Positive semi-definite variants of CBM4 and CBM05 chemistry schemes for atmospheric composition models

Risto Hänninen[1], Rostislav Kouznetsov[1], and Mikhail Sofiev[1]

[1]Finnish Meteorological Institute, P.O. BOX 503, FI-00101 Helsinki, Finland

Correspondence: Risto Hänninen (Risto.Hanninen@fmi.fi)

Abstract. Carbon bond chemistry models are widely used to describe tropospheric chemistry reactions in atmospheric chemistry transport models. The standard implementation of these carbon bond models includes a species for paraffin carbon bond which is positioned in chemistry reactions such that its concentration can be easily driven to negative unless additional mechanisms have been implemented around the scheme. Here we describe an alternative formulation which is positively semi-definite, while preserving the main structure of the carbon bond mechanism. We apply the formulation to CBM4 and CBM05 schemes. The new formulation substantially improves the numerical robustness of the schemes and relaxes the need for external mechanisms to enforce positive concentrations.

1 Introduction

Chemistry-transport models (CTMs) have been used to address a large set of problems related to climate change, air pollution, atmospheric physics etc. One of the core parts of a CTM is the chemistry scheme that is responsible for chemical transformations of transported admixtures. While there are complicated chemistry schemes with several thousand chemistry equations and hundreds of species, they are computationally too heavy for cases where one need to evaluate large areas (or the globe) over long time. In operational / forecasting context computation time constraints pose even stricter limits. Therefore, simplified schemes, like a set of carbon bond models (e.g., Yarwood et al., 2005a, b, 2010), have been used both in operational and research environment.

Inside a CTM chemistry scheme is a module that transforms a chemical state (a vector of the species concentrations) at the beginning of a time step to a state at the end of the time step. An important property of a scheme is that being initialized with physically realizable state its subsequent states also stay physically realizable. In particular it means that if the scheme is given all non-negative initial concentrations of chemicals the concentrations should remain non-negative during the integration of the scheme over arbitrary time interval. As we show below, several commonly used schemes, such as CBM4 (Yarwood et al., 2005a), CBM05 (Yarwood et al., 2005b), and CBM6 (Yarwood et al., 2010) do not satisfy this criterion. It is possible to formulate the initial state for the scheme with all non-negative concentrations that would drive the concentration of the paraffin carbon bond (PAR) negative. These negative values pose a technical problem for some formulations of convergence criteria for the scheme integrator, that should be able to distinguish convergence to a state with negative concentrations from failure to converge. Besides that, the resulting model state is not physically-realizable. In applications such situations are

relatively rare, so their handling has no significant effect on the overall model performance in real-case simulations. On the other hand, handling such negative concentrations, e.g. by simply trimming them to zero, explicitly breaks mass budgets, and poses additional overhead on model implementation and affects run-time stability and performance in large simulations.

In this paper, we describe a variable change that allows for reformulating CBM4 and CBM05 chemistry equations such that the chemical does not evolve to negative concentration. This updated definition has helped us, since the version 5.7 of SILAM CTM [https://silam.fmi.fi/], to get rid of many situations where the simulation otherwise would crash or become very slow.

The paper is organised as follows. The basic requirements to the chemistry schemes are formulated in the Problem description section. It is followed by the CBM schemes adaptation section, where we show how the equations can be changed in case of CBM4 model to make it comply to the requirements. In section 4 we write the full set of modified equations in case of CBM05 model where the changes required a bit more work. Section 5 is devoted to discussion of the proposed principles and features of the adjusted CBM schemes.

2 Problem description

2.1 Basic terms

Let a chemical scheme be formulated for a set of N species C_i, whose concentrations are $[C_i]$. Then the state of the scheme can be described by a vector of concentrations $[C_i]$, where $i = 1 \ldots N$.

Definition 1. A state of is called *physical* if concentrations of all species are non-negative: $[C_i] \geq 0, \forall i \in 1..N$.

Definition 2. A chemistry scheme is called *positive semi-definite* if a solution starting from any physical state stays physical, for arbitrary integration time.

Definition 3. A chemistry scheme is called *analytically positive semi-definite* if its solution with a solver of infinitely-high precision is positive semi-definite.

Definition 4. A chemistry scheme is called *numerically positive semi-definite* if its solution with a solver of finite precision is positive semi-definite.

Proposition. A scheme formulated as a set of binary-reaction equations

$$C_i + C_j \rightarrow \sum_k d_{ijk} C_k \tag{1}$$

is analytically positive semi-definite if all reaction rates $R_{ij} \geq 0$, and all yields $d_{ijk} \geq 0, \forall i, j, k \in 1..N$ and $k \neq i$, $k \neq j$.

A proof of this proposition is straightforward if one notices that this equation in differential form is equivalent to:

$$\frac{d[\mathrm{C}_i]}{dt} = -\sum_j R_{ij}[\mathrm{C}_i][\mathrm{C}_j] \tag{2}$$

$$\frac{d[\mathrm{C}_k]}{dt} = \sum_{i,j \neq k} R_{ij} d_{ijk}[\mathrm{C}_i][\mathrm{C}_j] \tag{3}$$

In that system of differential equations the first one describes exponential decay of the corresponding species (both C_i and C_j in Eq.(1)), whereas the second one distributes the products of the reaction to other species. As long as all the yields d_{ijk} and the

reaction rates R_{ij} are positive, every product of the reaction gets a positive increment, and the decrease of a reactant happens only until it is positive. Conversely, if a yield d_{ijk} is negative for some k, and additionally there are no large enough positive yields d_{iik} and d_{jjk}, a solution starting from a state with $C_k = 0$ will immediately drive C_k to negative values.

