Accelerating simulations using Direct Relation Graphs for atmospheric chemistry mechanism reduction

Zacharias Marinou Nikolaou [1], Jyh-Yuan Chen [2], Yiannis Proestos [3], Johannes Lelieveld [3,4], and Rolf Sander [4]

[1]Computation-based Science and Technology Research Center (CaSToRC), The Cyprus Institute, Nicosia, 2121, Cyprus.
[2]University of California at Berkeley, Department of Mechanical Engineering, 6163 Etcheverry Hall, Mailstop 1740, USA.
[3]Energy, Environment and Water Research Center (EEWRC), The Cyprus Institute, Nicosia, 2121, Cyprus.
[4]Max Planck Institute for Chemistry, Atmospheric Chemistry Department, 55128 Mainz, Germany.

Correspondence: Z.M Nikolaou (zacharias.nicolaou@cyi.ac.cy)

Abstract.

Chemical mechanism reduction is common practice in combustion research for accelerating numerical simulations, however there have been limited applications in atmospheric chemistry. In this study, we employ a powerful reduction method in order to produce a skeletal mechanism of an atmospheric chemistry code that is commonly used in air quality and climate modelling. The skeletal mechanism is developed using input data from a model scenario, and its performance is evaluated both a priori against the model scenario results, but also posteriori by implementing the skeletal mechanism in a chemistry-transport model namely the Weather Research and Forecasting code with Chemistry. Preliminary results, indicate a substantial increase in computational speedup for both cases, with a minimal loss of accuracy with regards to the spatio-temporal mixing ratio of the target species, which was selected to be ozone.

1 Introduction

Atmospheric chemical mechanisms which are typically used in air quality research and forecasting codes, typically contain a large number of species and reactions. This poses a significant computational workload, which in some cases may account for more than 80% of the total simulation time (Dunker, 1986). These mechanisms describe an important set of processes in the troposphere, for example the degradation of Volatile Organic Compounds (VOCs), and the formation of ozone (O_3) the latter being a major oxidant and pollutant. As a result, mechanisms with varying levels of complexity are included in regional and global atmospheric chemistry codes, the overall performance of which depends strongly on the choice of chemical mechanism.

Apart from the large number of species which require solving at every point in the computational domain and for every time-step, there is a large disparity in the chemical time-scales of the interacting species (Sandu et al., 1997b). This results in a stiff system of non-linear equations for the reaction rates, which is computationally expensive to integrate, and which adds to the computational cost (Sandu et al., 1997a). Similar issues are encountered in the field of combustion research: detailed mechanisms describing the combustion of a fuel, contain hundreds of species and thousands of reactions. From a practical point of view however, one is usually interested in only a handful of important variables-in combustion this includes quantities such

as ignition delay time, laminar flame speed etc. while in atmospheric chemistry this includes ozone/NOx mixing ratios and so on. In both cases, the sensitivity to quantities of interest to certain species and reactions, can be minor in relation to dominant species and reactions. As a result, solving for all species in the detailed chemical mechanism might not actually be required in order to obtain accurate estimates of the target quantities. To this end, chemistry reduction techniques have been developed

5   to reduce the dimensionality of the problem. In turn, this results in a reduction in the computational requirements associated with detailed-chemistry simulations and an acceleration of the simulation. Even though this is common practice in combustion research using a variety of methods (Turanyi et al., 1989; Peters and Rogg., 1993; Tomlin et al., 1997; Lam and Goussis, 1988; Mass and Pope., 1992; Pope, 1997; Lu and Law, 2005; Pepiot-Desjardins and Pitsch, 2008; Nikolaou et al., 2014, 2013; Niemeyer and Sung, 2011), chemistry reduction methods have seen limited use for atmospheric-chemistry applications with

10   only a few studies in the literature.

The usual reduction process of a detailed chemical mechanism begins with the identification of an accurate "skeletal" mechanism. The "skeletal" mechanism is a subset of the detailed mechanism, and is generated by eliminating un-important species and reactions from the detailed mechanism for the problem at hand. Further reduction of the skeletal mechanism is also possible. This can be achieved by a variety of time-scale analysis methods which are applied to the skeletal mechanism such

15   as Quasi-Steady-State Assumption (QSSA), Computational Singular Perturbation (CSP) (Lam and Goussis, 1988) etc. Time-scale methods are employed for finding species which are approximately in steady-state. Following this, a non-linear system of equations is solved for the steady-state species mixing ratios. As a result, time-scale analysis methods are most efficient when applied to relatively small skeletal mechanisms rather than the full detailed mechanism. Such an approach, using CSP, was used by Neophytou et al. (2004) in order to construct a reduced mechaninsm for the Carbon Bond mechanism (CBMIV)

20   (Gerry et al., 1989). In this study, our interest is in generating skeletal mechanisms which is the first step in the reduction process, and which can be used as a starting point for further reduction, and be applied to more comprehensive chemistry codes.

Sensitivity Analysis (SA) is perhaps the oldest and most straightforward of methods for identifying skeletal mechanisms (Turanyi et al., 1989). In SA, suitable sensitivity coefficients are defined which are usually reaction-based. The sensitivity of

25   a species with each reaction is calculated for a particular configuration (reaction mode), and reactions having sensitivity coefficients below a threshold value are identified as redundant, and are removed from the detailed mechanism. Such an approach was employed by Heard et al. (1998) for reducing the CBM-EX tropospheric chemical mechanism, and an SA-assisted tabulation method was used by Dunker (1986) for accelerating the species integration. SA was also employed by Whitehouse et al. (2004) as a first reduction step for generating a skeletal mechanism from the Master Chemical Mechanism (MCM) (Derwent

30   et al., 1998). Reaction-based sensitivity analysis, results in removing reactions from the detailed mechanism, but may not always significantly reduce the number of species which is the key factor controlling compute time in numerical simulations. To this end, a number of other techniques have been developed which are species-oriented rather than reaction-oriented. Direct Relation Graph (DRG), is a species-based method for the generation of skeletal mechanisms, originally proposed by Lu and Law (2005). In DRG, a suitable species Direct Interaction Coefficient (DIC) is defined. The DIC measures the importance

35   a particular species has on a pre-defined set of target species. DRG eventually results in removing species, in contrast with

classic reaction-based SA. In the original version of the DRG method, the target species set includes only species appearing in the same reaction as the target species. However, species not interacting directly with the target species through a reaction, may still be important indirectly for a target species of interest. An extension of the DRG method, namely DRG with Error Propagation (DRGEP) was therefore proposed to address this issue (Pepiot-Desjardins and Pitsch, 2008; Niemeyer and Sung,

5   2011). In DRGEP, the DIC is defined so that the effect of reaction path is also taken into account during the reduction process. DRGEP has been extensively used to generate skeletal mechanisms for combustion applications, with overall good results, and many variants of the method have henceforth been developed using different DIC definitions and search-finding algorithms (Niemeyer et al., 2010; Stagni et al., 2016; Chen and Chen, 2016).

In comparison to SA, DRGEP has seen limited use for reducing complex atmospheric chemical mechanisms, despite its

10   large potential. In addition, the majority of studies in the literature (which used SA) focused on generating sub-sets of very detailed chemical mechanisms such as the MCM. As a result, the skeletal mechanisms generated from MCM are still of a prohibitive size also for efficient forecasting purposes (Whitehouse et al., 2004). Our focus in this study instead, is on chemical mechanisms which are commonly used in atmospheric models. These mechanisms are already condensed mechanisms, which were typically developed from a bottom-up approach, and include a large number of surrogate/lumped species. It is thus

15   instructive to investigate whether DRGEP can be used for further reduction of these already condensed mechanisms, as a fist step in the reduction process.

Another important point to note, is that the majority of studies in the literature have only focused on a priori evaluation of the skeletal chemical mechanisms: their performance was evaluated against the model scenario results only, which usually involved 0-D box-model simulations. In a practical forecast simulation however, advection, diffusion, but also the more refined

20   calculation of photolysis rates, all affect the spatio-temporal concentration of the species. This affects the species production/destruction rates, and in turn the reduction process, and to date, at least to the knowledge of the authors, no a posteriori valiation has been conducted using actual forecasting simulations.

