Master Thesis
Program in Environmental Physics (PEP)

Enhancing chemical schemes accounted in the FLEXPART v10.4 transport model using a kinetic preprocessor

Ruben Sousse

Date of birth: 18.07.1993
Matriculation number: 3166789

1st examiner: Mihalis Vrekoussis
2nd examiner: Nikolaos Daskalakis

Laboratory for Modeling and Observation of the Earth System (LAMOS)
Institute of Environmental Physics (IUP)
University of Bremen

18 August 2020

to Andreas

Coupling *KPP* to FLEXPART 10.4

Ruben Sousse

Abstract

In this project, two main developments have been performed over the FLEXPART v10.4 atmospheric Lagrangian transport model. In the first place, TM5-MP OH fields have been made available as input for FLEXPART, replacing its default GEOS-Chem OH fields which are outperformed in terms of spatial and temporal resolution among other characteristics. Secondly, a box model produced with the kinetic preprocessor KPP have been coupled to FLEXPART as chemical solver based in Rosenbrock Rodas-3 numerical integration. It replaces the previous parametrized chemical reaction based on direct time-discretization.

These developments have been validated in terms of reproducibility with respect to the undeveloped FLEXPART v10.4 version in two different sets of experiments: in a detailed analysis of a froward-in-time global simulation of an ethane release and in a statistical analysis of similar simulations for each FLEXPART OH-reacting species (methane, ethane, PCB28 and γ-HCH). Additionally, a validation experiment against observational data has been attempted through a simulation emulating CMIP6 ethane emissions at European scale, although computational problems made it impossible to obtain relevant results from it. All these experiments were performed separately for each development stage, (i.e the implementation of TM5-MP OH fields and KPP coupling), making it possible to assess the relative effect of each development on the model.

The validation results show that the new developments cause an overall slight increase of the OH sink rate for all the species, specially for those short-lived like γ-HCH. The detailed analysis of the global ethane single release shows that most of the differences are regional-independent at the surface, and that they are caused mainly by the TM5-MP OH fields implementation rather than the change of chemical solver. No relevant conclusions can be obtained from the observational validation results, since not enough particles were possible to be released due to computational limitations.

Although the new developments could not be validated against observational data and therefore their effect could not be quantised, the reproducibility of the results seems to indicate that the new implementations should not affect simulations' results significantly in future experiments. On the other hand, the coupling of KPP-produced box model with FLEXPART signifies an increase in adaptability and flexibility on its chemical mechanism coefficients. Moreover, KPP opens the door to expand the FLEXPART chemical mechanism (currently limited to the OH oxidation of four species) to virtually any set of reactions required, setting the path for an enhancement in FLEXPART simulations' quality.

Contents

1. Introduction

1.1. Atmospheric transport and FLEXPART

Atmospheric pollution analysis is nowadays a critical study field in atmospheric sciences; the fast-growth of industrialized and urban areas added to an increase concern in understanding the reasons leading to breathing illnesses makes of air quality assessment and its causes a main concern of public interest. The correct assessment of air quality drivers, however, must take into account the transport of air pollutants from other areas than the one studied, like emissions from industrial areas moving to urban areas. In this scenario, the simulation of air transport dynamics and chemistry plays a key role in predicting and improving air quality studies [1].

When simulating atmospheric transport, FLEXPART (FLEXible PARTicle dispersion model) is a widely used tool and it has proofed reliable in different air quality assessments [2] [3] [4] [5] [6]. Among other features, it is able to account for the OH oxidation as a sink for the transported species. However, it is currently limited to this only reaction and cannot consider other chemical transformations of the transported species.

1.2. Main objectives

The final aim of this project is to couple FLEXPART to a chemical box model produced with the chemical preprocessor KPP. Such a coupling would lead to an enhanced chemical scheme in the model, hence improving the reliability of atmospheric transport simulations with FLEXPART.

In the first stage of this thesis, in order to obtain reliable results from the KPP coupling, high quality atmospheric tracers distribution data is obtained. Consequently, FLEXPART is developed to make concentration fields from the TM5-MP global chemistry model available for it as an input. These fields outweigh the default FLEXPART OH fields from GEOS-Chem model, both in terms of resolution and time frequency sampling.
In the second stage, the chemical reaction with OH in FLEXPART is developed. In this step, the default OH reaction conducted through a single equation based on time discretization is overridden by the numerical integration of the reaction in a chemical box model. This box model is created using KPP, which allows the chemical mechanism of the box model to be easily adapted to each simulation. This implementation can potentially make a broader variety of atmospheric chemical reactions available for FLEXPART. Extending the chemical mechanisms with the preprocessing tool *KPP* can imply a more realistic chemical scheme, highly flexible to specific situations, hence causing an enhancement of the transport simulation accuracy.

1.3. Previous work

As stated above, we have performed two main developments of FLEXPART in this project: 1) The implementation of TM5-MP OH fields and 2) the coupling of a KPP-produced box model.

In the first case, the implementation of offline concentration fields is a common situation among Lagrangian models, since they do not generate online chemistry fields. Therefore they usually are read from pre-generated data [5]. One can find most of the previous related work on models similar to FLEXPART, such asHYSPLIT [Draxler and Hess, 1997], NAME [Jones et al., 2007], or CLaMS [McKenna et al., 2002]. In the case of HYSPLIT, it has O_3 concentrations simulated through an Eulerian framework. In posterior developments of this model, also OH and SO_2 concentrations were estimated through model results and empirical data [7]. For example in the reaction of mercury emissions, these species were all retrieved from different empirical methods [8]. Another similar case is the NAME III model, which takes 3D fields of OH, HO_2, H_2O_2 and O_3 from the global chemistry model STOCHEM [9]. Finally, another more complete approach is provided by the CLaMS model, which includes 36 chemical species and 115 reactions for the stratosphere. It generates concentration fields from tracer-tracer and tracer-potential vorticity mapping using meteorological observations and model-generated data [10].

Regarding the coupling with a chemical preprocessor, there are some examples of Lagrangian models that incorporate chemical solvers. Chemical solvers are needed to numerically solve non-linear chemistry during the transport of species [5]. Transport and chemistry are normally simulated separately in Lagrangian models [1]; In general, the Lagrangian model simulates exclusively the transport and other physical processes (i.e. deposition, wet scavenging, etc.), while the coupled chemical solver computes the chemistry reactions at each time step of the simulation[7]. This coupling between the chemical solver and the Lagrangian model can be done directly or through an Eulerian model, which has its own chemical solver. Examples of this last type of arrangement are HYSPLIT or Geos-CHEM. In the first case, HYSPLIT uses an Eulerian model to compute the chemistry [7]. Meanwhile, the 1D-Eulerian model Geos-CHEM is directly coupled to the KPP chemistry preprocessor [11]. This last approach is the one taken as a referent in the current project.

2. Scientific background

2.1. Chemistry involved

The simulation of transport of chemical species in the atmosphere becomes a difficult problem when chemistry is considered. While being transported, species photolyze, deposit or chemically react with other species. To model these changes, the temperature, pressure and sunlight around the particle must be taken into account, together with the surrounding chemical species with which it might interact.

The atmospheric species used in this project have relatively simple chemistry. One of the simplest is ethane (C_2H_6), which reacts almost exclusively with hydroxyl radical (OH) and has a relatively short lifetime of 2.5 months in the atmosphere. These characteristics makes of ethane a good candidate to asses the sensibility of the simulation results to the model OH fields: Only the chemical sink of OH must be taken into account and simulations do not need to be run for long to appreciate differences in the results.

All the reaction rates in FLEXPART are calculated from the constants C, n and D following equation 1 [12]:

$$k = CT^n e^{-D/T} \tag{1}$$

In this section an overview of the chemical reactions involving OH in FLEXPART is presented.

2.1.1. Chemistry of atmospheric OH

OH is highly reactive in the low atmosphere. It reacts with most of the anthropogenic air pollutants or biogenic emissions from forests. The most common reactions include CO, NO_2, H_2S, SO_2, methane, other non-methane hydrocarbons (e.g. propane and ethane) and isoprene (C_5H_8) in forested areas. All these reactions cause the lifetime of OH very short, estimated in 1 second in the troposphere [13]. These reactions are summarized in the figure 1.

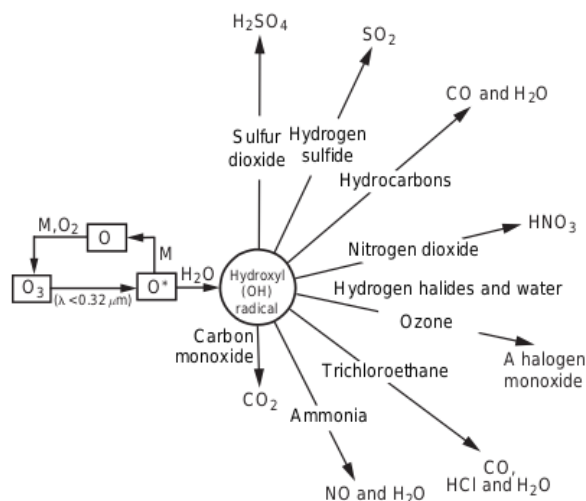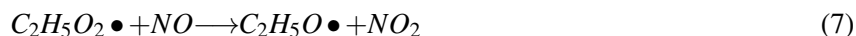OH is formed through the photolysis of ozone as follows [13]:

$$O_3 + hv \longrightarrow O \bullet + O_2 \tag{2}$$
$$O \bullet + H_2O \longrightarrow 2OH \bullet \tag{3}$$
$$O_3 + H_2O + hv \longrightarrow 2OH \bullet + O_2 \tag{4}$$

Fig. 1: Graphical summary of *OH* formation and afterwards reactions in the atmosphere (Source [13], p.163)

2.1.2. Chemistry of atmospheric C_2H_6

Ethane is an Alkane (C_nH_{2n+2}, with n being the number of carbon atoms) and classified as a VOC in the atmosphere. Its concentration and half-lifetime vary seasonally, being estimated in ranges of 0.07-3 ppbv and 2.5-3 months, respectively [14][15]. As mentioned before, in the atmosphere ethane reacts principally with OH (reaction 5), although its reaction with free Chlorine (Cl) also happens in the stratosphere and to a less extent with O and $O(^1D)$ in the troposphere [16][17]. The oxidation by OH produces methanol (reaction 7) in the presence of NO_X or C_2H_5OOH if no NO_X are present (reaction 6) [17]:

$$C_2H_6 + OH \bullet \longrightarrow C_2H_5 \bullet + H_2O \tag{5}$$
$$C_2H_5 \bullet + O_2 + M \longrightarrow C_2H_5O_2 \bullet + M \tag{6}$$
$$C_2H_5O_2 \bullet + NO \longrightarrow C_2H_5O \bullet + NO_2 \tag{7}$$

Where $\bullet$ refers to the excited mode of the molecule, i.e. the radical. As seen, main products are NO_2 and C_2H_5O, which leads to further reactions not concerning ethane.

By default, in FLEXPART v.10.4 model the rate of the first reaction is $0.254 10^{-12} = 1.52 10^{-17}T^2e^{498/T}$ $cm^3/molec^{-1}s^{-1}$ at 298K [12].

9

2.1.3. Chemistry of PCB-28

Another species that can be modelled by default in FLEXPART is Polychlorinated Biphenyl 28 (figure 2), whose emissions were forbidden in the 1970s because of its high persistence and toxicity for marine and terrestrial animals [18], classified as a Persistent Organic Pollutant (POP). Because of its danger, this species' transport has been of high interest for scientific studies, which concluded that OH oxidation was one of the main sinks of this compound [18].

It has been attributed a life time in the atmosphere of approximately 11 days, with a reaction rate of $1.10 10^{-12}$ $cm^3/molec^{-1}s^{-1}$, its main sink accuring in the lower troposphere [19].

By default, in FLEXPART v.10.4 model the rate of the reaction with OH is $1.00 10^{-11} e^{1203/T}$ $cm^3/molec^{-1}s^{-1}$ [12].
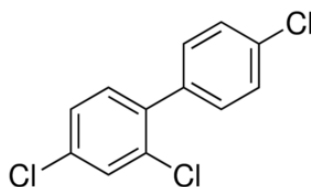

Fig. 2: PCB28 structure.

2.1.4. Chemistry of G-HCH

Also called Lindane, gamma-hexachlorocyclohexane (γ-HCH or G-HCH) was an insecticide used for agriculture and pharmacological purposes. Like PCB28, it is also classified as a POP and it was banned by the Stockholm Convention in 2009 in agriculture. Its structure can be appreciated in figure 3. Nowadays, it is only used for medical purposes as an alternative to most wide spread treatments [20] [21]. It's main sink in the atmosphere is wet deposition and oxidation by OH radical, for which it has been attributed a lifetime of approximately 20 days with a rate coefficient of $k_{OH} = 6.4e-13 cm^3 molecule^{-1}s^{-1}$ at 300K [21].

By default, in FLEXPART v.10.4 model the rate of the reaction with OH is $XXXXXXX = 6.00 10^{-11} T^2 e^{1203/T}$ $cm^3/molec^{-1}s^{-1}$ [12].
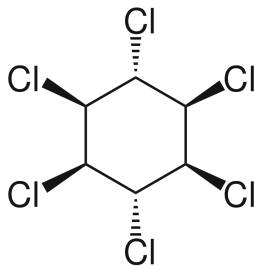

Fig. 3: G-HCH structure.

2.2. Simulating transport: Eulerian and Lagrangian models

Atmospheric pollutants are transported from their sources by wind fields. Taking this transport into account is essential to properly evaluate the air quality of a specific location, since pollutants might be transported from their original emission points. To accurately simulate species transport, atmospheric models have grown in complexity with time, taking into account more and more atmospheric processes [1].

There are two main atmospheric model schemes used for modelling chemical transport: *Eulerian scheme*, which makes use of static grid-cells to simulate the transport of atmospheric species and their chemical reactions, and *Lagrangian scheme*, which attempt to simulate their transport following them while they are driven by wind fields, without the need of a gridded pattern [1]. A graphical representation of the differences between Eulerian and Lagrangian models is shown in figures 4 and 5. All of them make use of *box modelling*, an approach focused exclusively on modelling chemistry, without spatial dimensions only temporal, along which species concentrations are simulated. Box modelling is applied by Lagrangian and Eulerian models at each grid-cell; Eulerian models do it consecutively in a set of grid-cells, while Lagrangian models do it in a single grid-cell transported with the species, with changing conditions.

Among Lagrangian models, a common approach are Lagrangian Particle Dispersion Models (LPDM). They divide the emission plume into homogeneous small blocks called *air parcels* [1] and use them to represent the simulated species particles. Along their paths, air parcels' identities are conserved and LPDM keep track of their properties while they interact with the rest of the atmosphere [5][3].

An example of LPDM is FLEXPART. In this project however, it has been combined with OH fields generated with TM5, an Eulerian model, and coupled with KPP, a chemical preprocessor used to produce a box model [22].

In this section we present an overview of these three models different in nature: FLEXPART, TM5 and KPP.

---

[1]The concept of *air parcel* relates to a parcel of air big enough to have homogeneous properties and composition, but small enough to be represented as a point in space. These are the basic blocks Lagrangian transport models are going to play with.
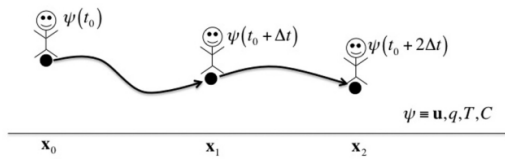
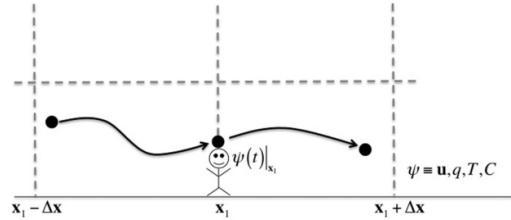**Fig. 4:** Lagrangian model scheme, where variables changes are tracked along the motion. (Source [?])

Fig. 5: Eulerian model scheme. Here variables are tracked only in a fixed location, and their changes are attributed to local and to advection factors. (Source [?])

2.2.1. FLEXPART

FLEXPART is a widely used LPDM, an open-source model developed by the community [2] [4], and precedented by FLEXTRA (FLEXible TRAjectory model). In the current project, FLEXPART 10.4 version has been used.

Wind fields - which are the main drivers of species transport - are taken from meteorological data which, in the case of FLEXPART, have two sources: European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) and the United States National Centers of Environmental Prediction (NCEP) Global Forecast System (GFS). The meteorology fields used by FLEXPART are from the ECMWF ERA-Interim reanalysis: global atmospheric meteorological data that covers the period from 1 January 1979 to 31 August 2019. ERA-Interim is preceded by ERA-40 and currently superseded by ERA5 reanalysis [23]. It provides FLEXPART with wind fields of approximately 80km, 60 vertical layers resolution and 3h frequency [24]. Also FLEXPART uses its cloud water content data for wet depositing of particles. [3].

As air parcels go through space during the simulation, different gridded information can be superimposed on the particle's location. Not only wind fields for its transport, but all the information regarding the different atmospheric processes that affect the air parcel in the location where it is found. These processes are normally implemented in a modular way. They include, besides advection and diffusion processes, chemical process such as [25]:

- Dry and wet deposition of particles. For the first one, surface type and orography are needed. For the second, it is defined using cloud masks, surface precipitation rates and temperature data from ECMWF meteorological data. These together are used in FLEXPART to determine where wet scavenging happens and in which conditions (if in-cloud, below-cloud and through rain or snow) [3].

- Gravitational settling of aerosols.

- Chemical reactions. By default only OH chemistry is considered in FLEXPART. Its chemical reaction with the transported species is calculated using the local concentration of OH in the particle's location and the reaction rates of each species. Because of this last factor, the effect of OH is very specific on the species transported.

- Radioactive decay, which is determined as a constant rate independent of the surrounding conditions of the particle.

Each of these processes is parametrized for each transported species. The parameters are predefined in a database, which can be extended virtually to any species if the proper coefficients are provided.

