006
HC3	0.025 ± 0.020	0.011 ± 0.007	0.027 ± 0.015	0.026 ± 0.014
HC5	0.021 ± 0.015	0.008 ± 0.004	0.040 ± 0.021	0.038 ± 0.019
HC8	0.045 ± 0.026	0.016 ± 0.010	0.062 ± 0.028	0.060 ± 0.026
OL2	0.021 ± 0.018	0.017 ± 0.007	0.041 ± 0.007	0.038 ± 0.006
OLT	0.041 ± 0.026	0.028 ± 0.015	0.069 ± 0.006	0.058 ± 0.006
OLI	0.073 ± 0.025	0.041 ± 0.028	0.073 ± 0.008	0.062 ± 0.008
ISO	0.042 ± 0.030	0.045 ± 0.036	0.077 ± 0.008	0.062 ± 0.008
TOL	0.091 ± 0.051	0.047 ± 0.022	0.067 ± 0.013	0.065 ± 0.011
XYL	0.086 ± 0.026	0.048 ± 0.025	0.128 ± 0.017	0.128 ± 0.015
CSL	0.227 ± 0.078	0.525 ± 0.105	0.053 ± 0.017	0.108 ± 0.021
HCHO	0.007 ± 0.005	0.005 ± 0.004	0.023 ± 0.007	0.021 ± 0.007
ALD	0.075 ± 0.031	0.046 ± 0.023	0.038 ± 0.008	0.052 ± 0.011
KET	0.029 ± 0.024	0.014 ± 0.008	0.034 ± 0.015	0.033 ± 0.013
GLY	0.007 ± 0.006	0.019 ± 0.008	0.026 ± 0.009	0.020 ± 0.008
MGLY	0.100 ± 0.060	0.070 ± 0.039	0.117 ± 0.038	0.102 ± 0.033
DCB	0.104 ± 0.047	0.059 ± 0.033	0.120 ± 0.031	0.119 ± 0.033

CSL, GLY, and ISO). The sensitivity ranking for simulations initialised during the nighttime is relatively independent of the specific initial night (see Fig. 4). Only species CSL, TOL, and OL2 show significant sensitivity changes for different initial nights, but their ranking with regard to importance remains unchanged. Again, simulation length is a more critical feature (see Fig. 4), leading to changes in importance from “high” to “medium” (MGLY), from “medium” to “little” (ALD), from “little” to “medium” (ISO, OLI), and from “negligible” to “little” (HC3, HC8, KET) between category C_{b1} and category C_{b4} . Although simulation length also changes the importance of DCB, XYL, and TOL, there is no persistent direction of change.

4.1.2 Scenario URBAN with final projection on O_3 and PAN

For simulations initialised during the daytime, XYL, DCB, and MGLY show the largest mean sensitivities within category C_a (see Table 6). Accordingly, these species are of high importance, while ISO, OLI, OLT, TOL, HC8, and CSL are of medium importance, and OL2, HC5, ALD, KET, HC3, GLY, and HCHO are of little importance for the formation of O_3 and PAN. As for scenario PLUME, O_3 and PAN are the least sensitive to ETH (negligible importance). In contrast to scenario PLUME, the sensitivity ranking of scenario URBAN is highly dependent upon the specific initial day (see Fig. 4). The later the initial day, the larger are the mean sensitivities of HC8, HC5, and HC3 (changes in importance from “little” to “medium” for HC8 and HC5 and from “negligible” to “little” for HC3), while the mean sensitivities of HCHO, GLY, MGLY, and DCB decrease (changes in im-

portance from “high” to “medium” for MGLY and DCB and no change in importance for HCHO and GLY). Simulation length changes the sensitivity ranking as well (see Fig. 4). Most prominent are the decreasing mean sensitivities of MGLY and DCB, altering the importance of MGLY and DCB from “high” to “medium” between category C_{a1} and category C_{a4} . Simultaneously, the importance of TOL, HC8, CSL, HC5, and ETH increases from “negligible” to “little” (ETH) and from “little” to “medium” for TOL, HC8, CSL, and HC5.

For simulations initialised during the nighttime, the mean sensitivities of category C_b (see Table 6) rank XYL, DCB, CSL, and MGLY as being of high importance. TOL, ISO, OLI, HC8, OLT, and ALD are of medium importance, followed by HC5, OL2, KET, HC3, HCHO, and GLY (little importance). Again, ETH is the species to which O_3 and PAN are least sensitive. In terms of ranking order, ALD and CSL show the most significant differences between simulations initialised during the nighttime and simulations initialised during the daytime. Both species are ranked higher for simulations initialised during the nighttime. The sensitivity ranking of simulations initialised during the nighttime is dependent upon the specific initial night and the simulation length. Here, the sensitivities of the VOCs show the same trends and changes in importance as they did for simulations initialised during the daytime, the only difference being that HC8 shows no change in importance due to simulation length, while the importance of CSL changes from “medium” to “high”.

4.1.3 Summary of results for sensitivities with final projection on O₃ and PAN

In summary the statistics for the VOCs for both scenarios indicate the following results:

- The O₃ and PAN formation is highly sensitive to DCB, MGLY, and XYL, whereas it is least sensitive to ETH and HCHO.
- With increasing pollution, the sensitivity of O₃ and PAN to alkanes and XYL increases, and the sensitivity to ALD and CSL decreases. Further, O₃ and PAN are very sensitive to CSL for scenario PLUME and of only medium sensitivity to CSL for scenario URBAN.
- With growing simulation length the O₃ and PAN sensitivity to ETH, HC3, HC5, and HC8 increases, while the O₃ and PAN sensitivity to DCB and MGLY decreases.
- For simulations initialised during the nighttime, the O₃ and PAN sensitivity to CSL is significantly higher than for simulations initialised during the daytime.

With regard to the final perturbations associated with the analysed sensitivities, the entries corresponding to O₃ are more than one order of magnitude higher than the entries corresponding to PAN (not shown), indicating that the sensitivity results are governed by the O₃ sensitivity to VOCs. The O₃-VOC chemistry is dependent on OH reactivity, carbon number, and carbon bonding of each individual VOC (a brief explanation of the O₃-VOC chemistry can be found at the beginning of Sect. 4). According to their bonding arrangements, VOCs are assigned to the functional group of alkanes (including ETH, HC3, HC5, and HC8), aromatics (including TOL, CSL, and XYL), carbonyls (including HCHO, ALD, KET, GLY, MGLY, and DCB) or alkenes (including OL2, OLT, OLI, and ISO). In the case of simulations of scenario URBAN initialised during the daytime, the VOC sensitivity ranking within one functional group is determined by OH reactivity and carbon number, assigning the lowest sensitivity to the smallest carbon number and the least OH reactivity. Here, the sensitivity ranking matches the expected results exactly. This agreement between expectations and results is due to the fact that scenario URBAN is VOC sensitive and the OH-VOC reaction is preferred to the OH-NO₂ reaction. For simulations initialised during the nighttime, however, OH concentrations are low and species HCHO, ALD, GLY, MGLY, DCB, CSL, OL2, OLT, OLI, and ISO react with NO₃. Since the NO₃ reactivity of VOCs is highest for CSL, the sensitivity to CSL is higher for simulations initialised during the nighttime. The exceptionally high CSL sensitivity of simulations of scenario PLUME initialised during the nighttime can be explained in the same way. Remarkably, the high CSL sensitivity for simulations initialised during the nighttime is still visible for the longest simulation interval (category C_{b4}). For simulations of scenario PLUME

initialised during the daytime, the mean sensitivities of ISO and CSL are significantly less and higher, respectively, than expected. This is due to the fact that scenario PLUME features only moderate pollution and is neither VOC nor NO_x limited. Accordingly, the OH-VOC reaction does not always predominate over other possible VOC reactions.