In this study, DRGEP is employed in order to examine whether sufficiently accurate skeletal mechanisms can be generated for a detailed mechanism which is commonly used in weather forecasting codes, namely the Regional Atmospheric Chemistry

25   Mechanism (RACM). This is an updated version of the Regional Acid Deposition Model (RADM2) mechanism by Stockwell et al. (1997a), and describes the degradation of a number of VOCs. It is a condensed chemical mechanism but it is relatively large, including 75 species and 237 reactions, and is therefore a good candidate for reduction using DRGEP.

In the text which follows, section 2 introduces the DRGEP method, and section 3 lists the process for generating the skeletal mechanism as well as the a priori validation. Details of the a posteriori validation using a popular weather-research and

30   forecasting code are given in section 4.

2   Mathematical background: DRGEP

Direct Relation Graph (DRG) is a method for generating subsets of detailed chemical mechanisms by removing species having a negligible effect on a pre-defined set of target species. In the original version of the DRG method developed by Lu and Law (2005), the DIC r_{TB} between a target species T and a non-target species B is defined as,

$$r_{TB} = \frac{\sum_{i=1}^{N_r} |\dot{w}_{iT}\delta_{iB}|}{\sum_{i=1}^{N_r} |\dot{w}_{iT}|} \tag{1}$$

where $\dot{w}_{iT}$ is the net rate of species T from reaction i, and δ_{iB} is an index specifying the existence of B in reaction i. This index is equal to 1 if B exists in reaction i, and 0 otherwise. Clearly, $0 \leq r_{TB} \leq 1$, and r_{TB} is generally not equal to r_{BT}. A large value of r_{TB} implies that species B is important in the evaluation of the rate of T, while a low value implies it is not as important. A threshold ϵ is introduced, and provided $r_{TB} > \epsilon$, species B is added to the set of dependent species of T, otherwise it is deemed un-important and removed. This relation is denoted as a directed path $T \rightarrow B$. An example of a DRG involving four species A,B,C and D, is given in Fig. 1 with the numbers indicating the values of the DICs for each pair. The process is repeated for all target species T_i, and the final set of species in the skeletal mechanism is constructed from the union of all target species sets. Species not included in the union set are eliminated, as are reactions involving any of the eliminated species.

DRG has been applied with overall good results in a number of studies in combustion research, yet the simple definition of interaction coefficient in Eq. 1 has some important limitations. Consider the model situation depicted in Fig. 1. If A is the target species and C is the species in question, then with a threshold value ϵ of say 0.1, C will be added to the dependent set of A. However, it is clear that there exist "stronger" paths from A to C, for example $A \rightarrow B \rightarrow C$. Thus, the notion of "path" becomes important, and a suitable DIC definition is required able to describe this. In addition, there are also alternative paths e.g. $A \rightarrow D \rightarrow C$, or $A \rightarrow D \rightarrow B \rightarrow C$ which should also be checked if they are "stronger".

The DRGEP method aims to account for the above points by using an improved DIC definition and reduction strategy. In DRGEP (Pepiot-Desjardins and Pitsch, 2008; Niemeyer and Sung, 2011), the DIC is first defined using,

$$r_{TB} = \frac{|\sum_{i=1}^{N_r} \dot{w}_{iT}\delta_{iB}|}{max(P_T, C_T)} \tag{2}$$

where the production P_T and consumption terms C_T of T, are defined as,

$$P_T = \sum_{i=1}^{N_r} max(0, \dot{w}_{iT}) \tag{3}$$

$$C_T = \sum_{i=1}^{N_r} max(0, -\dot{w}_{iT}) \tag{4}$$

The DIC as defined above is calculated for all species. The Path Interaction Coefficient (PIC) r^p_{TB} for a given path p connecting target species T and B is then defined as,

$$r^p_{TB} = \prod_{i=1}^{n-1} r_{S_i S_{i+1}}|_p \tag{5}$$

i.e. is the product of all DICs along that path. The PIC is calculated for all possible paths connecting T to B, and an Overall

5   Path Interaction Coefficient (OIC), $r^o{}_{TB}$ is then calculated using,

$$r^o{}_{TB} = max\left(\prod_{i=1}^{n-1} r_{S_i S_{i+1}}|_p\right) = max(r^p_{TB}) \tag{6}$$

i.e. the strongest path from T to B is identified based on the product of DICs of connected nodes across all paths linking T and B. In the example of Fig. 1, the strongest path is $A \rightarrow B \rightarrow C$ since for this path $r^o{}_{AC}$=0.9 · 0.7=0.63 which is the largest value, and both B and C will be included in the set of A. The process is repeated for all target species of interest, and species

10   with overall interaction coefficients less than a pre-defined threshold value are removed.

The identification of the stongest path is a common problem in computational science, and a number of different search-finding algorithms have been developed for this task. In this study, we employ a classic algorithm for searching through the connected nodes and obtaining $r^o{}_{TB}$ and hence the "strongest" path (Dijkstra, 1959). An in-house code was specifically developed for the DRGEP method, and for all associated functions including the search-finding subroutines.

3   Skeletal mechanism development

Reduction methods require input data, and these data should be representative of the actual reaction scenario. This translates to using actual weather-forecast simulation data as input to the DRGEP method. However, this is hardly ever done in practice since there is a large computational overload for conducting such simulations in the first place. As a result, a computationally more efficient initial-value problem (box-model) is used as a model scenario for the reduction, which is common practice in

20   chemical mechanism reduction studies (Dunker, 1986; Heard et al., 1998; Neophytou et al., 2004). The species mixing ratios C_i evolve according to,

$$\frac{dC_i}{dt} = G_i(C_k) \tag{7}$$

where G_i are non-linear functions of the species rates, with initial conditions $C_i{}^0$ for the species mixing rations, and T^0 for temperature. The pressure is kept fixed at 1.0 atm, and the temperature at 298.0 K. The Kinetic Pre-Processor library (Damian

25   et al., 2002; Sandu et al., 2003; Daescu et al., 2003) is used for the numerical integration of Eq. 7. The KPP library includes a number of different solvers, and in this study a 5-stage Runge-Kutta method (Hairer et al., 1993; Hairer and Wanner, 1993) is used from the package (Radau5). This method is stiffly accurate and robust, and is often used for benchmarking purposes.

Six different initial scenarios are considered. The species mixing ratios for each model scenario are given in Table 1. These model scenarios were used by Heard et al. (1998) for reducing the CBMEX mechanism using sensitivity analysis. In this work, the same mixing ratios are used as per Table 1 of Heard et al. (1998), but are adapted here for the chemical mechanism used in this study. Here we group important alkane, alkene, and aromatic species for each mechanism, and initialise their mixing ratios

5 based on the relevant alkane, alkene and aromatic species found in the study of Heard et al. (1998). Cases A-C correspond to increasing VOC/NOx ratios not including isoprene, while cases D-F correspond to about the same VOC/NOx ratios with isoprene included. The VOC/NOx ratios are relatively high, and this ensures that a large number of VOC-relevant reactions are activated-in this process a relatively large region of the composition space is thus covered.

The J values for the photolysis rates are based on parameterisations as developed in the MCM (Derwent et al., 1998). These

10 are given by,

$$J = I cos^M(\theta) e^{-N sec(\theta)} \qquad (8)$$

where θ is the solar azimuth angle, and I, M and N are reaction-specific constants. A 48-hour run is conducted for each case, which results in a total of 496 datasets. These datasets are used as input for DRGEP and the reduction process is done on a dataset-basis. Important species are retained for each dataset, and the process is repeated for the next dataset. Any new species

15 not already included in the dataset are added. Once all datasets are considered, a species union set is formed, and any reactions involving species other than those included in the species union set are removed. The target species is set to be O_3 which is an important pollutant of interest. In addition, O_3 is eventually produced by the degradation of the VOCs through a large number of reaction pathways, and is therefore a good target for the reduction.

