**Geoscientific
Model Development**

Development cycle 2 of the Modular Earth Submodel System (MESSy2)

P. Jöckel[1,*], **A. Kerkweg**[2], **A. Pozzer**[3], **R. Sander**[1], **H. Tost**[1], **H. Riede**[1], **A. Baumgaertner**[1], **S. Gromov**[1], **and B. Kern**[1]

[1]Max-Planck Institute for Chemistry, Air Chemistry Department, P.O. Box 3060, 55020 Mainz, Germany
[2]Institute for Atmospheric Physics, University Mainz, 55128 Mainz, Germany
[3]Cyprus Institute, EEWRC, P.O. Box 27456, 1645 Nicosia, Cyprus
[*]now at: Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82230 Wessling, Germany

Abstract. The Modular Earth Submodel System (MESSy) is an open, multi-institutional project providing a strategy for developing comprehensive Earth System Models (ESMs) with highly flexible complexity. The first version of the MESSy infrastructure and process submodels, mainly focusing on atmospheric chemistry, has been successfully coupled to an atmospheric General Circulation Model (GCM) expanding it into an Atmospheric Chemistry GCM (AC-GCM) for nudged simulations and into a Chemistry Climate Model (CCM) for climate simulations.

Here, we present the second development cycle of MESSy, which comprises (1) an improved and extended infrastructure for the basemodel independent coupling of process-submodels, (2) new, highly valuable diagnostic capabilities for the evaluation with observational data and (3) an improved atmospheric chemistry setup. With the infrastructural changes, we place the headstone for further model extensions from a CCM towards a comprehensive ESM. The new diagnostic submodels will be used for regular re-evaluations of the continuously further developing model system. The updates of the chemistry setup are briefly evaluated.

1 Introduction

The Modular Earth Submodel System (MESSy, Jöckel et al., 2005) defines a strategy for building comprehensive Earth System Models (ESMs) from process based modules, the so-called *submodels*. Technically, MESSy comprises standard interfaces to couple the different components, a simple coding standard and a set of submodels coded accordingly.

The basic idea is to organise the code into 4 different layers: the basemodel layer (BML) ultimately consists only of a central clock and run-time control, currently however, typically of a general circulation model (GCM) or a box model. The basemodel interface layer (BMIL) comprises the basemodel specific implementation of the MESSy infrastructure; it can be regarded as a multiple socket outlet for the communication between the basemodel and the submodels. The submodel interface layer (SMIL) represents the connector of a specific process to the infrastructure (BMIL). And last but not least, the submodel core layer (SMCL) comprises the basemodel independent implementation of a specific process in the Earth System, or of a diagnostic tool of the model system. It can be regarded as an operator using the data provided via its SMIL and providing data back via its SMIL to other submodels and/or the basemodel. The currently available, published MESSy submodels with references are listed in Table 1. More, yet unpublished, submodels are listed on the project page[1].

The MESSy user interface is based on the namelist concept of the Fortran95 standard (ISO/IEC-1539-1). Each submodel is controlled by (at least) two namelists: the CTRL-namelist (for the SMCL) contains all parameters and switches affecting the internal complexity and flow control of a specific submodel, whereas the CPL-namelist (for the SMIL) contains all parameters and switches for the coupling

---

Table 1. List of MESSy submodels. Submodels marked with an asterisk have been used in the evaluation simulation presented in Sect. 9.