The time direction simulated in FLEXPART can be set by the user. Setting a reference release point and time, simulations can go forward or backward temporally from it. Computationally, forward and backward simulations are very different. In the forward case, chemical reactions and atmospheric transport follow the natural reaction rates and dynamics. Physics and chemistry are simply as equations express them, and the transport path of each particle is unique at each time step. In the backward case, transport and chemical reactions are reversed in time: deposition becomes emission, decomposition turns into production and wind-fields produce a reversed transport. Because of this, the result of back-time simulation is expressed in terms of probabilities: since multiple combinations of factors (emission, chemical reactions, deposition or transport) may produce the same concentration of a given species at the current point, it cannot be known for sure which combination is the actual cause of it. When expressing these probabilities in a spatial and temporal distribution they are named *sensitivities*: the potential mass contribution of each location to the final concentration, in s^{-1}.

2.2.2. TM5-MP

The Tracer Model 5 Massively Parallel version (TM5-MP) is an Eulerian global Chemistry Transport Model (CTM), based on the TM5-chem-v3.0 model [26]. Like FLEXPART, TM5-MP uses ECMWF ERA-Interim reanalysis as meteorological data. Also, in the present project, TM5 is used with the modified Carbon Bond 05 chemical reaction mechanism (mCB05). Its emission databases are CMIP6 for anthropogenic emissions [27] and MEGAN-MACC for biogenic ones [28].

Global distributions of OH are generated for the period from 01.06.2017 to 31.08.2017 with TM5-MP in order to integrate them to FLEXPART. The OH concentrations are given in volume mixing ratio ($molmol^{-1}$), hourly and with a resolution of 1×1 degrees and 34 vertical layers, extending from surface to approximately 50km (0.001hPa pressure).

2.2.3. KPP

In this project, the Kinetic PreProcessor (KPP) has been chosen to replace the default chemistry reactions scheme of FLEXPART (which only considers reactions with OH). Given a chemical mechanism, KPP assists in the implementation of its differential equations and in the numerical integration along the simulation. The numeral integration scheme to be used can be chosen by the user [22]. These concepts are developed below.

- Chemical mechanism

  The chemical species considered in a model react between them or the environment (i.e. photolysis). These reactions are expressed like differential equations, conforming a system of ordinary differential equations (ODE) called *chemical mechanism* of the model. It calculates the concentration time change of all the species. This change in time is determined by the loss and production rates of each species [1] (equation 8).

$$\frac{dq_i}{dt} = p_i - l_i q_i \tag{8}$$

  Where q_i is the concentration of species i, while l_i and p_i represent its loss and production rates, respectively.

  These reaction rates are not fixed normally, but depend on the concentration of the species involved in the reaction (equation 9). However, at the same time each species concentration is affected by all the reactions where the species takes part. Therefore, all the reactions involving a specific species affect the reaction rates. This crossed dependency between equations turns the ODE system into a coupled system of ODE.

  To manipulate these dependencies in a mathematical form, commonly reaction coefficients are expressed as matrices while concentrations are vectors . Their product turns the change rate of a given species concentration (dq_i/dt) in a function of the other species concentrations and time ($f_i(t, \mathbf{q})$), as shown in equation 9 [29].

$$\frac{dq_i}{dt} = \mathbf{P}(\mathbf{q}) - \mathbf{L}(\mathbf{q})q_i = f_i(t, \mathbf{q}) \tag{9}$$

  Where bold characters represent matrices of production ($\mathbf{P}$) and loss ($\mathbf{L}$) rates and the vector of concentrations of all species ($\mathbf{q}$). $f_i(t, \mathbf{q})$ is the reaction rate of the species i as a function of the species concentrations vector and time t. This dependency with the vector $\mathbf{q}$ turns the ODE system into a coupled system.

  Solving a coupled system is not a simple task, specially in very complex chemical mechanisms comprising a large number of reactions. Such systems can have thousands of equations with equal number of

coefficients dependent on each other. This generally causes the system not having an analytical solution and making a numerical integration necessary [1].

Given a chemical mechanism, KPP assists in the implementation of its differential equations and in the numerical integration along the simulation. Different numeral integration schemes are available, to be chosen by the user [22].

- Chemical solvers
  The numeral integration of coupled ODE systems is carried out by numerical integration schemes called *chemical solvers*. These solvers are mathematical algorithms that define how to approximate temporal derivatives of the ODEs by finite temporal differences (see equation 10). This way, the differential equations are integrated at each time step *n* along the simulation.

$$\mathbf{q}(t_{n+1}) = \mathbf{q}(t_n) + \mathbf{f}(t_n, \mathbf{q}(t_n))\Delta t \tag{10}$$

Where $\mathbf{f}(t_n, \mathbf{q}(t_n))$ is the vector of reaction rates for all species and Δt the time between time steps of the integration.

Two chemical solvers are included in KPP: Runge-Kutta and Rosenbrock. Each of them consists of different integration methods, namely different groups of coefficients used in the integration. In this project the solver Rosebrock with a method called Rodas-3 have been used.

- Rosenbrock Rodas-3 numerical integrator
  Rosenbrock is a chemical solver that can be considered a generalization of the Runge-Kutta solver. It is defined through the implementation of equation 11, which implement *s-stages* to compute each time step of the vector of species concentrations [1]. Different integration methods can be used inside Rosenbrock. In KPP, methods from Rosenbrock that can be implemented are Ros-2, Ros-3, Ros-4, Rodas-2 and Rodas-3 [22].

$$\mathbf{q}(t_{n+1}) = \mathbf{q}(t_n) + \sum_{i=1}^{s} b_i \mathbf{k}_i \Delta t \tag{11}$$

Where b_i is a coefficient defined by each integration method depending on the aimed accuracy, and $\mathbf{k}_s$ is a redefinition of the reaction rate vector $\mathbf{f}$ in equation 10. It is calculated for a number of *s* stages, and its definition changes if the system is autonomous (time as an independent variable) or non-autonomous

(not time dependent) (equations 12 and 13, respectively).

$$\mathbf{k}_i = \left[f\left(\mathbf{q}(t_n) + \Delta t \sum_{j=1}^{i-1} a_{ij}\mathbf{k}_j \right) + \mathbf{J}\Delta t \sum_{j=1}^{i} \gamma_{ij}\mathbf{k}_j \right] \tag{12}$$

$$\mathbf{k}_i = \left[f\left(t_n + \Delta t a_i, \mathbf{q} + \Delta t \sum_{j=1}^{i-1} a_{ij}\mathbf{k}_j \right) + \gamma_i \mathbf{J}(\Delta t)^2 + + \mathbf{J}\Delta t \sum_{j=1}^{i} \gamma_{ij}\mathbf{k}_j \right] \tag{13}$$

J being the Jacobian of the reaction rate f, while i is the stage calculated, and a_{ij}, γ_{ij} are coefficients of the used integration method to achieve a certain stability and accuracy [1][29].

Rosenbrock is an implicit solver, meaning that its functions are always evaluated at the next time step (t_{n+1}) from the current time. It is also a one-step algorithm, so the calculations only depend on the immediate previous time step ($q(t_n)$), and not on states before that ($q(t_{n-1})$, $q(t_{n-2})$, $q(t_{n-3})$,etc.). These two characteristics makes of Rosenbrock a very computationally-efficient solver, able to conserve mass along the integration if the Jacobian is not simplified (i.e. it is expressed in its full form, without simplifying any term). However, three main drawbacks arise: at each time step the Jacobian has to be computed and a linear system has to be solved (the vector $\mathbf{k}$), while it consists of several matrix multiplications. It does not ensure positivity either [1].

The method RODAS-3 from Rosenbrock consists in 4 stages for k, determining their coefficients (equation 14) and their definitions (equations 15-18) form:

$$\mathbf{q}(t_{n+1}) = \mathbf{q}(t_n) + \frac{5}{6}\mathbf{k}_1\Delta t - \frac{1}{6}\mathbf{k}_2\Delta t - \frac{1}{6}\mathbf{k}_3\Delta t + \frac{1}{2}\mathbf{k}_4\Delta t \tag{14}$$

$$\mathbf{k}_1 = \left[f\left(\mathbf{q}(t_n)\right) + \frac{1}{2}\mathbf{J}\Delta t \mathbf{k}_1 \right] \tag{15}$$

$$\mathbf{k}_2 = \left[f\left(\mathbf{q}(t_n)\right) + \mathbf{J}\Delta t \mathbf{k}_1 + \frac{1}{2}\mathbf{J}\Delta t \mathbf{k}_2 \right] \tag{16}$$

$$\mathbf{k}_3 = \left[f\left(\mathbf{q}(t_n) + \mathbf{k}_1\Delta t\right) - \frac{1}{4}\mathbf{J}\Delta t \mathbf{k}_1 - \frac{1}{4}\mathbf{J}\Delta t \mathbf{k}_2 + \frac{1}{2}\mathbf{J}\Delta t \mathbf{k}_3 \right] \tag{17}$$

$$\mathbf{k}_4 = \left[f\left(\mathbf{q}(t_n) + \frac{3}{4}\mathbf{k}_1\Delta t - \frac{1}{4}\mathbf{k}_2\Delta t + \frac{1}{2}\mathbf{k}_3\Delta t + \right) + \frac{1}{12}\mathbf{J}\mathbf{k}_1\Delta t + \frac{1}{12}\mathbf{J}\mathbf{k}_2\Delta t - \frac{2}{3}\mathbf{J}\mathbf{k}_3\Delta t - \frac{1}{2}\mathbf{J}\mathbf{k}_4\Delta t \right] \tag{18}$$

RODAS-3 has proved to be accurate for stiff problems, although less stable for long time steps and in non linear problems [1]. Nevertheless, for the purpose of this project, where rather simple chemistry mechanisms are used, RODAS-3 has been adequate.

3. Methodology

3.1. Default FLEXPART OH fields and reaction

The OH field data for FLEXPART version 10.4 is taken from GEOS-Chem model simulations (Bey et al., 2001). They are read from 12 different netCDF files that content monthly averaged data. Their resolution is 4×5 (lat $\times$ lon) degrees and 17 vertical layers, distributed from surface to 15km altitude.

To obtain hourly OH fields, the OH daily cycle is computed using O_3 photolysis rate data [3], following:

$$OH(t,x,y,z) = OH(x,y,z) \frac{j_{O_3}(t,x,y,z)}{j_{O_3}^{avg}(x,y,z)} \tag{19}$$

Where x,y,z is longitude, latitude, and altitude of the current location respectively, j_{O_3} is the ozone photolysis rate at each day specific hour t and location, $j_{O_3}^{avg}$ the O_3 monthly average value and $OH(x,y,z)$ the monthly average hydroxyl radical concentration value at that location.

The expression 19 is directly proportional on the solar zenith angle of each latitude and day time, which is the main driver of OH concentration.

The reading of OH fields is implemented in modules. In FLEXPART 10.4, the overall reading and processing of GEOS-Chem simulation OH fields are found in two modules named `readOHfield` and `gethourlyOH`. While the first one gets all the variables from netCDF files, the second one interpolates for each particle the OH field value to the exact current time of the simulation from its hourly values (which are computed as in equation 19).

By default, the reaction between OH and the transported species is approximated by expression (20) in FLEXPART. Reaction constants (α, β and δ) are parametrised for each transported species. All concentrations in the reactions are given in $ngcm^{-3}$.

$$[X](t_{n+1},x,y,z) = [X](t_n,x,y,z)e^{-k_{OH}\Delta t} \tag{20}$$

$$k_{OH} = \alpha T^\beta e^{-\delta/T}[OH] \tag{21}$$

3.2. Implementation of TM5-MP OH fields in FLEXPART

This section focuses on the methodology to read TM5-MP OH fields from FLEXPART. Since FLEXPART is built to read GEOS-Chem netCDF files, with their specific format, FLEXPART code must be adapted to

TM5-MP netCDF output.

After investigating the differences between GEOS-Chem and TM5-MP files format with Python (table 1), FLEXPART modules that read input concentration fields of OH are adapted from GEOS-Chem to TM5-MP format.

Table 1: Differences between GEOS-Chem and TM5-MP OH data sets.

	GEOS-Chem	**TM5-MP**
Horizontal resolution	4 lat × 5 lon	1 × 1
Number of horizontal grid-cells	46 lat × 72 lon	180 lat × 360 lon
Vertical layers	17	34
Altitude definition	Mid-level altitude	Top-level altitude
Conversion to metric altitude	*Eta* factor, to pressure levels	Geopotential data
Concentration magnitude	$ngcm^{-3}$	$molmol^{-1}$
Diurnal cycle definition	Parametrized[a]	Hourly data provided

[a] through O_3 photolysis rate (equation 19)

The development of the FLEXPART modules, to adapt them to read TM5-MP netCDF files, is carried out as follows:

- Units: since GEOS-Chem OH concentrations are in $moleccm^{-3}$ and TM5-MP in $molmol^{-1}$, it is enough to convert TM5-MP output with a conversion factor from the equation of state, taking into account the gridded temperature and pressure data from the same TM5-MP model during the simulated period.

- Altitudes: GEOS-Chem provides a vector of *eta* factors that are used to turn each layer to altitudes in meters through a conversion factor named *scaleheight* (*H = 7000*) and the expression:

$$altitude = \log\left(\frac{1010}{1010\eta}\right) \cdot H \tag{22}$$

Where η takes values from 0.12 to 0.99 for the characteristics of the simulations of this project (months, year and regions evaluated).

Meanwhile, TM5-MP output have 34 vertical layers, with horizontal geopotential (G_p) data for each one of them. So the altitude in meters for each grid-cell is computed from the definition of geopotential height (Z):

$$G_p = \int_0^Z gdz \longrightarrow Z = \frac{G_p}{g} \tag{23}$$

where $g = 9.81ms^{-1}$ is the average surface gravity.

Table 2: Simulation and release temporal and spatial data for the two simulations setting.

	Begin	End
Simulation date/time	01 June 2006 / 12:00:00	31 August 2006 / 12:00:00
Release date/time	01 June 2006 / 12:30:00	02 June 2006 / 12:30:00
Release point	+2.2E., +41.38N., 100m alt.	

- Horizontal resolution: it is not necessary to correct it since, as it is going to be seen later, FLEXPART simply takes the nearest grid to the position of each particle. By having a higher resolution, the advantage that TM5-MP data provide is more accurate OH concentration data for each particle's location.

Since TM5-MP already provides hourly data - it does not have to be computed as in equation 19 - the `gethourlyOH` subroutine is not needed any more. And, instead of storing in memory all the hourly OH fields (which might be memory intensive), just two hourly fields are stored in memory, representing the hour before and after the current model time. In the same `readOHfield` module, it is included the time linear interpolation of these two hourly fields to the exact current time simulated.

A third module, `OHreaction`, which computes the chemical reaction of the simulated species with the OH field, is changed to turn TM5-MP OH fields to *molec cm*$^{-3}$ before the reaction computation. In this computation, the temperature data for the calculation of the reaction rate is taken from TM5-MP fields instead of the default FLEXPART input meteorology data.

3.3. TM5-MP OH fields validation methodology

In order to evaluate the different results obtained between using GEOS-Chem and TM5-MP OH fields, a simulation of ethane (C_2H_6) transport is performed, because of ethane's relative high reactivity with OH. Two releases of this gas at the same time period and location are set, each using one of the OH fields. The location is Barcelona, Spain (+2.2E., +41.38N.), as a source of pollution being an urban area. Final ethane concentrations in both cases are then compared to assess the correctness of the implementation.

These simulations are set globally, with the parameters for simulation run and release stated in table 2. The outgrid for the ethane concentrations is as stated in table 3.

The release consists in 24kg of ethane spread along one-day period, beginning 30min after the simulation start.

Additionally, decline rates of ethane from both simulations are computed. Then from their subtraction, the difference in OH concentrations between both OH fields is estimated and compared to the actual

Table 3: Out-grid of the simulations: they are set globally and reaching upper stratosphere.

	Minimum(deg)	Maximum(deg)	Resolution(deg)
Latitude	-90	90	0.5
Longitude	-180	180	0.5
Top altitude of layers (m)	100, 200, 500, 1000, 15000, 60000		

comparison between GEOS-Chem and TM5-MP OH fields.

3.4. Coupling of FLEXPART and KPP

The second aim of this project is to couple the kinetic preprocessor KPP to FLEXPART. KPP produces a box model used to compute the chemistry at each time step and location of a FLEXPART simulation.

The coupling comprises the following steps: 1) a box model is generated with KPP on the basis of FLEXPART chemical scheme (i.e. OH reaction), and 2) replacing the default chemical reaction in FLEXPART (equation 20) by the integrator of this box model.

The box model is rather simple to generate with KPP, since there are not more than 4 species that react with OH in FLEXPART, and their chemical mechanism do not extend beyond 10 equations. Also, these equations are not coupled, because transported species do not interact between them in FLEXPART and only one species can be transported per simulation. Since there is only one species transported, only its reaction with OH will take place. Nevertheless, all OH reactions have been integrated in the chemical mechanism of KPP, in order no to change the KPP chemical mechanism every time we change the transported species.

The default FLEXPART chemical reaction (equation 20) is replaced by the integrator through the Fortran file *integrator_kpp.f90*, which has been developed in this project. This file links the transported species' and OH concentration from FLEXPART to the KPP-generated box model. Also other local FLEXPART variables such as temperature, pressure and time are introduced as input to the box model. With all these variables, the box model runs for a period equal to the FLEXPART simulation time step (normally 300s). The internal time steps of the box model numerical integrator do not need to be determined if Rosenbrock integration method is used. At the end of the box simulation, the species concentrations are provided. Only the transported species concentration is saved again back to FLEXPART (*xmass*). A mid-step in this process is the reordering of species variables inside KPP, since KPP reorders the species in terms of sparsity

properties of the Jacobian and Hessian of the chemistry equations. This task is made straight forward thanks to built-in functions in KPP. More technical information regarding this development can be found in appendix 8.2.