As an example, Table 2 lists the OIC values for target O_3 as obtained for case A at midday (t=12 hours) and midnight (t=24

20 hours). Clearly, there is a difference in the OIC values for each species since the rate constants depend on the solar azimuth angle which determines the rate constants of the photolytic reactions. Top-scoring species for O_3 include third-body species M, and oxygen species such as O3P (ground-state oxygen atom) and O1D (excited state oxygen atom). This is expected since these species readily react to produce O_3 through the reactions $O3P + M => O_3$ and $O1D + M => O_3$. Nitrogen oxides NO and NO_2 also score high as they too react both directly and indirectly with ozone. Direct paths include the reactions

25 $O_3 + NO => NO_2 + O_2$, and $O_3 + NO_2 => NO_3 + O_2$ while indirect paths include reactions with O3P. The methyl-peroxy radical MO_2 also ranks high since it is involved in numerous reactions with VOCs. The hydroxyl radical HO also scores high since it is the main oxidation path of the VOCs which eventually end up producing ozone. With a threshold of 9.0E-3 (which results in a 54-species subset as explained later on), 18 species will be included in the set for t=12 hours, and 11 species will be included for t=24 hours, and the process is repeated for all other data sets to form the overall species union set for the target

30 O_3.

In order to quantify the quality of the skeletal mechanisms generated using DRGEP, a percentage error is defined based on the target species of interest i.e. ozone. For a mixing ratio obtained using the skeletal mechanism $C_i{}^s$, and a mixing ratio using the detailed mechanism $C_i{}^d$, this is defined as,

$$e = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \int 100 \cdot \frac{|C_i{}^s(t) - C_i{}^d(t)|}{|C_i{}^d(t)|} dt \tag{9}$$

5   where N is the total number of cases. Note that zero values are not taken into account for the error calculation.

Figure 2 shows this error against the number of species in the skeletal mechanism. As expected, the error increases as more species are removed. With about 10 species removed (75 to 65) the error is negligibly small. This error increases once more than about 10 species are removed, however the error remains small down to about 55 species (20 species removed) at less than 10%. The huge spike in the error below about 55 species, results from the removal of important intermediates.

10   Figure 3 shows the percentage in speedup (cpu-time) gained, against the number of species in the skeletal mechanisms. This is done for case A, and similar results were observed for the rest of the cases. The speedup is calculated both for the total simulation time, and for integrating Eq. 3 alone. The threshold line where the error in O_3 prediction spikes as per the results in Fig. 2 is also shown. Clearly, as the number of species is reduced, compute time drops and there is an increase for both speedups. It is interesting to note that going from 64 to 58 species there is no increase in integration speedup. This implies that

15   the stiffness of the remaining species and equations is unchanged. Nevertheless, there is an increase in total speedup of almost 10% going from 64 to 58 species, and this is only due to simulation overheads alone which is quite significant. The average speedup due to overhead computations is found to be 6.6%. At the threshold error e of 10%, the smallest possible skeletal mechanism contains 54 species and 150 reactions (threshold for OIC at this point is 9.0E0-3), which is significantly smaller than the detailed mechanism.

20   For this mechanism, the total speedups and integration speedups are 54.4% and 43.7% respectively. In general, the CPU time scales with the total number of species in the system due to the evaluation of the Jacobian when dealing with stiff systems. For integration of the system alone, the expected speedup percentage scales as speedup (%)=$100 \cdot |(N_{sp}/N_{sp,det})^2 - 1|$, and this is shown as a blue line in Fig. 3. Clearly, there is a good qualitative agreement with the theoretical result.

In order to visualise the errors in ozone mixing ratio more clearly, Fig. 4 shows the solution profiles for cases A-C for the

25   48-hour runs, using the detailed and worst-performing skeletal mechanism (having 54 species). It is clear that the agreement with the detailed mechanism is very good for all six cases. In the isoprene cases D-F, similarly good results were obtained and are not presented here. It is also important to note that the agreement is particularly good early on. Given that typical time-steps used in forecasting simulations are of the order of minutes, this is very encouraging.

4   A posteriori validation

30   The results of the previous section constitute an a priori validation. In an actual simulation, the photolysis rates but also the species mixing ratios due to the effects of advection, diffusion, and deposition processes, can be substantially different from

the conditions in a box-model. As a result, the species rates will also differ which affects the reduction process. The aim of this section is to examine the performance of the skeletal mechanism generated in the previous section, by implementing it in an actual atmospheric-chemistry simulation, and comparing with results using the detailed chemical mechanism.

4.1 WRF-Chem simulation setup

5   The Weather Research and Forecasting system with Chemistry (WRF-Chem/version 3.9.1.1) has been employed in this work. WRF-Chem, which has been developed jointly between several research institutes (https://www2.acd.ucar.edu/wrf-chem), is a state-of-the art, open source, limited-area atmospheric model, featuring a highly parallelised code. WRF-Chem is used both for research applications and for operational numerical weather and air-quality predictions, and is an online, fully coupled model, which integrates and calculates meteorology, gas-phase chemistry and aerosols simultaneously (Grell et al., 2005).

10 WRF-Chem utilizes the Advanced Research WRF (ARW) solver (Skamarock et al., 2005), where the transformation, mixing-phase, and transport of chemical species and aerosols, are calculated following the same prognostic equations, time-step, and spatial configuration with the meteorology, physics and other transport constituents of the ARW dynamical core.

In this study, the model is configured over a single domain using the LatLon geographical projection, with about 0.15° (∼16 km) horizontal grid spacing, and a domain of 165 (E–W) by 165 (N–S) grid points as shown in Fig. 5. Thirty vertical

15 model levels were used, which correspond to a maximum height of about 20 km (∼50 hPa). Owing to its modular design, WRF-Chem provides several choices of chemical mechanisms and physics parameterisations. In this study, RACM is used for the gas-phase chemistry (Stockwell et al., 1997b; Geiger et al., 2003), and KPP for the integration of the species mixing ratios. The full RACM mechanism as implemented in WRF-Chem includes 75 species and 237 reactions. Table 3 summarizes the major model features and physical parameterizations as used in the simulations.

The meteorological fields were forced by initial and lateral boundary conditions obtained from the National Centers for the

20 Environmental Prediction/Global Forecast System (NCEP/GFS) at a spatial resolution of 0.25°, and updated at a three-hour interval. MODIS-based geo-terrestrial data, including land categories, soil types, and terrain height, have been used. Our aim, as a first step, is to examine ozone mixing ratio predictions between the detailed mechanism and the skeletal mechanism using DRGEP, without the influence of any external source terms. This will give a clearer picture as to the effect of transport terms

25 only on the spatio-temporal distribution of ozone. In order to do that, neither anthropogenic/biogenic emissions have been utilised, nor any chemical initial and boundary conditions have been applied to the chemistry fields. For the latter, the model uses instead, idealized climatologically based values to initialise the chemical species.

For the purposes of this study, two separate model simulation runs were conducted for the period of 12 to 28 July, 2017, a time of year during which ozone photochemistry is particularly active. The first five days of the model output are considered

30 as model spin-up time, and have been excluded from our analysis. The model instantaneous, grid-cell averaged mixing ratios, were set to be written out (at the beginning of) every hour. The first run, used the complete (unmodified) RACM mechanism as implemented in the WRF-Chem package, while the second one utilised the skeletal (via DRGEP algorithm) mechanism. For a fair comparison between the two runs, both setups shared the same namelist, which is included in the Supplement.

The implementation of a new chemical mechanism in WRF-Chem is a rather tedious process. This includes creating new reaction and species files, compiling KPP with the new mechanism, but also writing new mechanism-specific driver routines, and initialisation routines. A work-around, is to instead modify the existing chemical mechanism file (in this case RACM), so that it accounts for the reduced chemistry. This simple method implies that neither driver routines need to be re-written nor

5 calls to subroutines need to change to account for the reduction in species etc. This is achieved by setting dummy reactions for all species which are removed in the skeletal mechanism. The corresponding KPP reactions in the skeletal mechanism, are included in the Supplement. From a computational perspective, the skeletal mechanism was found to be more efficient than the detailed mechanism, as expected: on 40 MPI processes, the wall-clock times using the detailed and skeletal mechanisms were 959 minutes and 730 minutes respectively. This translates to an overall gain in CPU time of 24.6%.