submodel	description	reference(s)
generic (infrastructure) submodels:		
BLATHER*	standard output to log-files	Sect. 8
CHANNEL*	memory and meta-data management and data export	Sect. 2
CONSTANTS*	constants shared between submodels	Jöckel et al. (2005), Sect. 1
GRIDTRAFO	grid transformations	Pozzer et al. (2011, in prep. for GMD), Sect. 8
QTIMER*	optimal use of queue limits and run-time diagnostics	Sect. 4
RND	random number generators	Sect. 8
SWITCH* & CONTROL*	switch and call individual submodels	Jöckel et al. (2005)
TIMER*	time control	Sect. 3
TOOLS*	tools shared between submodels	Jöckel et al. (2005), Sect. 1
TRACER*	management of data and meta-data of constituents	Jöckel et al. (2008)
NCREGRID*	rediscretisation on different grids	Jöckel (2006)
process and diagnostic submodels:		
AIRSEA*	air-sea exchange of trace gases	Pozzer et al. (2006)
CLOUD*	ECHAM5 cloud scheme as MESSy submodel	Roeckner et al. (2006, and references therein)
CONVECT*	convection parameterisations	Tost et al. (2006b)
CVTRANS*	convective tracer transport	Tost (2006)
D14CO*	^{14}CO as diagnostic tracer	Sect. 6.2
DRADON*	^{222}Rn as diagnostic tracer	Sect. 6.1
DRYDEP*	dry deposition of trace gases and aerosols	Kerkweg et al. (2006a)
E4CHEM	simplified (fast) stratospheric chemistry	Baumgaertner et al. (2010b)
GMXE	aerosol microphysics and gas aerosol partitioning	Pringle et al. (2010)
H2O*	stratospheric water vapour and feedback	Jöckel et al. (2006)
HETCHEM	heterogeneous reaction rates	Jöckel et al. (2006)
JVAL*	photolysis rates	based on Landgraf and Crutzen (1998)
LNOX*	lightning NO_x production	Tost et al. (2007), Sect. 7.3
M7*	aerosol microphysics	Vignati et al. (2004)
MADE	aerosol microphysics	Lauer et al. (2007)
MECCA1	atmospheric chemistry	Sander et al. (2005)
MECCA1_AERO	aerosol chemistry sub-submodel	Kerkweg et al. (2007)
MECCA*	revised MECCA1 (incl. MECCA_AERO)	Sander et al. (2011, in prep. for GMD), Sect. 7.1
MECCA_KHET*	heterogeneous reaction rates	Sect. 7.1.2
MECCA_TAG/DBL	kinetic chemistry tagging, isotopic tagging	Gromov et al. (2010)
MSBM*	multi-phase stratospheric box model	Sect. 7.2
OFFLEM*	prescribed emissions of trace gases and aerosols	Kerkweg et al. (2006b)
ONLEM*	on-line calculated emissions of trace gases and aerosols	Kerkweg et al. (2006b)
PSC	polar stratospheric clouds	Kirner et al. (2010)
PTRAC	define additional prognostic tracers via namelist	Jöckel et al. (2008)
QBO*	Newtonian relaxation of quasi-biennial oscillation	Giorgetta and Bengtsson (1999); Jöckel et al. (2006)
RAD4ALL*	ECHAM5 radiation scheme as MESSy submodel	Roeckner et al. (2006); Jöckel et al. (2006)
RAD4ALL_FUBRAD*	high-resolution short-wave radiation sub-submodel	Nissen et al. (2007)
SCAV*	scavenging and wet deposition of trace gases and aerosol	Tost et al. (2006a)
SCOUT*	stationary column output	Sect. 5.2
SEDI*	sedimentation of aerosol particles	Kerkweg et al. (2006a)
SORBIT*	sampling along sun-synchronous satellite orbits	Sect. 5.4
SPE	solar proton events parameterisation	Baumgaertner et al. (2010a)
SPACENOX	energetic particle precipitation	Baumgaertner et al. (2009)
S4D*	sampling in 4 dimensions	Sect. 5.3
TNUDGE*	Newtonian relaxation of species as pseudo-emissions	Kerkweg et al. (2006b)
TREXP	tracer release experiments from point sources	Sect. 6.3
TROPOP*	tropopause and other diagnostics	Jöckel et al. (2006)
VISO*	iso-surfaces and maps	Sect. 5.1

of a specific submodel to the basemodel and to other submodels via the MESSy infrastructure (see Jöckel et al., 2005).

The BMIL of MESSy is based on so-called generic (or infrastructure) submodels, where the term *generic* indicates that these submodels are also coded as basemodel indepen-

dent submodels, i.e., organised into an interface and a core layer. Each generic submodel serves a specific, superordinate purpose (see also Jöckel et al., 2005), among them

– a central submodel control interface (SWITCH + CONTROL) for switching the individual submodels on/off and for providing the main entry points for calling the submodels from the basemodel,

– a submodel (TOOLS) providing common subroutines and functions shared between different submodels (e.g., sorting algorithms, interpolation methods),

– a data import interface (NCREGRID) for the automatic re-discretisation of gridded geoscientific data from netCDF[2] files to the actual rectangular (e.g., Gaussian) model grid (Jöckel, 2006).

Another important generic submodel, the data transfer and export interface introduced by Jöckel et al. (2005) comprises mainly three parts:

– a module for the definition of constants (CONSTANTS) shared between the basemodel and/or different submodels,

– a memory and meta-data management interface (TRACER, Jöckel et al., 2008) specialised for the representation of constituents in the different domains of the Earth System, for instance water in different phases in the atmosphere, chemical compounds in the atmosphere and the ocean, and aerosol in the atmosphere,

– a memory and meta-data management interface for the data exchange between the submodels and between the submodels and the basemodel, and for the data export to files.

In development cycle 1 of MESSy (MESSy1), the third part is entirely based on the stream interface (unpublished) of the atmospheric GCM ECHAM5 (Roeckner et al., 2006), which has been extended (see supplementary material of Jöckel et al., 2005) for the ECHAM5/MESSy Atmospheric Chemistry (EMAC) model (Jöckel et al., 2006). Since this stream interface is rather specific for ECHAM5, the application of the MESSy infrastructure was so-far limited to the basemodel ECHAM5. To overcome this limitation and to allow for different basemodels, we present in Sect. 2 a completely new, basemodel independent implementation of a memory and meta-data management and data export interface (named CHANNEL) with much enhanced flexibility (e.g., w.r.t. the output capabilities and control) and modularity following an object oriented approach.

We further extended the MESSy infrastructure by two new generic submodels: TIMER for the time control and QTIMER for the optimal use of scheduler run-time limits and run-time diagnostics. These are described in Sects. 3 and 4, respectively.

In addition, new diagnostic (Sect. 5) and process (Sect. 6) submodels have been developed. And finally, modifications

---

[2]http://www.unidata.ucar.edu/software/netcdf

of the chemistry setup compared to previous versions (Jöckel et al., 2006) have been introduced (Sect. 7) together with minor structural changes (Sect. 8). With the updated model system a comprehensive re-evaluation simulation has been performed (Sect. 9). Throughout the text, the ECHAM5/MESSy Atmospheric Chemistry (EMAC) model based on MESSy1 is denoted as EMAC1 and the one based on MESSy2 as EMAC2, respectively.