Remark. In the above proposition, the yields of the species present in both sides of the equation can have any sign, i.e. d_{iji} and d_{ijj} can have any sign.

Validity of this remark is easily seen if one considers the following reaction:

$$A + B \rightarrow D - \alpha A \tag{4}$$

where $\alpha > 0$, i.e. the reaction describes a chain that produces a product D additionally consuming α molecules of A for each molecule of A that enters the chain. Its differential form reads as:

$$\frac{d[\text{A}]}{dt} = -(k + \alpha)[\text{A}][\text{B}] \tag{5}$$

The equation 5 describes an exponential decay with the rate $k + \alpha$, which does not lead to unphysical results.

Generally, it might be possible to derive a more relaxed condition for second-order reaction yields d_{ijk} (for $i, j \neq k$) which keeps the scheme analytically positively semi-definite. A weaker condition could be found using the same principles than is used to determine whether a symmetric matrix is positive semi-definite, or not. However, one easily notices that a negative yield d_{ijk} is allowed only if there are large enough positive yields d_{iik} and d_{jjk}. In case of original CBM4 and CBM05 formulations, both of these are zero when $d_{ijk} < 0$. Therefore, in case of CBM4 and CBM05, a simple test of non-negativity is enough. Also, for an analytically positively defined scheme a negative yield d_{ijk} (for $i, j \neq k$) might result that the scheme is not numerically positively semi-definite any more. This originates from the fact that now the production rate for $[C_k]$ in Eq.(3) contains a negative number. Due to a finite relative precision used to solve the system, this sum might become negative.

2.2 CBM schemes are not positive semi-definite

The problem that occurs in the carbon bond schemes is illustrated here for a case of CBM4 model. Here the naming of the species and the equation numbers are taken from Yarwood et al. (2005a) which lists all the reactions for the OTAG version of CBM4. Here we describe the method and only list the affected reactions.

Comparing the CBM4 reactions (Table 1) with the above Proposition, one can see that several reactions violate its requirements. Namely, the negative factor -1 for PAR in the reaction 58 means that not only O3 (ozone) and OLE (olefin carbon bond) are consumed in reaction, but also PAR (paraffin carbon bond) is consumed in equal amount, while not affecting the reaction. In the reaction 53 the paraffin bonds are consumed 2.1 times more than ROR (secondary alkoxy radical). Writing the reaction 58 in differential form, we get:

$$\frac{d[\text{OLE}]}{dt} = -k_{58}[\text{OLE}][\text{O3}] \tag{6}$$

$$\frac{d[\text{PAR}]}{dt} = -k_{58}[\text{OLE}][\text{O3}] \tag{7}$$

Here k_{58} is the reaction rate. Assuming for simplicity that ozone is in large excess and its concentration can be considered constant, the system of equations can be solved analytically:

$$[\text{OLE}](t) = [\text{OLE}]_0 \exp(-k_{58}[\text{O3}]t) \tag{8}$$

$$[\text{PAR}](t) = [\text{PAR}]_0 - [\text{OLE}]_0 \left(1 - \exp(-k_{58}[\text{O3}]t)\right), \tag{9}$$

where t is time and subscript 0 denotes concentrations at $t = 0$. From Eq. (9), if the initial molar concentrations $[\text{PAR}]_0 < [\text{OLE}]_0$, the concentration $[\text{PAR}]$ will become negative at:

$$t_b = -\frac{1}{k_{58}[\text{O3}]} \ln \left(1 - \frac{[\text{PAR}]_0}{[\text{OLE}]_0}\right) \tag{10}$$

In majority of real-life cases, the concentration of PAR is much larger than OLE and ROR species and even if their concentrations decrease the PAR concentrations will remain positive. But for practically meaningful applications in large-scale chemistry transport models strictly positive semi-definite schemes are required.

The implications of the reaction representation with a negative term on the right-hand-side depend on the reaction itself. In particular, the yield positiveness can be relaxed if the same term appears in left- and right-hand side of Eq. (1), i.e. if $k = i$ in some term. Such term can have negative sign without breaking down the scheme positiveness (e.g. reaction 52 in Table 1). Indeed, writing the reaction as a differential equation for PAR and OH, one gets:

$$\frac{d[\text{PAR}]}{dt} = -1.11 k_{52}[\text{PAR}][\text{OH}] \tag{11}$$

$$\frac{d[\text{OH}]}{dt} = -k_{52}[\text{PAR}][\text{OH}] \tag{12}$$

The first equation just has a 11% higher rate than the second one, not causing any problem from the general standpoint.

Table 1. Original CBM4 reactions affected by the variable changes described in the text. The reaction numbers correspond to number used in Yarwood et al. (2005a), which lists the reactions of the CBM4 model where so-called OTAG update is included.

React. #	Reactants	Products
52	PAR + OH	0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR - 0.11 PAR
53	ROR	0.96 XO2 + 1.1 ALD2 + 0.94 HO2 + 0.04 XO2N - 2.10 PAR
54	ROR	HO2
55	ROR + NO2	NTR
56	O + OLE	0.63 ALD2 + 0.38 HO2 + 0.28 XO2 + 0.3 CO + 0.2 FORM + 0.02 XO2N + 0.22 PAR + 0.2 OH
57	OH + OLE + M	FORM + ALD2 + XO2 + HO2 - PAR + M
58	O3 + OLE	0.5 ALD2 + 0.74 FORM + 0.22 XO2 + 0.1 OH + 0.33 CO + 0.44 HO2 - PAR
59	NO3 + OLE	0.91 XO2 + FORM + 0.09 XO2N + ALD2 + NO2 - PAR

2.3 Can the problem be patched?

Despite not being positive semi-definite, the CBM schemes are among the most-widely used in chemistry transport modeling.