10 This speedup is of the same order of magnitude as in the box-model runs. However due to the implementation of the skeletal mechanism in WRF-Chem (all species kept), this speedup does not include over-heads. In case a new mechanism is written (with fewer number of species), and all relevant subroutine calls are suitably modified in order to include only the species in the skeletal mechanism, an even further gain in speedup is expected from simulation overheads (input/output, calls to subroutines etc.).

15 **4.2   Comparison of mixing ratios**

In order to warrant a more quantitative evaluation of the performance of the skeletal mechanism, we additionally calculate the volume-averaged error (based on ozone mixing ratio), in time, between the skeletal and detailed mechanisms. This is defined as,

$$e'(t) = \frac{1}{V} \int_{\mathbf{x}} 100 \cdot \frac{|C^s(\mathbf{x},t) - C^d(\mathbf{x},t)|}{|C^d(\mathbf{x},t)|} d\mathbf{x} \qquad (10)$$

20 where V is the sample-space volume and C^d, C^s, are the predictions of scalar field C using the detailed and skeletal mechanisms respectively. The sample volume V is taken to include all points in the longitudinal and latitudinal directions, and also the nine lower levels in the vertical direction, thus covering a significant part of the troposphere-including the planetary boundary layer. Mixing ratios having zero values are not considered for the error calculation.

Figure 6 shows a direct comparison between the ozone mixing ratios as predicted using the full mechanism (left) and the

25 skeletal mechanism (right). The instant depicted, corresponds to the case of maximum error $e'(t)$. From a visual inspection alone, it is clear that there is very good agreement for the spatial ozone concentration prediction using the skeletal mechanism.

The corresponding error for the ozone mixing ratio field is depicted on the left plot of Fig. 7. Over the 265 time steps (hours) included in the analysis, it is found that e' varies between 2.52% and 4.21%. These small errors confirm the good agreement observed for the instantaneous ozone predictions shown in Fig. 6. The right plot of Fig. Fig. 7 shows the distribution of error,

30 averaged in the vertical layers only, for the time instant of maximum e' i.e. at 90 hours. This helps to elucidate the actual spatial distribution of the error in terms of latidude and longitude. The error distribution is also within reasonable bounds at this instant,

not exceeding 10 %. Note, that this distribution applies at the instant of maximum volume-averaged error as calculated using Eq. 10. This error is actually transported during the simulation i.e. is not specific to a particular region.

Figures 8, 9 and 10, 11 show the corresponding results for carbon monoxide which is a slow-reacting species, and formaldehyde which is a relatively faster-reacting species. Both these species were not included as targets during the reduction, therefore

5   it is instructive to examine how their mixing ratio predictions compare to ozone which was the target species. Figures 8 and 9 show that the CO predictions are also in good agreement with relatively small percentage errors. The maximum volume-weighted error e' for carbon monoxide during the simulation does not exceed 2.6 %. The instantaneous error averaged in the vertical layers only at the time of maximum e' also remains low as one may observed from the right plot of Fig. 9. The errors for formaldehyde on the other hand are relatively larger as one may observe from the results in Figs. 10 and 11. The maximum

10   volume-weighted error is about 7% while the instantaneous error for the time of maximum e' is in the region of 20% as one may observe from the right plot of Fig. 11. Formaldehyde, which is an important intermediate species (Lelieveld et al., 2016), is involved in many oxidation reactions including a number of VOCs hence the relatively larger errors. The same applies for the hydroxyl radical HO, whose errors where also found to be relatively large. In particular, for the subset mechanism including 54 species, higher alkanes such as HC3 but also TOL (Toluene) were identified as redundant species from the DRGEP. These

15   species constitute an important HO consumption pathway, and excluding them leads to an over-estimation of the HO mixing ratio. Much better results may be obtained by either reducing the OIC threshold (hence including more species) or by including more targets during the reduction. Both these approaches however, will lead to larger skeletal mechanisms and a reduction in speedup, in other words, careful selection of the targets is required to obtain both an accurate but also computationally fast mechanism.

20   ## 5   Conclusions

A Direct Relation Graph approach for generating skeletal chemical mechanisms from more detailed mechanisms has been employed, in order to produce a computationally more efficient mechanism for accelerated atmospheric chemistry simulations. A code has been developed for the task, and the method has been applied to a commonly used mechanism, namely RACM, with the target species being ozone which is a major pollutant.

25   The skeletal mechanism was developed using input from a 0-D initial-value problem, and was validated both a priori against the 0-D problem results, but also a posteriori. The a posteriori validation involved implementing both the detailed and the skeletal mechanisms in an actual forecasting code, namely WRF-Chem, and running simulations to compare the spatio-temporal ozone mixing ratio profiles. The skeletal mechanism is found to perform well, with relatively low percentage errors. A speedup of 24.6% was achieved for the total simulation time, which does not yet include any speedup due to overheads such as input-

30   output computations.

The method is general, and can be applied to any chemical mechanism in the WRF-Chem package or other chemistry-transport codes, for producing computationally more efficient air quality and climate simulations. Since a significant speedup

has been achieved with the already optimized chemical mechanism used in this study, it is expected that future application to more comprehensive chemistry mechanisms may lead to significant gains in computational efficiency.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. ZMN acknowledges funding through the VI-SEEM project which receives funding from the European Union's H2020
5   research and innovation program under grant agreement no. 675121.

Code and data availability. The WRF-Chem package used for the numerical simulations is available from the National Center for Atmospheric Research (NCAR): https://www2.acom.ucar.edu/wrf-chem. The WRF-Chem namelist file, and the skeletal RACM mechanism are given as supplements. The code used for the DRGEP can be obtained from the authors upon request.

References

Chen, Y. and Chen, J.: Application of Jacobian defined direct interaction coefficient in DRGEP-based chemical mechanism reduction methods using different graph search algorithms, Combust. Flame, 174, 77–84, 2016.

Daescu, D., Sandu, A., and Carmichael, G.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: II – Validation and Numerical Experiments, Atmos. Env., 37, 5097–5114, 2003.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G.: The Kinetic PreProcessor KPP–A Software Environment for Solving Chemical Kinetics, Comput. and Chem. Eng., 26, 1567–1579, 2002.

Derwent, R., Jenkin, M., Saunders, S., and Pilling, M.: Photochemical ozone creation potentials for organic compounds in northwest Europe calculated with a master chemical mechanism, Atmos. Env., 32, 2429–2441, 1998.

Dijkstra, E.: A note on two problems in connexion with graphs, Numer. Math., 1, 261–271, 1959.

Dunker, A.: The reduction and parameterisation of chemical mechanisms for inclusion in atmospheric reaction-transport models, Atmos. Env., 20, 479–486, 1986.

Geiger, H., Barnes, I., Bejan, I., Benter, T., and Spittler, M.: The tropospheric degradation of isoprene: an updated module for the regional atmospheric chemistry mechanism, Atmospheric Environment, 37, 1503–1519, 2003.

Gerry, M., Whitten, G., Killus, J., and Dodge, M.: A photochemical kinetics mechanism for urban and regional scale computer modelling, J. Geoph. Res., 94, 12 925–12 956, 1989.

Grell, G. A. and Dévényi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophysical Research Letters, 29, 2002.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmospheric Environment, 39, 6957–6975, 2005.

Hairer, E. and Wanner, G.: Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems, Springer-Verlag, Berlin, 1993.

Hairer, E., Norsett, S., and Wanner, G.: Solving Ordinary Differential Equations I. Nonstiff Problems, Springer-Verlag, Berlin, 1993.

Heard, A., Pilling, M., and Tomlin, A.: Mechanism reduction techniques applied to tropospheric chemistry, Atmos. Env., 32, 1059–1073, 1998.

Hong, S.-Y., Dudhia, J., and Chen, S.-H.: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation, Monthly Weather Review, 132, 103–120, 2004.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly weather review, 134, 2318–2341, 2006.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, Journal of Geophysical Research: Atmospheres, 113, 2008.