The singular-vector-based sensitivity results are compared with two previous studies concerning the ozone formation potential of VOCs (Butler et al., 2011; Derwent et al., 1998). Derwent et al. (1998) used the Master Chemical Mechanism for their study of the photochemical ozone creation potential (POCP), while Butler et al. (2011) based their study of the tagged ozone production potential (TOPP) on a subset of the MCM. In contrast to the RADM2 where most of the VOC species are lumped, the MCM employs no empirical lumping. For the purpose of comparison, the POCP and TOPP results associated with alkenes, alkanes and aromatics are aggregated according to Middleton et al. (1990) to be compatible with our sensitivity results. Furthermore, only category C_{a2} (associated with a simulation length between 25 h and 39 h) of scenario URBAN is considered due to the fact that the simulations of Butler et al. (2011) and Derwent et al. (1998) are based on polluted environments and are initialised during the day. Comparing the sensitivity order of VOCs for TOPP (first day, Los Angeles model run) and POCP with the one calculated with singular vector analysis, all three measures of sensitivity show that the sensitivity to aromatics is highest, followed by alkenes and then alkanes. Furthermore, the three measures provide the same sensitivity ranking for the individual VOC species of RADM2, with the one exception that the singular vector analysis ranks the species OL2 lower than the measure of POCP and the species TOL higher than the TOPP approach (the singular-vector-based sensitivity ranks ETH lowest, followed by HC3, HC5, HC8, OL2, TOL, OLT, OLI, ISO, and XYL). Butler et al. (2011) also calculated the time dependence of the TOPP sensitivities. In order to compare the time dependence of the singular-vector-based sensitivities with those of Butler et al. (2011), the unit-length singular vector sensitivities are retransformed to their original length through multiplication with the associated singular value. Since the associated singular values increase between category C_{a1} and category C_{a3} and decrease afterwards, the retransformed singular vectors show that the sensitivities of aromatics and alkenes increase between category C_{a1} and category C_{a3} and decrease afterwards, while the sensitivities to the alkanes are persistently increasing. These differences between time dependence of aromatics, alkenes and alkanes are in good agreement with the TOPP results. Remarkably, the retransformed singular vectors of scenario PLUME show the same time dependence of aromatics, alkenes and alkanes. Here, the period of increase is shorter for all functional groups.

The analysed projected singular vectors include little information about the PAN sensitivity to different VOCs. In order to get a general picture of both PAN and O₃ sensitivities,

the specific PAN sensitivity is calculated. Considering only the one final variable (PAN), projected singular vectors and adjoint sensitivities yield the same sensitivity results, yet adjoint sensitivity calculations are computationally much faster than singular vector calculations. Therefore, the PAN sensitivities are determined with the adjoint method. Figure 5 and Table 7 display the statistical results of the introduced categorisations for both scenarios. The most important findings per scenario are described briefly in Sect. 4.1.4 and Sect. 4.1.5.

4.1.4 Scenario PLUME with final projection on PAN

For simulations initialised during the daytime, the mean sensitivities of category C_a (see Table 7) indicate that the PAN formation is mostly influenced by uncertainties in the initial concentrations of MGLY (high importance), followed by OLI, CSL, ALD, ISO, HC8, OLT, DCB, XYL, and HC5 (medium importance). While species TOL, HC3, KET, and OL2 are of little importance for the formation of PAN, the importance of ETH, HCHO, and GLY is negligible. Increasing the initial time alters the sensitivity to CSL, changing its importance from “medium” to “high”, while the sensitivities of the other VOCs remain more or less unchanged (see Fig. 5). Varying the simulation length alters the VOCs’ sensitivities more significantly. Here, the importance of CSL, DCB, MGLY, ALD, and OLI decreases with increasing simulation length (from “high” to “medium”). At the same time, the importance of TOL, XYL, and KET increases from “little” to “medium”. The importance of the alkanes also increases with simulation length (from “little” to “high” for HC8, from “negligible” to “medium” for HC5, and from “negligible” to “little” for HC3 and ETH).

For simulations initialised during the nighttime, Table 7 shows that PAN is most sensitive to CSL and MGLY (high importance). Species of medium importance are ISO, ALD, OLI, HC8, XYL, OLT, and DCB, followed by species of little importance, HC5, TOL, HC3, KET, and OL2. Again, PAN is least sensitive to ETH, GLY, and HCHO (negligible importance). Comparing these results to sensitivities of category C_a , the increase of the mean sensitivity to CSL is most noticeable, while the mean sensitivities of the other VOCs changed very little. CSL is as well the only VOC that shows significant changes for different initial nights, with influence increasing for later initial times. The importance of CSL is furthermore very sensitive to simulation length (see Fig. 5), decreasing from “high” (category C_{b_1}) to “medium” (category C_{b_2}) and inversely from “medium” (category C_{b_3}) to “high” (category C_{b_4}). The sensitivities of the other VOCs also show significant changes with simulation length. Here, the importance of MGLY and ALD decreases (from “high” to “medium”), while the importance of DCB, XYL, TOL, HC3, HC5, KET, HC8, and ETH increases (for DCB, XYL, TOL, HC3, HC5, and KET from “little” to “medium”, for

ETH from “negligible” to “little”, and for HC8 from “little” to “high”).

4.1.5 Scenario URBAN with final projection on PAN

For simulations initialised during the daytime, the calculated initial perturbations show largest mean values for HC8 (high importance), followed by XYL, MGLY, HC5, OLI, DCB, CSL, ISO, KET, OLT, TOL, and HC3 (medium importance). While ALD, OL2, ETH, GLY, and HCHO have smaller mean sensitivities (little importance), none of the VOCs have negligible mean sensitivities (see Table 7). The mean sensitivities of all VOCs are highly dependent upon the specific initial day and the simulation length (see Fig. 5). The later the initial point in time, the higher the importance of CSL (altered from “little” to “medium”) and of the alkanes (changing from “negligible” to “little” for ETH, from “little” to “medium” for both HC3 and HC5, and from “medium” to “high” for HC8). At the same time, the importance of species HCHO, GLY, MGLY, XYL, and DCB increases (from “little” to “negligible” for HCHO and GLY, and from “high” to “medium” for MGLY, DCB, and XYL). With an increase in simulation length, the importance of XYL, DCB, MGLY, OLI, ISO, OLT, and ALD decreases (for MGLY, OLI, and ISO from “high” to “little”, for XYL from “high” to “medium”, and for DCB, OLT, and ALD from “medium” to “little”), whereas the importance of TOL, HC8, CSL, HC5, KET, HC3, and ETH increases (for HC8 and HC5 from “little” to “high” and for TOL, CSL, KET, and HC3 from “little” to “medium”).

For simulations initialised during the nighttime, species CSL and HC8 are of high importance (see Table 7). PAN is furthermore sensitive to XYL, HC5, MGLY, DCB, KET, TOL, HC3, ISO, OLI, and OLT (medium importance), as well as to ALD, OL2, ETH, GLY, and HCHO (little importance). Again, none of the VOCs have negligible importance. The most apparent difference between simulations initialised during the daytime and simulations initialised during the nighttime is a significant increase in the mean sensitivity to CSL. As for simulations initialised during the daytime, the mean sensitivities are highly dependent upon the specific initial time and simulation length (see Fig. 5). Here, the mean sensitivities of the VOCs show the same trends as for simulations initialised during the daytime, but the associated importance is not always the same. By increasing the initial time, the importance of HC8, MGLY, and DCB is altered from “medium” to “high” (HC8) and from “medium” to “little” (MGLY and DCB), while the importance of CSL remains unaltered. Increasing simulation length alters the importance of XYL, OLI, and ISO from “medium” to “little”. Simultaneously, the importance of ETH increases from “negligible” to “little”, whereas CSL remains of high importance.