An scheme of the model workflow after the coupling of KPP using TM5-MP OH fields is presented in figure 6.

3.4.1. Further coupling programming: multiple reacting species and photolysis

Also, in this project the coupling has been performed not only for a single reacting species. In order to include more than one environmental species (e.g. *OH* and *NO$_x$*), the current development allows multiple reacting species to be defined, always that their reactions with the transported species are properly defined in the chemical mechanism of the box model. This has been achieved extending the dimensions of the environmental reacting species' concentration matrix. Every additional dimension stores the concentrations of each additional reacting species. The numerical integrator then, computes the concentration sink of the transported species in terms of the chemical mechanism it has been defined and with the concentrations specified in this matrix. However, this development has not been possible to be tested in this project due to time limitations and is subject for future work.

Additionally, also photolysis reactions has been included for the first time in FLEXPART. The insolation factor has been made available from FLEXPART to the box model. In the box model's chemical mechanism, the photolysis rate can be included in the reactions, becoming an additional sink like any other reacting species. However, again due to time limitations, this development has been tested but not validated in this project.

3.5. KPP-coupling validation methodology

3.5.1. Single ethane release simulation

To compare the effect that the new chemical scheme (based on the KPP generated box model) has on the simulations, the same experiment as in section 3.3 is performed. The specifications regarding the outgrid and time period are the same as in tables 2 and 3. This experiment is used to assess the effect that the KPP coupling has on FLEXPART performance, since KPP is used together with TM5-MP OH fields and their effect has already been assessed in section 3.3.
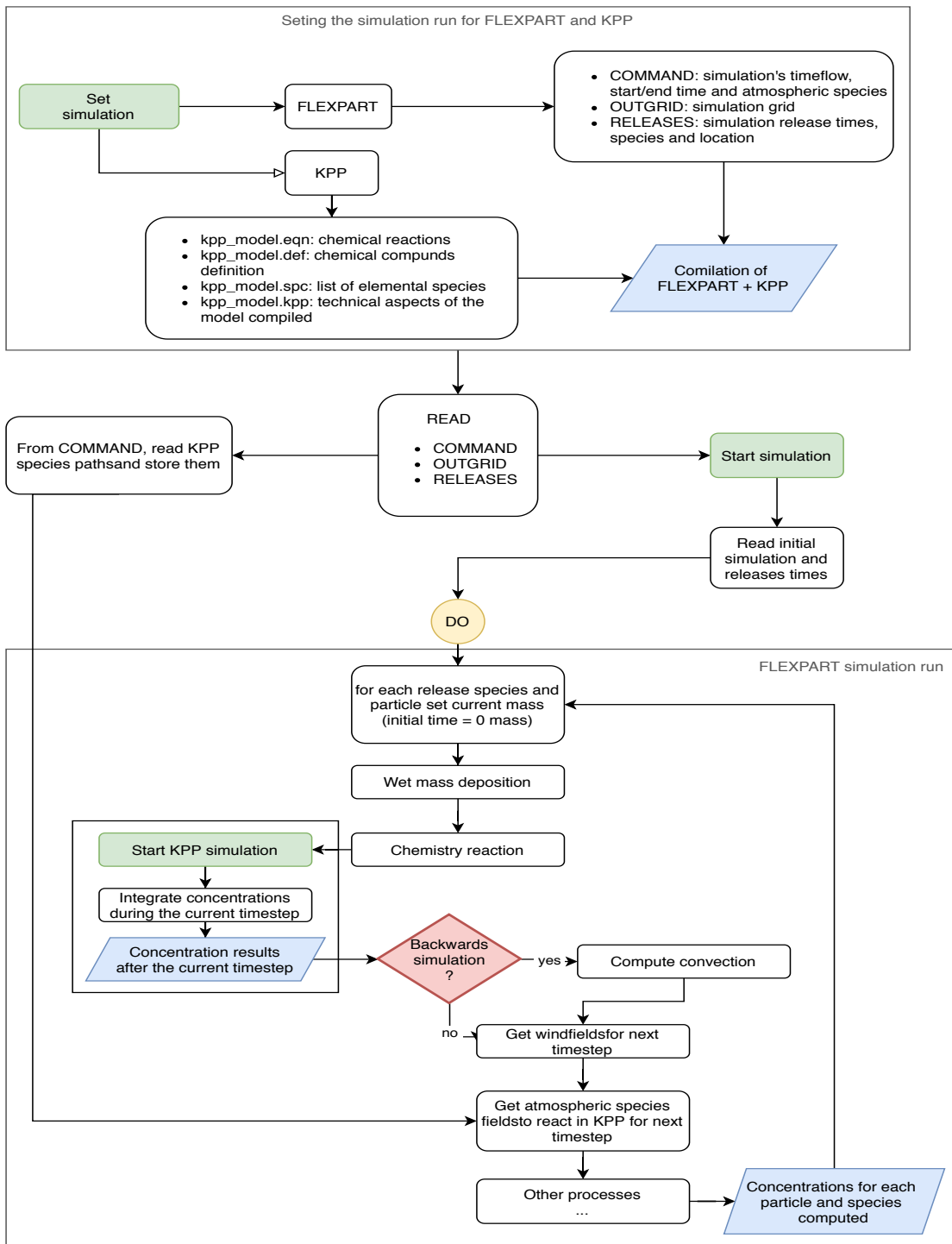
Fig. 6: Flowchart of FLEXPART after TM5-MP OH fields and KPP have been coupled.

3.5.2. Validation per species

The simulations using KPP coupled with FLEXPART are validated in two ways: comparing them with previous FLEXPART simulations not using KPP and with observations.

By default in FLEXPART there are four species that react with OH: methane (CH_4), ethane (C_2H_6), polychlorinated biphenyl 28 ($PCB28$) and gamma-hexachlorocyclohexane ($G-HCH$). The simulations transporting these species are directly affected by the developments on the OH chemical scheme. To validate these developments we perform three simulations for each species, one for each development stage, using: 1) the default FLEXPART version (with GEOS-Chem OH fields), 2) FLEXPART with TM5-MP OH fields and keeping the default chemical reaction and 3) FLEXPART with TM5-MP OH fields and KPP coupled computing the chemical reactions. These three simulations have been performed with the configuration indicated in tables 3 and 2.

To compare these three simulations, the time evolution of the total mass for each transported species during the simulation time period has been used. This variable is useful for assessing the impact of each development in the OH chemistry sink mechanism. Furthermore, performing a simulation for each development stage makes possible to separate the effects that each development (TM5-MP OH fields and KPP coupling) has on the simulations. It must be noticed, though, that a direct comparison between simulations is not fully possible, because of the FLEXPART turbulent component when calculating the particle's velocities. It uses a Gaussian distribution that adds a random component for each run. The evaluation is statistical and qualitative [3].

3.5.3. Validation with observations

Finally, in order to assess the viability of the results, a validation using ground-based sensor data have been performed. This validation consists in simulating ethane emissions for a two-month period (June and July 2014) over Europe. Ethane is used because of its relative atmospheric short half life-time and simple chemistry (only reacting with OH).

In order to recreate ethane emissions, historical Climate Model Intercomparison Project 6 (CMIP6) ethane emission data has been used. This emission data has been emulated with FLEXPART using multiple simultaneous releases, like it has been done before in [30]. CMIP6 emission data comprises biomass burning and anthropogenic emissions with the following characteristics:

- Biomass burning CMIP6 emissions: Version 1.2 of historical emissions from 1750 to 2015, produced by Vrije Universiteit Amsterdam. It has a resolution of 0.25×0.25 (approximately a nominal resolution of

25km) and it is monthly sampled.

- Anthropogenic CMIP6 emissions: Version 2017-05-18 of historical emissions from January 2000 to December 2014. Produced by Pacific Northwest National Laboratory - Joint Global Change Research Institute (PNNL-JGCRI). With a resolution of 0.5×0.5 (50km nominal resolution) and it is also monthly sampled. It comprises the ethane emission sectors: Agriculture, energy, industrial, transportation, residential, Commercial, solvents production and application, waste, international shipping and others. All sectors are added up, since in the current project we do not aim at distinguishing from emission source.

The date selected for both datasets is June 2014, which for the sake of simplicity is considered representative for June and July 2014. Biomass burning emission data is averaged to the anthropogenic dataset resolution. Then, both datasets are summed, therefore obtaining a bulk ethane emission dataset of 0.5×0.5 resolution. The merged emission data for June 2014 can be observed in figure 7 for the global case, and in figure 8 for the European region.
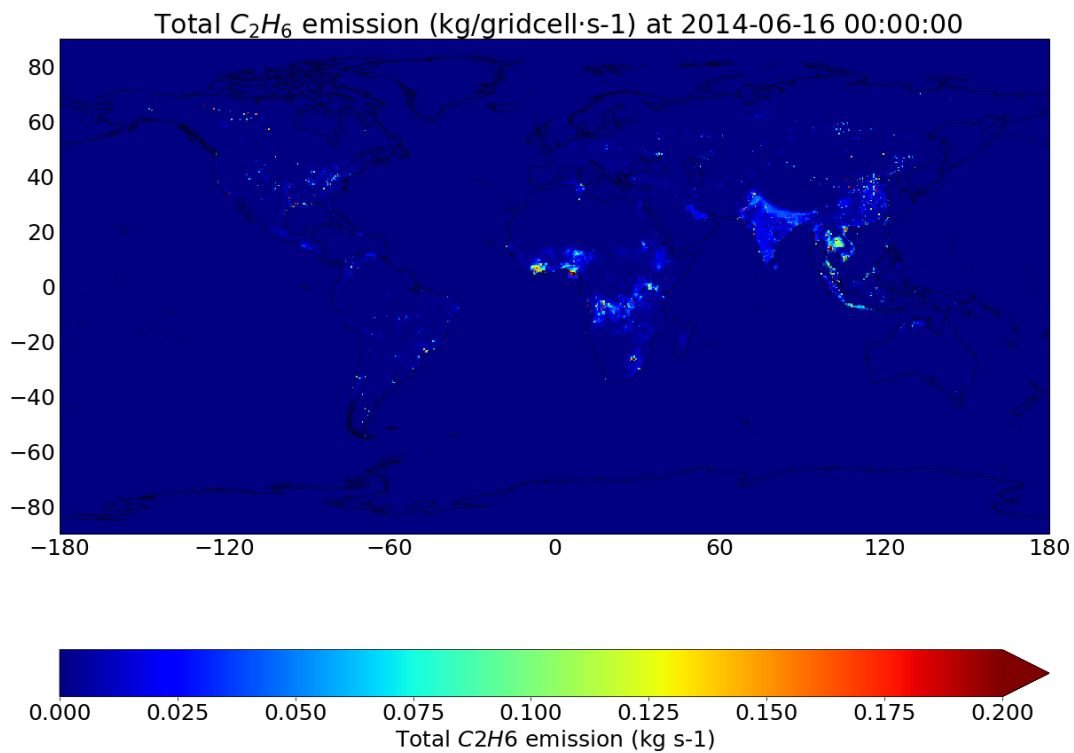
Fig. 7: CMIP6 global ethane emissivities data. The components of biomass burning and anthropogenic have been added up in this figure, to obtain the total ethane emission data. Each grid cell emissivity has been multiplied by the respective grid cell area to obtain kg per second.
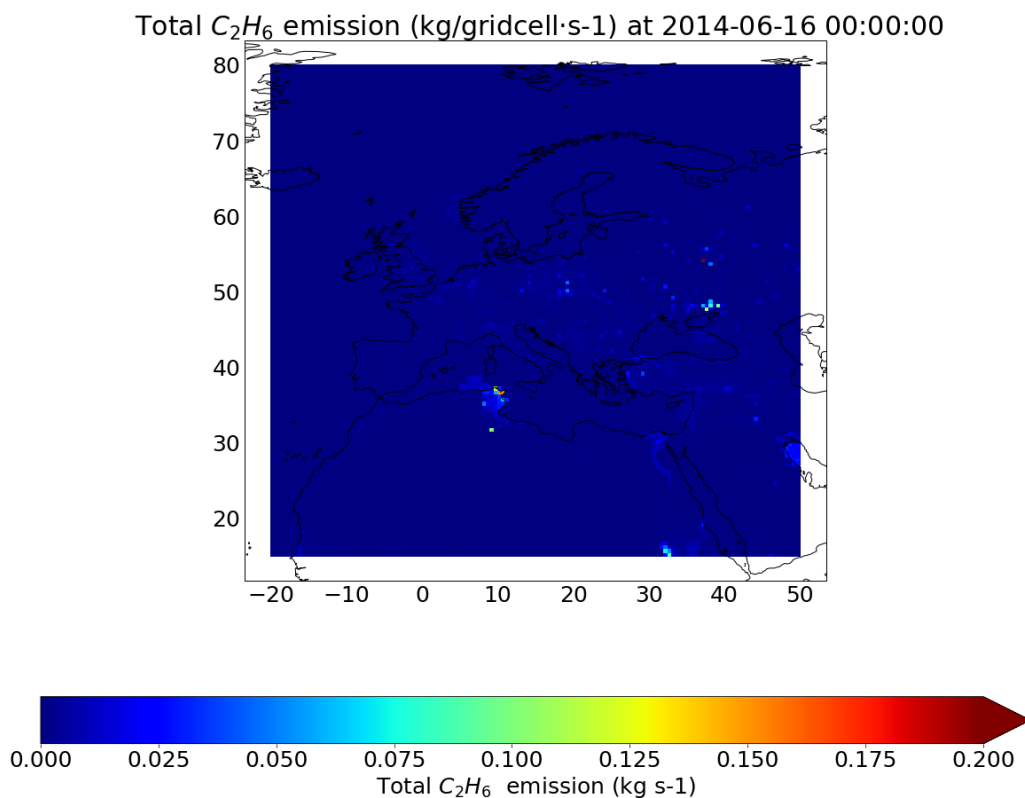
Fig. 8: CMIP6 European ethane emission data used for the validation, after trimming the global emissions for the European region.

The emission data is cut-off using a threshold, in order to select the most important emission points. This cutting prevents reaching a large amount of particles transported, which on one hand is more computationally efficient and it allows attributing a smaller mass to each particle released, which improves accuracy on the transport simulation. On the other hand, this simplification is made on behalf of losing emission points. Therefore a balance between both effects should be found, mostly in terms of computational performance. The threshold is set using a histogram of emissivities (figure 9) where the amount of grid cells emitting per emissivity value can be appreciated in the horizontal axis. From this histogram, it is possible visualize the amount of emission points lost for a certain threshold value chosen. In this study, a threshold of 0.0002 kg/s has been chosen, which takes into account the mostly all the emission points (62 %) without substantially
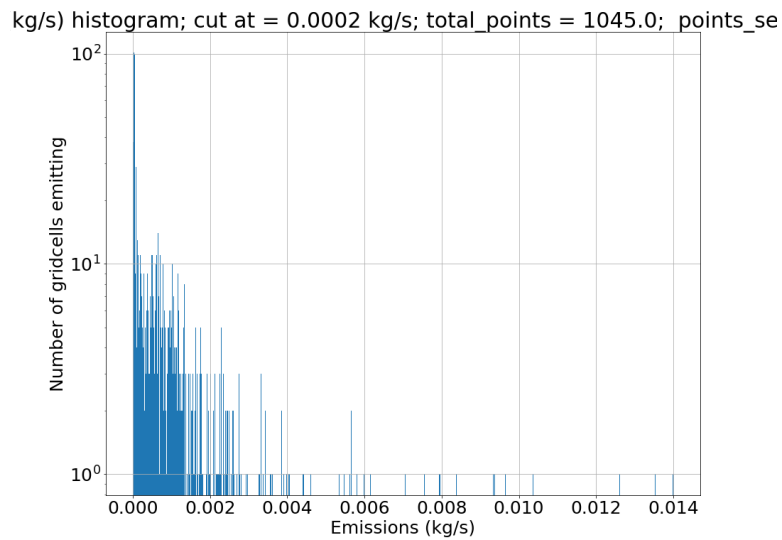
Fig. 9: Ethane emissivities histogram for the CMIP6 European database. From this histogram, a cutting level is set to trim the amount of emission points and, thus, reducing the amount of particles released. This takes into account only the most important points.

increasing the computation time. After trimming the emission rates, the cut emission data base used is shown in figure 10. After multiplied by the number of seconds simulated (2 months), the absolute emission values per grid cell are shown in figure 11.

C_2H_6 emission (kg/s); cut at = 0.0002 kg/s; points = 657.0 (62%); total_particles = 8000000 (7.7
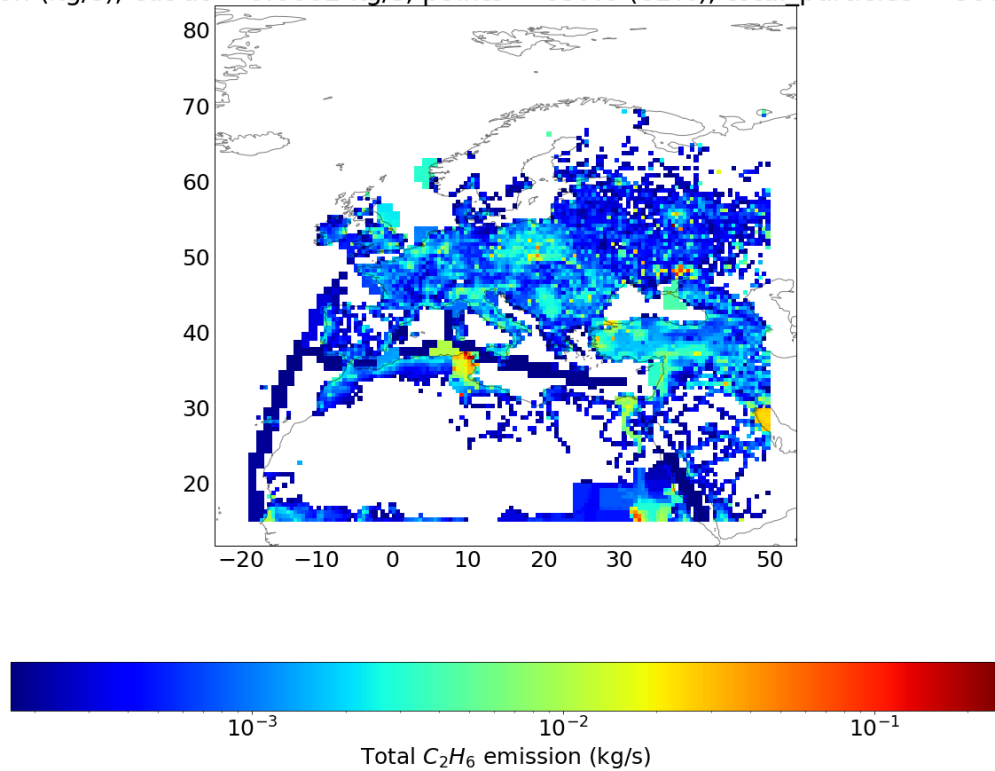
Total C_2H_6 emission (kg/s)

Fig. 10: CMIP6 European selected ethane emission data in kg per second, after filtering the low emissivity values.