Lam, S. and Goussis, D.: Understanding Complex Chemical Kinetics with Computational Singular Perturbation, 22nd Sympt. Int. Combust., pp. 931–941, 1988.

Lelieveld, J., Gromov, S., Pozzer, A., and Taraborelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmos. Chem. Phys., 16, 12 477–12 493, 2016.

Lu, T. and Law, C.: A directed relation graph method for mechanism reduction, Proc. Combust. Inst., 30, 1333–1341, 2005.

Mass, U. and Pope., S.: Simplifying Chemical Kinetics: Intrinsic Low-Dimensional Manifolds in Composition Space, Combust. Flame, 88, 239–264, 1992.

Neophytou, M., Goussis, D., van Loon, M., and Mastorakos., E.: Reduced chemical mechanisms for atmospheric pollution using Computational Singular Perturbation analysis, Atmos. Env., 38, 3661–3673, 2004.

Niemeyer, K. and Sung, C.: On the importance of graph search algorithms for DRGEP-based mechanism reduction methods, Combust. Flame, 158, 1439–1443, 2011.

Niemeyer, K., Sung, C., and Raju, M.: Skeletal mechanism generation for surrogate fuels using direct relation graph with error propagation and sensitivity analysis, Combust. Flame, 157, 1760–1770, 2010.

Nikolaou, Z., Chen, J., and Swaminathan, N.: A 5-step reduced mechanism for combustion of CO/H2/H2O/CH4/CO2 mixtures with low hydrogen/methane and high H2O content, Combust. Flame, 160, 56–75, 2013.

Nikolaou, Z., Swaminathan, N., and Chen, J.: Evaluation of a reduced mechanism for turbulent premixed combustion, Combust. Flame, 161, 3085–3099, 2014.

Paulson, C. A.: The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer, Journal of Applied Meteorology, 9, 857–861, 1970.

Pepiot-Desjardins, P. and Pitsch, H.: An efficient error-propagation-based reduction method for large chemical kinetic mechanisms, Combust. Flame, 154, 67–81, 2008.

Peters, N. and Rogg., B.: Reduced Reaction Mechanisms for Applications in Combustion Systems, Notes in Physics: Springer-Verlag, 15, 1993.

Pope, S.: Computationally efficient implementation of combustion chemistry using in situ adaptive tabulation, Combust. Theory Model., 1, 41–63, 1997.

Sandu, A., Verwer, J., Blom, J., Spee, E., Carmichael, G., and Potra, F.: Benchmarking of stiff ODE solvers for atmospheric chemistry problems II: Rosenbrock solvers, Atmos. Env., 31, 3459–3472, 1997a.

Sandu, A., Verwer, J., van Loon, M., Carmichaels, G., Potra, F., Dabdub, D., and Seinfeld., J.: Benchmarking of stiff ODE solvers for atmospheric chemistry problems I: implicit vs explicit, Atmos. Env., 31, 479–486, 1997b.

Sandu, A., Daescu, D., and Carmichael, G.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: I–Theory and Software Tools, Atmos. Env., 37, 5083–5096, 2003.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., and Powers, J. G.: A description of the advanced research WRF version 2, Tech. rep., National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div, 2005.

Stagni, A., Frassoltadi, A., Cuoci, A., Faravelli, T., and Ranzi, E.: Skeletal mechanism reduction through species-targeted sensitivity analysis, Combust. Flame, 163, 382–393, 2016.

Stockwell, W. R., Kirchner, F., and Kuhn, M.: A new mechanism for regional atmospheric chemistry modeling, J. Geoph. Res., 102, 25 847–25 879, 1997a.

Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, Journal of Geophysical Research: Atmospheres, 102, 25 847–25 879, 1997b.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, in: 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, vol. 1115, 2004.

Tomlin, A., Turanyi, T., and Pilling, M.: Chapter 4: Mathematical tools for the construction, investigation and reduction of combustion mechanisms, Compr. Chem. Kinetics, 35, 293–247, 1997.

Turanyi, T., Berces, T., and Vajda, S.: Reaction rate analysis of complex kinetic systems, Int. J. Chem. Kinetics, 21, 83–99, 1989.

Webb, E. K.: Profile relationships: The log-linear range, and extension to strong stability, Quarterly Journal of the Royal Meteorological Society, 96, 67–90, 1970.

Whitehouse, L., Tomlin, A., and Pilling, M.: Systematic reduction of complex tropospheric chemical mechanisms, Part II: Lumping using a time-scale based approach, Atmos. Chem. Phys., 4, 2057–2081, 2004.

Wild, O., Zhu, X., and Prather, M. J.: Fast-J: Accurate simulation of in-and below-cloud photolysis in tropospheric chemical models, Journal of Atmospheric Chemistry, 37, 245–282, 2000.

Species-RACM	A	B	C	D	E	F
NO	163.0	65.1	72.1	163.0	65.1	72.1
NO2	6.9	2.2	3.3	6.9	2.2	3.3
HCHO	0.9	0.9	1.7	0.9	0.9	1.7
Alkanes:						
ALD	4.2	4.2	8.4	4.2	4.2	8.4
CH4	155.0	155.0	310.0	155.0	155.0	310.0
ETH	155.0	155.0	310.0	155.0	155.0	310.0
HC3	155.0	155.0	310.0	155.0	155.0	310.0
HC5	155.0	155.0	310.0	155.0	155.0	310.0
HC8	155.0	155.0	310.0	155.0	155.0	310.0
Alkenes:						
OLT	13.0	13.0	26.0	13.0	13.0	26.0
OLI	13.0	13.0	26.0	13.0	13.0	26.0
DIEN	13.0	13.0	26.0	13.0	13.0	26.0
ETE	30.0	30.9	62.0	30.0	30.9	62.0
ISO	0.0	0.0	0.0	28.0	28.0	55.8
Aromatics:						
TOL	14.6	14.6	29.2	14.6	14.6	29.2
XYL	10.6	10.6	21.2	10.6	10.6	21.2

Table 1. Initial species mixing ratios (ppbv) for RACM as deduced from Heard et al. (1998). Water content is 10^7 ppbv, CO content is 2310 ppbv. Pressure is fixed at 1 atm, and temperature at 298 K. VOC/NOx ratios (methane not included) for cases A-F: 4.2, 10.7, 19.1, 4.4, 11.1, 19.8.

Index	Species name	OIC, t=12 hours	Species name	OIC, t=24 hours
1	O3	1	O3	1
2	M	0.99965	M	0.99023
3	O3P	0.96864	O3P	0.95424
4	NO2	0.25278	NO2	0.10845
5	NO	0.24826	NO	0.10544
6	O1D	4.30E-02	O1D	5.06E-02
7	HO2	3.94E-02	HO2	2.01E-02
8	HO	1.83E-02	MO2	1.48E-02
9	MO2	1.78E-02	H2O	1.32E-02
10	ACO3	1.51E-02	HO	1.32E-02
11	ALD	1.41E-02	ACO3	1.10E-02
12	HCHO	1.29E-02	HCHO	8.71E-03
13	ETHP	1.27E-02	CO	8.60E-03
14	CO	1.18E-02	CO2	7.15E-03
15	H2O	1.15E-02	ALD	7.10E-03
16	KET	1.13E-02	ETHP	5.41E-03
17	NO3	1.12E-02	NO3	5.11E-03
18	XO2	1.06E-02	KET	4.63E-03
19	ONIT	8.90E-03	OP1	4.11E-03
20	CO2	6.20E-03	XO2	3.47E-03
21	HC8	5.35E-03	OP2	2.85E-03
22	HKET	5.35E-03	HC3P	2.84E-03
23	HC8P	5.35E-03	ONIT	2.81E-03
24	OP2	5.00E-03	HC5	1.93E-03
25	PAN	4.72E-03	HC5P	1.93E-03
26	HC5	4.56E-03	GLY	1.80E-03
27	HC5P	4.56E-03	H2O2	1.79E-03
28	HC3P	3.92E-03	ORA1	1.63E-03
29	OP1	3.38E-03	HC3	1.63E-03
30	ORA1	2.32E-03	HKET	1.13E-03
31	HC3	2.32E-03	ORA2	1.06E-03
32	GLY	2.32E-03	HC8	1.06E-03
33	PHO	2.09E-03	HC8P	1.06E-03
34	CSL	2.09E-03	KETP	1.05E-03
35	HNO3	2.07E-03	PAA	7.95E-04

Table 2. Overall Interaction Coefficients (OICs) for target O_3 (top 35) for case A at t=12 and t=24 hours.