2 CHANNEL: a memory and meta-data management, data transfer and export interface

A common task in ESMs is the storage (in memory) and output (to files) of information (data), which represent the state of the simulated system, for instance the temperature of air, the ocean salinity, or the abundances of ozone and water vapour in the atmosphere. During the simulation, this information needs in general to be shared between different processes, thus providing the coupling between them. A complete description of such data comprises information about

– the quantity (meta-data), such as the measuring unit,

– the underlying geometry, e.g., the mathematical representation, the discretisation, and the corresponding dimensions,

– the arrangement of values in memory (array shape),

– the layout in the output file, depending on the file format,

– the parallel decomposition, if the ESM runs in a parallel environment,

– the values.

The generic submodel CHANNEL described here, provides a powerful application programming interface (API) to handle such data for the flexible and efficient data exchange/data sharing between different processes (submodels). It is written in Fortran95 (ISO/IEC-1539-1) following an object-oriented approach to the extent possible. The basic entities, implemented as Fortran95 structures, of CHANNEL are

– *attributes*, representing time independent, scalar characteristics, e.g., the measuring unit,

– *dimension variables*, representing specific coordinate axes, e.g., the latitude in degrees north, the zonal wave number, the trajectory number,

– *dimensions*, representing the basic geometry in one dimension, e.g., the number of latitude points, the number of trajectories,

– *representations*, describing multidimensional geometric structures (based on *dimensions*), e.g., Eulerian (or gridpoint), spectral, Lagrangian,

– *channel objects*, representing data fields including their meta information (*attributes*) and their underlying geometric structure (*representation*), e.g., the 3-D vorticity in spectral representation, the ozone mixing ratio in Eulerian representation, the pressure altitude of trajectories in Lagrangian representation,

– *channels*, representing sets of "related" *channel objects* with additional meta information. The "relation" can be, for instance, the simple fact that the *channel objects* are defined by the same submodel.

CHANNEL further serves the input/output (IO) of data from/into files. The implemented IO features comprise

– a complete control (user interface) via two Fortran95 namelists,

– a powerful restart facility for simulation chains,

– output redirection for tailor-made output files,

– a flexible choice of the output file format, of the output method, of the output precision, of the output frequency, and

– the capability to conduct basic statistical analyses w.r.t. time on-line, i.e., to output in addition (or alternative) to the instantaneous data (i.e., at a specific model time step) the average, standard deviation, minimum, maximum, event counts, and event averages for the output time interval.

The current IO implementation comprises the netCDF output format, either by serial or parallel access (parallel netCDF[3]). Entry points for alternative output formats or methods are provided in the code, which can be expanded easily.

In order to enable the application of the IO features also for *tracers* defined by the generic submodel TRACER (Jöckel et al., 2008), CHANNEL also provides subroutines to elegantly associate the tracer memory with CHANNEL meta-information without additional memory requirements.

A reference manual with more detailed information on CHANNEL is part of the Supplement of this article. The source code of CHANNEL including a simple example application is provided on request.

3   TIMER: a generic submodel for time control

A central part of a comprehensive ESM is the control of the timing information. In development cycle 1 of MESSy (MESSy1), this is entirely based on the timer (unpublished)

of the atmospheric GCM ECHAM5 (Roeckner et al., 2006). To enable the application of the timer also for other basemodels, we extracted the timer relevant code and re-implemented it as the generic MESSy submodel TIMER, by keeping largely its original functionality and namelist syntax. TIMER comprises three SMCL Fortran95 modules, which serve different purposes:

– `messy_main_timer.f90` provides

  – the basic type structure to store date and time information,

  – the basic variables to store date and time information,

  – tools (functions and subroutines) for date and time format conversions, time span calculations, etc.

– `messy_main_timer_manager.f90` provides the internal clock for the model simulation. It manages the time stepping and the date and time information during the simulation.

– `messy_main_timer_event.f90` provides data types and routines to schedule processes at specific (regular) time intervals (so-called events), e.g., to trigger regular output or input.

In addition TIMER comprises the basemodel interface layer (BMIL) module `messy_main_timer_bi.f90`, which establishes the connection to the basemodel, and thus contains the basemodel specific settings. The Supplement contains a detailed documentation of TIMER including all definitions, subroutines and a description of the functionality of the event manager.

4   QTIMER: optimal use of queue limits and run-time diagnostics

Complex simulations with comprehensive GCMs, Chemistry Climate Models (CCMs) or ESMs are usually expensive in terms of the computational effort. These simulations are mostly carried out on high performance computers hosted by specialised computing centres serving a larger number of users. To facilitate an evenhanded distribution of computer resources among the different users, for accounting the used CPU time, and to guarantee an optimal usage of the hardware, computational tasks are usually organised and scheduled by resource management tools, such as for instance IBM Load Leveler[4], Sun Grid Engine[5], Network Queuing System, Load Sharing Facility[6], or Portable Batch System[7].

---

[3]http://www.mcs.anl.gov/parallel-netcdf

[4]http://www.redbooks.ibm.com/abstracts/SG246038.html

[5]http://gridengine.sunsource.net/

[6]http://www.platform.com/Products/platform-lsf

[7]http://www.pbsgridworks.com

```
&CTRL
QTIME  = 24,0,0, ! QUEUE TIME LIMIT (hh,mi,se) (NOTE: 0,0,0 TO SWITCH OFF)
QCLOCK = 'wall', ! QUEUE CLOCK TYPE ('wall','cpu ','user','sys ')
QFRAC  = 0.95    ! USABLE FRACTION OF QUEUE TIME LIMIT
/
&CPL
QMETHOD = 'max'  ! METHOD FOR PARALLEL PROCESSING ('max', 'ave', 'sum')
L_DIAG  = F      ! DIAGNOSTIC OUTPUT TO LOG-FILE ?
/
```