Scenarios PLUME and URBAN, Adjoint Sensitivity with Final Projection on PAN

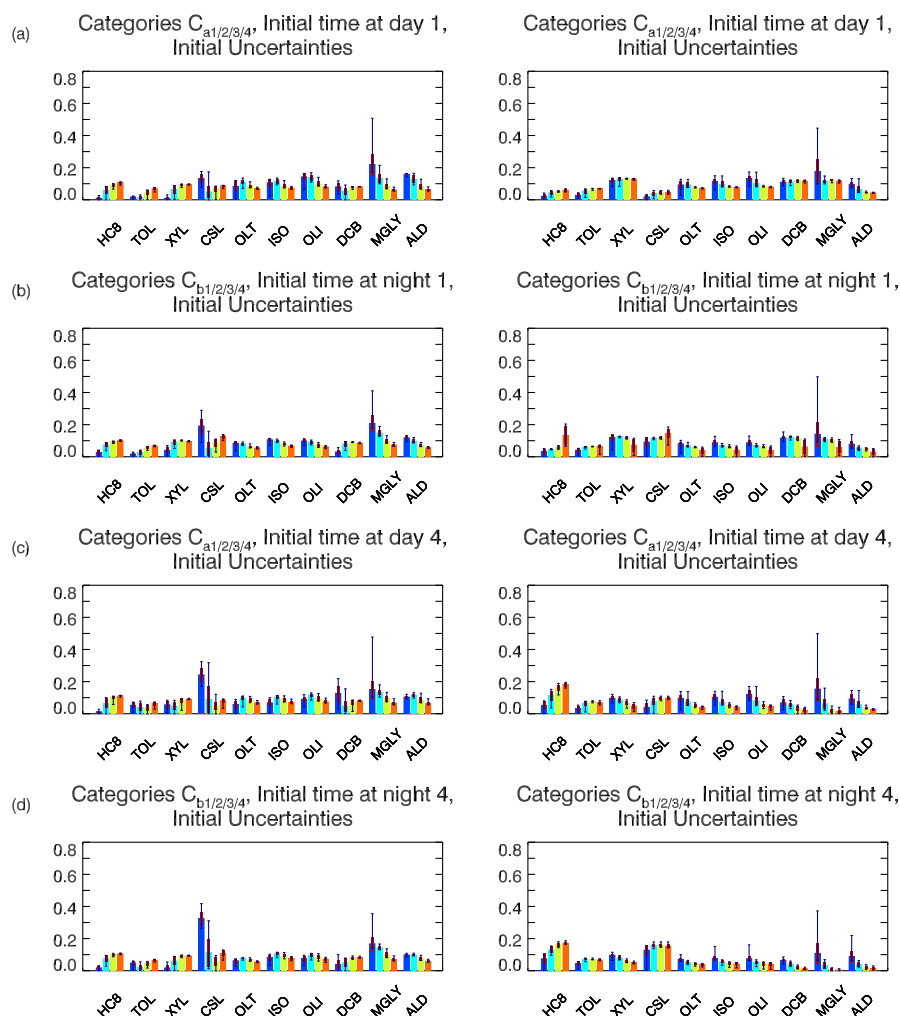


Fig. 5. Statistics of adjoint sensitivities with respect to initial uncertainties for scenario PLUME (left panel column) and scenario URBAN (right panel column) for (a) day 1, (b) night 1, (c) day 4, and (d) night 4. Displayed are results for categories $C_{a1/2/3/4}$ and $C_{b1/2/3/4}$ for the VOCs with the most influence. All panels depict mean impact ($C_{a/b}$: blue bars; $C_{a1/b1}$: blue bars; $C_{a2/b2}$: turquoise bars; $C_{a3/b3}$: green bars; $C_{a4/b4}$: orange bars), minimum/maximum value (dark blue lines) and standard deviation (red bars).

4.1.6 Summary of results for sensitivities with final projection on PAN

The analysis of both scenarios yields the following results:

- The sensitivity of PAN to VOCs is dominated by HC8 (scenario URBAN), MGLY (scenario PLUME), and CSL (both scenarios, for simulations initialised during the nighttime). For both scenarios, PAN is least sensitive to OLT, ETH, HCHO, and GLY and of medium sensitivity to OLI, CSL (only for simulations initialised during the daytime), ISO, HC8 (only for scenario PLUME), MGLY (only for scenario URBAN), OLT, DCB, XYL, HC5, TOL, HC3, and KET.

- By increasing the pollution, the sensitivity of PAN to alkanes and XYL increases, whereas the sensitivity of PAN to ALD and CSL decreases.
- The PAN sensitivity to ETH, HC3, HC5, and HC8 increases with growing simulation length.

The PAN formation potential of the VOCs is dependent on their OH reactivity, carbon number, and carbon bonding. For VOC-sensitive regimes, where the VOC-OH reaction is preferred to the NO_2 -OH reaction, it is to be expected that the sensitivity ranking within a functional group (alkenes, alkanes, aromatics, and carbonyls) is determined by these quantities. For simulations initialised during the daytime, scenario URBAN is VOC sensitive, and VOC degradation starts with

Table 7. Mean impact and standard deviation of VOCs for categories C_a and C_b . Considered are adjoint sensitivities with respect to initial uncertainties and final projection on PAN.

Species	Scenario PLUME, Category C_a	Scenario PLUME, Category C_b	Scenario URBAN, Category C_a	Scenario URBAN, Category C_b
ETH	0.009 ± 0.005	0.009 ± 0.004	0.026 ± 0.026	0.026 ± 0.024
HC3	0.043 ± 0.018	0.040 ± 0.015	0.058 ± 0.035	0.058 ± 0.032
HC5	0.052 ± 0.025	0.049 ± 0.022	0.075 ± 0.047	0.075 ± 0.044
HC8	0.080 ± 0.039	0.074 ± 0.034	0.109 ± 0.063	0.108 ± 0.059
OL2	0.027 ± 0.011	0.023 ± 0.007	0.029 ± 0.011	0.027 ± 0.009
OLT	0.079 ± 0.020	0.068 ± 0.012	0.062 ± 0.027	0.052 ± 0.019
OLI	0.096 ± 0.024	0.082 ± 0.015	0.071 ± 0.033	0.055 ± 0.022
ISO	0.084 ± 0.019	0.089 ± 0.015	0.066 ± 0.029	0.055 ± 0.021
TOL	0.046 ± 0.022	0.042 ± 0.021	0.061 ± 0.018	0.060 ± 0.015
CSL	0.094 ± 0.066	0.131 ± 0.098	0.067 ± 0.028	0.135 ± 0.030
XYL	0.073 ± 0.027	0.073 ± 0.033	0.088 ± 0.037	0.084 ± 0.035
HCHO	0.002 ± 0.001	0.002 ± 0.002	0.013 ± 0.009	0.013 ± 0.009
ALD	0.090 ± 0.028	0.084 ± 0.021	0.046 ± 0.026	0.043 ± 0.031
KET	0.041 ± 0.015	0.039 ± 0.013	0.063 ± 0.036	0.063 ± 0.033
GLY	0.002 ± 0.001	0.003 ± 0.003	0.014 ± 0.010	0.013 ± 0.008
MGLY	0.106 ± 0.051	0.127 ± 0.050	0.081 ± 0.064	0.069 ± 0.060
DCB	0.076 ± 0.031	0.066 ± 0.029	0.071 ± 0.040	0.065 ± 0.043

the VOC-OH reaction. Here, the VOC sensitivities within a functional group are generally ranked according to OH reactivity and carbon number. Yet, the sensitivity to MGLY is highest within the group of carbonyls, even though both OH reactivity and carbon number suggest that the sensitivity to MGLY should be ranked second. Upon regarding the degradation paths of the VOCs, it is seen that MGLY and ALD are the only VOCs in the RADM2 mechanism that form the acetyl radical instantaneously via their reaction with OH (the acetyl radical is needed for the formation of PAN). Combining this characteristic with the high OH reactivity of MGLY leads to its higher-than-anticipated sensitivity. Furthermore, the PAN sensitivity to alkenes is larger than the O_3 and PAN sensitivity to alkenes. This is caused by the fact that alkenes form both ALD and MGLY during their oxidation path (which then form the acetyl radical). For simulations of scenario URBAN initialised during the nighttime, the PAN sensitivity to CSL is highest due to its nighttime reaction with NO_3 . Scenario PLUME shares most of the mentioned characteristics with scenario URBAN. Here, only species ALD, CSL, and ISO show unexpected sensitivity rankings within their functional group. This is caused by the fact that scenario PLUME is not VOC sensitive, and hence the VOC-OH reaction is not always preferred.

The adjoint PAN sensitivities are compared with the aggregated PPCP (photochemical PAN creation potential) of Derwent et al. (the study of Derwent et al., 1998, and the aggregation of POCP/PPCP results are briefly explained in Sect. 4.1.3). Due to the fact that the simulations of Derwent et al. (1998) are based on polluted environments and are initialised during the day, category C_{a2} of scenario URBAN is considered for comparison. Here, PPCP results and

adjoint sensitivities show that PAN has a higher sensitivity to alkenes than ozone. Moreover, the sensitivity ranking of PPCP is in good agreement with the one of adjoint sensitivities (the adjoint sensitivities rank ETH the lowest, followed by OL2, HC3, HC5, TOL, HC8, OLT, ISO, OLI, and XYL), with the one exception that the adjoint sensitivity approach ranks species OL2 lower and species TOL higher than the PPCP approach.

4.2 Relative error growth of initial uncertainties

So far, the investigation of the scenarios has been carried out without any weighting of the perturbations. Hence, current results give insight into the influence of the kinetics without considering typical mixing ratios. This leads to the error growth being predominated by more abundant species. In order to evaluate relative influences of chemical compounds the previously considered scenarios of Sect. 4.1 are re-investigated with relative error growths.