Feature	Description	Details
Horizontal grid resolution	0.15° (∼16 km)	Geographic LatLon
Number of vertical layers	30 terrain following sigma coordinates	
Simulation time step	60 s	
Land-surface scheme	Noah land-surface unified model (Tewari et al., 2004)	sf_surface_physics = 2
Cloud microphysics	WRF Single-Moment (WSM) 3-class simple ice scheme (Hong et al., 2004)	mp_physics = 3
Surface layer	Monin-Obukhov similarity theory (MM5) with Carlson-Boland viscous sublayer and similarity functions from look-up tables (Paulson, 1970; Webb, 1970)	sf_sfclay_physics = 1
Shortwave radiation	Rapid Radiative Transfer Model (RRTMG) (Iacono et al., 2008)	ra_sw_physics = 4
Longwave radiation	Rapid Radiative Transfer Model (RRTMG) (Iacono et al., 2008)	ra_lw_physics = 4
Planetary boundary layer	Yonsei University (YSU) PBL (Hong et al., 2006)	bl_pbl_physics = 1
Cumulus convection	Grell 3D ensemble scheme (Grell and Dévényi, 2002)	cu_physics = 5
Gas-phase chemistry	RACM-KPP (Stockwell et al., 1997b)	chem_opt = 103
Photolysis parameterization	Fast-J (Wild et al., 2000)	phot_opt = 2