Fig. 1. The CTRL and CPL namelists of QTIMER in `qtimer.nml`. QTIME denotes the maximum available scheduler time specified as hours, minutes, seconds. If all three are zero, QTIMER is switched off. QCLOCK denotes the time to be measured for the restart trigger, which is either the wall-clock time ('wall'), the CPU time ('cpu'), the user time ('user'), or the system time ('sys'). QFRAC is the fraction of QTIME to be actually consumed by the executable. The rest is left to the finalising phase of the model and to outside tasks, such as for instance cleaning up within the wrapper script which called the executable, etc. In parallel environments, the time on each CPU is measured separately, however, the restart needs to be triggered for all parallel processes simultaneously. Depending on the accounting policy of the computing centre, the times are averaged ('ave') or summed ('sum') over all CPUs, or the maximum ('max') over all CPUs is used (specified by QMETHOD). The L_DIAG switch is to turn on (T) diagnostic summary output in every time step to standard output. The default is to avoid this output (F).

In most cases, these systems are configured to dedicate and account specific resources (number of nodes, number of CPUs, CPU time, memory, etc.) to the tasks. The user has to request the required resources and submit a task (job) to the queue manager. The resource limitations configured by the high performance computing centre, in particular the maximum allowed CPU time per task, is usually not sufficient to perform a comprehensive or long term climate simulation. Therefore, the simulation needs to be split into pieces, so-called chain elements of a simulation chain, which are processed in sequence. The prerequisite for this method is that the model is able to dump its complete state (in full numeric precision) to files on disk, from which the simulation can be unambiguously continued. Such a facility is provided by CHANNEL (see Sect. 2).

Moreover, such a file dump and simulation interrupt needs to be triggered from within the model itself. This can for instance be achieved after a fixed number of model time steps, or by defining an event in the generic submodel TIMER (see Sect. 3) after a specific simulation time interval. The drawback of both methods is that they do not automatically allow an optimal usage of the maximum reserved CPU time dedicated by the job scheduler. The user has to estimate the required CPU time for the given model configuration and setup and set the trigger manually. This method bears the risk that either the trigger comes too late and the scheduler terminates the job before the restart files are dumped, or that the trigger comes too early giving away valuable CPU time already reserved by the scheduler and thus decreasing the turn-around by unnecessary increasing the waiting times (in the queue) between the sequenced tasks.

The generic submodel QTIMER measures on-line the time consumed by the simulation (accumulated for each model time step) and triggers the restart just before the maximum time reserved by the scheduler is reached. The corresponding queue time limit needs to be specified in the CTRL namelist of QTIMER (see Fig. 1).

As added value, QTIMER provides a performance monitoring tool, since the time consumed by each model time step is measured on-line. The resulting timing information is stored in *channel objects* of the *channel* "qtimer" (see Sect. 2) and can therefore elegantly be output, e.g., to netCDF files. An example is shown in Fig. 2.

5 New diagnostic submodels

The evaluation of the simulation output of comprehensive ESMs usually requires additional post-processing steps (statistics and visualisation). The output data for this post-processing, i.e., the model output, is limited, however, by the available disk-storage and by the performance decrease caused by input/output latencies of the disk access. For both reasons, it is normally not feasible to output the entire model state in every model time step. Depending on the scientific task, output of the 3-D model state variables is triggered in regular intervals (5-hourly, 6-hourly, 12-hourly, etc.). This method, however, has several disadvantages. In case the output frequency is an integer divisor of 24 h (such as 6 h), the model output at a specific geographic location is always for the same local times (more precisely for the same mean solar times, e.g., 00:00, 06:00, 12:00, 18:00 at 0° E). Averaging such output (e.g., monthly) can cause unintended localised biases, since the same mean solar time implies the same solar zenith angle. Such biases occur for instance for photochemically active, short lived species and are in general not desirable. Another drawback of the simple output scheme is a twofold relative disproportion of the spatio-temporal data density: whereas data in "smooth" regions (i.e., with small gradients, e.g., in remote areas) is comparably dense, data in regions with large gradients is comparably sparse. Other
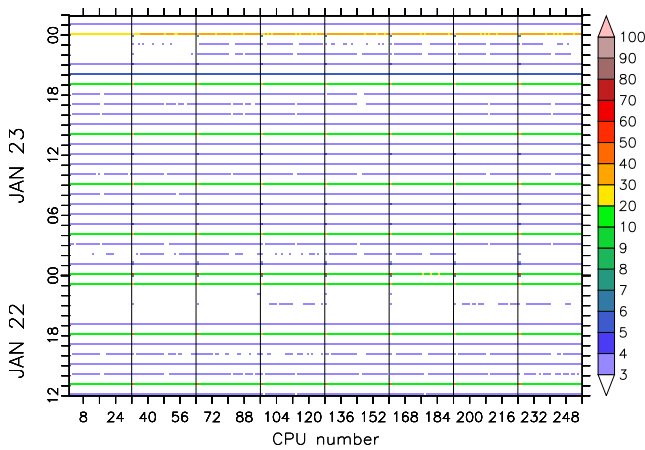
Fig. 2. Example of QTIMER output: the figure shows the CPU time (in seconds) of a T42L90MA EMAC2 simulation with a comprehensive chemistry setup and additional diagnostic netCDF output on an IBM Power6 using 8 nodes à 32 CPUs (the node boundaries are indicated by the vertical black lines). The netCDF output in this case is serial, i.e., the data is gathered and output to files by the first CPU, only. The vertical axis shows the simulation time from 22 January (12:00 UTC) to 24 January (02:00 UTC) of an arbitrary year. The time step of the model simulation is 12 min. Model time steps without output require less than 3 s (white), whereas the serial netCDF output increases the CPU time considerably: in the example, the output of the vertical column of various data fields at 86 locations (with the submodel SCOUT, see Sect. 5.2) every hour increases the CPU time to up to 4 s (light blue). The full, 3-D (standard) output every 5 h requires up to 20 s (green), also the additional output along 12 different polar satellite orbits (with submodel SORBIT, see Sect. 5.4) every 24 h at 00:00 UTC. If the 5-hourly 3-D output and the SORBIT output coincide (e.g., on 24 January, 00:00 UTC in the example) the required CPU times add up to a total of 30 to 40 s (orange). The first CPU on each node requires a factor of 2 to 3.5 increased CPU time for the serial netCDF output, since all the communication between the (shared memory) nodes is executed here. All numbers strongly depend on the amount of output data.