Since grouped relative singular vectors weight the influence of chemical compounds by their typical concentrations, the main questions raised in Sect. 4.1 are modified to:

- To what extent does the relative uncertainty in the initial concentration of an individual VOC contribute to relative uncertainties in the O_3 and PAN evolution?
- What is the degree of dependence of these results on the scenario selected and the time interval chosen, respectively?

The current section aims to answer these questions by analysing projected relative singular vectors. In order to get a

comprehensive picture of the time dependence of these relative changes, a complete TSVD of optimal projected relative singular vectors has been calculated for scenarios PLUME and URBAN. The categorisations and statistics introduced in Sect. 4.1 are applied to analyse the TSVD results. The associated results for scenarios PLUME and URBAN are collected in Fig. 6 and Table 8. At a glance they reveal, that the relative O₃ and PAN sensitivity to the VOCs is seemingly independent of the initial time of the simulation (in terms of simulations initialised during the daytime versus simulations initialised during the nighttime). In the following subsections, the most important findings per scenario are described and subsequently integrated into the associated measurement strategy.

4.2.1 Scenario PLUME

For scenario PLUME, most of the O₃ and PAN sensitivity to VOCs is caused by HC3, KET, and ALD (high importance). Other relevant VOCs are HC5 (medium importance) as well as TOL, HC8, ETH, and OL2 (little importance). The O₃ and PAN sensitivity to HCHO, DCB, MGLY, OLI, and GLY is of negligible importance (see Table 8). Due to its zero concentration, the species ISO has no relative influence on the O₃ and PAN formation. XYL, CSL, and OLT have no fixed positions in this sensitivity ranking due to significant differences between their sensitivities of simulations initialised during the daytime and their sensitivities of simulations initialised during the nighttime. Nevertheless, the influence of XYL may be ranked as being of medium importance, and the importance of CSL and OLT may be classed as “negligible”. The mean sensitivities of all VOCs show little change for different initial days or nights. Only the influence of KET changes significantly when altering the initial days or nights (increasing influence with time). In contrast, the O₃ and PAN sensitivity to most of the VOCs is strongly dependent on simulation length (see Fig. 6). While HC3, KET, HC5, HC8, ETH, and OL2 gain influence with increasing simulation length, species ALD, HCHO, CSL, MGLY, and OLI simultaneously lose influence. Associated changes in importance lead from little to medium importance for HC5 and HC8, from negligible to less importance for ETH and OL2, from high to medium importance for ALD, and from little to negligible importance for HCHO and CSL.

4.2.2 Scenario URBAN

The most prominent VOCs of scenario URBAN are HC3 and KET (high importance), followed by HC5, TOL, and HCHO (medium importance). Even though O₃ and PAN are less sensitive to species HC8, ETH, and OL2 (little importance), their effect is still considerable. VOCs of negligible importance are OLT, MGLY, OLI, and GLY (see Table 8). Again, ISO has no relative influence on the O₃ and PAN formation. The positions of species ALD, XYL, DCB, and CSL are not

fixed within this sensitivity ranking, since their mean sensitivities show significant differences between category C_a and category C_b. Still, the importance of these four species is constant for both categories (ALD is of medium importance, XYL is of little importance, and DCB and CSL are of negligible importance). The VOC sensitivities of scenario URBAN are highly dependent upon both the specific initial day or night of the simulation and the simulation length (see Fig. 6). With an increase in initial time, species KET and HC3 gain influence, while species ALD, HCHO, TOL, XYL, HC5, HC8, and OL2 lose influence. Here, ALD and HCHO change importance from “high” to “little”, TOL, XYL, HC5, and HC8 from “medium” to “little”, and OL2 from “little” to “negligible”. Increasing simulation length, on the other hand, leads to increased influence of HC3, KET, HC5, HC8, and ETH and decreased influence for ALD, HCHO, DCB, OLT, MGLY, OLI, and GLY. In the process, the importance of HC5, ETH, ALD, HCHO, and DCB changes (from “little” to “medium” for HC8, from “negligible” to “little” for ETH, from “high” to “little” for ALD, from “medium” to “little” for HCHO, and from “little” to “negligible” for DCB).

4.2.3 Summary of sensitivity results and associated measurement strategy

Examination of the statistical results of both scenarios, for all days and nights, yields the following notable findings:

- Within the VOC group, the O₃ and PAN sensitivity to species HC3 and KET is high. O₃ and PAN are of medium sensitivity to HC5 and of little sensitivity to HC8, ETH, and OL2. The mean sensitivity of O₃ and PAN to GLY, MGLY, OLI, CSL, DCB, and OLT may be neglected (whereas this is not the case for the mean sensitivity to DCB and CSL for short simulation intervals).
- The O₃ and PAN sensitivity to species ALD, XYL, TOL, and HCHO is dependent on the degree of pollution. For scenario URBAN, O₃ and PAN are highly sensitive to ALD, of medium sensitivity to XYL, and of little sensitivity to TOL. The sensitivity of O₃ and PAN to HCHO is negligible. In less-polluted air (scenario PLUME), the O₃ and PAN sensitivity to ALD, TOL, and HCHO is of medium importance, while the sensitivity to XYL is of little importance.
- The O₃ and PAN sensitivity to all alkanes (ETH, HC3, HC5, and HC8) increases with increasing simulation length.

Considering the final perturbations of the calculated sensitivities, the entries corresponding to PAN are found to be about one magnitude higher than the entries corresponding to O₃. Thus, the sensitivity results are dominated by the relative PAN sensitivity. Regarding the absolute sensitivity results of the adjoint calculations with final projection on PAN

Table 8. Mean impact and standard deviation of VOCs for categories C_a and C_b . Considered are optimal projected relative singular vectors with respect to initial uncertainties and final projection on O_3 and PAN.

Species	Scenario PLUME, Category C_a	Scenario PLUME, Category C_b	Scenario URBAN, Category C_a	Scenario URBAN, Category C_b
ETH	0.013 ± 0.006	0.013 ± 0.005	0.020 ± 0.017	0.022 ± 0.017
HC3	0.320 ± 0.095	0.325 ± 0.071	0.308 ± 0.070	0.305 ± 0.076
HC5	0.070 ± 0.029	0.074 ± 0.025	0.077 ± 0.022	0.076 ± 0.022
HC8	0.032 ± 0.015	0.041 ± 0.017	0.039 ± 0.015	0.041 ± 0.013
OL2	0.011 ± 0.004	0.012 ± 0.005	0.020 ± 0.015	0.019 ± 0.015
OLT	0.003 ± 0.002	0.008 ± 0.005	0.004 ± 0.004	0.004 ± 0.004
OLI	$8 \times 10^{-4} \pm 6 \times 10^{-4}$	$4 \times 10^{-4} \pm 5 \times 10^{-4}$	$7 \times 10^{-4} \pm 8 \times 10^{-4}$	$2 \times 10^{-4} \pm 6 \times 10^{-4}$
TOL	0.041 ± 0.023	0.042 ± 0.020	0.063 ± 0.034	0.061 ± 0.032
CSL	0.006 ± 0.007	$2 \times 10^{-4} \pm 9 \times 10^{-4}$	0.002 ± 0.001	$2 \times 10^{-5} \pm 1 \times 10^{-4}$
XYL	0.011 ± 0.007	0.033 ± 0.018	0.030 ± 0.032	0.047 ± 0.040
HCHO	0.007 ± 0.007	0.007 ± 0.009	0.053 ± 0.047	0.052 ± 0.048
ALD	0.207 ± 0.153	0.183 ± 0.122	0.081 ± 0.086	0.068 ± 0.091
KET	0.271 ± 0.062	0.259 ± 0.057	0.294 ± 0.121	0.303 ± 0.122
GLY	$1 \times 10^{-5} \pm 2 \times 10^{-4}$	$2 \times 10^{-5} \pm 6 \times 10^{-5}$	$2 \times 10^{-4} \pm 2 \times 10^{-4}$	$7 \times 10^{-5} \pm 7 \times 10^{-5}$
MGLY	0.002 ± 0.003	0.001 ± 0.001	0.002 ± 0.004	$5 \times 10^{-4} \pm 9 \times 10^{-4}$
DCB	0.004 ± 0.004	$0.002 \pm 9 \times 10^{-4}$	0.006 ± 0.005	0.002 ± 0.002

(see Sect. 4.1.4, Sect. 4.1.5, and Sect. 4.1.6), it becomes apparent that the sensitivity results of the projected relative singular vectors are reproducible through multiplication of the adjoint sensitivities with the initial concentrations of the associated simulation. Consequently, the projected relative singular vectors are dependent on the underlying kinetics, the initial concentrations and the final concentrations of the associated simulation.