Table 3. Settings and physical parameterization schemes selected during the WRF-Chem simulations
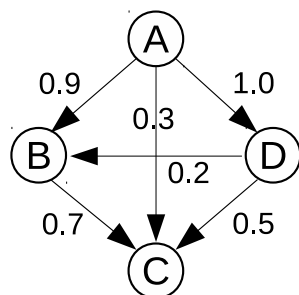
Figure 1. Example of a Direct Relation Graph involving four species.
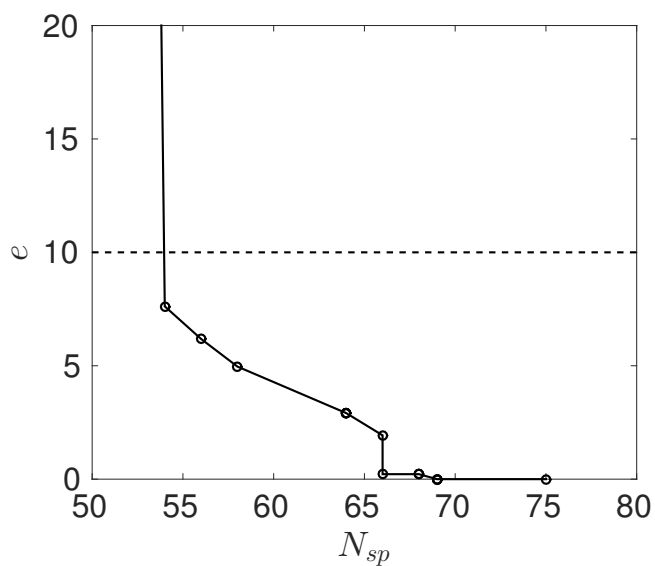
Figure 2. Ozone mixing ratio percentage error against number of species in skeletal RACM mechanism.
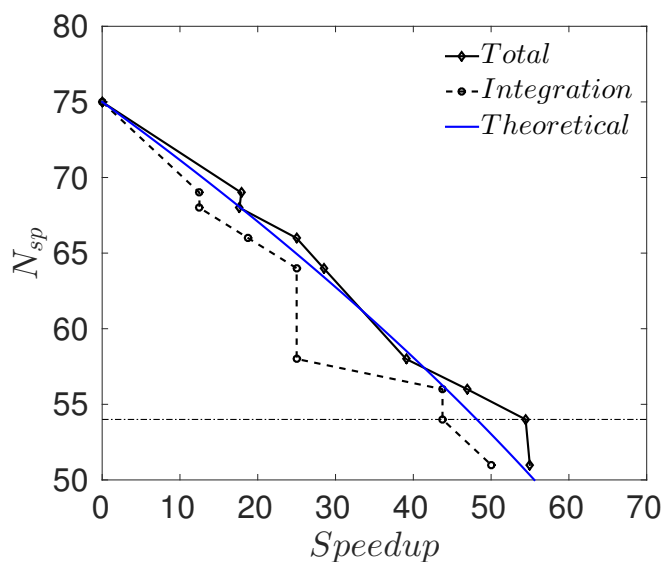
Figure 3. Number of species against percentage speedup for total simulation time, and for integrating Eq. 7 alone.
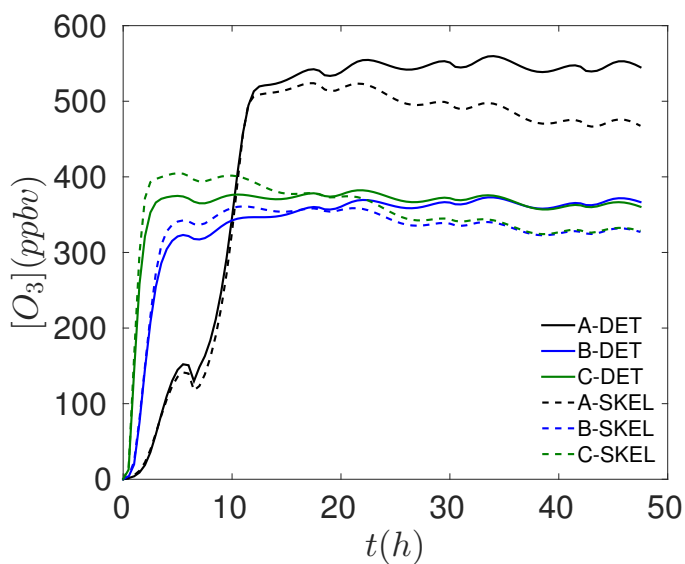
Figure 4. Detailed and skeletal ($N_{sp} = 54$) ozone profiles for cases A-C.
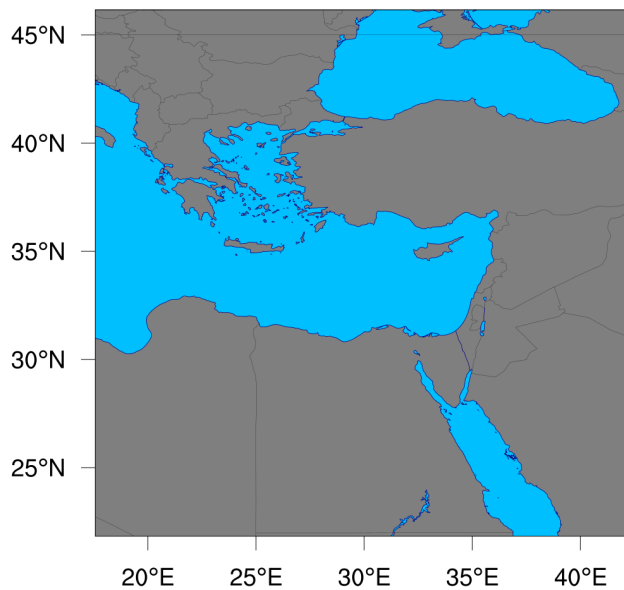
Figure 5. The geographic domain utilized during the WRF-Chem simulations conducted in this study. The domain extends between 17.6°E to 42.4°E in the longitude direction, and between 21.9°N to 46.1°N in the latitude direction.
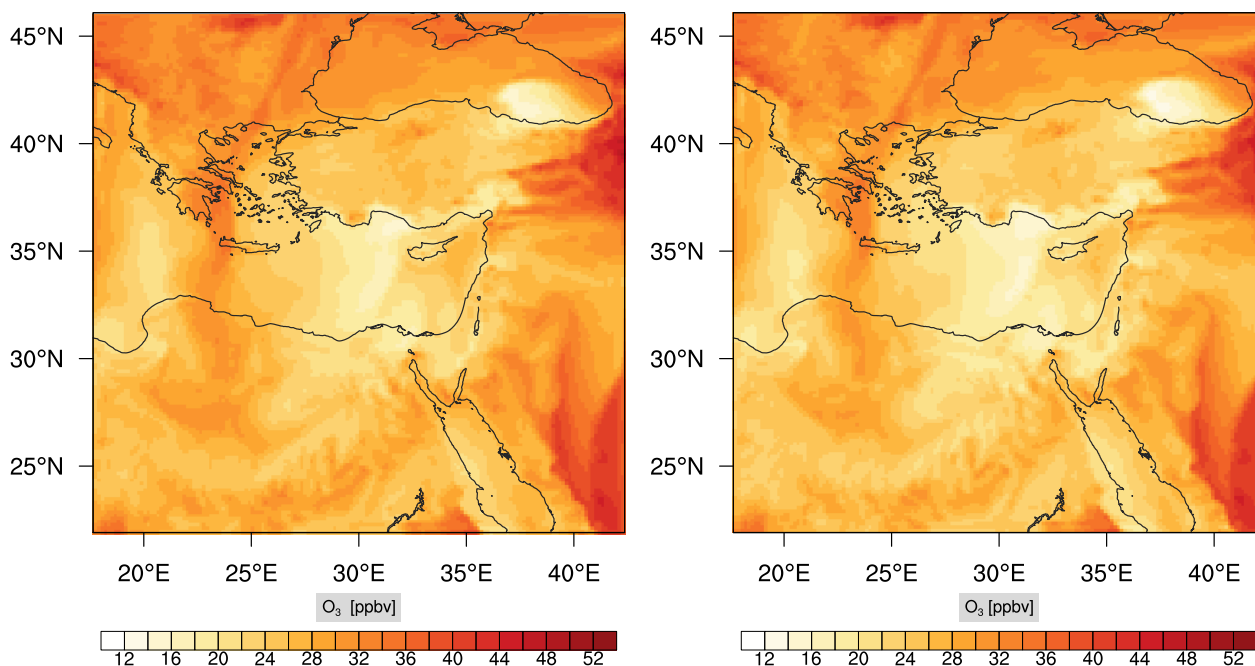
Figure 6. Instantaneous comparison of ozone spatial mixing ratio, averaged over the first nine vertical layers, using: the full mechanism (left), and skeletal mechanism (right).
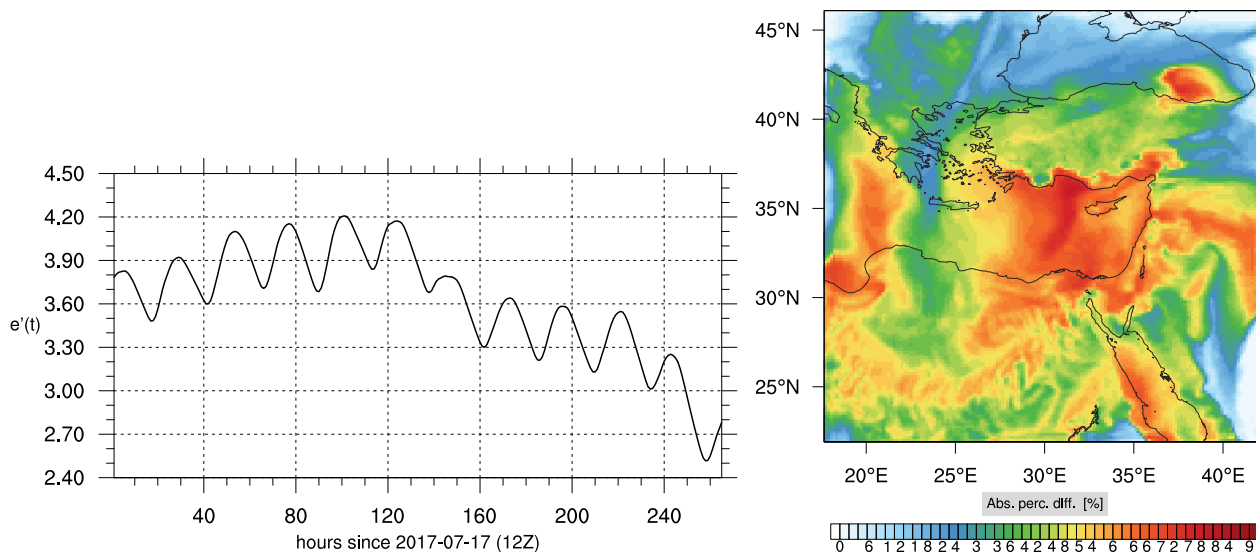
Figure 7. The line plot on the left shows the volume-weighted average of the absolute percentage difference between the full and skeletal mechanisms, $e'(t)$, for the ozone mixing ratio. The plot on the right depicts the spatial distribution of the absolute percentage difference between the reduced and the full mechanisms, w.r.t. full mechanism, for the ozone mixing ratio when $e'(t)$ is maximum.