105 The implementations try to alleviate the non-positiveness problem by a variety of patches, depending on numerical implementation of the chemistry solver and stability of other model components. For instance, the solver can (i) allow for negative-masses, (ii) cut out the negative mass to zero, (iii) start reducing the integration time step under (false) assumption that negative mass is just a numerical issue, (iv) recognize the problem and break down the simulations. None of these solutions is satisfactory: (i) is unphysical and has high chances to break the run later, (ii) breaks down the mass conservation, (iii) makes the run stuck

110 with the underflow error, (iv) breaks the run.

A popular recommendation seemingly eliminating the issue is to restrict the possible relations of initial concentrations of some species. Indeed, from the Eq. (10) one can see that if initially [PAR] > [OLE] the solution will remain positive. From the Reaction 53 of CBM4, the relation must be [PAR] > 2.1 [ROR], etc. For the first look, imposing these restrictions at the start of the chemistry solver can still be sufficient to avoid unphysical state after the time step. True for analytical

115 solutions, it still does not work in numerical simulations, which are performed with a finite relative precision floating-point representations. Due to finite relative precision, the absolute uncertainty of the concentrations of each species is proportional to their absolute concentrations. Then the negative result can be obtained just due to loss of precision on subtraction: in case of high and close concentrations of PAR and OLE, the numerical uncertainty of their values can be higher than their difference, thus allowing for negative values. Our experience with both CBM4 and CBM05 in SILAM model v.5.7 (System for Integrated

120 modeLling of Atmospheric coMposition [https://silam.fmi.fi/], Sofiev et al. (2015)) showed that problems appear already at $[PAR]_0 \sim [OLE]_0$ causing many-fold reduction of the integration time step and, eventually, negative masses.

Table 2. Affected CBM4 reactions after variable changes. Additionally, one simply needs to replace PAR with PAR4 in the remaining CBM4 reactions 72, 75, 77, 78, 92-95 in Yarwood et al. (2005a).

React. #	Reactants	Products
52	PAR4 + OH	0.87 XO2 + 0.13 XO2N + 0.11 HO2 + 0.11 ALD2 + 0.76 ROR4 - 1.706 PAR4
53	ROR4	0.96 XO2 + 1.1 ALD2 + 0.94 HO2 + 0.04 XO2N
54	ROR4	HO2 + 2.1*PAR4
55	ROR4 + NO2	NTR + 2.1*PAR4
56	O + OLE4	0.63 ALD2 + 0.38 HO2 + 0.28 XO2 + 0.3 CO + 0.2 FORM + 0.02 XO2N + 1.22 PAR4 + 0.2 OH
57	OH + OLE4 + M	FORM + ALD2 + XO2 + HO2 + M
58	O3 + OLE4	0.5 ALD2 + 0.74 FORM + 0.22 XO2 + 0.1 OH + 0.33 CO + 0.44 HO2
59	NO3 + OLE4	0.91 XO2 + FORM + 0.09 XO2N + ALD2 + NO2

3 Resolving the problem: positive semi-definite modification of CBM4

The suggested solution to the above problem is redefining the OLE and ROR species to include some PAR's. Let us define OLE4 = OLE+PAR and ROR4 = ROR+2.1PAR. This means that we do not count all the free olefin carbon bonds (C=C), but

125 make a combination of olefin and paraffin bond (C=C-C), and that on average the secondary alkoxy radical (ROR) contains also 2.1 paraffin bonds. The left-over of PAR we denote PAR4. When these new species, OLE4 or ROR4, break or react with other species they release PAR4. Table 2 illustrates the reactions after the above changes.

As one may note, the negative signs at the reaction yields have disappeared, except for equation 52, which, as shown above, is not a problem: PAR4 appears on both sides of the equation. In addition to the changes denoted in Table 2, one should replace

130 PAR with PAR4 all equations not mentioned in the table.

4 Applying the modification to CBM05

Since CBM05 model is more modern and currently used in many atmospheric chemistry transport models, we list here the transformations needed to modify the chemistry equations and the full list of equations after the transformations. The original names for species, equations, and their numbering, are taken from Yarwood et al. (2005b). The principle is very similar to

135 the CBM4 case, but here one needs more redefined species where PAR is included. The new variables are defined in Table 3 and the resulting new equations are listed in Table 4 and Table 5. The affected reaction numbers are marked in bold in the corresponding tables.

Table 3. Variable changes done for the CBM05 chemistry scheme. The naming of the species originates from publication Yarwood et al. (2005b). The last line means that PAR5 denotes remaining free PAR's after these transformations.

New species		Original CBM05 species
ROR5	=	ROR + ((2.1+0.04*0.66)/0.98)*PAR = ROR + 2.1698*PAR
OLE5	=	OLE + (1+0.09*0.66)*PAR = OLE + 1.0594*PAR
NTR5	=	NTR + 0.66*PAR
CRO5	=	CRO + 0.66*PAR
CRES5	=	CRES + 0.66*PAR
TO25	=	TO2 + 0.66*PAR
TOL5	=	TOL + (0.56+0.36)*0.66*PAR = TOL + 0.6072*PAR
XO2N5	=	XO2N + 0.66*PAR
IOLE5	=	IOLE + 0.3*0.09*0.66*PAR = IOLE + 0.01782*PAR
ETHA5	=	ETHA + 0.009*0.66*PAR = ETHA + 0.00594*PAR
ISOP5	=	ISOP + 0.088*0.66*PAR = ISOP + 0.05808*PAR
TERP5	=	TERP + 0.78*0.66*PAR = TERP + 0.5148*PAR
PAR5	=	PAR

Table 4. All main CBM05 equations [from Yarwood et al. (2005b)] after variable changes denoted in Table 3.