criteria for this "sensible" data density are regions of special interest or points in space and time with a high availability of observational data for model evaluation; more specific, high frequency stationary observations, measurement campaigns, and in particular satellite observations. In many cases, it is therefore desirable to increase the output data frequency for specific variables at specific locations, or to provide tailor-made output appropriate for very specific purposes (e.g., to improve the statistics); in any case, however, without increasing the amount of output data much further beyond the standard output data.

In addition to the statistical (w.r.t. time) diagnostic capabilities of the generic submodel CHANNEL (see Sect. 2), we therefore implemented four diagnostic output submodels, which are all utilising the generic submodel CHANNEL.

5.1 VISO: iso-surfaces and maps

The first diagnostic submodel, VISO, serves two purposes. First, it is used to diagnose vertically layered, 2-D iso-surfaces in 3-D scalar fields in Eulerian (grid-point) representation (see Sect. 2). The search algorithm determines in every vertical column of the field the level index of the box with the specified value and, by linear interpolation, the fraction of the vertical box length below the specified value. The submodel is entirely controlled via its CPL namelist (Fig. 3), and an iso-surface is defined by (values in parentheses correspond to the first example in Fig. 3)

- the keyword ISO with an arbitrary but unique number in parentheses (ISO(1)),

- a unique name of the iso-surface (isent340),

- the name of the *channel* containing the 3-D scalar field (PHYS),

- the name of the *channel object* representing the 3-D scalar field (tpot),

- the value of the iso-surface in units of the 3-D scalar field (340.0),

- a switch to calculate the level index only (F), or the level index and the vertical box fraction below the iso-value (T),

- a switch to search from the top to the lowest layer (F) or from the lowest to the top layer (T),

- the number of levels to skip for the search from the top layer (default is zero), and

- the number of levels to skip for the search from the lowest layer (default is zero).

In the example, isentropes of 340 (ISO(1)), 380 (ISO(2)) and 420 K (ISO(3)) are defined, and further surfaces of constant potential vorticity at 2 (ISO(4)) and 3.5 PVU (ISO(5)). For PV2 the search is from the top layer down, skipping the 4 lowest model layers, and for PV3.5 the search is reversed, skipping the 3 lowest layers.

The second application of VISO is for mapping 3-D scalar fields in Eulerian (grid-point) representation on surfaces defined by a level index (and optionally by a fraction of the box), as for instance an iso-surface defined by the same submodel. Such a map is defined in the CPL namelist by (values in parentheses correspond to the first example in Fig. 3)

- the keyword MAP with an arbitrary but unique number in parentheses (MAP(1)),

- a unique name of the map (pth340),

- the name of the *channel* containing the surface (viso),

```
&CPL
!####################################
!# A. DEFINITION OF ISO-SURFACES
!####################################
!# NOTES:
!# - channel objects <name>_i, and <name_f> are added for
!#   'index' and 'fraction below'
!# SYNTAX:
!#     ISO-SURFACE name, channel, object, iso-value, index + fraction ?,
!#        reverse search ?, skip levels from top, skip levels from sfc
!# ISENTROPES (THETA=CONST.)
ISO(1) = 'isent340',  'PHYS',  'tpot',    340.0, T, T,  ,  ,
ISO(2) = 'isent380',  'PHYS',  'tpot',    380.0, T, T,  ,  ,
ISO(3) = 'isent420',  'PHYS',  'tpot',    420.0, T, T,  ,  ,
!# CONST. POTENTIAL VORTICITY
ISO(4) = 'PV2',         'tropop',  'PV',        2.0, T, T, 1, 4,
ISO(5) = 'PV3.5',       'tropop',  'PV',        3.5, T, F, 1, 3,
!#
!####################################
!# B. FIELDS MAPPED TO (ISO-)SURFACES
!####################################
!# NOTES:
!# - '_i' and '_f' are internally appended to SURFACE(object) name
!#   for 'index' and 'fraction below'; availability of '_f' determines
!#   the mapping method
!# SYNTAX:
!#     MAP name, ISO-SURFACE(channel), ISO-SURFACE(object),
!#             FIELD(channel), FIELD(object)
!#
!# PRESSURE ON ISENTROPES
MAP(1)  = 'pth340',   'viso',  'isent340', 'PHYS', 'press',
MAP(2)  = 'pth380',   'viso',  'isent380', 'PHYS', 'press',
MAP(3)  = 'pth420',   'viso',  'isent420', 'PHYS', 'press',
!#
!# PRESSURE AND TEMPERATURE AT TROPOPAUSE
MAP(10) = 'ptp',      'tropop', 'tp',     'PHYS', 'press',
MAP(11) = 'ttp',      'tropop', 'tp',     'PHYS', 'tm1',
!#
/
```

Fig. 3. Example CPL namelist of VISO in `viso.nml`.