With the help of the sensitivity results above, an optimal measurement strategy is established. Measurement strategies where at least one species can be omitted from the measurement, for all considered measurement times and forecast intervals, are desirable. Since desirable measurement strategies are advantageous to the limiting of instrumentation, they are hereinafter referred to as *profitable measurement strategies*. It is of further advantage to find not only profitable but also scenario-independent measurement strategies, since scenario-independent measurement strategies allow for a single measurement set-up and are therefore, e.g., suitable for campaign measurements. In this work, a profitable and scenario-independent measurement strategy omits measurement of VOCs that are negligible for all categories of both scenarios. Considering initial values measurement, focus is placed on measurement of HC3, KET, HC5, HC8, ETH, OL2, ALD, XYL, TOL, HCHO, DCB, and CSL; measurement of ISO, GLY, OLI, MGLY, and OLT may be omitted. For both scenarios, the summarised mean sensitivities of the prioritised species account for more than 99% of the O_3 and PAN sensitivity to VOCs. Since the criterion for omission from measurement is carefully determined, it may be expanded to include furthermore the omission of species that are i) of negligible importance for categories C_a and C_b and

ii) of negligible or little importance for all other categories. This allows for omission of measurement of species CSL and DCB. Here, despite omission of the aforementioned species, the summarised mean sensitivities of the prioritised species still account for more than 96% of the O_3 and PAN sensitivity to VOCs.

4.3 Relative error growth of emission factors

In this section, the VOC dependence of the O_3 and PAN evolution is analysed with respect to emissions factors. Since uncertainties in emission factors are typically given as relative errors, only relative error growths are used here. More precisely, the main questions to be faced are the following:

- To what extent does the relative uncertainty in the emission strength of an individual VOC contribute to relative uncertainties in the O_3 and PAN evolution?
- What is the degree of dependence of these results on the scenario selected and the time interval chosen, respectively?

In order to provide insight into these aspects, the TSVD of grouped relative singular vectors is applied to scenarios PLUME and URBAN. The results of the TSVD calculations are categorised and analysed according to the categories and statistics introduced in Sect. 4.1. Figure 7 and Table 9 display the results of the categorisation for both scenarios. In the following subsections, scenarios PLUME and URBAN are analysed separately to provide a more comprehensive picture of the results. Since the emission strengths of GLY, DCB, CSL, ISO, and MGLY are zero for both scenarios,

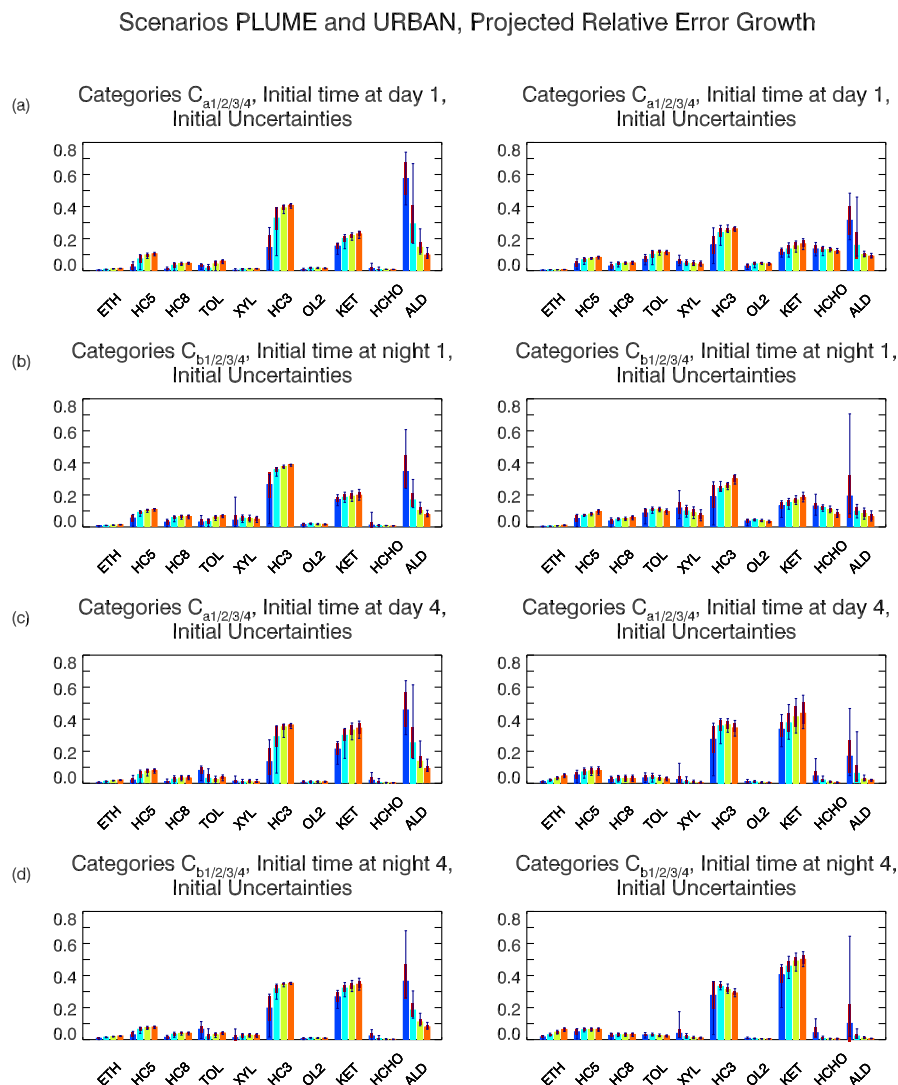


Fig. 6. Statistics of optimal projected relative singular vectors with respect to initial uncertainties for scenario PLUME (left panel column) and scenario URBAN (right panel column) for (a) day 1, (b) night 1, (c) day 4, and (d) night 4. Displayed are results for categories $C_{a1/2/3/4}$ and $C_{b1/2/3/4}$ for the VOCs with the most influence. All panels depict mean impact ($C_{a/b}$: blue bars; $C_{a1/b1}$: blue bars; $C_{a2/b2}$: turquoise bars; $C_{a3/b3}$: green bars; $C_{a4/b4}$: orange bars), minimum/maximum value (dark blue lines) and standard deviation (red bars).

these compounds have no emission impact on the formation of O_3 and PAN.

4.3.1 Scenario PLUME

Within scenario PLUME the O_3 and PAN sensitivity to emission factors of VOCs is almost entirely determined by HC3, OLI (high importance), XYL, HC5 (medium to high importance), TOL, HC8, and OLT (medium importance). To KET, OL2, ALD (little importance), ETH, and HCHO (negligible importance), O_3 and PAN show little sensitivity (see Table 9). The sensitivity ranking of scenario PLUME is only slightly dependent upon the specific initial day or night of the simulation, but shows a high dependence upon simulation

length (see Fig. 7). Only the mean sensitivity to TOL shows significant changes with increasing initial time (altering its importance from “medium” to “high”), whereas the mean sensitivities to species HC3, OLI, XYL, HC5, TOL, HC8, ALD, ETH, and HCHO show significant changes with increasing simulation length (increasing mean sensitivities for all alkanes and decreasing sensitivities for OLI, XYL, TOL, ALD, and HCHO). Here, changes only occur in the importance of OLI, XYL (from “high” to “medium” for both), HC5 (from “little” to “high”), TOL (from “high” to “little”), HC8 and ALD (from “medium” to “little”).