C_2H_6 emission (kg/s); cut at = 0.0002 kg/s; points = 657.0 (62%); total_particles = 8000000 (7.7
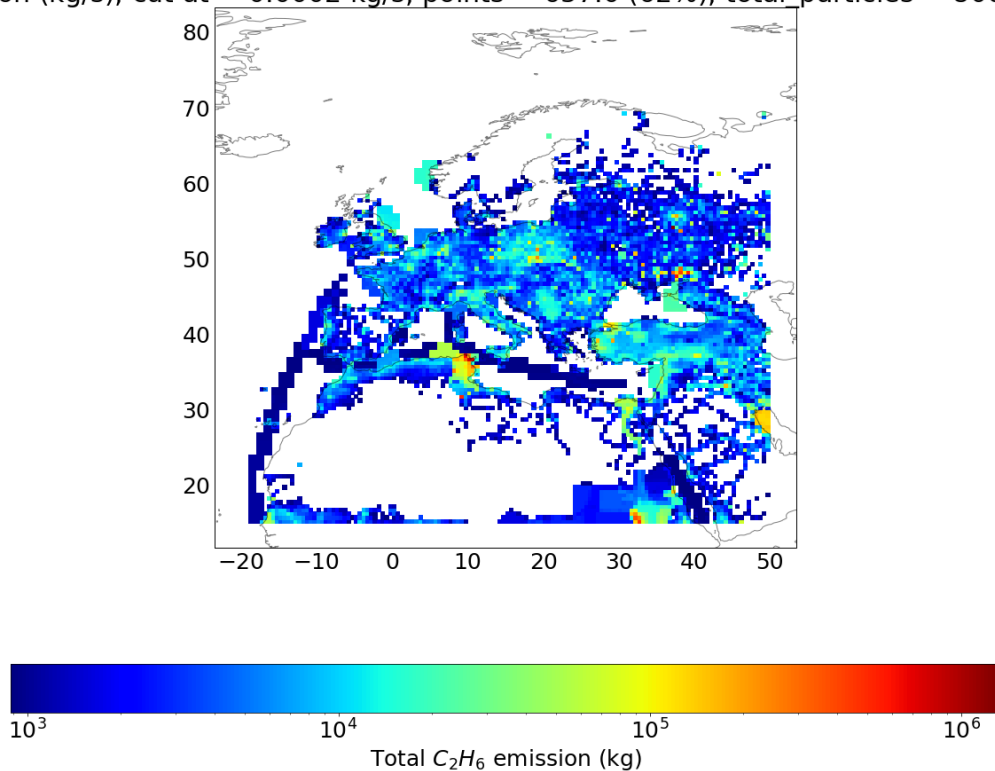
Fig. 11: CMIP6 European ethane emission data used for the validation in kg, after multiplying by the total number of seconds of the 2 months simulated period.

Finally, the observational data is obtained from the European Environment Agency[2]. The sensors' geographical distribution is shown in figure 12. Each sensor has different time sampling frequency and dates. Therefore, data is linearly interpolated to obtain a continuous time plot for the simulations period (June and July 2014). The stations' data for the studied period is shown in figure **??**.

During the two months simulated, emissions are considered constant (no daily variation). The simulated evolution of ethane concentration is compared to the respective sensor's ethane data at each station's location.

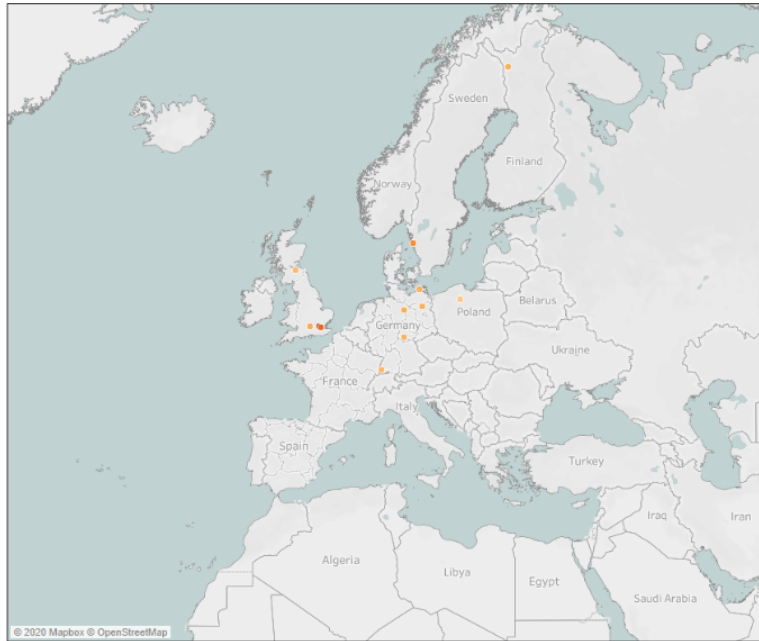[2]www.eea.europa.eu/data-and-maps/dashboards/air-quality-statistics-expert-viewer

Fig. 12: Stations providing the ethane concentration data used for the observational validation. Yellow dots represent the rural stations and red the urban ones.

Time evolution of ethane permits us to take into account the whole evolution at each station location, in spite of the high variability of the data. The ethane's concentration change along the simulated period is shown in figure 13.
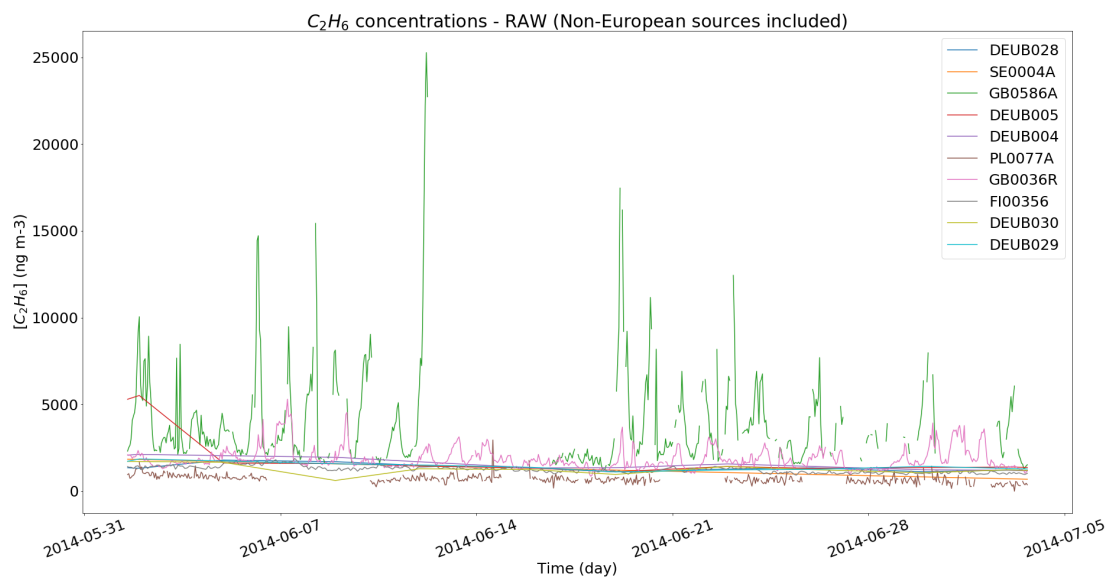
Fig. 13: The considered stations time series of ethane concentration.

4. Results

4.1. Comparison of OH fields: GEOS-Chem versus TM5-MP

Using Python, GEOS-Chem and TM5-MP OH fields are represented to compare their values and OH distributions.

In figure 14, the two OH fields are shown for each noon local time (as stated before, in GEOS-Chem field a diurnal cycle parametrization is applied through equation 19).

It can be observed how resolution is enhanced with TM5-MP output. Also, overall, OH concentration values are higher for TM5-MP. While GEOS-Chem shows highest values in mid-latitudinal coastal areas and in Himalayas, TM5-MP resolve high OH concentration from ship tracks in open ocean, as well as from small scale industrialized inland and coastal regions.

Figure 15 shows the difference between the two OH fields from figure 14 (GEOS-Chem - TM5-MP), computed through the interpolation of GEOS-Chem data to adapt it to the higher resolution of TM5-MP data.

Finally, the horizontally-averaged vertical profiles of both datasets are plotted in figure 16, where TM5-MP data is shown to provide a more complete representation of the vertical distribution reaching stratospheric altitudes (until 65km), while GEOS-Chem top layer remains in the troposphere (around 15km).

4.2. Validation: ethane concentration with each OH field

Results for ethane concentrations along the two simulations, one reacting with OH fields from GEOS-Chem and the second from TM5-MP, are shown in figure 17. In them, the ethane plume is plotted at approximately 1.5 weeks after the release time. It can be observed how the plume splits in two main components along mid-latitudes: one over Europe and one over the Atlantic. In both cases, the spatial pattern is almost identical. It is expected, because the only intrinsic difference between them is the OH fields implemented, thus the OH chemical loss rate of the plume. The driving meteorology remains unchanged.

Differences between both ethane distributions can be better appreciated by calculating the difference between the datasets. The relative difference is shown in figure 18, which is computed with regard to the average values of both data. With TM5-MP data, ethane concentration is shown to be approximately 20% lower than using GEOS-Chem data.

To validate the results obtained using TM5-MP OH fields, the time evolution of the total ethane mass in the atmosphere is shown in the second plot of figure 21. Here, it can be observed how the total mass is similar in both cases, and slightly lower in the case of TM5-MP OH fields simulation. This last is in agreement with the fact that TM5-MP fields have a little bit higher OH concentration values than GEOS-Chem, as it has been
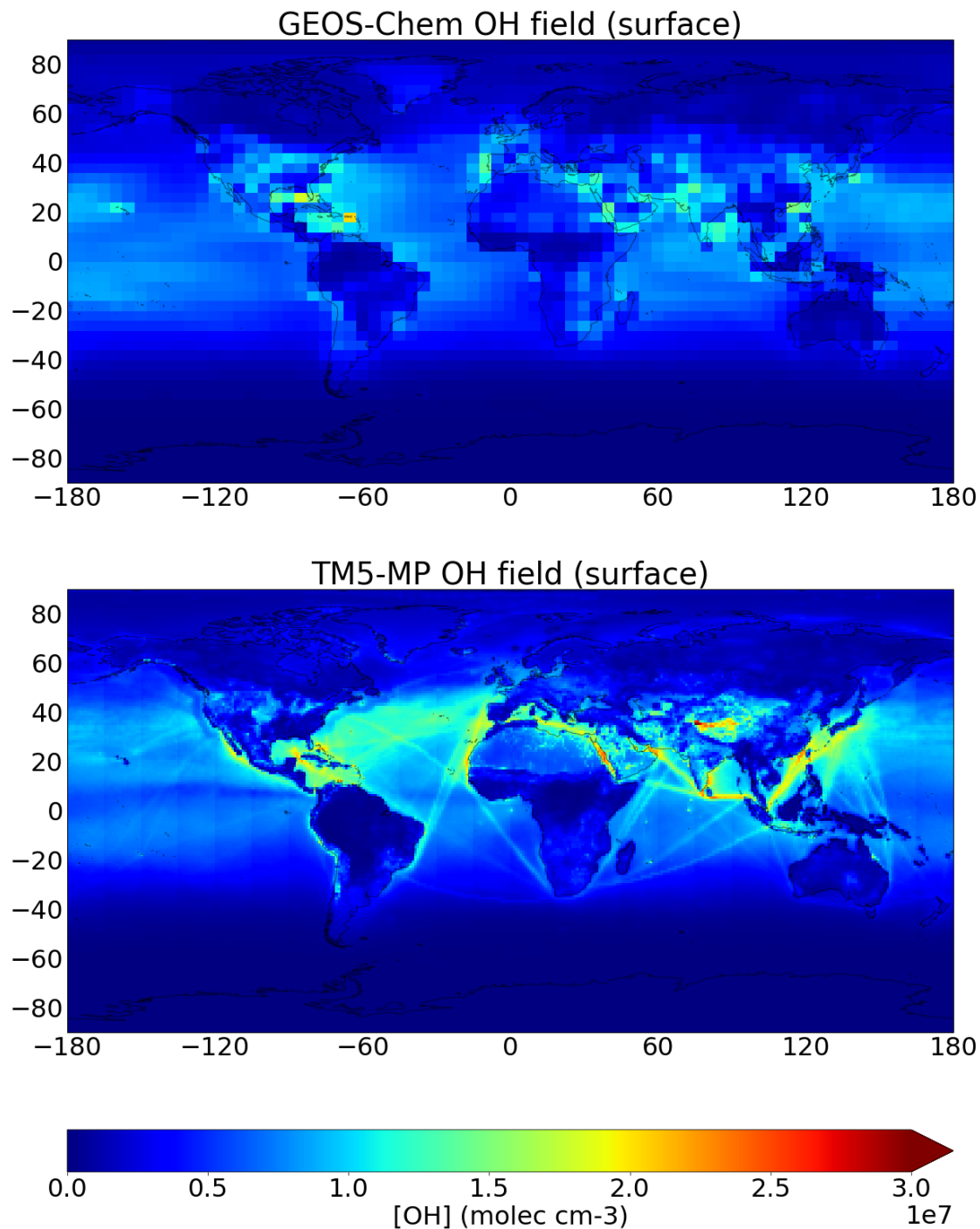
Fig. 14: Monthly averaged surface OH concentration fields from GEOS-Chem simulation used by default by FLEXPART (above) and from the newly implemented TM5-MP simulation (bottom).
The surface average concentration of three months (June, July and August) of 2005 for GEOS-Chem and 2017 for TM5-MP is shown at noon local time.
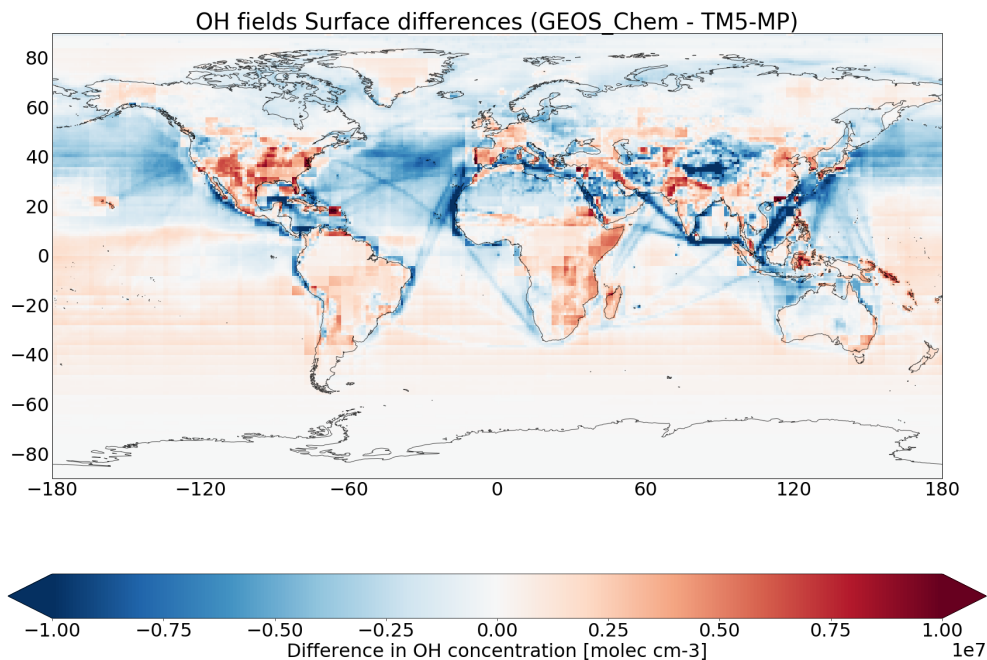
Fig. 15: Difference of OH surface concentration fields from figure 14 (*GEOS − Chem − TM5 − MP*). GEOS-Chem data has been interpolated to adapt the resolutions between the two data sets.

observed in figure 18: thus ethane is depleted slightly faster in presence of TM5-MP OH fields. This fact might also explain why when using TM5-MP OH fields, the model not reach the 24kg one day after the release time, when its curve peaks.

Finally, the average altitude of the ethane plume during the simulation period is plotted in figure 19. In this plot the close similarities between the two simulations can be observed.

4.2.1. Correspondence between dissimilarity in ethane trends and differences in OH fields

From time-series in second plot of figure 21, the ethane decline trend can be computed for each simulation and compared to the values from figure 15, to study if the drift in ethane decline trends approximately correspond to the difference in OH concentration between the two OH data sets used.