- the name of the *channel object* representing the surface (`isent340`),

- the name of the *channel* containing the 3-D scalar field in Eulerian representation (`PHYS`), and

- the name of the *channel object* representing the 3-D scalar field in Eulerian representation (`press`).

In the examples the pressure altitude of the isentropes of 340 (`MAP(1)`), 380 (`MAP(2)`) and 420 K (`MAP(3)`) are defined, and further the pressure (`MAP(10)`) and temperature (`MAP(11)`) at the tropopause. The tropopause information (vertical level index and fraction of the tropopause box below the tropopause) are provided as *channel objects* `tp_i` and `tp_f`, respectively, in *channel* `tropop` by the submodel

TROPOP (see Jöckel et al., 2006). The suffixes `_i` and `_f` for level index and fraction are automatically appended internally, and the presence of the *channel object* for the fraction determines the mapping algorithm. If the fraction is present, the values are linearly interpolated in vertical direction, if the fraction is not defined, the value at the level index of the corresponding surface is selected.

The iso-surfaces and maps of VISO are defined as 2-D *channel objects* in Eulerian representation in the *channel* `viso`. For the index and fraction of an iso-surface, the corresponding channel object names are internally generated by appending `_i` and `_f`, respectively, to the iso-surface name specified in the CPL namelist. For the maps, the name specified in the namelist is used directly. Operating with channel objects enables automatically all namelist controllable
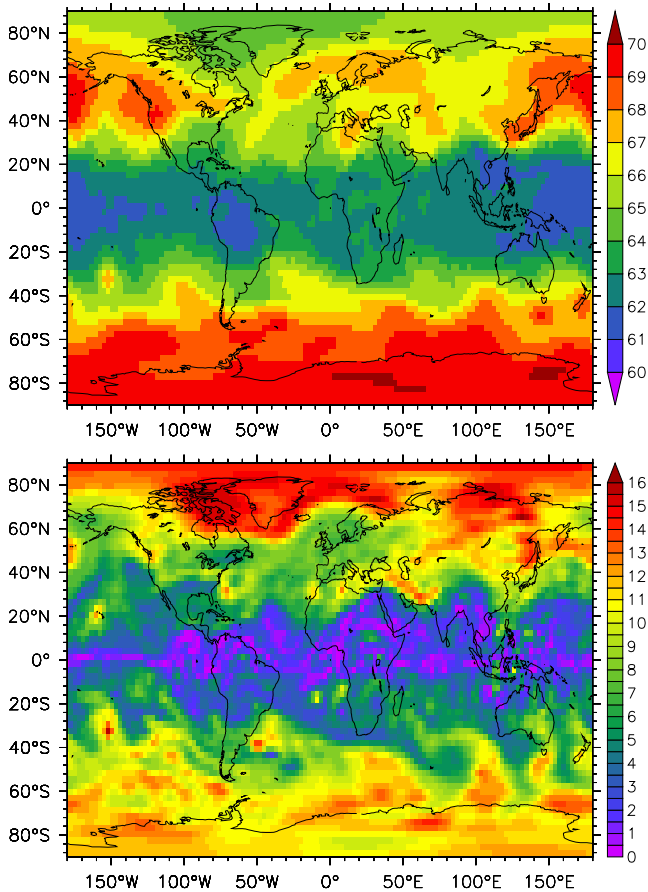
Fig. 4. Example of VISO output: the upper panel shows the level index (counted from the top of the atmosphere to the surface) of the 380 K isentrope from a T42L90MA EMAC2 simulation on 1 January 2005 at 02:00 UTC. The lower panel shows the potential vorticity (absolute value in PVU) on the 380 K isentrope at the same time.

output features for the iso-surface and map information of VISO (see Sect. 2), such as for instance time averaging, output redirection, choice of the output frequency, etc. Figure 4 shows typical examples of VISO applications.

With the on-line diagnosis of iso-surfaces and maps with VISO, grid-point data are reduced in one (the vertical) dimension. This data reduction can be exploited to increase the output frequency (i.e., the time resolution) without blowing up the storage requests. Obtaining the same information via post-processing of the standard output requires the output of 3-D data (one field for iso-surfaces, two fields for a map) with the desired output frequency, instead.

5.2 SCOUT: stationary column output

A proper evaluation of ESMs, CCMs and GCMs requires a thorough comparison of the model results with observations. Large databases of observations at stationary, ground-based observatories exist for meteorological data and also for the

chemical composition of the atmosphere. These observations often comprise also vertical information obtained for instance with radar, lidar, balloon sondes, and other techniques. Moreover, specifically in case of continuous or quasi-continuous measurement techniques, the time resolution of these observations is usually much higher than the typical data output frequency of a 3-D, global (or regional) model simulation. The standard methods for the comparison of model results with such observations, is either the time averaging of the observations or a time interpolation of the model output. In both cases, valuable information about the variability is lost. To overcome this limitation with the classical, off-line post-processing approach, requires the output of large amounts of 3-D data implying the need for large storage capacities and slowing down the model simulation considerably, due to the comparably slow input/output.

To enable the high-frequency output of data from the model at the position of stationary observatories on-line, i.e., from within the model simulation, we implemented the submodel SCOUT.