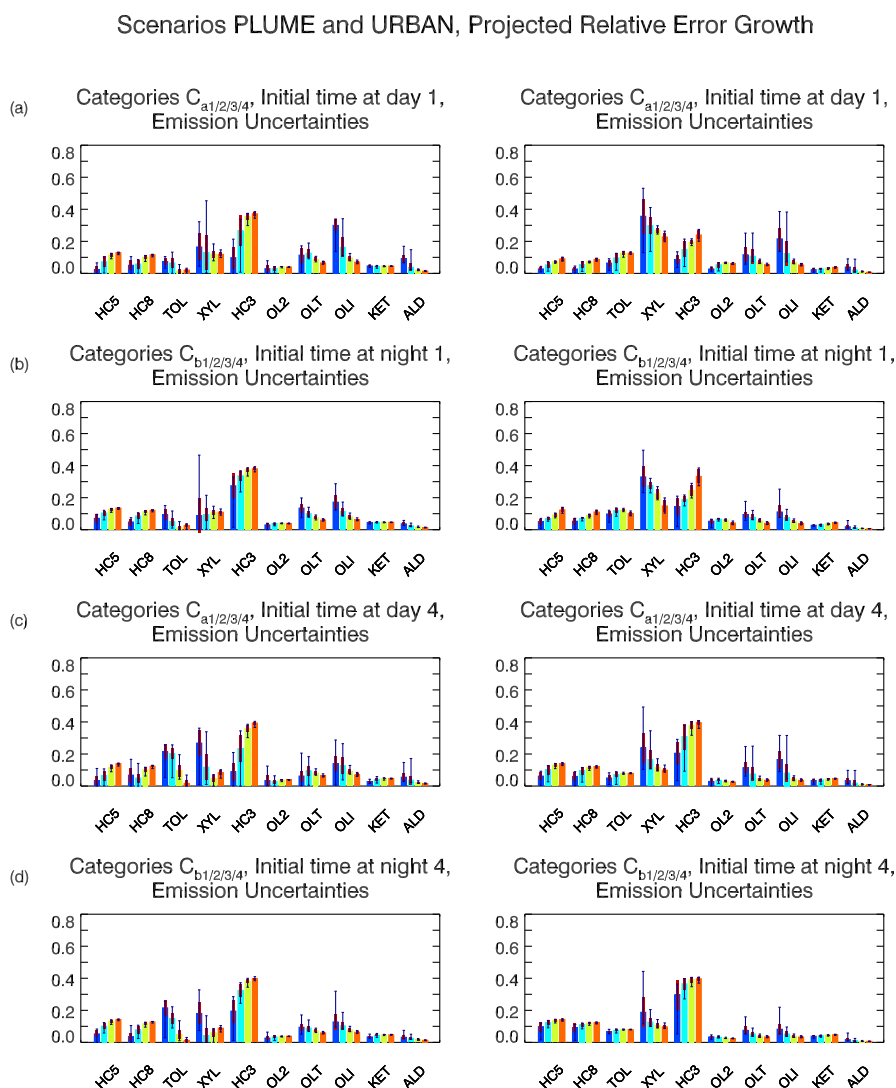


Fig. 7. Statistics of optimal projected relative singular vectors with respect to emission uncertainties for scenario PLUME (left panel column) and scenario URBAN (right panel column) for (a) day 1, (b) night 1, (c) day 4, and (d) night 4. Displayed are results for categories $C_{a_{1/2/3/4}}$ and $C_{b_{1/2/3/4}}$ for the VOCs with the most influence. All panels depict mean impact ($C_{a/b}$: blue bars; C_{a_1/b_1} : blue bars; C_{a_2/b_2} : turquoise bars; C_{a_3/b_3} : green bars; C_{a_4/b_4} : orange bars), minimum/maximum value (dark blue lines) and standard deviation (red bars).

4.3.2 Scenario URBAN

For scenario URBAN, the mean sensitivities of categories C_a and C_b (see Table 9) suggest that the O_3 sensitivity to emission factors of VOCs is dominated by species HC3, XYL, and HC5 (high importance) and HC8, TOL, OLI, and OLT (medium importance). VOCs of little importance are OL2, KET, and ALD, while ETH and HCHO have the least sensitivity among the VOCs (negligible importance). The results of scenario URBAN are dependent upon both the specific initial time and the simulation length (see Fig. 7). With increasing initial time, the mean sensitivities of HC5, HC8, and HC3 increase (HC5 and HC8 change from having medium importance to having high importance), and the mean sensi-

ivities of TOL, XYL, OL2, OLT, and OLI decrease (changes in importance from “high” to “medium” for TOL and from “medium” to “little” for OLT and OLI). The mean sensitivities of these species are furthermore influenced by simulation length, but here the influence of HC3, HC5, HC8, ETH, and TOL increases with simulation length (changes in importance from “medium” to “high” for HC5 and HC8, and from “negligible” to “little” for ETH). The influence of XYL, OLI, OLT, ALD, and HCHO simultaneously decreases (OLI and OLT change their importance from “high” to “little”, and ALD and HCHO from “little” to “negligible”).

Table 9. Mean impact and standard deviation of VOCs for categories C_a and C_b . Considered are optimal projected relative singular vectors with respect to emissions uncertainties and final projection on O_3 and PAN.

Species	Scenario PLUME, Category C_a	Scenario PLUME, Category C_b	Scenario URBAN, Category C_a	Scenario URBAN, Category C_b
ETH	0.005 ± 0.002	0.006 ± 0.002	0.007 ± 0.004	0.008 ± 0.004
HC3	0.298 ± 0.115	0.334 ± 0.076	0.286 ± 0.107	0.305 ± 0.096
HC5	0.097 ± 0.044	0.108 ± 0.034	0.100 ± 0.039	0.108 ± 0.034
HC8	0.089 ± 0.038	0.090 ± 0.036	0.092 ± 0.032	0.099 ± 0.027
OL2	0.035 ± 0.015	0.036 ± 0.008	0.038 ± 0.009	0.041 ± 0.015
OLT	0.082 ± 0.028	0.086 ± 0.026	0.066 ± 0.039	0.056 ± 0.024
OLI	0.112 ± 0.058	0.101 ± 0.042	0.078 ± 0.062	0.058 ± 0.032
TOL	0.095 ± 0.081	0.081 ± 0.073	0.086 ± 0.023	0.091 ± 0.019
XYL	0.111 ± 0.079	0.090 ± 0.059	0.188 ± 0.096	0.179 ± 0.089
HCHO	0.002 ± 0.003	0.001 ± 0.001	0.006 ± 0.006	0.006 ± 0.006
ALD	0.030 ± 0.025	0.023 ± 0.011	0.014 ± 0.014	0.010 ± 0.007
KET	0.043 ± 0.007	0.045 ± 0.005	0.039 ± 0.015	0.040 ± 0.007

4.3.3 Summary of sensitivity results and associated measurement strategy

The most notable features of scenarios PLUME and URBAN are as follows:

- HC3 is of high importance for the O_3 and PAN sensitivity to emission factors, whereas species HC8, TOL, and OLT are of medium importance. The O_3 and PAN sensitivity to emission factors of KET, OL2, and ALD is of little importance, and the emission factors of HCHO and ETH are of negligible importance. Species ISO, CSL, GLY, MGLY, and DCB are not emitted and therefore have no emission impact on the formation of O_3 and PAN.
- The O_3 and PAN sensitivity to emission factors of XYL and HC5 is of high importance in polluted air (scenario URBAN), but only of medium importance in moderately polluted air (scenario PLUME). Contrarily, emissions of OLI are of high importance within scenario PLUME and of medium importance within scenario URBAN.
- The O_3 and PAN sensitivity to alkane emissions (ETH, HC3, HC5, and HC8) increases with increasing simulation length.

As before, the PAN entries of the final perturbations of the calculated sensitivities show highest values (about one magnitude higher than the entries corresponding to O_3). Hence, the sensitivity results are predominated by the relative PAN sensitivity. Therefore, it is possible to approximately reproduce the presented sensitivity ranking by multiplying the adjoint sensitivities (see Sect. 4.1.4, Sect. 4.1.5, and Sect. 4.1.6) with the associated VOC emission strengths (see Table 3).

Since both scenarios list the same species as negligible, a profitable and scenario-independent measurement strategy is

readily attainable. The associated measurement strategy includes measurement of HC3, XYL, OLI, HC5, TOL, HC8, OLT, KET, OL2, and ALD. Measurement of HCHO and ETH may be omitted. The summarised mean sensitivities of the prioritised species account for more than 98 % of the O_3 and PAN sensitivity to VOC emissions in both scenarios. When KET and OL2 are additionally omitted from emission measurement (both are of little importance in all considered categories), more than 90 % of the O_3 and PAN sensitivity to VOCs is still covered.

5 Summary and discussion

In the present work, singular vector analysis is applied to atmospheric chemical modelling in order to optimise the measurement configuration of chemical compounds. Initial chemical values and emissions are investigated as target variables. The application of emissions as target variables is a novel feature in the field of targeted observations. Here, the emission factor approach of Elbern et al. (2007) is adopted, in which the amplitudes of the diurnal profile of emission rates are analysed. In contrast to emissions, these amplitudes are time invariant and their consideration allows for performing only a single singular vector calculation per time interval.