Ethane reaction kinetics with OH from reaction 5 can be defined:

$$\frac{d[C_2H_6]}{dt} = k[OH][C_2H_6] \tag{24}$$

Where k is the reaction rate of ethane with OH. As a first approximation - and just for the sake
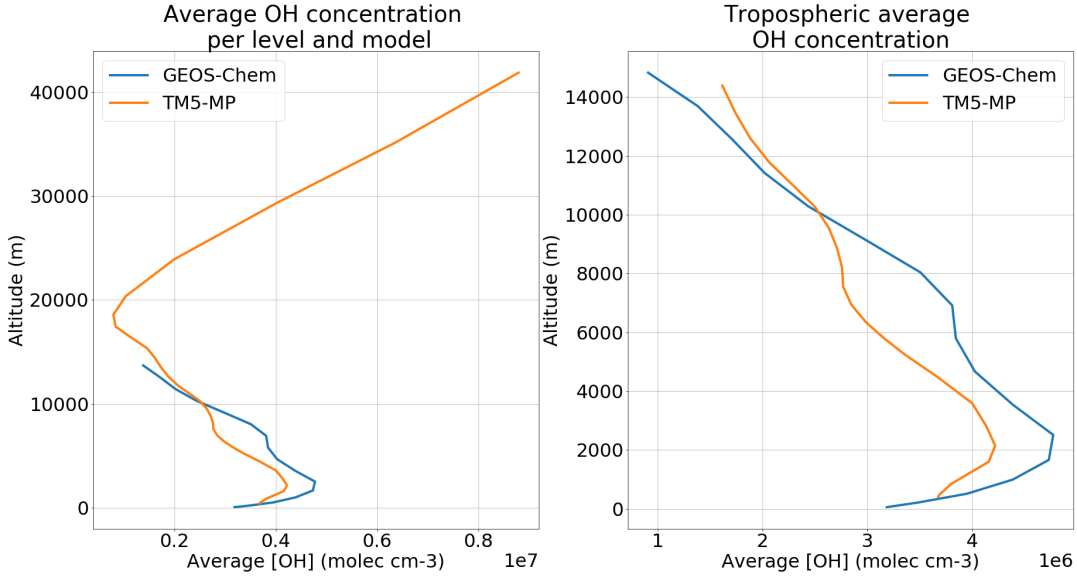
35

Fig. 16: Horizontally and monthly averaged vertical OH profile for GEOS-Chem OH fields in 2005 June, July and August months. The totality of the available data is plotted on the left, while on the right hand side the troposphere is selected (approximately 15km altitude).

of simplicity in this validation analysis - we could consider it temperature independent and equal to $k = 290 \cdot 10^{-15} cm^3 molec^{-1} s^{-1}$ (from [28], at 296K).

Assuming all the concentrations are limited to the troposphere (from 0m to 10km altitude), since from figure 19 it is observed that the average altitude of the cloud is around 1.4km, the volume over which to compute the concentration of ethane ($[C_2H_6]$) is: $V = 4 \cdot \pi \cdot R_E^2 \cdot 10^6 \ cm^3$ ($R_E = 6.371 \cdot 10^8$ cm).

Therefore, considering the period from the peaking concentration after the first day release (June the 2nd, ethane mass = 23kg for TM5 and 24kg for GEOS-Chem) to 13 days later (until June 15th, ethane mass = 13.5kg for TM5 and 16kg for GEOS-Chem), the difference in ethane decline trend can be computed as follows:

$$GEOS-Chem: \frac{d[C_2H_6]}{dt} = \frac{24-16}{13 \cdot 24 \cdot 60 \cdot 60} \cdot \frac{1}{V} \tag{25}$$

$$TM5-MP: \frac{d[C_2H_6]}{dt} = \frac{23-13.5}{13 \cdot 24 \cdot 60 \cdot 60} \cdot \frac{1}{V} \tag{26}$$

These trends are expressed in $kg \cdot cm^{-3} s^{-1}$. To compare them with figure 15 concentrations, they are converted to $molec \cdot cm^{-3} \cdot s^{-1}$, through the molar mass of ethane ($M_{C_2H_6} = 30g$):
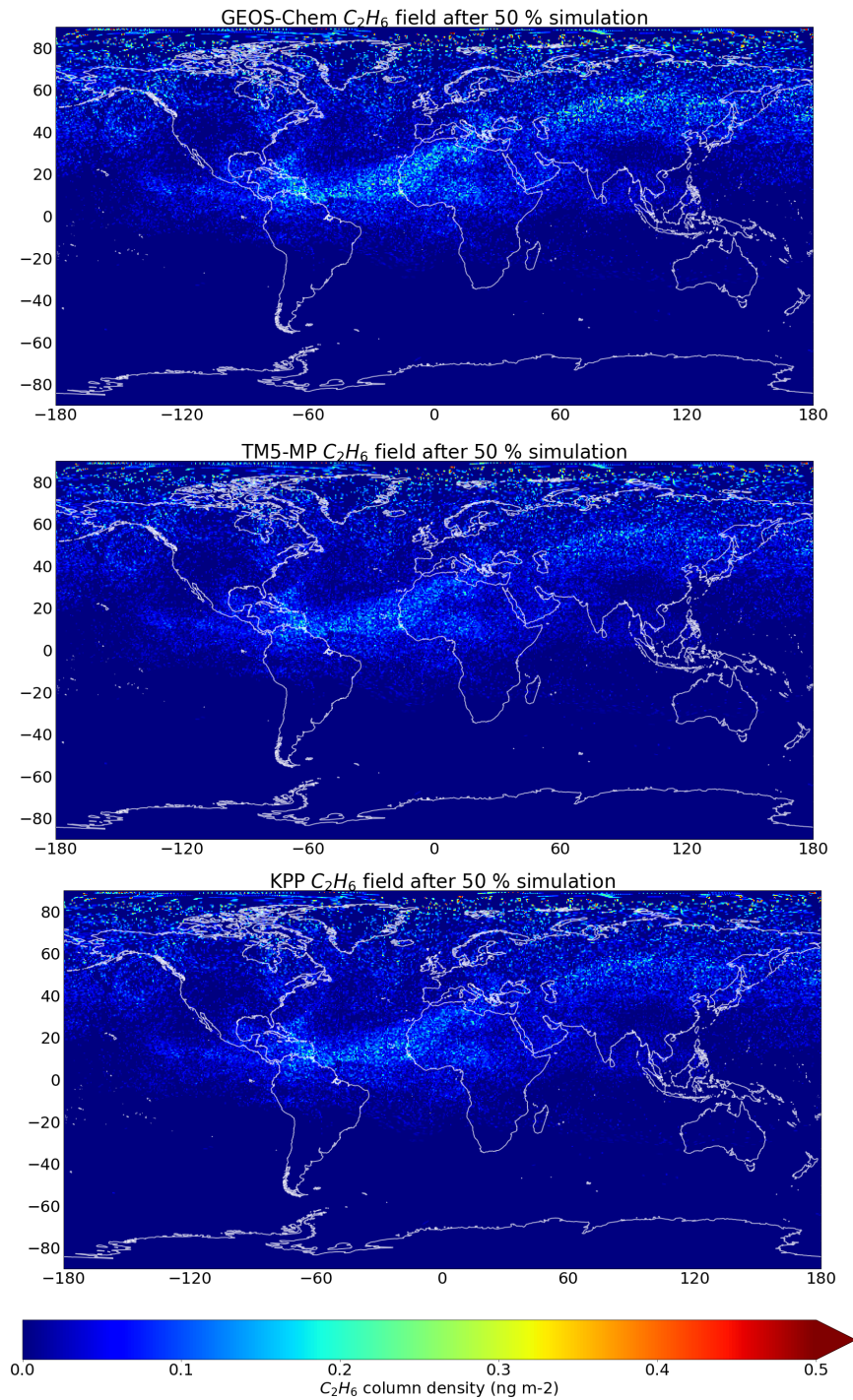
36

Fig. 17: Simulated column density of ethane 1 month after the release, with FLEXPART using GEOS-Chem (top), TM5-MP (middle) OH fields, and TM5-MP OH fields with KPP-box model as chemical scheme (bottom).
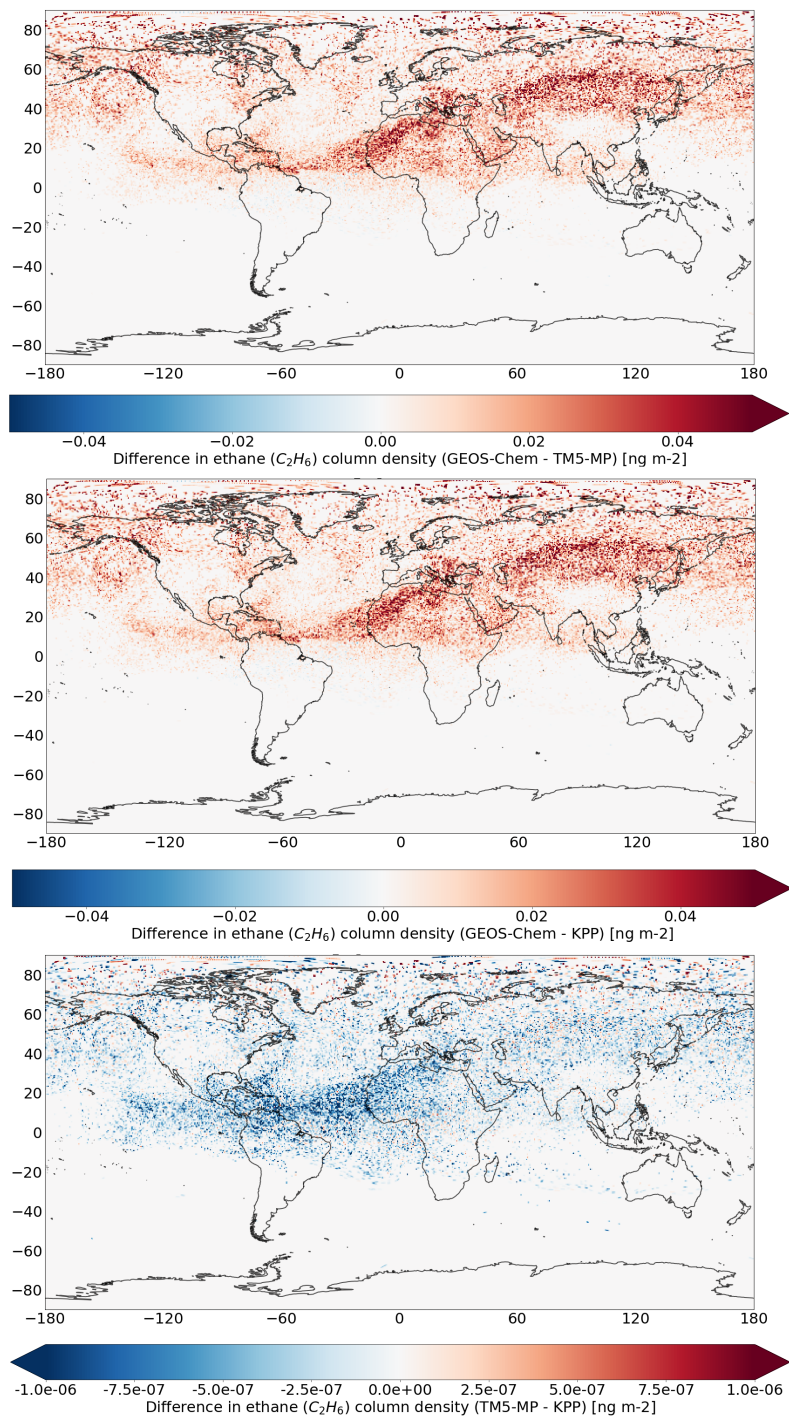
Fig. 18: Differences in ethane column density of the three simulations after 1 month from the release (figure 17). Subtraction of ethane fields using: default chemical scheme with GEOS-Chem and TM5-MP OH fields (top), default chemical scheme using GEOS-Chem and KPP chemical scheme using TM5-MP OH fields (middle) and default chemical scheme and KPP chemical scheme both using TM5-MP OH fields (bottom). It is important to note the change of scale in the last plot.
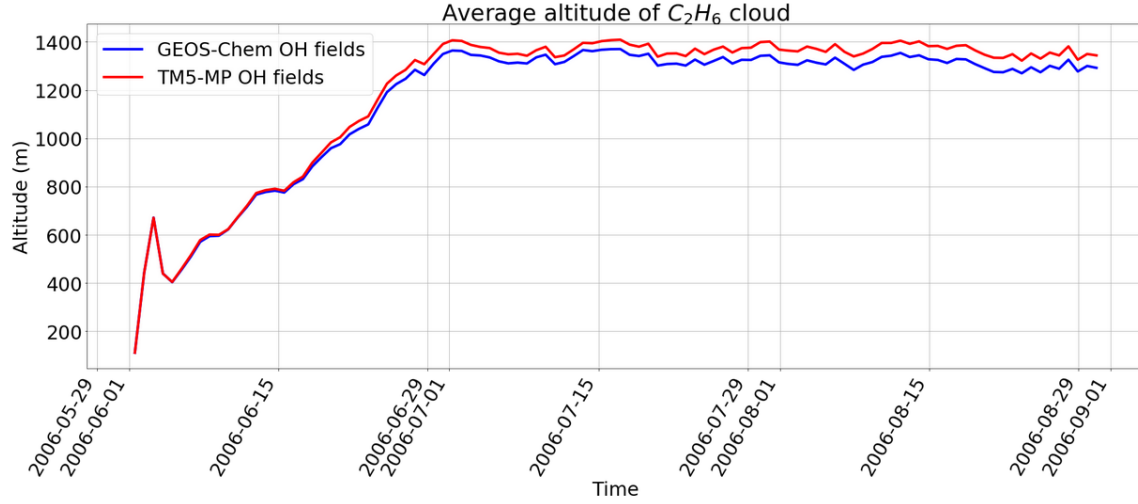
Fig. 19: Time evolution of the average altitude of the ethane cloud in the atmosphere during the simulation. It can be observed how its evolution using TM5-MP OH fields is similar to the GEOS-Chem case.

$$GEOS-Chem: \frac{d[C_2H_6]}{dt}\bigg|_{GEOS} = \frac{8 \cdot 1000 \cdot N_A}{13 \cdot 24 \cdot 60 \cdot 60 \cdot 30} \cdot \frac{1}{V} \tag{27}$$

$$TM5-MP: \frac{d[C_2H_6]}{dt}\bigg|_{TM5} = \frac{9.5 \cdot 1000 \cdot N_A}{13 \cdot 24 \cdot 60 \cdot 60 \cdot 30} \cdot \frac{1}{V} \tag{28}$$

Where N_A is Avogadro's number.

Thus, the difference between the two ethane decrease trends is:

$$\Delta\left(\frac{d[C_2H_6]}{dt}\right) = \frac{d[C_2H_6]}{dt}\bigg|_{GEOS} - \frac{d[C_2H_6]}{dt}\bigg|_{TM5} = -5.30 \cdot 10^{-6} molec \cdot cm^{-3} \cdot s^{-1} \tag{29}$$

To cause this decrease trend dissimilarity, the OH concentration difference between the two fields should be (from equation 24):

$$\Delta[OH] = \frac{\Delta\left(\frac{d[C_2H_6]}{dt}\right)}{k[C_2H_6]} \tag{30}$$

As a very simple first approximation: assuming an average ethane mass of 20kg (average between 24 and 16kg in the period considered), distributed evenly throughout the troposphere's volume (from 0m to 10km altitude), its concentration becomes $[C_2H_6] = 78.71 molec \cdot cm^{-3}$. Which leads to a OH difference (from equation

39

30) of:

$$\Delta[OH] = [OH]_{GEOS} - [OH]_{TM5} = \frac{-8.78 \cdot 10^{-7}}{290 \cdot 10^{-15} \cdot 13.11} = -0.231 \cdot 10^{7} molec \cdot cm^{-3} \tag{31}$$

In figure 15 the subtraction of the two OH fields from figure 14 (GEOS-Chem - TM5-MP) is shown. As it can be appreciated, in average the range of values obtained are of the same order as the result from equation 31. Moreover, most of the values - specially over the ocean - are negative, being in agreement with the result from 31, despite some scattered localized positive values inland.

Overall, the result from equation 31 and figure 15 indicate that the ethane trend differences from the second plot of figure 21 are fairly consistent with the differences observed between the two OH fields used.

4.3. KPP coupling results

4.3.1. Single ethane release simulation

The experiment described in section 3.5.1 is performed and the result is shown in figure 17. The comparison of this results against the default chemical scheme using GEOS-Chem and TM5-MP OH fields can be found in the figures 18 and 20 .