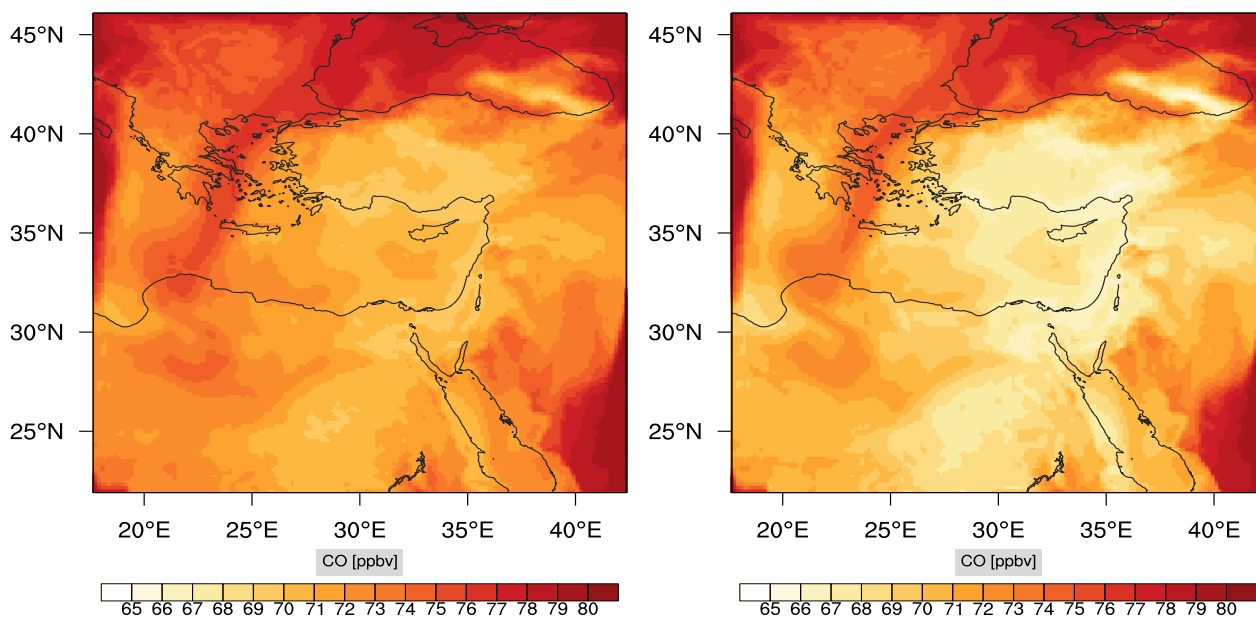
Figure 8. Instantaneous comparison of carbon monoxide spatial mixing ratio, averaged over the first nine vertical layers, using: the full mechanism (left), and skeletal mechanism (right).
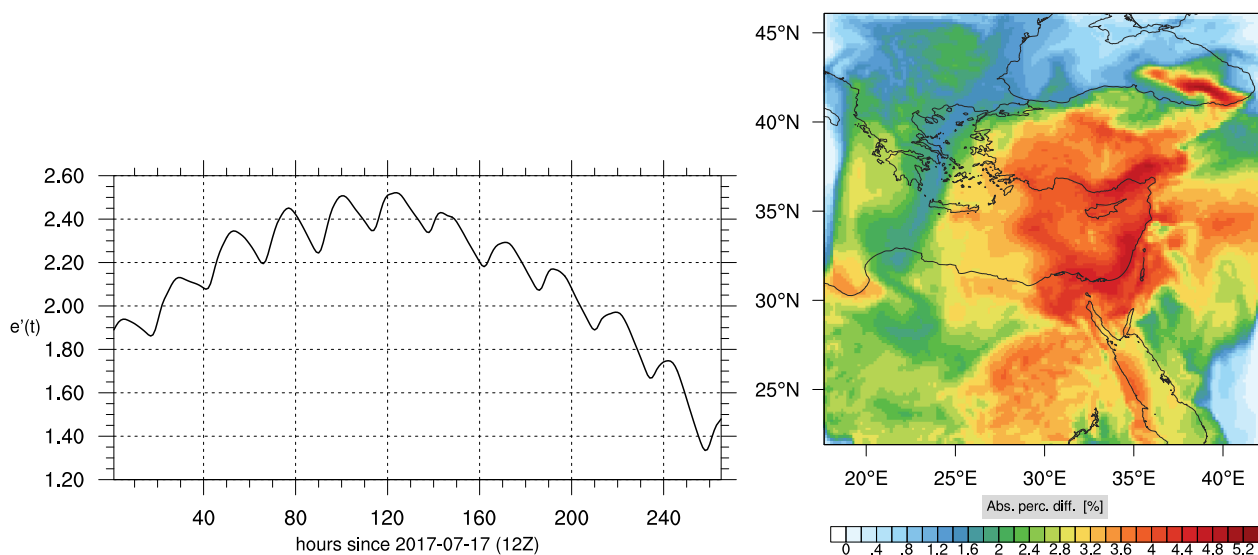
Figure 9. The line plot on the left shows the volume-weighted average of the absolute percentage difference between the full and skeletal mechanisms, $e'(t)$, for carbon monoxide mixing ratio. The plot on the right depicts the spatial distribution of the absolute percentage difference between the reduced and the full mechanisms, w.r.t. full mechanism, for the CO mixing ratio when $e'(t)$ is maximum.
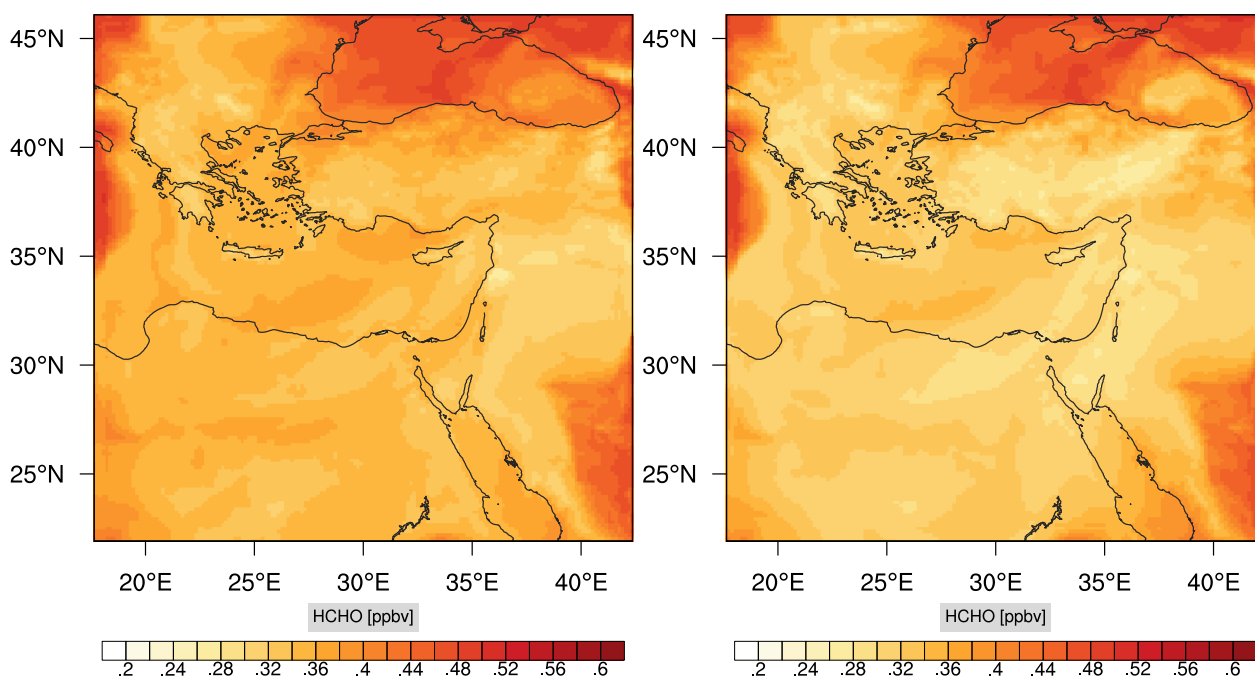
Figure 10. Instantaneous comparison of formaldehyde spatial mixing ratio, averaged over the first nine vertical layers, using: the full mechanism (left), and skeletal mechanism (right).
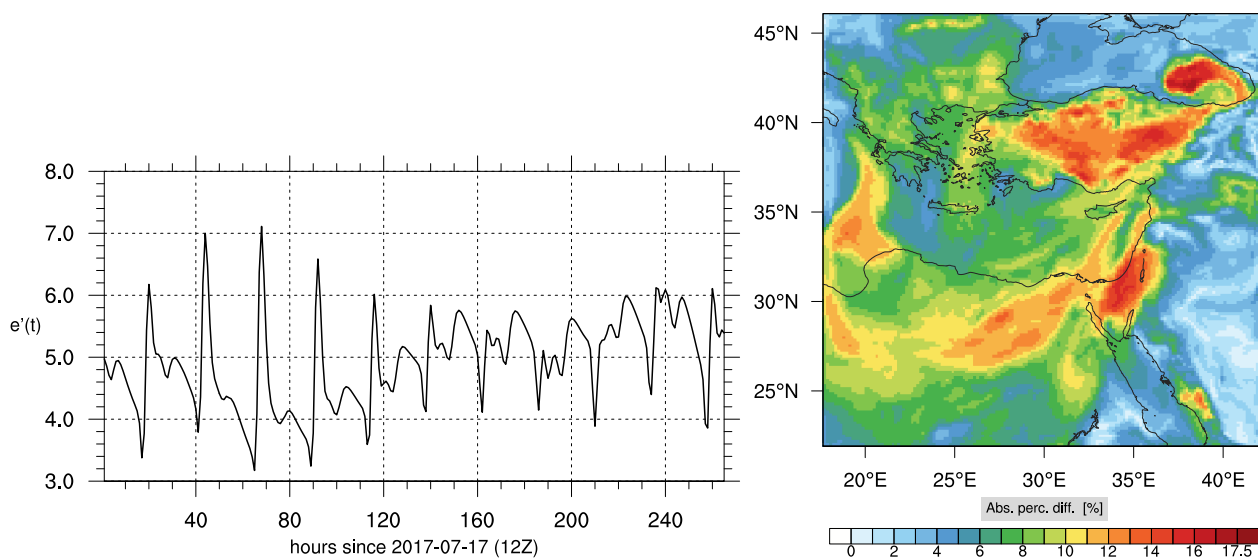
Figure 11. The line plot on the left shows the volume-weighted average of the absolute percentage difference between the full and skeletal mechanisms, $e'(t)$, for formaldehyde mixing ratio. The plot on the right depicts the spatial distribution of the absolute percentage difference between the reduced and the full mechanisms, w.r.t. full mechanism, for the formaldehyde mixing ratio when $e'(t)$ is maximum.