The key algorithm is a function (`locate_in_decomp`)

$$(x, y) \rightarrow (p, i, j) \tag{1}$$

to calculate the process (or CPU) identifier (p) and the array indices (i and j) of the corresponding grid-box (i.e., the nearest grid-point) from the geographic longitude (x) and latitude (y) in a parallel environment. This function depends on the *representation* of the *channel object*, in particular on its parallel decomposition in distributed memory and on the memory layout of the corresponding data arrays on each process (or CPU). Since this function serves a general purpose and is also used by other submodels (e.g., S4D, see Sect. 5.3), it is part of the *generic submodel* TRANSFORM (see Sect. 8). The submodel SCOUT is entirely controlled by its CPL namelist as shown in Fig. 5.

For 3-D fields (i.e., with a vertical dimension) the vertical column is sampled, for 2-D, horizontally oriented fields, such as for instance the tropopause pressure or temperature, or iso-surfaces and maps from VISO (see Sect. 5.1), the corresponding scalar value is sampled and output. Internally, the sampled data are stored as *channel objects* in column (for 3-D fields) or scalar (for 2-D fields) representation, respectively. Operating with channel objects enables automatically all namelist controllable output features for sampled data (see Sect. 2), such as for instance time averaging, output redirection, choice of the output frequency, etc. With SCOUT, high frequency output of model data for comparison with stationary observations can be provided, examples are shown in Fig. 6.

5.3 S4D: sampling in 4 dimensions

In addition to measurement data from ground based, stationary observatories, data from moving platforms like aircraft, ships, and trains are well suited for the model evaluation.

```
&CPL
!# NAME, LATITUDE, LONGITUDE, LIST OF CHANNEL OBJECTS
!# NOTES:
!#   - NAME <= 5 CHARACTERS
!#   - SYNTAX FOR CHANNEL OBJECT LIST:
!#       "channel:object,object,object;channel:object;"
!#       (in object-names wildcards (*,?) can be used)
LOC(1) = 'MAINZ', 49.98, 8.23, "tracer_gp:Rn,CO,H2O;tropop:tp*,PV;",
/
```

Fig. 5. Example CPL namelist of SCOUT in `scout.nml`: up to 500 locations of observatories can be defined with the keyword `LOC` and an arbitrary, but unique number (between 1 and 500). The definition consists of an up to 5 character long, unique name (`MAINZ`), followed by the latitude (`49.98`) in degrees north, the longitude (`8.23`) in degrees east (between −180 and 360), and a string with the list of *channel objects* in Eulerian (grid-point) representation to be sampled. This list comprises semicolon-separated blocks, one for each *channel* starting with the name of the *channel* and followed by colon and a comma-separated list of *channel objects*. If a channel object name contains wildcards (*,?), all channel objects with matching names are selected. In the example, the objects Rn, CO and H2O from the channel `tracer_gp` are sampled, and the objects PV and those starting with tp from the `tropop` channel.

This kind of observational data stem either from dedicated campaigns targeting on specific scientific questions, but also from regular or quasi regular investigations (e.g., CARIBIC[8], Brenninkmeijer et al., 2007; or MOZAIC[9], Marenco et al., 1998). A direct comparison of observations from moving platforms with model simulation results is challenging in many aspects. The common approach is to sample the 3-D model output off-line (i.e., as a post-processing step) to the position and time of the moving platform. In addition to the time resolution (similar to stationary observatories), for moving platforms also the spatial resolution of the data is usually much finer than that of the model output. This implies that the model output needs to be interpolated in space and time. In particular along the time dimension a lot of information is lost, since usually the data output frequency is much lower than the model time stepping frequency.

In order to retrieve the maximum information out of the model simulation for comparison with observations from moving platforms, we implemented the submodel S4D, which interpolates the model data to the platform track on-line, i.e., during the model simulation. The platforms and the requested data are specified in the CPL-namelist (see Fig. 7) by

– the keyword TRACK followed by an arbitrary, but unique number (between 1 and 100) in parentheses,

– an up to 8 character long, unique name,

– a string with the path and file name base to the ASCII file containing the track positions,

– a switch indicating how the track position files are partitioned (either monthly (1), or daily (0)), with −1 the track is deactivated,

– a switch to produce either time series output along the track of the complete model column (T), or to perform a vertical interpolation onto the track (F),

– a switch to output all model time steps along the track (T), or only those dates and times listed in the position file (F); the latter potentially causes non-equidistant time intervals in the output files; this switch has only an effect, if the the frequency of the track position information is lower than the model time stepping frequency,

– a fill value used for indicating missing data, in case the previous switch is T and the model time step is shorter than the distance between the track position way points,

– a list of *channel objects* to be sampled along the track; the list needs to be specified as a semicolon-separated list of channel names, each followed by a colon and a comma-separated list of channel object names; channel objects can be summarised by using wildcards (* and ?).

The ASCII track position files must contain 9 columns with year, month, day, hour, minute, second (UTC), geographic longitude (in degrees east), geographic latitude (in degrees north) and pressure altitude (in hPa). If the specified pressure altitude at a given horizontal position (longitude and latitude) is out of the range of the hybrid pressure grid at this position (and time), the value at the nearest vertical model boundary (either lowest or uppermost layer, respectively) is sampled.

For each defined track, S4D produces an additional output *channel* named `s4d_` followed by the track name specified in the namelist (second item in the list above). The sampled *channel objects* are in scalar or column representation, depending on the 5th item in the namelist above and on the object to be sampled (2- or 3-D grid-point representation). The *channel objects* in the new S4D *channel* are named as the original *channel*, followed by an underscore and the name of the original *channel object*.