It can be observed how the simulation using the box model produce similar results as the simulation using the default chemical scheme with TM5-MP OH fields, obtaining a relative difference below 0.001% (figure 20 bottom). This similarity leads to obtain alike differences when comparing the KPP-coupled simulations using TM5-MP OH fields against the default chemical scheme with GEOS-Chem OH fields (figure 20 middle) as with the comparison between GEOS-Chem and TM5-MP OH fields without any development in the chemical scheme where approximately a $0.1 \ ng/m^{-2}$ difference is obtained (figure 18 top).
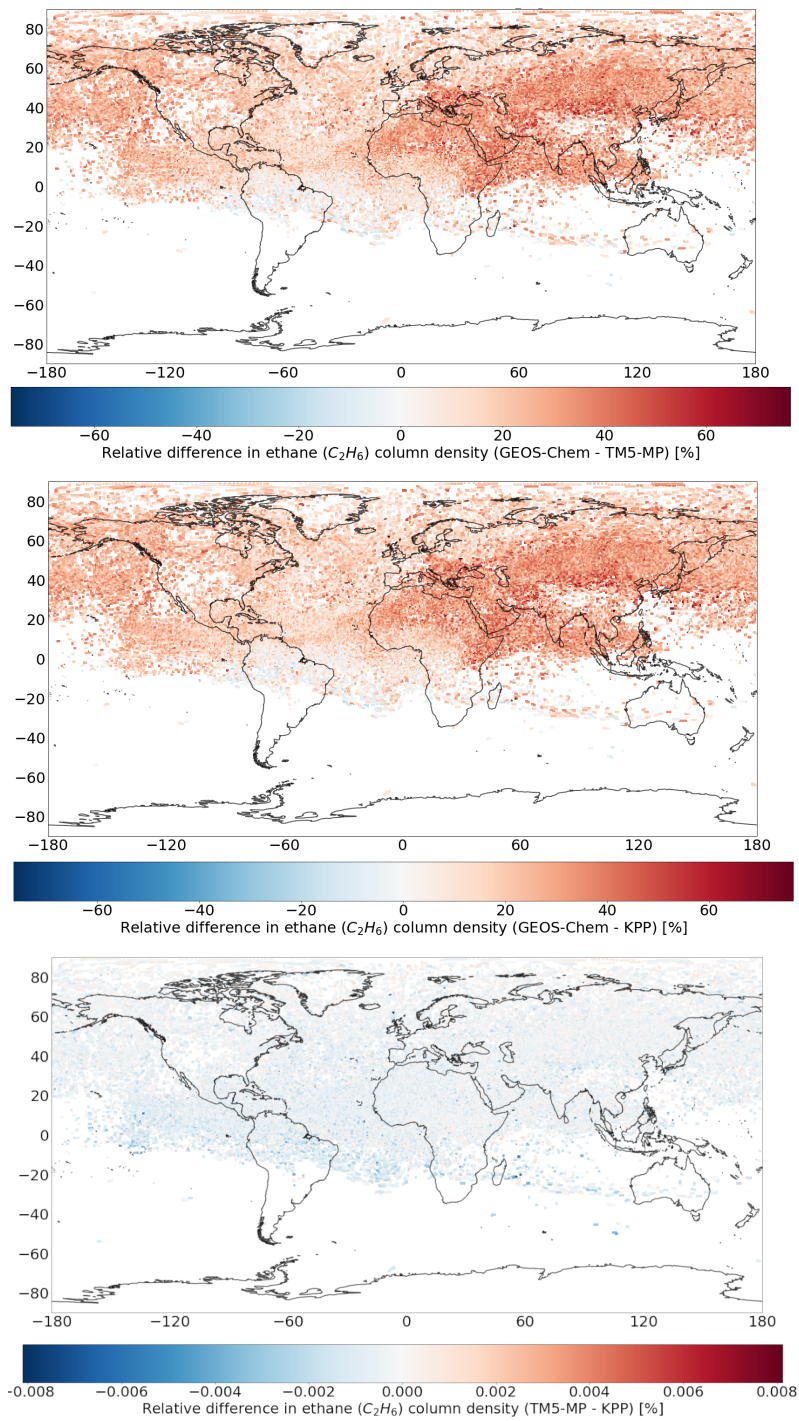
Fig. 20: Relative differences of figure 18: (top) the default chemical scheme with GEOS-Chem OH fields and TM5-MP OH fields, (middle) the default chemical scheme with GEOS-Chem OH fields and KPP chemical scheme with TM5-MP OH fields and (bottom) the default and KPP chemical schems both using TM5-MP OH fields. All after 1 month from the release (figure 17). The differences have been calculated through: $100\frac{GEOS-TM5}{\frac{GEOS+TM5}{2}}$

4.3.2. Species validation

The performance evaluation of the developed version of FLEXPART with coupled KPP is done only for those species that, by default, react with OH in FLEXPART. We run the same simulation for each of them and for each development stage made in the OH chemical mechanism in this project. The results of these simulations can be appreciated in the figure 21.

Each species' validation results are represented by the time evolution of each species' total mass. This variable is the most useful to assess the performance of the OH sink for each transported species. The evolution of the mass of each species differ from each other. While methane's (CH_4) mass remains almost unchanged along the simulation, due its long life-time in the atmosphere, the other species' masses have more diverse evolution: Ethane (C_2H_6) total mass decreases at a slower pace compared to polychlorinated biphenyl 28 ($PCB28$), again due to their different life-times in the atmosphere. Finally, gamma-hexachlorocyclohexane's ($G-HCH$) behaviour reflects this species' short lifetime, since its total mass only increases during night-time when OH concentration is low and completely plunges during day-time, when OH is produced through photolysis. Even at night its concentration is low, not reaching the levels of the other species (note the ordinates axis difference due to the high release in figure 21).

Parallelly, one can compare tn the same plots the different FLEXPART development stages performance for each species (the different lines in each plot). In this case, one can observe that the time evolution of each species' total mass is similar: FLEXPART together with TM5-MP OH fields and KPP coupled produces similar results as using FLEXPART with its default chemical scheme, and any GEOS-Chem OH fields or TM5-MP OH fields. This similarity between the development stages can be better assessed through the correlation scatter plots in figure 22 for each species case.
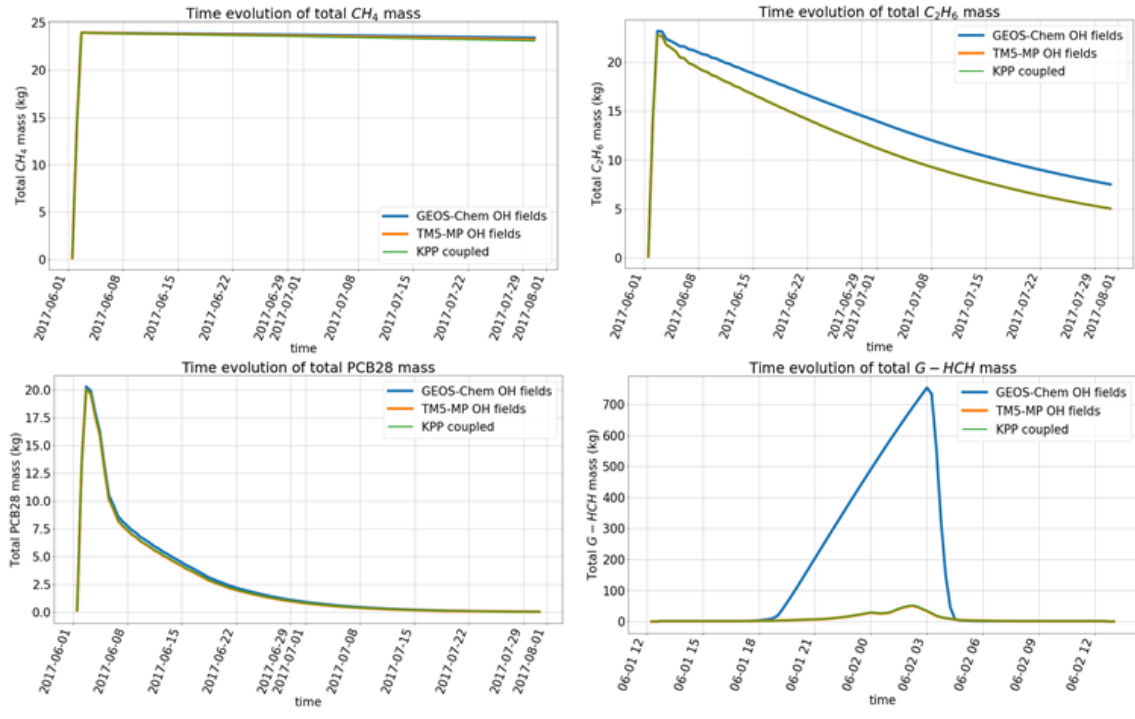
Fig. 21: Validation of the KPP coupling for each FLEXPART species reacting with OH. In each figure, three plots can be observed corresponding to the three validation simulations: default FLEXPART, FLEXPART using TM5-MP OH fields and FLEXPART using TM5-MP OH fields coupled with KPP. In each case, 24kg of species has been released during the first 24h of the simulation. For $G-HCH$, where $24e3$ kg were released to make the atmospheric concentration observable. Note the change of time scale for the $G-HCH$ case because of this effect.

Fig. 22: Cross-correlation between the species simulations with the default chemical scheme using GEOS-Chem and TM5-MP OH fields (left) and with the default chemical scheme and KPP-produced box model using both TM5-MP OH fields (right). Note that there is no plot 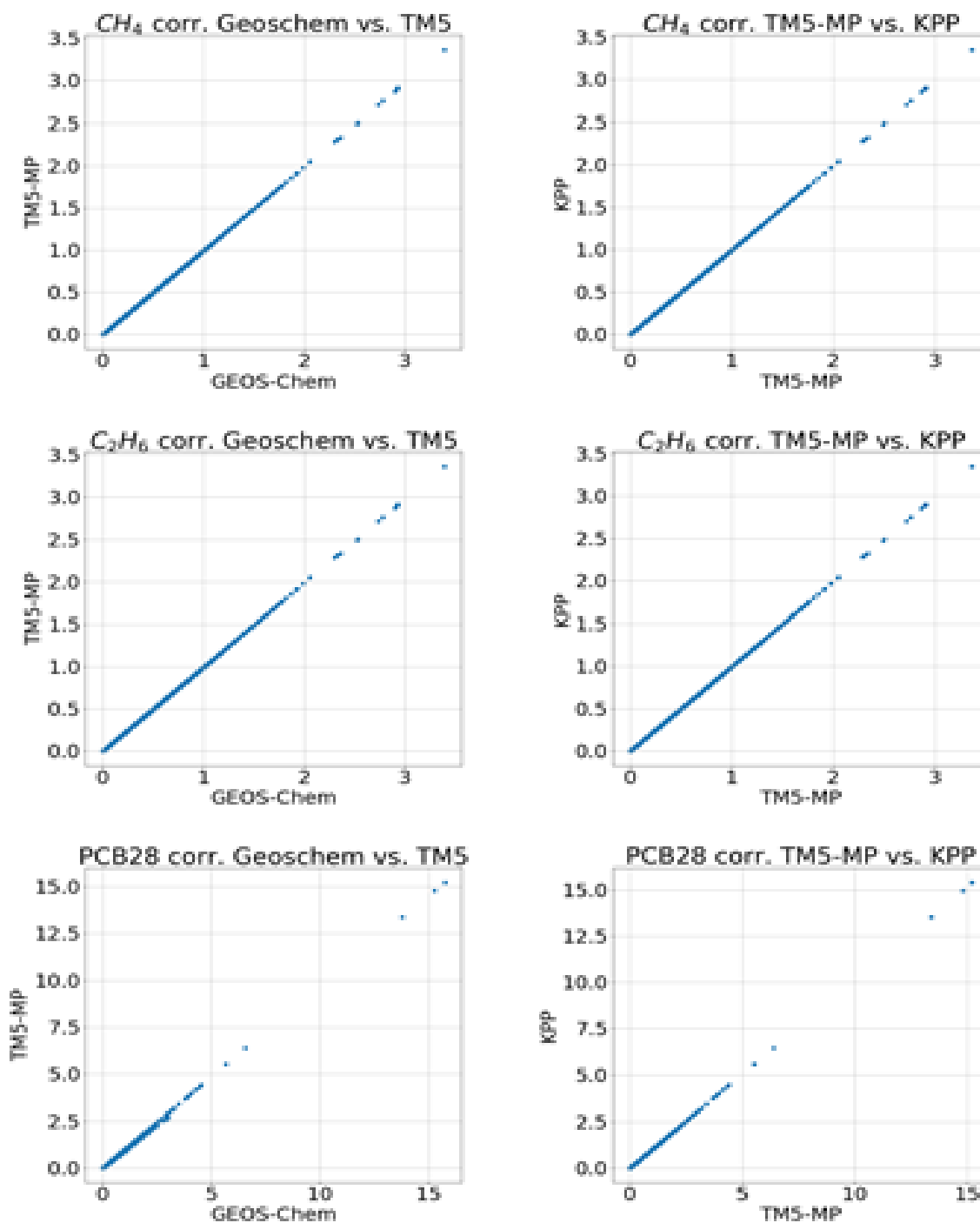for $G - HCH$ because of the limited amount of data for this specific species, consequence of its short atmospheric lifetime as shown in figure 21.

4.3.3. Observational data validation

Additionally to each species' validation and each development stage's validation, a simulation to compare the results against observational data has been performed.

The simulation in this case is ethane at , emulating with FLEXPART CMIP6 emissions for ethane. Ethane is released continuously during the simulation and its concentration is compared at each station's location during the time period of the simulation. The time plots of the ethane's concentration at some representative locations are shown in figure 23. From them, it can be appreciated that despite some dissimilarities, the general behaviour of the simulated ethane concentration is close to the observational data. Some station's data has a low time sampling, which obliges to interpolate the concentration along the days. In this cases the daily (or hourly) fluctuations of C_2H_6 concentrations cannot be appreciated. Some other stations, despite providing data with high time sampling frequency, this cannot be reproduced by the model, although the order of magnitude is maintained. Finally, at specific locations the model and the observations are very close even in short-period fluctuations.
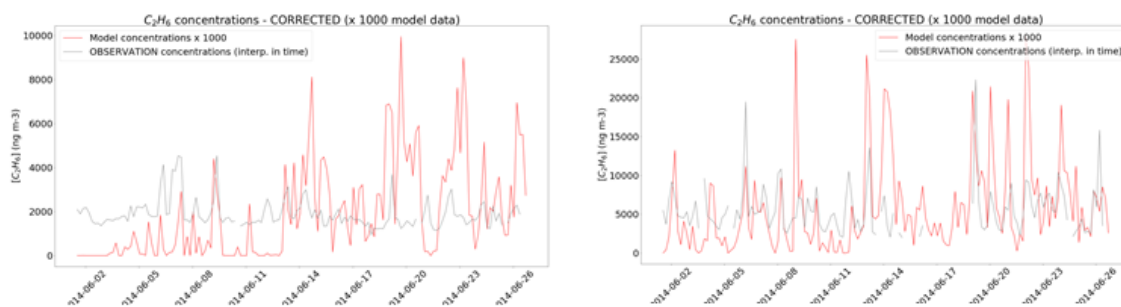

Fig. 23: Validation of the KPP coupling with observational data. In each plot the observational timeseries (red), the simulation with default FLEXPART (gray) and the simulation with TM5-MP OH fields and coupled KPP (black) can be observed.

5. Discussion

5.1. Discussion on data

The data used in this project has been obtained from two main sources: TM5-MP simulations (OH fields) and the CMIP6 ethane emission data. All the other data was generated by the own FLEXPART model or the KPP preprocessor.

Regarding the data coming from TM5-MP, the project was constrained to the data available previously generated with TM5-MP. Therefore, the only requirement of all the simulations of this project to evaluate the new OH fields to FLEXPART was that the simulation time period had to fit with the OH fields dates available (01.06.2006 to 31.08.2006). The format of this data was similar to the default OH fields (GEOS-Chem) of FLEXPART, and only the structure of this data required FLEXPART to be adapted.

In the case of CMIP6 emission data, it can be highlighted the complete coverage of ethane sources that this dataset offers. From biomass burning to all sectors of anthropogenic sources were taken into account. The only lacking component is the biogenic ethane emissions. However, it has been shown the relative irrelevance of this component [31], thus it has been neglected from the considered emissions.

The data obtained through FLEXPART simulations provide an insight on the quality of the implementation of TM5-MP OH fields and KPP. The only variable used to evaluate them has been the total mass of the transported species, neglecting any importance of its spatial distribution. This variable is an approximate good approach to the general behaviour of the OH fields in a global/European scale (where the simulations have been performed). Nevertheless, the total mass transported is not a proper approach if small-scale (i.e. regional and local) simulations are to be evaluated. This is because the OH spatial and temporal distribution differs a lot from GEOS-Chem to TM5-MP, since the second one provides a higher resolution and different values as shown in section 4.2.

5.2. Discussion on methodology

The methodology has consisted in two main stages of development. The first one, the implementation of TM5-MP generated OH fields, and the second the coupling of KPP.

Regarding the order of the stages, it makes the TM5-MP OH fields readable from FLEXPART before

implementing the KPP. This is done because of the quality of the data needed by. Technically KPP could be coupled without having TM5-MP fields available beforehand. However, TM5-MP OH fields present a higher resolution and time accuracy than the default fields. They provide a more accurate base for, afterwards, applying the KPP chemical schemes which is able to adapt to the quality of the fields. Also, the task of implementing TM5-MP OH fields is simpler than the KPP coupling. This approach resulted to be appropriate when aiming the final objective of this project, that is to improve the FLEXPART. Although the opposite order would have been also possible to be followed.

The methodology for the TM5-MP OH fields implementation was to adapt the original FLEXPART code to the specific structure of the new netcdf files from TM5-MP. This task was more complex due to the fact that TM5-MP was not providing all the variables in a single file, but instead dividing them in two different files. Nevertheless, all the complications were overcome after duplicating the reading scripts from FLEXPART and joining the variables read from the two files. Other methodologies form the default FLEXPART reading process of OH fields were kept the same, like the interpolation of the fields to the centre of each grid cell. Others were even improved, like reading directly only the next two time values of the field, instead of reading all the OH fields first - keeping them in memory - and loading the two next values into variables for each time step.

The methodology regarding the KPP coupling has been similar to other reference projects, like in [11]. The coupling is based in developing scripts that patch the FLEXPART environment with the box model numerical integrator. After adapting the variables to be shared from one model to the other, a make file is generated to compile them both at the same time. Nevertheless, the current methodology requires the chemical mechanism of the box model to be set before the compilation of both models.

5.3. Discussion on the results

The initial aim of the project was met successfully; Combining the capabilities of the three tools, an enhanced product was created. This product generates results that reproduce those from the default FLEXPART v10.4, as it is thoroughly validated in sections 4.2 and 4.3.

The results at every stage of the development are consistent and prove reproducibility. Simulations' results show that the change of the OH fields used has a stronger effect on the results than the change on the chemical reaction mechanism. This can be appreciated from the comparison of ethane simulations using different OH fields and different chemical schemes from sections 4.2 and 4.3.1. The results of these simulations show that stronger differences arise when GEOS-Chem OH fields are shifted to TM5-MP ones. On the other hand, results remain almost the same between using the default chemical reaction and the box

model. This can be interpreted as a successful coupling of the KPP to FLEXPART, since the coupling of the two models on its own do not rise simulation differences, and these can be attributed to differences in the quality of the OH fields used.