React. #	Reactants	Products
R1	NO2 + hv	NO + O
R2	O + O2 + M	O3 + M
R3	O3 + NO	NO2
R4	O + NO2	NO
R5	O + NO2 + M	NO3 + M
R6	O + NO + M	NO2 + M
R7	NO2 + O3	NO3
R8	O3 + hv	O
R9	O3 + hv	O1D
R10	O1D + M	O + M
R11	O1D + H2O	2*OH
R12	O3 + OH	HO2
R13	O3 + HO2	OH
R14	NO3 + hv	NO2 + O
R15	NO3 + hv	NO
R16	NO3 + NO	2*NO2
R17	NO3 + NO2	NO + NO2
R18	NO3 + NO2 + M	N2O5 + M
R19	N2O5 + H2O	2*HNO3
R20	N2O5 + 2*H2O	2*HNO3 + H2O
R21	N2O5 + M	NO3 + NO2
R22	NO + NO + O2	2*NO2
R23	NO + NO2 + H2O	2*HONO
R24	NO + OH + M	HONO + M
R25	HONO + hv	NO + OH
R26	OH + HONO	NO2
R27	HONO + HONO	NO + NO2
R28	NO2 + OH + M	HNO3 + M
R29	OH + HNO3 + M	NO3 + H20 + M
R30	HO2 + NO	OH + NO2
R31	HO2 + NO2 + M	PNA + M
R32	PNA + M	HO2 + NO2 + M
R33	OH + PNA	NO2
R34	HO2 + HO2	H2O2
R35	HO2 + HO2 + H2O	H2O2 + H2O

React. #	Reactants	Products
R36	H2O2 + hv	2*OH
R37	OH + H2O2	HO2 + H20
R38	O1D + H2	OH + HO2
R39	OH + H2	HO2
R40	OH + O	HO2
R41	OH + OH	O + H2O
R42	OH + OH + M	H2O2 + M
R43	OH + HO2	H2O
R44	HO2 + O	OH
R45	H2O2 + O	OH + HO2
R46	NO3 + O	NO2
R47	NO3 + OH	HO2 + NO2
R48	NO3 + HO2	HNO3
R49	NO3 + O3	NO2
R50	NO3 + NO3	2*NO2
R51	PNA + hv	0.61*HO2 + 0.61*NO2 + 0.39*OH + 0.39*NO3
R52	HNO3 + hv	OH + NO2
R53	N2O5 + hv	NO2 + NO3
R54	XO2 + NO	NO2
R55	XO2N5 + NO	NTR5
R56	XO2 + HO2	ROOH
R57	XO2N5 + HO2	ROOH + 0.66*PAR5
R58	XO2 + XO2	
R59	XO2N5 + XO2N5	1.32*PAR5
R60	XO2 + XO2N5	0.66*PAR5
R61	NTR5 + OH	HNO3 + HO2 + 0.33*FORM + 0.33*ALD2 + 0.33*ALDX
R62	NTR5 + hv	NO2 + HO2+ 0.33*FORM + 0.33*ALD2 + 0.33*ALDX
R63	ROOH + OH	XO2 + 0.500*ALD2 + 0.500*ALDX
R64	ROOH + hv	OH + HO2 + 0.500*ALD2 + 0.500*ALDX
R65	OH + CO	HO2
R66	OH + CH4	MEO2
R67	MEO2 + NO	FORM + HO2 + NO2
R68	MEO2 + HO2	MEPX
R69	MEO2 + MEO2	1.37*FORM + 0.74*HO2 + 0.63*MEOH
R70	MEPX + OH	0.7*MEO2 + 0.3*XO2 + 0.3*HO2

React. #	Reactants	Products
R71	MEPX + hv	FORM + HO2 + OH
R72	MEOH + OH	FORM + HO2
R73	FORM + OH	HO2 + CO +H20
R74	FORM + hv	2*HO2 + CO
R75	FORM + hv	CO
R76	FORM + O	OH + HO2 + CO
R77	FORM + NO3	HNO3 + HO2 + CO
R78	FORM + HO2	HCO3
R79	HCO3	FORM + HO2
R80	HCO3 + NO	FACD + NO2 + HO2
R81	HCO3 + HO2	MEPX
R82	FACD + OH	HO2
R83	ALD2 + O	C2O3 + OH
R84	ALD2 + OH	C2O3
R85	ALD2 + NO3	C2O3 + HNO3
R86	ALD2 + hv	MEO2 + CO + HO2
R87	C2O3 + NO	MEO2 + NO2
R88	C2O3 + NO2 + M	PAN + M
R89	PAN + M	C2O3 + NO2
R90	PAN + hv	C2O3 + NO2
R91	C2O3 + HO2	0.8*PACD + 0.2*AACD + 0.2*O3
R92	C2O3 + MEO2	0.9*MEO2 + 0.9*HO2 + FORM + 0.1*AACD
R93	C2O3 + XO2	0.9*MEO2 + 0.1*AACD
R94	C2O3 + C2O3	2*MEO2
R95	PACD + OH	C2O3
R96	PACD + hv	MEO2 + OH
R97	AACD + OH	MEO2
R98	ALDX + O	CXO3 + OH
R99	ALDX + OH	CXO3
R100	ALDX + NO3	CXO3 + HNO3
R101	ALDX + hv	MEO2 + CO + HO2
R102	CXO3 + NO	ALD2 + NO2 + HO2 + XO2
R103	CXO3 + NO2	PANX
R104	PANX + M	CXO3 + NO2 + M
R105	PANX + hv	CXO3 + NO2