Special arrangements are made to adequately address problems particular to atmospheric chemistry. Error growths are extended to allow for projected target variables, not only at final time but also at initial time. This grants a choice between different operators and, moreover, between different projections. A flexible algorithm is obtained.

As a first step, the adapted theory of singular vector analysis is applied to the RADM2 mechanism in a box model context. The influence of individual VOCs on the O_3 and PAN formation is investigated. VOCs are directly emitted precursors of the secondary pollutants O_3 and PAN. By forming the peroxy radicals, RO_2 , and the acetyl radical, CH_3CO , in their

oxidation sequence, the VOCs contribute to the O₃ and PAN formation, respectively. This oxidation sequence is initiated by the VOC-OH reaction. Since the VOCs compete against each other for available OH radicals, it is expected that those VOCs with higher OH reactivity, higher accessibility (associated with their carbon bonding), and larger carbon numbers will have a larger impact on the O₃ and PAN formation. With the help of singular vector analysis, this hypothesis is tested for two photochemical regimes that feature moderately polluted and polluted air. The chosen box model context gives the opportunity to validate the newly adapted singular vector technique and furthermore allows for the calculation of a comprehensive set of simulations with different metrics. Since chemical simulations are strongly dependent on the location of the start time within the diurnal cycle, as well as on the length of the simulation, it is of special interest to conduct singular vector calculations for different simulation intervals. In order to capture the time-dependent features of the results, the calculations are categorised into simulations initialised during the daytime and simulations initialised during the nighttime. With a further subcategorisation into four different simulation length intervals, the development of the chemical sensitivities can be adequately covered.

The absolute influence of each VOC on the O₃ and PAN formation is calculated with the projected error growth. The conducted singular vector analyses reveal that the resulting VOC sensitivities can, for the most part, indeed be explained by OH reactivity, carbon bonds and carbon number. This finding is most pronounced in regimes where the OH reaction with VOCs is preferred (i.e. simulations of the highly polluted regime that are initialised during the daytime), proving that the singular vector algorithm performs well. In the case of the PAN sensitivity to VOCs, the results are, moreover, dependent on each VOC's ability to form the acetyl radical CH₃CO. In regimes where the VOC-OH reaction is not clearly preferred, the reactivities of the VOCs with other species may as well play an important role. Here, the NO₃ reactivity is of importance for simulations initialised during the nighttime (for both chemical regimes). In the moderately polluted regime, several other reactivities may play an important role in determining the O₃ and PAN sensitivity to an individual VOC. Overall, the singular vectors identify DCB, MGLY, and XYL as the species with the most absolute influence, followed by the species CSL, TOL, ALD, OLI, HC8, ISO, OLT, KET, HC3, and HC5. Species ETH and HCHO are of negligible importance. Based on OH reactivity, carbon bonds, and carbon number, some of these sensitivity rankings may be obvious, but particularly the high importance of CSL and ALD is not a trivial result. The identified sensitivities are not static, but depend on the underlying simulation lengths. Especially alkanes show increasing sensitivities with increasing simulation length. Both the time dependence of the sensitivities and the sensitivity order of the VOCs are in good agreement with previous studies concerning the ozone and PAN formation potential of VOCs (Butler et al., 2011;

Derwent et al., 1998). The studies of Butler et al. (2011) and Derwent et al. (1998) are, however, based on the Master Chemical Mechanism, which is a highly detailed chemical mechanism that contains no empirical lumping. Hence, their results reveal the possible range of sensitivity results within one lumped VOC species of RADM2 (e.g. the results of Derwent et al., 1998, show variations of the calculated photochemical ozone creation potentials from 29.0 to 49.4 for species associated with HC8). This kind of information is lost when conducting studies with mechanisms that employ empirical lumping.

Singular vector analysis is furthermore applied to calculate relative sensitivities. Here, the sensitivity results determine HC3 and KET to be the species with largest relative influence on the O₃ and PAN formation, followed by HC5, HC8, ETH, and OL2. While the mean sensitivity of O₃ and PAN to GLY, MGLY, OLI, CSL, DCB, and OLT is of negligible importance, the relative sensitivity to species ALD, XYL, TOL, and HCHO is dependent on the underlying chemical regime. In highly polluted environments, the relative importance of ALD is high, XYL is of medium importance, and species TOL and HCHO are of little and of negligible importance, respectively. Contrarily, the sensitivities of ALD, TOL, and HCHO are of medium importance in moderately polluted air, and the sensitivity to XYL is of little importance. These rankings of relative sensitivities may be roughly approximated by multiplying the absolute sensitivities with the initial species concentrations. In the same manner, multiplying the absolute sensitivities with the emission strength yields an approximate ranking of the sensitivities with respect to emissions. Accordingly, the projected singular vectors with respect to emissions reveal that emissions of HC3 are of high importance, whereas the emissions of HC8, TOL, and OLT are of medium importance for the O₃ and PAN formation. Emissions of XYL and HC5 are of high importance in polluted air, but only of medium importance in moderately polluted air. Contrarily, emissions of OLI are of high importance in moderately polluted air and of medium importance in polluted air. O₃ and PAN are less sensitive to the emissions of KET, OL2, and ALD, and the O₃ and PAN sensitivity to emission factors of HCHO and ETH is negligible. Species ISO, CSL, GLY, MGLY, and DCB are not emitted and therefore have no impact.

Furthermore, the error growths of the calculated singular vectors are compared with the error growths of adjoint sensitivities. Results show that the error growths of the optimal singular vectors are superior to those of the adjoint sensitivities for each conducted simulation. Table 10 gives a brief overview of the summarised results. Due to a very limited number of degrees of freedom at initial and final time (determined by the number of VOCs and O₃ and PAN, respectively), the differences between singular vectors and adjoint sensitivity are quite small. Based on numerous grid points, chemical transport models present an increased number of degrees of freedom, allowing for larger differences between

Table 10. Summarised error growths of scenario PLUME and URBAN. Considered are relative error growths (REGs) with respect to initial and emission uncertainties (emission uncertainties are denoted with a tilde). The error growths are obtained by inserting perturbation vectors of unit length resulting from singular vector analysis (SVA) and adjoint sensitivity (ADJ). The chosen final projections are denoted in brackets. Simulation intervals were chosen according to the temporal configuration of the TSVD, and the results of all simulations were summed up.

Scenario PLUME		
Initial perturbations are obtained by ↓	Sum of REGs ($\sum g_r$)	Sum of REGs ($\sum \tilde{g}_r$)
SVA (O ₃ & PAN)	735.976	2326.60
ADJ (O ₃ & PAN)	732.417	2325.66
ADJ (PAN)	735.924	2323.66
ADJ (O ₃)	617.783	2294.96

Scenario URBAN		
Initial perturbations are obtained by ↓	Sum of REGs ($\sum g_r$)	Sum of REGs ($\sum \tilde{g}_r$)
SVA (O ₃ & PAN)	3559.56	6870.76
ADJ (O ₃ & PAN)	3557.26	6859.35
ADJ (PAN)	3554.00	6796.09
ADJ (O ₃)	3383.35	6296.24

singular vector analysis and adjoint sensitivity. While singular vector calculations promise more benefits when applied to a full 3-dimensional model, the validation of results becomes more difficult due to complex interactions between chemistry and meteorology. Therefore, the present study is carried out with a box model, allowing for validation of the algorithm.