When comparing simulations transporting different species that take part in reactions with OH (section 4.3.2), despite some differences, the general behaviour of the species is similar between development stages. It is important to note the high dissimilarities between results from each development stage in the case of $G - HCH$ simulations, which are attributed to the species' relative short lifetime compared to the other species: The fast oxidation of $G_H CH$ in the atmosphere makes the evolution of its concentration very sensible to the spatial distribution of OH. This high sensibility to the small-scale values of the OH field can justify the differences observed in the evolution of $G - HCH$ between simulations with the GEOS-Chem and TM5-MP OH fields, and also explains why its evolution does not change between FLEXPART with and without KPP coupled, both using the same OH fields. However, it should be discussed why the half-lifetime of $G - HCH$ in the simulations do not match the bibliography value of 20 days from [21].

Regarding the chemical scheme, the box model integrator reproduces very similar results to the default chemical reaction based on equation 20, as shown in section 4.2. This results show that the main differences arise from the change of OH fields, and not from the shift in the chemical scheme: Time evolution of the concentrations are close in all the simulations performed between the stage with and without KPP coupled.

Some drawbacks arise from the current development of FLEXPART though. In the first place, the chemical mechanism has not been expanded as originally intended, which would be useful to further evaluate the enhancement of the simulations' quality compared to the default FLEXPART v10.4. Secondly, a most user friendly interface could have been developed in the KPP coupling stage, because with the current development the box model chemical mechanism must be changed and compiled manually with KPP every time that the transported species is changed. Finally, it would have been useful to obtain more reliable results regarding the OH chemistry when validating the simulations against the observable data. Maybe using small-scale study cases with OH-reacting species or performing more complete global simulations, with more releases and not filtering the emissions database.

Nevertheless, the coupled model is likely to provide enhanced simulations regarding OH chemistry in the atmosphere and a tool with a potentially easy to expand chemical mechanism, depending on the requirements.

6. Summary and conclusions

In this project, FLEXPART v10.4 has been developed in two ways: 1) making TM5-MP output fields readable as input data by the model and 2) coupling a box model created with KPP as chemical scheme of the Lagrangian model.

As a first test case, in this preparatory project the GEOS-Chem OH fields that FLEXPART has by default have been replaced by TM5-MP OH data. The TM5-MP OH fields are shown to represent an improvement regarding the horizontal and vertical resolution compared to the GEOS-Chem fields. Additionally, they represent the OH distribution at the specific time of the simulation, not using a representative year and a simulated diurnal cycle to obtain the actual OH concentration. Two simulations of an ethane emission from Europe with FLEXPART have been used to validate the results: One using the default OH fields of the model and the other implementing the new TM5-MP OH fields. The differences are regionally important, reaching a 30% difference compared to the average of both concentrations. Nevertheless, the evolution of the total ethane mass present in the atmosphere along the considered time period is consistent between the two simulations. This indicates that 1) the overall ethane loss due to OH reaction follows the same pattern in both cases, and 2) that high relative differences between the two ethane simulations can be attributed to localized high OH concentration parcels in TM5-MP data, which are not observed in GEOS-Chem OH data possibly due to its lower horizontal resolution. Finally, the OH differences between data sets is compared to the drift in ethane decline trends, evaluating the consistency between both.

The second development stage has been the coupling of FLEXPART with a KPP-generated box model. The KPP numerical integrator was implemented into the FLEXPART code through an amid script between the two models, in order to use each time step variables from FLEXPART (particles' position and mass, as well as local temperature and pressure) as input to the box model. The box model integrates the chemistry for each grid cell during the time period of each time step. The results obtained using KPP and TM5-MP OH fields prove a strong reproducibility of the results obtained with the undeveloped version of FLEXPART and with the developed version that only includes TM5-MP OH fields.

These developments represent a first approach to extend the chemical mechanism of FLEXPART, in order to obtain more accurate results in the future and make FLEXPART a useful tool for studies with complex chemistry. Additionally, other developments have been performed in this project, including multiple reacting agents and photolysis reaction for the transported species. Although, due to time and resources limitations, they could not be tested and could not be validated, respectively, they open the door to simulate even more complex chemical mechanisms in the future.

7. Future developments

As a first step in future developments of this project, it is necessary to perform the validation of the current advances against observational data. Although, as stated in the previous section, so far this project reproduces the results of the undeveloped version of FLEXPART from which it started, long-term behaviour of the chemical solver needs to be assessed.

To achieve this proper validation, it is needed to tackle the errors that halt simulations with a large number of release points and particles. These kind of simulations are needed specially for ethane emissions simulations, where large regions of emissions are to be emulated and long time periods simulated (at least longer than the half-lifetime of ethane, i.e. 2.5 months). The error that prevent the model to complete these kind of simulations is almost surely due to a lack of memory error (see appendix 8.3).

If that is solved, further developments are possible, like the extension of FLEXPART chemical mechanism to other species that currently do not react chemically in the model (such as CO, SO_2, nitrates and ammonia). This task would be fairly easy with KPP, and no further changes in FLEXPART code would be needed beyond including these species in the chemical reaction.

A fully developed version of the KPP-FLEXPART model would be useful for small scale studies in, for example, industrial emissions of nitrates or sulfates. Even with short lived species, the model would have become adaptable to virtually any situation.

References

[1] G. Brasseur and D. Jacob, *Modeling of Atmospheric Chemistry*. Cambridge University Press, 2017.

[2] A. Stohl, M. Hittenberger, and G. Wotawa, "Validation of the lagrangian particle dispersion model flex-part against large-scale tracer experiment data," *Atmospheric Environment*, vol. 32, no. 24, pp. 4245 – 4264, 1998.

[3] I. Pisso, E. Sollum, H. Grythe, N. I. Kristiansen, M. Cassiani, S. Eckhardt, D. Arnold, D. Morton, R. L. Thompson, C. D. G. Zwaaftink, N. Evangeliou, H. Sodemann, L. Haimberger, S. Henne, D. Brunner, J. F. Burkhart, A. Fouilloux, J. Brioude, A. Philipp, P. Seibert, and A. Stohl, "The lagrangian particle dispersion model FLEXPART version 10.4," *Geoscientific Model Development*, vol. 12, pp. 4955–4997, dec 2019.

[4] A. Stohl, C. Forster, A. Frank, P. Seibert, and G. Wotawa, "Technical note: The lagrangian particle dispersion model FLEXPART version 6.2," *Atmospheric Chemistry and Physics*, vol. 5, pp. 2461–2474, sep 2005.

[5] J. C. Lin, "Lagrangian modeling of the atmosphre: An introduction," in *Lagrangian Modeling of the Atmosphere* (geopress, ed.), pp. 1–11, American Geophysical Union, mar 2013.

[6] J. Lin, *Lagrangian modeling of the atmosphere*. Washington, DC: American Geophysical Union, jan 2012.

[7] A. F. Stein, R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, and F. Ngan, "NOAA's HYSPLIT atmospheric transport and dispersion modeling system," *Bulletin of the American Meteorological Society*, vol. 96, pp. 2059–2077, dec 2015.

[8] M. Cohen, R. Artz, R. Draxler, P. Miller, L. Poissant, D. Niemi, D. Ratté, M. Deslauriers, R. Duval, R. Laurin, J. Slotnick, T. Nettesheim, and J. McDonald, "Modeling the atmospheric transport and deposition of mercury to the great lakes," *Environmental Research*, vol. 95, pp. 247–265, jul 2004.

[9] A. Jones, D. Thomson, M. Hort, and B. Devenish, "The u.k. met office's next-generation atmospheric dispersion model, NAME III," pp. 580–589.

[10] D. S. McKenna, "A new chemical lagrangian model of the stratosphere (CLaMS) 2. formulation of chemistry scheme and initialization," *Journal of Geophysical Research*, vol. 107, no. D15, 2002.

[11] "Interfacing geoschem with kpp."

[12] R. Atkinson, "Gas-phase tropospheric chemistry of volatile organic compounds: 1. alkanes and alkenes," *Journal of Physical and Chemical Reference Data*, vol. 26, pp. 215–290, mar 1997.

[13] J. M. Wallace and P. V. Hobbs, "5 - atmospheric chemistry," in *Atmospheric Science (Second Edition)* (J. M. Wallace and P. V. Hobbs, eds.), pp. 153 – 207, San Diego: Academic Press, second edition ed., 2006.

[14] D. R. Blake and F. S. Rowland, "Global atmospheric concentrations and source strength of ethane," *Nature*, vol. 321, pp. 231–233, may 1986.

[15] J. Rudolph, "The tropospheric distribution and budget of ethane," *Journal of Geophysical Research*, vol. 100, no. D6, p. 11369, 1995.

[16] A. C. Aikin, J. R. Herman, E. J. Maier, and C. J. McQuillan, "Atmospheric chemistry of ethane and ethylene," *Journal of Geophysical Research*, vol. 87, no. C4, p. 3105, 1982.

[17] R. Koppmann, ed., *Volatile Organic Compounds in the Atmosphere.* Blackwell Publishing Ltd, 2007.

[18] M. Mandalakis and E. G. Stephanou, "Study of atmospheric PCB concentrations over the eastern mediterranean sea," *Journal of Geophysical Research: Atmospheres*, vol. 107, pp. ACH 18–1–ACH 18–14, dec 2002.

[19] P. N. Anderson and R. A. Hites, "OH radical reactions: the major removal pathway for polychlorinated biphenyls from the atmosphere," *Environmental Science & Technology*, vol. 30, pp. 1756–1763, jan 1996.

[20] J. Ma, S. Daggupaty, T. Harner, and Y. Li, "Impacts of lindane usage in the canadian prairies on the great lakes ecosystem. 1. coupled atmospheric transport model and modeled concentrations in air and soil," *Environmental Science & Technology*, vol. 37, pp. 3774–3781, sep 2003.

[21] T. Vera, E. Borrás, J. Chen, C. Coscollá, V. Dale, A. Mellouki, M. Ródenas, H. Sidebottom, X. Sun, V. Yusá, X. Zhang, and A. Muñoz, "Atmospheric degradation of lindane and 1,3-dichloroacetone in the gas phase. studies at the EUPHORE simulation chamber," *Chemosphere*, vol. 138, pp. 112–119, nov 2015.

[22] A. Sandu and R. Sander, "Technical note: Simulating chemical systems in fortran90 and matlab with the kinetic PreProcessor KPP-2.1," *Atmospheric Chemistry and Physics*, vol. 6, pp. 187–195, jan 2006.

[23] P. Berrisford, D. Dee, P. Poli, R. Brugge, M. Fielding, M. Fuentes, P. Kållberg, S. Kobayashi, S. Uppala, and A. Simmons, "The era-interim archive version 2.0," p. 23, 11 2011.

[24] D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Khler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart, "The ERA-interim reanalysis: configuration and performance of the data assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 137, pp. 553–597, apr 2011.

[25] A. Badia and O. Jorba, "Gas-phase evaluation of the online NMMB/BSC-CTM model over europe for 2010 in the framework of the AQMEII-phase2 project," *Atmospheric Environment*, vol. 115, pp. 657–669, aug 2015.

[26] J. E. Williams, K. F. Boersma, P. L. Sager, and W. W. Verstraeten, "The high-resolution version of TM5-MP for optimized satellite retrievals: description and validation," *Geoscientific Model Development*, vol. 10, pp. 721–750, feb 2017.

[27] R. M. Hoesly, S. J. Smith, L. Feng, Z. Klimont, G. Janssens-Maenhout, T. Pitkanen, J. J. Seibert, L. Vu, R. J. Andres, R. M. Bolt, T. C. Bond, L. Dawidowski, N. Kholod, J.-I. Kurokawa, M. Li, L. Liu, Z. Lu, M. C. P. Moura, P. R. O'Rourke, and Q. Zhang, "Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the community emissions data system (ceds)," *Geoscientific Model Development*, vol. 11, no. 1, pp. 369–408, 2018.

[28] K. Sindelarova, C. Granier, I. Bouarar, A. Guenther, S. Tilmes, T. Stavrakou, J.-F. Müller, U. Kuhn, P. Stefani, and W. Knorr, "Global data set of biogenic voc emissions calculated by the megan model over the last 30 years," *Atmospheric Chemistry and Physics*, vol. 14, no. 17, pp. 9317–9341, 2014.

[29] V. Damian, A. Sandu, M. Damian, F. Potra, and G. R. Carmichael, "The kinetic preprocessor KPP-a software environment for solving chemical kinetics," *Computers & Chemical Engineering*, vol. 26, pp. 1567–1579, nov 2002.

[30] N. Evangeliou, H. Grythe, Z. Klimont, C. Heyes, S. Eckhardt, S. Lopez-Aparicio, and A. Stohl, "Atmospheric transport is a major pathway of microplastics to remote regions," *Nature Communications*, vol. 11, jul 2020.

[31] G. Etiope and P. Ciccioli, "Earth's degassing: A missing ethane and propane source," *Science*, vol. 323, pp. 478–478, jan 2009.

[32] Unidata, "Network common data form (netcdf)," 2019.

8. Appendix

In the following appendix sections technical information regarding TM5-MP OH fields implementation and the coupling programming of FLEXPART with the KPP chemical preprocessor is provided.

It must be taken into account that the current information is specific for the development branch of FLEXPART v10.4 from LAMOS department in Bremen University, which is used in the HPC *Aether* compiled with `ifort/2019a`. Therefore, the following implementations are liable of not being reproducible under other branches of FLEXPART or using other compilers such as GFortran.

8.1. TM5-MP OH fields implementation

TM5-MP model has been used to provide the OH fields for this project. In this section we provide an overview of the characteristics of these files and some information on how FLEXPART has been adapted to read them as input.

TM5-MP output fields are provided in `netCDF` file format. `NetCDF` provide several advantages [32]:

- Self-describing: information about the data contained in the `netCDF` file is also included in it.

- Network-transparent: it does not matter how the computer stores the different numbers formats (integers, floating-point, double, etc.), `netCDF` files can be accessed regardless the computing format.

- Direct-access: it is not necessary to read all the file to access a specific part of the data it contains.

- Extendible: different `netCDF` files can be combined without needing a change in their structure or format.

The TM5-MP OH fields used in this thesis are daily outputs, with hourly time frequency. Their main characteristics (i.e. spatial resolution, temporal extension and magnitudes) have been exposed in section 3.2. Moreover, output variables are distributed over two files, *gas* and *TP* files. The content for each of them is summarized in table 4. The time independent variables are 1D coordinates, while the time independent ones are 4D (i.e. longitude, latitude, vertical level and time). The treatment of the vertical layer coordinate is already exposed in section 3.2.

As stated in the main text, just two time steps for each variable field are stored in memory at each time step of the simulation, in the variable `OH_hourly`, that is 4D with just 2 time coordinates (current and next

time step). This improves the computational performance when reading the OH fields from TM5-MP with respect to the GEOS-Chem implementation that stores all fields. The time dependent variables in table 4 are read using the subroutine `read_variables`. This subroutine takes as input the variables: `month`, `day`, `start4D`,`count4D`, `spec`, `timestep`. Where `month` and `day` are the current simulation time, `start4D` and `count4D` is the starting point and `count4D` the amount of data to read. Finally, `timestep` is the field to be stored in memory (either 1 for the current field, or 2 for the next one). This treatment using subroutines might get the reading more complex, but it has the advantage of allowing a better management of day change. Since the `read_variables` subroutine can be specified which day and how many time steps to read at each time step, making it flexible to read the last field of a day (at 23h) for the current field and the first field for the next day (00h) as the second field to be stored.

Table 4: TM5-MP files structure. The time independent variables are 1D coordinates, while the time independent ones are 4D, and just two times are stored for them at each simulation time step.

	TP files	*Gas* files
Time dependent	Pressure, geopotential and temperature	OH concentrations
Time independent		Longitudes, latitudes, levels and time

The FLEXPART files structure has been changed in different aspects with respect to the default v10.4 configuration. In FLEXPART, all the variables are centralized by the `par_mod.f90` (for general variables) `oh_mod.f90` (for OH-related variables) modules. Meanwhile, the file `timemanager.f90` centralizes all the subroutine callings. In the present development, the calling of the `gethourlyOH` subroutine has been replaced by the calling of the `readOHfield` module, which contains the subroutines that carry out the tasks specified in the precedent paragraph. After storing the two OH fields for the next time steps, the `timemanager` calls the `ohreaction` subroutine that computes the reaction of the OH field, taking the OH fields from the `OH_hourly` variable stored. In this last subroutine, all the computations to adapt the units from TM5-MP output files to the FLEXPART input units are performed before the computation of the OH reaction.

With all these changes, OH fields from TM5-MP has been made available for FLEXPART, and most of the remaining developments to be performed for coupling KPP to FLEXPART is going to centralize in linking the `ohreaction` subroutine to a coupling script between the two models.

It is also important to note that all these modifications have been also performed in the MPI version of the files, such as `timemanager_mpi.f90`.

8.2. KPP coupling

The KPP produced box model coupling to FLEXPART has been performed principally on the `ohreaction.f90` and `integration_kpp.f90` files.

The first subroutine (`ohreaction.f90`) performs the OH reaction with the transported species, which by default is computed from equation 20. The coupling programming is based on replacing this equation by the integrator of the box model created with KPP. To achieve this, a linking subroutine is needed, which performs the migration of variables from FLEXPART to the integrator and back to FLEXPART. This subroutine is developed in the `integration_kpp.f90` file that imports the variables:

- `itime` Being the current simulation time.

- `ltsample` The time-step length in FLEXPART (normally 300s).

- `loutnext` The next time for an output.

- `SPCconc` The transported species concentration.

- `OHconc` The OH concentration field, for the next two time steps (`itime` and `itime + ltsample`).

- `Temperature` The temperature field at `itime`.

- `Pressure` The pressure field at `itime`.

- `ldirect` The direction of the simulation, if forward (1) or backward (-1).