React. #	Reactants	Products
R106	PANX + OH	ALD2 + NO2
R107	CXO3 + HO2	0.8*PACD + 0.2*AACD + 0.2*O3
R108	CXO3 + MEO2	0.9*ALD2 + 0.9*XO2 + HO2 + 0.1*AACD + 0.1*FORM
R109	CXO3 + XO2	0.9*ALD2 + 0.1*AACD
R110	CXO3 + CXO3	2*ALD2 + 2*XO2 + 2*HO2
R111	CXO3 + C2O3	MEO2 + XO2 + HO2 + ALD2
R112	PAR5 + OH	0.87*XO2 + 0.13*XO2N5 + 0.11*HO2 + 0.06*ALD2 - 1.8448*PAR5 + 0.76*ROR5 + 0.05*ALDX
R113	ROR5	0.96*XO2 + 0.6*ALD2 + 0.94*HO2 + 0.04*XO2N5 + 0.02*ROR5 + 0.5*ALDX
R114	ROR5	HO2 + 2.1698*PAR5
R115	ROR5 + NO2	NTR5 + 1.5098*PAR5
R116	O + OLE5	0.2*ALD2 + 0.3*ALDX + 0.3*HO2 + 0.2*XO2 + 0.2*CO + 0.2*FORM + 0.01*XO2N5 + 1.2528*PAR5 + 0.1*OH
R117	OH + OLE5 + M	0.80*FORM + 0.33*ALD2 + 0.62*ALDX + 0.80*XO2 + 0.95*HO2 + 0.3594*PAR5
R118	O3 + OLE5	0.18*ALD2 + 0.74*FORM + 0.32*ALDX + 0.22*XO2 + 0.10*OH + 0.33*CO + 0.44*HO2 + 0.0594*PAR5
R119	NO3 + OLE5	NO2 + FORM + 0.91*XO2 + 0.09*XO2N5 + 0.56*ALDX + 0.35*ALD2
R120	O + ETH	FORM + 1.7*HO2 + CO + 0.7*XO2 + 0.3*OH
R121	OH + ETH + M	XO2 + 1.56*FORM + 0.22*ALDX + HO2 + M
R122	O3 + ETH	FORM + 0.63*CO + 0.13*HO2 + 0.13*OH + 0.37*FACD
R123	NO3 + ETH	NO2 + XO2 + 2*FORM
R124	IOLE5 + O	1.24*ALD2 + 0.66*ALDX + 0.1*HO2 + 0.1*XO2 + 0.1*CO + 0.11782*PAR5
R125	IOLE5 + OH	1.3*ALD2 + 0.7*ALDX + HO2 + XO2 + 0.01782*PAR5
R126	IOLE5 + O3	0.65*ALD2 + 0.35*ALDX + 0.25*FORM + 0.25*CO + 0.5*O + 0.5*OH + 0.5*HO2 + 0.01782*PAR5
R127	IOLE5 + NO3	1.18*ALD2 + 0.64*ALDX + HO2 + NO2 + 0.01782*PAR5
R128	TOL5 + OH	0.44*HO2 + 0.08*XO2 + 0.36*CRES5 + 0.56*TO25
R129	TO25 + NO	0.9*NO2 + 0.9*HO2 + 0.9*OPEN + 0.1*NTR5 + 0.594*PAR5
R130	TO25	CRES5 + HO2
R131	OH + CRES5	0.4*CRO5 + 0.396*PAR5 + 0.6*XO2 + 0.6*HO2 + 0.3*OPEN
R132	CRES5 + NO3	CRO5 + HNO3
R133	CRO5 + NO2	NTR5
R134	CRO5 + HO2	CRES5
R135	OPEN + hv	C2O3 + HO2 + CO

React. #	Reactants	Products
R136	OPEN + OH	XO2 + 2*CO + 2*HO2 + C2O3 + FORM
R137	OPEN + O3	0.03*ALDX + 0.62*C2O3 + 0.70*FORM + 0.03*XO2 + 0.69*CO + 0.08*OH + 0.76*HO2 + 0.2*MGLY
R138	OH + XYL	0.7*HO2 + 0.5*XO2 + 0.2*CRES5 + 0.8*MGLY + 0.77*PAR5 + 0.3*TO25
R139	OH + MGLY	XO2 + C2O3
R140	MGLY + hv	C2O3 + HO2 + CO
R141	O + ISOP5	0.75*ISPD + 0.50*FORM + 0.25*XO2 + 0.25*HO2 + 0.25*CXO3 + 0.30808*PAR5
R142	OH + ISOP5	0.912*ISPD + 0.629*FORM + 0.991*XO2 + 0.912*HO2 + 0.088*XO2N5
R143	O3 + ISOP5	0.65*ISPD + 0.6*FORM + 0.2*XO2 + 0.066*HO2 + 0.266*OH + 0.2*CXO3 + 0.15*ALDX + 0.40808*PAR5 + 0.066*CO
R144	NO3 + ISOP5	0.2*ISPD + 0.8*NTR5 + XO2 + 0.8*HO2 + 0.2*NO2 + 0.8*ALDX + 1.93008*PAR5
R145	OH + ISPD	1.565*PAR5 + 0.167*FORM + 0.713*XO2 + 0.503*HO2 + 0.334*CO + 0.168*MGLY + 0.252*ALD2 + 0.210*C2O3 + 0.250*CXO3 + 0.120*ALDX
R146	O3 + ISPD	0.114*C2O3 + 0.15*FORM + 0.85*MGLY + 0.154*HO2 + 0.268*OH + 0.064*XO2 + 0.02*ALD2 + 0.36*PAR5 + 0.225*CO
R147	NO3 + ISPD	0.357*ALDX + 0.282*FORM + 0.721*PAR5 + 0.925*HO2 + 0.643*CO + 0.85*NTR5 + 0.075*CXO3 + 0.075*XO2 + 0.15*HNO3
R148	ISPD + hv	0.333*CO + 0.067*ALD2+ 0.9*FORM + 0.832*PAR5+ 1.033*HO2 + 0.7*XO2+ 0.967*C2O3
R149	TERP5 + O	0.15*ALDX + 5.6348*PAR5
R150	TERP5 + OH	0.75*HO2 + 1.25*XO2 + 0.25*XO2N5 + 0.28*FORM + 2.0098*PAR5 + 0.47*ALDX
R151	TERP5 + O3	0.57*OH + 0.07*HO2 + 0.76*XO2 + 0.18*XO2N5 + 0.24*FORM + 0.001*CO + 7.4554*PAR5 + 0.21*ALDX + 0.39*CXO3
R152	TERP5 + NO3	0.47*NO2 + 0.28*HO2 + 1.03*XO2 + 0.25*XO2N5 + 0.47*ALDX + 0.53*NTR5
R153	SO2 + OH	SULF + HO2
R154	OH + ETOH	HO2 + 0.9*ALD2 + 0.05*ALDX + 0.1*FORM + 0.1*XO2
R155	OH + ETHA5	0.991*ALD2 + 0.991*XO2 + 0.009*XO2N5 + HO2
R156	NO2 + ISOP5	0.2*ISPD + 0.8*NTR5 + XO2 + 0.8*HO2 + 0.2*NO + 0.8*ALDX + 1.93008*PAR5