In a second step, the identified sensitivities are applied to address the problem of targeted observations. Sensitivity-based measurement strategies are proposed, determining which species may be omitted from measurement and which species need to be measured to reduce the uncertainty of the O₃ and PAN forecast. Here, a measurement strategy is considered profitable if measurement of at least one species can be omitted for all considered measurement times and forecast intervals. Such profitable measurement strategies allow for reduced instrumentation and are therefore desirable. The sensitivity results of the projected relative error growths suggest that a scenario-independent and profitable measurement strategy may be achieved by omitting ISO, GLY, OLI, MGLY, OLT, CSL, and DCB from initial value measurement, and HCHO, ETH, KET, and OL2 from emission measurement. Accordingly, species HC3, KET, HC5, HC8, ETH, OL2, ALD, XYL, TOL, and HCHO and species HC3, XYL, OLI, HC5, TOL, HC8, OLT, and ALD should be prioritised for initial value and emission measurement, respectively. The summarised mean sensitivities of the prioritised species still account for more than 96 % and 90 % of the O₃ and PAN sen-

sitivity to VOCs, respectively. The rather high profitability of a regime-independent measurement strategy for the projected relative error growth with respect to initial values is caused by the relatively stable concentration ratios of species. While the analysed chemical regimes (scenario PLUME and scenario URBAN as proposed by Poppe et al. (2001)) consider idealised conditions, they are still able to represent the main differences between the VOC chemistry in polluted and in moderately polluted chemical environments. Thus their application for identifying the key VOCs is worthwhile. Comparing the proposed singular-vector-based measurement configuration with the common real-world technique of measuring OVOCs (HCHO, KET, GLY, MGLY, DCB), it can readily be seen that the OVOC-measurement configuration prioritises species for measurement that are found to be less important or negligible for polluted environments. However, according to the singular vector calculations, observations of both the alkanes and the aromatics cover nearly all of the prioritised species.

6 Conclusions

In summary, it can be stated that singular vector analysis is a powerful tool that can be used effectively to identify sensitive chemical species. In the present work, the method is applied to analyse the time and regime dependence of the O₃ and PAN sensitivity to VOCs for two different scenarios. With the help of three different types of error growth, different chemical and temporal regimes are characterised. The multitude of indicators that determine the sensitivity of O₃ and PAN highlights the value of the singular vector technique. With the help of mean sensitivities, each analysed VOC is ranked to be of high, medium, little or negligible importance for the O₃ and PAN formation. This priority ranking of species remains relatively stable, irrespective of scenario. It allows for the effective application of time-invariant targeted observations by omission of measurement of several less sensitive compounds. The resulting measurement strategies illustrate the effectiveness of singular vectors, as well as their potential for identifying optimal instrumentation strategies.

Supplementary material related to this article is available online at: <http://www.atmos-chem-phys.net/13/5063/2013/acp-13-5063-2013-supplement.pdf>.

Acknowledgements. We thank Dirk Poppe for providing the tropospheric chemistry scenarios, his helpful advice on their implementation as well as for numerous discussions. We also thank Tim Butler, Pierre de Wet, Jörg Schwinger and three anonymous referees for their valuable and appreciated feedback on the manuscript.

This study was funded by the virtual institute for Inverse Modelling of the Atmospheric Chemical COMposition (IMACCO) in

the framework of the Helmholtz-Impuls- und Vernetzungsfonds under grant VH-VI-117. This support is gratefully acknowledged.

All simulations utilised the Kinetic PreProcessor (KPP) developed by Valeriu Damian, Adrian Sandu, Mirela Damian, Florian A. Potra and Gregory R. Carmichael. Many thanks to Adrian Sandu for his technical assistance.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: T. Butler

References

- Barkmeijer, J., Gijzen, M. V., and Bouttier, F.: Singular vectors and estimates of the analysis-error covariance metric, *Q. J. Roy. Meteorol. Soc.*, 124, 1695–1713, 1998.
- Buizza, R. and Montani, A.: Targeting Observations Using Singular Vectors, *J. Atmos. Sci.*, 56, 2965–2985, 1999.
- Buizza, R. and Palmer, T. N.: The Singular-Vector Structure of the Atmospheric Global Circulation, *J. Atmos. Sci.*, 52, 1434–1456, 1993.
- Buizza, R., Cardinali, C., Kelly, G., and Thepaut, J. N.: The value of targeted observations, *ECMWF Newsletter*, 111, 11–20, 2007.
- Butler, T. M., Lawrence, M. G., Taraborelli, D., and Lelieveld, J.: Multi-day ozone production potential of volatile organic compounds calculated with a tagging approach, *Atmos. Environ.*, 45, 4082–4090, 2011.
- Carter, W. P. L.: Development of Ozone Reactivity Scales for Volatile Organic Compounds, *J. Air Waste Manage.*, 44, 881–899, 1994.
- Derwent, R. G., Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: Photochemical Ozone Creation Potentials for Organic Compounds in Northwest Europe calculated with a Master Chemical Mechanism, *Atmos. Environ.*, 32, 2429–2441, 1998.
- Dunker, A. M., Ortman, J. P., Wilson, G. M., and Yarwood, G.: The decoupled direct method for sensitivity analysis in a three-dimensional air quality model—implementation, accuracy, and efficiency, *Environ. Sci. Technol.*, 36, 2965–76, 2002.
- Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007.
- Hairer, E. and Wanner, G.: *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer, Berlin, Germany, 1991.
- Hakami, A., Harley, R. A., Milford, J. B., Odman, M. T., and Russell, A. G.: Regional, three-dimensional assessment of the ozone formation potential of organic compounds, *Atmos. Environ.*, 38, 121–134, 2004.
- Jenkin, M. E., Saunders, S. M., and Pilling, M. J.: The tropospheric degradation of volatile organic compounds: A protocol for mechanism development, *Atmos. Environ.*, 31, 81–104, 1997.
- Khattatov, B. V., Gille, J., Lyjak, L., Brasseur, G., Dvortsov, V., Roche, A., and Waters, J.: Assimilation of photochemically active species and a case analysis of UARS data, *J. Geophys. Res.*, 104, 18715–18738, 1999.
- Kirchner, F., Jeanneret, F., Clappier, A., Krüger, B., van den Bergh, H., and Calpini, B.: Total VOC reactivity in the planetary boundary layer: 2. A new indicator for determining the sensitivity of the ozone production to VOC and NO_x, *J. Geophys. Res.*, 106, 3095–3110, 2001.
- Kuhn, M., Bultjes, P. J. H., Poppe, D., Simpson, D., Stockwell, W. R., Anderson-Sköld, Y., Baart, A., Das, M., Fiedler, F., Hov, Ø., Kirchner, F., Makar, P. A., Milford, J. B., Roemer, M. G. M., Ruhnke, R., Strand, A., Vogel, B., and Vogel, H.: Intercomparison of the gas-phase chemistry in several chemistry and transport models, *Atmos. Environ.*, 32, 693–709, 1998.
- Lehoucq, R. B., Sorensen, D. C., and Yang, C.: *ARPACK Users' Guide: Solution of Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, Philadelphia, 1998.
- Liao, W., Sandu, A., Carmichael, G. R., and Chai, T.: Singular Vector Analysis for Atmospheric Chemical Transport Models, *Month. Weather Rev.*, 134, 2443–2465, 2006.
- Lorenz, E. N.: A study of the predictability of a 28 variable atmospheric model, *Tellus*, 17, 321–333, 1965.
- Middleton, P., Stockwell, W. R., and Carter, W. P. L.: Aggregation and analysis of volatile organic compound emissions for regional modeling, *Atmos. Environ.*, 24A, 1107–1133, 1990.
- Mises, R. V. and Pollaczek-Geiringer, H.: *Praktische Verfahren der Gleichungsaufösung*, *Z. Angew. Math. Mech.*, 9, 58–77, 1929.
- Montani, A., Thorpe, A. J., Buizza, R. and Undén, P.: Forecast skill of the ECMWF model using targeted observations during FAS-TEX, *Q. J. R. Meteorol. Soc.*, 125, 3219–3240, 1999.
- Navon, I. M., Zou, X., Derber, J., and Sela, J.: Variational data assimilation with an adiabatic version of the NMC spectral model, *Mon. Weather Rev.*, 120, 1433–1446, 1992.
- Palmer, T. N.: Predictability of the atmosphere and oceans: from days to decades, in: *Proceedings of the ECMWF seminar on Predictability – Volume I*, 83–141, 1995.
- Poppe, D., Aumont, B., Ervens, B., Geiger, H., Herrmann, H., Röth, E.-P., Seidl, W., Stockwell, W. R., Vogel, B., Wagner, S., and Weise, D.: Scenarios for Modeling Multiphase Tropospheric Chemistry, *J. Atmos. Chem.*, 40, 77–86, 2001.
- Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, *Atmos. Chem. Phys.*, 6, 187–195, doi:10.5194/acp-6-187-2006, 2006.
- Sorensen, D. C.: *Implicitly restarted Arnoldi/Lanczos methods for large scale eigenvalue calculations*, ICASE Report No. 96–40, Institute for Computer Applications in Science and Engineering, Houston, Texas, 1996.
- Stockwell, W. R., Middleton, P., and Chang, J. S.: The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, 95, 16343–16367, 1990.