The integrator is given the species concentrations and other variables (e.g. temperature and pressure) for location (x,y,z) and time step, going through all the simulation domain at each time step. Then it performs the numerical integration. The integrator is a Rosenbrock Rodas-3, which does not need an internal time step to be specified. The integration takes place between `itime` and `itime + ltsample` in a `for` loop, since the integrator runs for automatically-specified time steps:

```
T = TSTART
kron: DO WHILE (ldirect*T < ldirect*TEND)
    TIME = T
    CALL GetMass( C, DVAL )

    !CALL SaveData() !Not saving at every reaction time step
    CALL Update_SUN()
```

```
       CALL Update_RCONST()

       CALL INTEGRATE( TIN = T, TOUT = T+DT, RSTATUS_U = RSTATE, &
       ICNTRL_U = (/ 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0 /) )
       T = RSTATE(1)

       write(20,350) 'T/TEND = ', T,'/', TEND
       write(20,991) (T-TSTART)/(TEND-TSTART)*100, T,     &
             ( TRIM(SPC_NAMES(MONITOR(i))),            &
             C(MONITOR(i))/CFACTOR, i=1,NMONITOR )

END DO kron

CALL GetMass( C, DVAL )
```

The concentrations of the transported species and OH are passed from/to FLEXPART to/from the the box model through the `GetMass` KPP subroutine. The `GetMass` subroutine at the same time takes the concentrations at the current time from the `VAR` vector, that includes the concentrations of all the species taking part in the box model. In this project, the products' concentrations are assumed to be zero before the integration starts, since we are not interested in their values after the integration:

```
use kpp_model_Model !includes t_Global & t_Integrator
use kpp_model_Initialize, ONLY: Initialize

C2H5conc = 0.0
H2Oconc = 0.0
VAR_user = (/SPCconc, OHconc, H2Oconc, C2H5conc/)
```

By default in this `VAR` vector, species' concentrations are reordered to make use of the sparsity structure of the Jacobian and Hessian to increase efficiency. To match them with the order read from FLEXPART, the subroutine `Shuffle_user2kpp` have to be used. Already defined in the KPP model, this subroutine reorders the vector of species' concentrations from FLEXPART according to the order imposed by the sparsity structure. The opposed subroutine `Shuffle_kpp2user` is also used to migrate the concentrations from the box model back to FLEXPART. Alternatively, the option `REORDERING` can be turned off in the KPP model, at the expense of computational efficiency:

```
CALL Shuffle_user2kpp(VAR_user, VAR)
```

```
... integration ...

CALL Shuffle_kpp2user(VAR,VAR_user)
```

Regarding the other variables, the linking script `integration_kpp.f90` replaces the default values coming into the KPP model by the current values of the FLEXPART time step. KPP global values are called from the module `kpp_model_Model`, which includes `kpp_Global` subroutine where values are declared. Variables such as temperature or pressure of the specific grid cell's location and time of the simulation are updated for each integration:

```
FIX = FIX
TIME = itime
TEMP = Temperature
TSTART = itime ! TIN = itime in the function
TEND = TIME + ldirect*ltsample !(Or itime + loutstep ???? )
DT = ldirect*ltsample !DT = end of the simulation of kpp, so lsynctime of FLX.
```

8.3. Issues with the observational validation

A persistent problem have been experienced in this project during simulations with a large number of particles and release points. Such was the case for the simulation to validate the coupling of KPP-produced box model with FLEXPART, in which CMIP6 ethane emission data was emulated with FLEXPART and this species concentration was compared against surface sensor data after 2 months.

Concretely, the model was unable to release all the particles required. They were emitted at a slow rate for the first half of the simulation and, at this point, the model tried to release all of them at once at the middle of the simulation period (see figure 24). This caused a segmentation error (SIGSEGV error), reported in the file *advance.f90* in the line 261 of the LAMOS FLEXPART branch:

```
forrtl: severe (174): SIGSEGV, segmentation fault occurred
Image              PC                Routine            Line       Source
libifcoremt.so.5   00002B6740C83522  for__signal_handl  Unknown    Unknown
libpthread-2.17.s  00002B67407F0100  Unknown            Unknown    Unknown
FLEXPART_MPI       00000000004C2813  advance_              266      advance.f90
FLEXPART_MPI       000000000044B32D  timemanager_          761      timemanager_mpi.f90
FLEXPART_MPI       000000000040783A  MAIN__                482      FLEXPART_MPI.f90
```

```
FLEXPART_MPI          0000000000405D02   Unknown                    Unknown    Unknown
libc-2.17.so          00002B6742FDDB15   __libc_start_main          Unknown    Unknown
FLEXPART_MPI          0000000000405C19   Unknown                    Unknown    Unknown
```

When plotting the total mass of ethane present in the atmosphere, this inhomogeneous release can also be observed (figure 25).
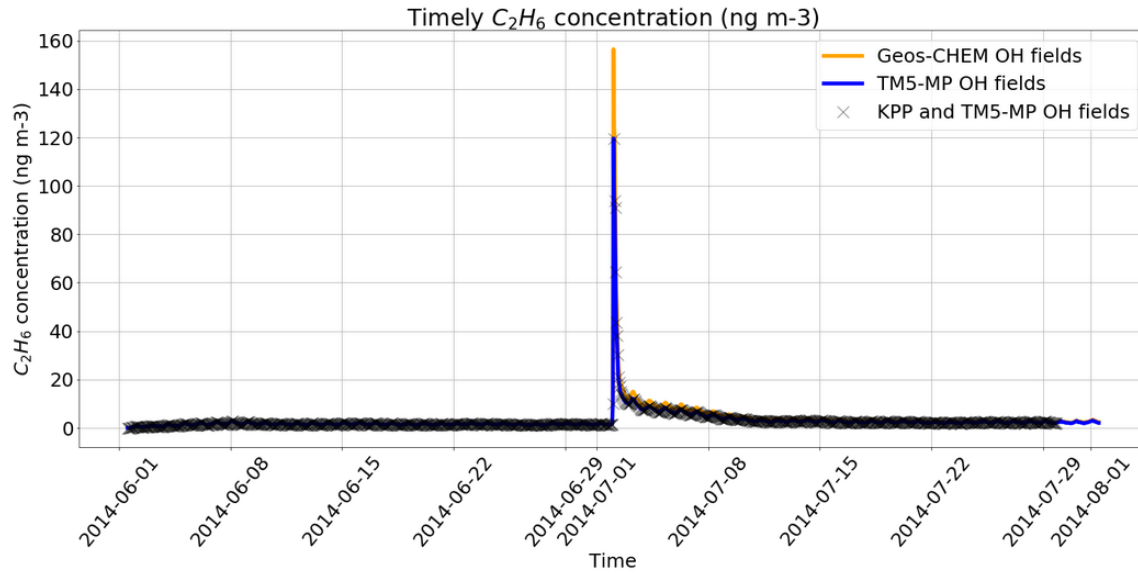

Fig. 24: Average concentration of surface ethane during the simulation emulating CMIP6 ethane emissions with FLEXPART. Inhomogeneous particle release during the simulation can be observed, with a sudden peak in the middle of the simulation period that normally arises a SIGSEGV error that interrupts the simulation (see figure **??**). In this figure, the amount of particles and release points are just the needed for the error not to be met, but the ethane concentration is highly affected by the abnormal release distribution.
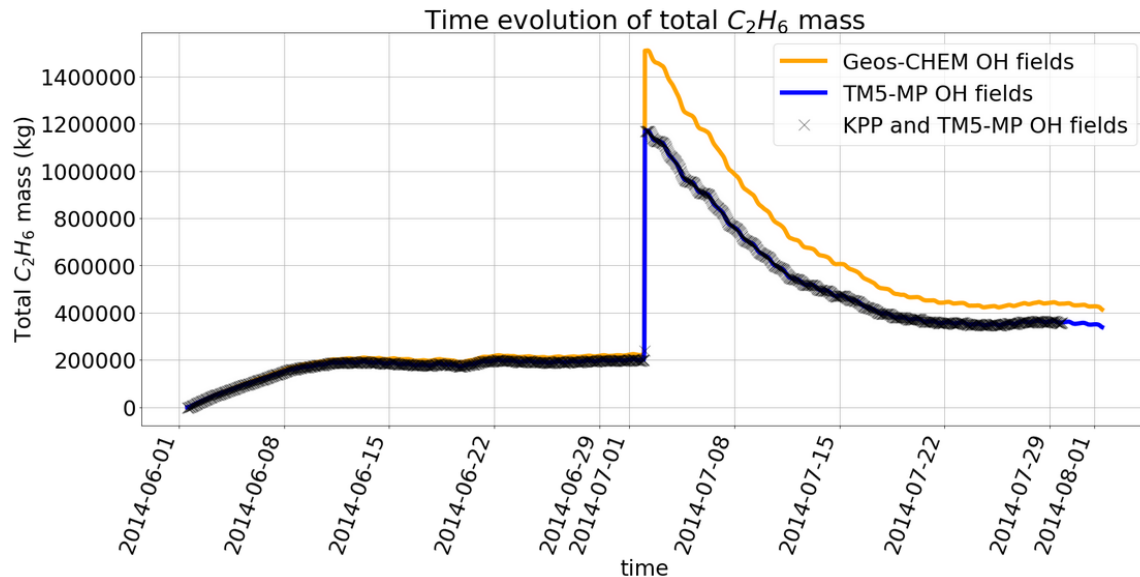
Fig. 25: Total ethane mass present in the atmosphere during the simulation emulating CMIP6 ethane emissions with FLEXPART. Here, like in figure 24, inhomogeneous particle release can also be observed.

If the concentrations of ethane are then compared to the observational data available, one can observe a two orders of magnitude difference between both, remaining the model data always below the observational data. One can appreciate that in figures 26 and 28.
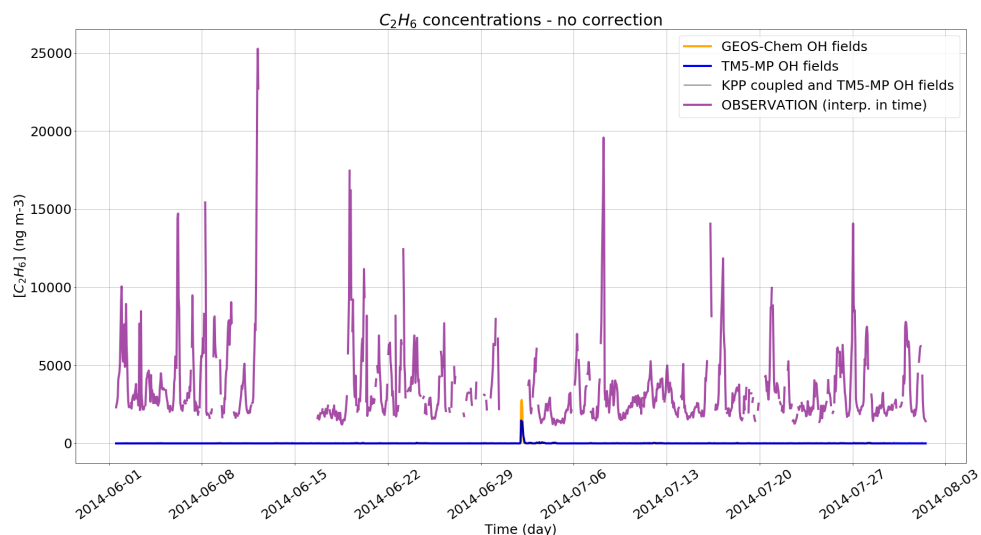
Fig. 26: Ethane concentration time series over one of the observational stations when using the concentrations until the middle of the simulation period. The orange, blue and black lines at the bottom of the plot represents the simulation concentrations, clearly below the observational levels (purple line).
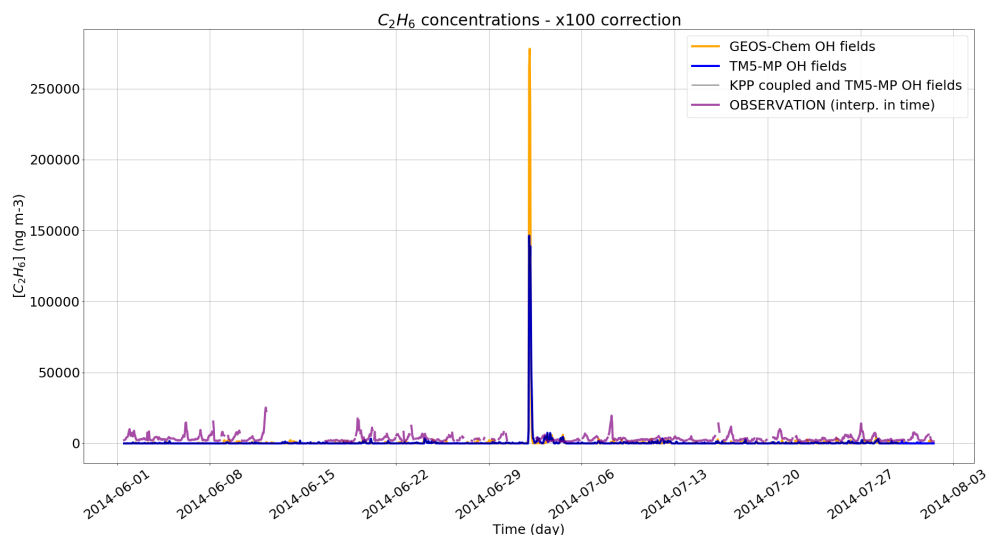

Fig. 27: Ethane concentration time series over one of the observational stations when the simulation concentrations are multiplied by 100 (orange, blue and black lines). In this case the concentrations are fairly close to the observational data (purple line).
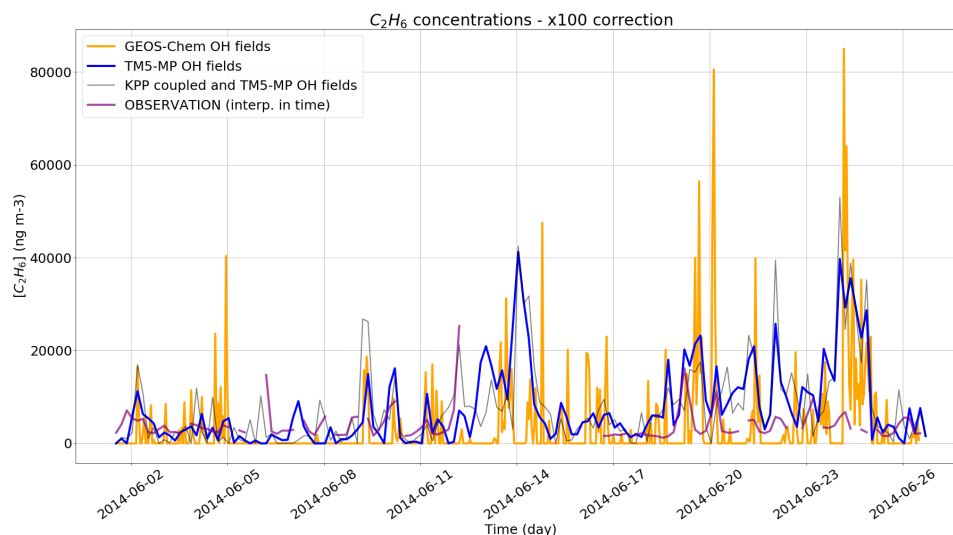
Fig. 28: Ethane concentration half-time series over one of the observational stations when the simulation concentrations are multiplied by 100.

If the amount of particles and release points are dropped (to approximately 10 million particles and 20 release points), the error is not encountered. Then the simulation keeps going until the end of the set period, like in figure 29. However, this is at the expense of the ethane concentration levels, which are highly affected by the drop in particles and release points number. This can be appreciated when comparing the average concentration levels in figures 29 and 29. Thus, if a successful simulation is to be conducted, the number of particles nor the release points can be high enough to represent accurately the whole European emissions of ethane and, consequently, not useful as for the validation in this project.
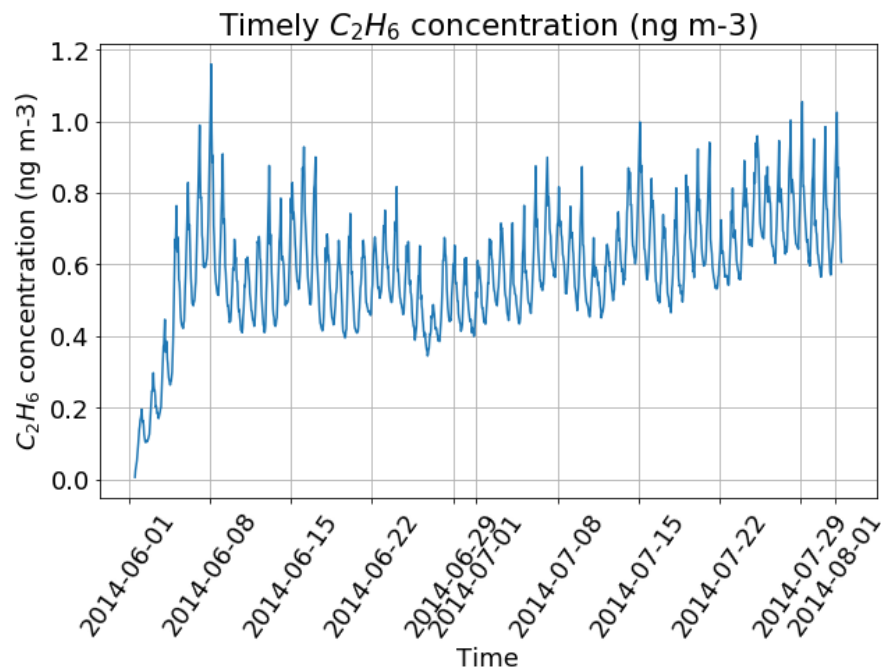
Fig. 29: Average concentration of surface ethane during a simulation with 10 million particles and 20 release points in the European region. If the time period is compared to figure 24, it can be appreciated that in this case the simulation reached the end of the period (01.08.2014), at the price of dropping the ethane concentration in the atmosphere.

Due to time and resources limitations, it was not possible to assess the nature of this error during this project. Nevertheless, it was not been identified in other FLEXPART development branches. Hence, it is most probable a version-related error, associated with the specific adaptation of FLEXPART to the *Aether* HPC environment.