5 Discussion

As it follows from the above considerations, the condition of a comparatively safe application of both CBM4 and CBM05 is
140 that the PAR concentrations are much larger than those of several other reactants. For the original CBM4 scheme it means that
one has to have [PAR] ≥ [OLE] and [PAR] ≥ 2.1[ROR] at the beginning of a time step to ensure that [PAR] would stay positive
by the end of the time step. For the original CBM05 formulation the requirements are similar but more numerous: [PAR] ≥
1.0594[OLE], etc (totally, 12 restrictions, Table 3). We could not find any criteria that would *guarantee* that an unphysical
solution would not appear due the finite relative precision of floating-point numbers during the integration. The numerical

Table 5. Additional CBM05 reactive chlorine equations [from Yarwood et al. (2005b)] after variable changes denoted in Table 3.

React. #	Reactants	Products
CL1	Cl2 + hv	2*Cl
CL2	HOCl + hv	OH + Cl
CL3	Cl + O3	ClO
CL4	ClO + ClO	0.3*Cl2 + 1.4*Cl
CL5	ClO + NO	Cl + NO2
CL6	ClO + HO2	HOCl
CL7	OH + FMCL	Cl + CO
CL8	FMCL + hv	Cl + CO + HO2
CL9	Cl + CH4	HCl + MEO2
CL10	Cl + PAR5	HCl + 0.87*XO2 + 0.13*XO2N5 + 0.11*HO2 + 0.06*ALD2
		- 1.8448*PAR5 + 0.76*ROR5 + 0.05*ALDX
CL11	Cl + ETHA5	HCl + 0.991*ALD2 + 0.991*XO2 + 0.009*XO2N5 + HO2
CL12	Cl + ETH + M	FMCL + 2*XO2 + HO2 + FORM + M
CL13	Cl + OLE5 + M	FMCL + 0.33*ALD2 + 0.67*ALDX + 2*XO2 + HO2 + 0.0594*PAR5 + M
CL14	Cl + IOLE5	0.3*HCl + 0.7*FMCL + 0.45*ALD2 + 0.55*ALDX + 0.3*OLE5 + 1.7*XO2 + HO2
CL15	Cl + ISOP5	0.15*HCl + XO2 + HO2 + 0.85*FMCL + ISPD + 0.05808*PAR5
CL16	Cl + FORM	HCl + HO2 + CO
CL17	Cl + ALD2	HCl + C2O3
CL18	Cl + ALDX	HCl + CXO3
CL19	Cl + MEOH	HCl + HO2 + FORM
CL20	Cl + ETOH	HCl + HO2 + ALD2

145   uncertainties in different chemical equations at any prescribed relative precision can flip the relation if the concentrations are high enough and close enough. In any chemistry transport model there are also other modules, like advection and emission parts, where the numerical errors can brake any of these requirements too. Noteworthy, CBM05 is much more prone to the numerical problem because of larger number of restrictions.

     The suggested modifications eliminate the problem by re-defining the species and modifying the affected reactions. Im-
150 portantly, the scheme becomes not only analytically positively defined but, as follows from the proposition, also numerically positively defined. Indeed, comparing two large and close numbers has high numerical uncertainty due to limited mantissa length, whereas comparison of any number with zero has low uncertainty: fixed relative precision means that small numbers are represented with high absolute precision.

     One can argue that after transformations the equations, especially for the paraffin bond, are not exactly what was intended in
155 the original model since the amount of "free PAR" (PAR4 or PAR5) is less after transformation and the rate for the PAR+OH reaction may require tuning. This is correct and should be accounted for during development of new schemes. However, in

majority of cases the amount of PAR is much larger than those of other lumped species or OLE. In this limit the new and the original formulations are equivalent. Therefore, redefinition to PAR4 (or PAR5) only slightly modifies its amount and the PAR+OH reaction rate can be left untouched. This shortcut, however, can be invalid for other schemes.

160   The redefinition of the variables also affects emissions. Generally, the PAR emissions should be decreased by the corresponding amount in the new variables. For example, in the CBM4 case, the amount of emitted PAR's should be decreased by the amount of OLE plus 2.1 times the amount of ROR's emitted in the original scheme. In principle, this may cause that in some places the PAR4 emissions become negative, which is unphysical, but is easily avoidable in the numerical scheme. This problem originates from the construction of the carbon bond models in general. These models use lumped species and negative

165   coefficient on some PAR products (right hand side of the chemistry equations). They are not intended to accurately model any other real organic species that are not already in the model, like ethene. For example, 1,2-propadiene (or any other species with more OLE's than PAR's), according to Table 1 in the Appendix of (Yarwood et al., 2005b), is mapped to 1.5 OLE, and assume that no other species are emitted. If such a species emission is dominant in some grid cell, the original carbon bond scheme turns unphysical at the first model time step. Transformation to the adjusted scheme will formally require negative

170   PAR4 emission in the cell manifesting the problem in the emission dataset. Once it is corrected in the emission input, the CTM run will pass.

    Evaluation of the above formulations were performed with the SILAM CTM v.5.7 [https://silam.fmi.fi/], Sofiev et al. (2015), where the original chemistry equations of CBM4 caused rare (about once per 1–3 computed years for Europe-wide simulations with 0.1 degree resolution) hanging of the chemistry integrator when it tried to keep the concentrations positive by reducing

175   the time-step. With CBM05 these problems became over 10 times more frequent, prohibiting any large scale simulations without major modifications to the chemistry time integration module. The problem was most-frequently caused by: (i) random perturbations of the concentration fields in generation of ensemble for Ensemble Kalman Filter data assimilation, (ii) numerical errors in the advection-diffusion scheme, which transports the species independently and thus cannot guarantee any prescribed relation between them at every time step. With the above changes, the problem has completely disappeared and we could use

180   the standard integrator without additional handling of exceptions. Importantly, when the original scheme worked, its results for all standard air pollutants, like ozone, NOx, SO2, and PM's, were identical to those of the modified scheme. Therefore, a practical added value of the development is a drastically improved stability of the simulations.

    SILAM v.5.7 and v.5.8 with the updated chemistry scheme has been operationally evaluated within the scope of several international frameworks: Copernicus Atmospheric Monitoring Service, for Europe: https://regional.atmosphere.copernicus.

185   eu/ (last access: 10 January 2023), Environment and Climate Change Canada within the scope of WMO GAFIS, for Northern America: https://hpfx.collab.science.gc.ca/~svfs000/na-aq-mm-fe/dist/ (last access: 10 January 2023), Chinese Meteorological Administration within the WMO GAFIS, for Asia: http://www.asdf-bj.net/gafis/index.html (last access: 10 January 2023). Additional evaluation of NO2, O3, CO, and SO2 is also visible in the paper by Fatahi et al. (2021), which uses SILAM v.5.7 with the updated chemistry scheme.

190   As an illustrative example, we provide a box model results when using the CBM4 chemistry scheme and applying our variable changes to CBM4 equations and reaction rates provided by the KPP (Kinetic PreProcessor) version 2.2.3, see Damian
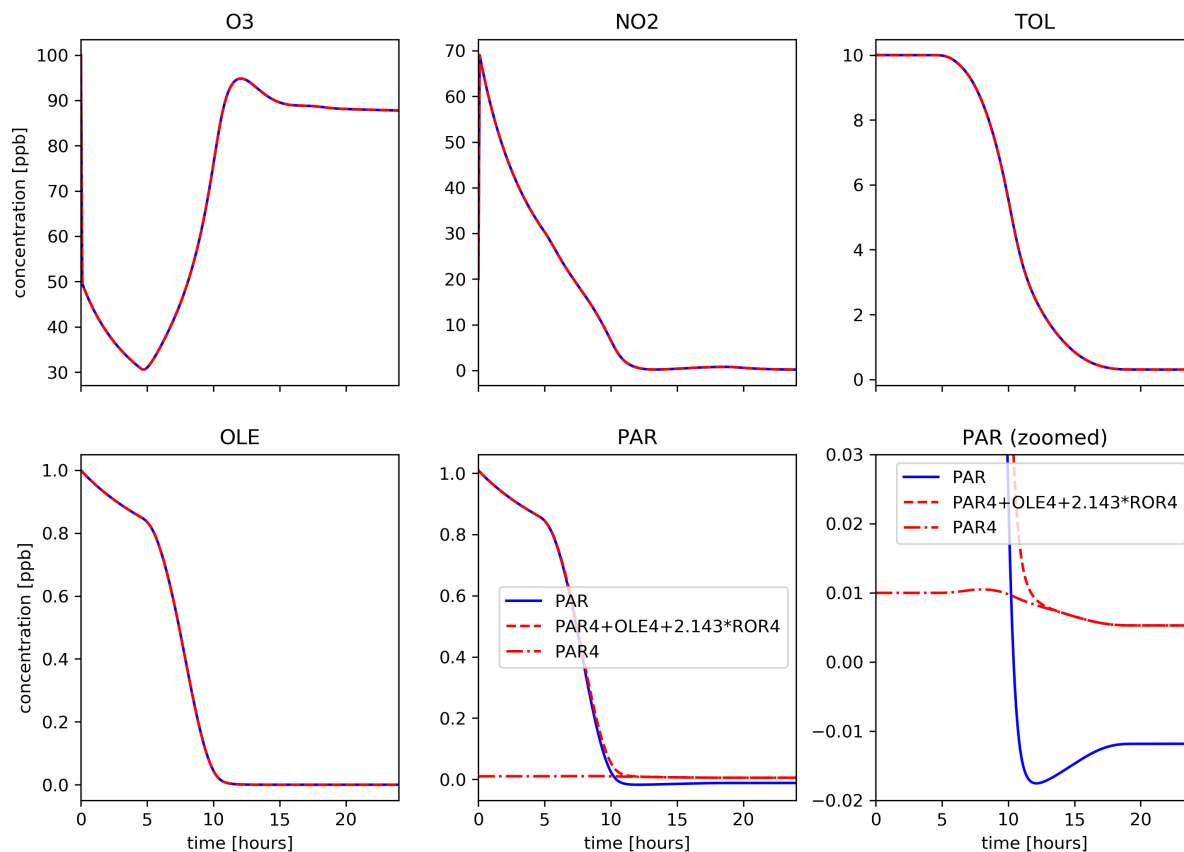
Figure 1. Box-model comparison of the original and modified CBM4 formulations when using the KKP v2.2.3 for integrating the concentrations. The blue lines are with the original CBM4 scheme, while the red dashed lines are with the update scheme. In case of PAR the red dash-dotted line is for PAR4, while the dashed line corresponds to the original PAR where the contributions from OLE4 and ROR4 are taken into account. The left-bottom panel illustrates negative concentration of PAR when using the original non positively defined CBM4 scheme.

et al. (2002); Daescu et al. (2003); Sandu et al. (2003). Slightly modifying the original initial configuration provided by KPP, one may obtain a situation where the original CMB4 formulation drives the PAR concentration into negative values. The figure 1 illustrates the concentrations of O3, NO2, TOL, OLE, and PAR when using the standard KKP setup for the integrator and

195   tolerances. The blue lines are with the original CBM4 equations and the red lines are with the modified formulation. In case of PAR, the red dash-dotted line is new variable PAR4, while the red dashed line describes the original PAR where the contributions from OLE4 and ROR4 are also counted in. Here the initial concentrations are such that [PAR]=1.01[OLE], still satisfying nicely the original requirement [PAR] ≥ [OLE]. As one may observe, both formulations provide very similar results, but the original formulation drives PAR concentrations into small negative values. This may cause that some other concentrations may

200   also get negative without additional handling of exceptions. The updated formulation keeps all concentrations positive.

6 Conclusions

The paper formulated a simple and sufficient criterion for positive definition of chemical schemes. Namely, yields of all chemical reactions must be non-negative:

$$C_i + C_j \rightarrow \sum_k d_{ijk} C_k \tag{13}$$

205 is positively defined if $d_{ijk} \geq 0$, $\forall i,j,k \in 1..N$ and $k \neq i$, $k \neq j$. Species present on the left-hand-side of equation can be additionally consumed as a result of the reaction, i.e. d_{iji} and d_{ijj} can be negative.

It has been shown that the popular carbon-bond schemes CBM4 and CBM05 are not positively defined. The problem was traced down to specific equations in each scheme, which violate the criterion: several equations can lead to negative concentrations of paraffin carbon bonds (PAR).

210 Simple modifications were suggested, such no concentration can be driven to negative values from any physically realizable initial state. The stability of the new formulations has been confirmed with the long-term global simulations of the SILAM chemistry transport model since version 5.7.

Similar modifications can be made also for, e.g., CBM6 scheme, which, does not satisfy the criterion of positive definition and thus can produce unphysical results.

215 *Code availability.* SILAM is an open-code system and can be obtained from the GitHub open repository: https://github.com/fmidev/silam-model (last access: 9 January 2023, https://doi.org/10.5281/zenodo.5713863). Numerical code used to make the box model illustration with CBM4 is available at Zenodo: https://doi.org/10.5281/zenodo.7763872.

Author contributions. RH formulated the modified chemistry scheme and implemented it in SILAM CTM with the help of all co-authors. RH prepared the manuscript with contributions from all co-authors.

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

Acknowledgements. We acknowledge support from EU Horizon 2020 projects EXHAUSTION (Grant No. 820655) and EMERGE (Grant No. 874990). Support of Academy of Finland research project GLORIA (Grant No. 310372) is appreciated.

References

Daescu, D., Sandu, A., and Carmichael, G.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: II – Validation and Numerical Experiments, Atmospheric Environment, 2003.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G.: The Kinetic PreProcessor KPP – A Software Environment for Solving Chemical Kinetics, Computers and Chemical Engineering, 2002.

Fatahi, Y., Kouznetsov, R., and Sofiev, M.: The effect of accounting for public holidays on the skills of the atmospheric composition model SILAM v.5.7, Geoscientific Model Development, 14, 7459–7475, https://doi.org/10.5194/gmd-14-7459-2021, 2021.

Sandu, A., Daescu, D., and Carmichael, G.: Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: I – Theory and Software Tools, Atmospheric Environment, 2003.

Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., and Genikhovich, E.: Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin, Geoscientific Model Development, 8, 3497–3522, https://doi.org/10.5194/gmd-8-3497-2015, 2015.

Yarwood, G., Whitten, G., and Rao, S.: Updates to the Carbon Bond 4 Photochemical Mechanism. Report prepared for the Lake Michigan Air Directors Consortium, Des Plains, IL, https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6a85e7478d0fb8f239053f31401ca355ec5516b1, (accessed 3 January, 2023), 2005a.

Yarwood, G., Yocke, S. R. M., and Whitten, G.: Updates to the Carbon Bond Mechanism: CB05. U.S. Environmental Protection Agency, Final Report, 161 pp., https://camx-wp.azurewebsites.net/Files/CB05_Final_Report_120805.pdf, (accessed 3 January, 2023), 2005b.

Yarwood, G., Whitten, G. Z., Jung, J., Heo, G., and Allen, D. T.: Development, Evaluation and Testing of Version 6 of the Carbon Bond Chemical Mechanism (CB6). Texas Commission on Environmental Quality, Final Report, 119 pp., https://www.eoas.ubc.ca/courses/atsc507/ADM/cmaq/cb6.pdf, (accessed 3 January, 2023), 2010.