[8] http://www.caribic-atmospheric.com
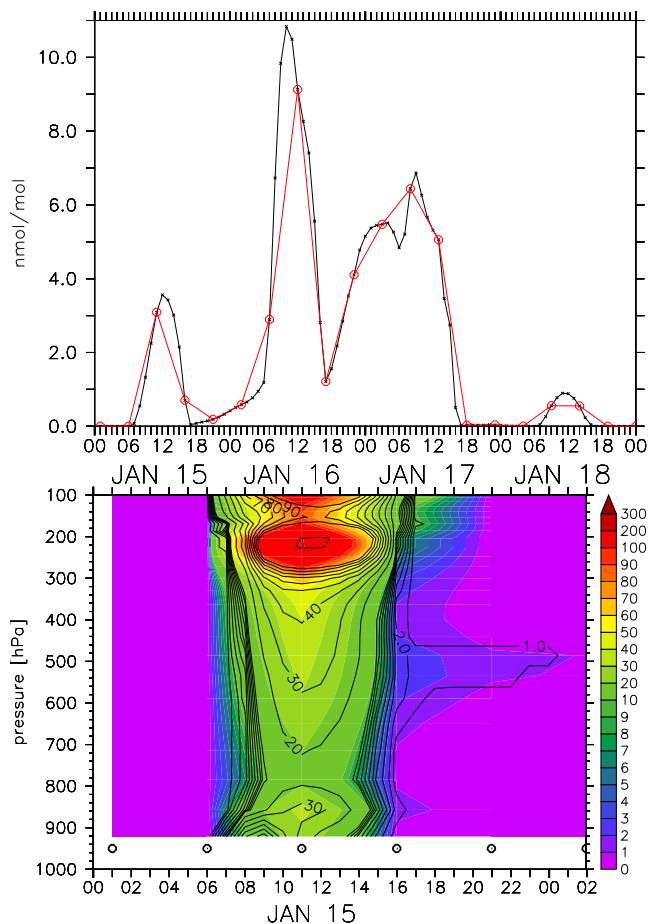[9] http://mozaic.aero.obs-mip.fr

Fig. 6. Example of SCOUT output: the upper panel shows the simulated near surface NO mixing ratio (in nmol/mol) at 49.98° N and 8.23° E for 4 arbitrary days. The red line and symbols depict the off-line interpolated data based on the 5-hourly, 3-D output; the black line and symbols show hourly sampled output with SCOUT. The latter shows – as to be expected – more details, the maximum on 16 January is for instance clearly underestimated by the off-line post-processing method. The lower panel shows the OH mixing ratio (in 10^{-15} mol/mol) versus pressure altitude at the same location for the first day. The colour shaded area is for the time series sampled hourly with SCOUT, the contour lines (with the same levels) are based on the 5-hourly 3-D output. The black symbols (at 950 hPa) denote the interpolation points of the 5-hourly output.

For the horizontal interpolation, the subroutine `locate_in_decomp` already introduced for the submodel SCOUT (see Sect. 5.2, Eq. 1) is overloaded

$$(x, y) \rightarrow (p, i, j)_n \tag{2}$$

to return the grid point index triples of the $n \in \{1, 2, 3, 4\}$ surrounding grid-points, instead of the index triple of the nearest neighbouring grid-point, only. This information is used for a horizontal bi-linear interpolation to the track position. The vertical interpolation, if requested, is a linear interpolation in pressure altitude.

For each track, S4D checks in every model time step, if the time span of the track lies in the future or in the past. In both cases, the track is not active and no calculations are performed. In case the current model time step falls within the time interval of the track, however, S4D searches the position of the track, of which the date and time information lies closest to $T + \Delta t/2$, where T is the current model date and time and Δt is the model time step length. At this position the horizonal (and, if requested vertical) interpolation(s) are performed and the results are stored in the corresponding S4D *channel object*. For S4D *channels*, output is automatically triggered for every model time step in which an interpolation took place (or in other words the track was active). This guarantees the highest possible output frequency along the tracks, which implies that no information from the model along the track is lost. The data volume added by this tailor-made output along a track is negligible compared to the regular output of 2- and 3-D data fields. Note that for retrieving the same information off-line (through post-processing) all desired 2- and 3-D fields must be output at every model time step first.

Figure 8 shows an example of the added value achieved with the application of S4D. The upper left panel shows the OH mixing ratio along a flight path sampled from an EMAC2 simulation with different techniques. The on-line sampling with S4D (red boxes) delivers the maximum information available, i.e., with one value every model time step (12 min in the example). Off-line sampling from 5-hourly, 3-D model output without interpolation in space and time (blue line) clearly shows the grid-box structure (due to the nearest neighbour selection at the time dependent aircraft position). A severe sampling artefact (of 2 orders of magnitude absolute error) is clearly visible on 9 September, between 00:30 and 01:00 UTC, i.e., right in the middle between two output time steps. The reason for this artefact is explained by the lower right panel of Fig. 8: for 9 September, 00:45 UTC, the nearest available output time in the 5-hourly 3-D output data is for 9 September, 03:00 UTC. The aircraft is at 00:45 UTC at a position (blue circle), where it is still night. At the same position at 03:00 UTC, however, (as shown in the figure) the sun is rising. The availability of sunlight is directly reflected in the abundance of the photochemically produced OH radical. Linear interpolation in time even drastically increases the problem with the sampling artefact (black line in the upper left panel of Fig. 8), since by linear interpolation in time, the sunlit region is artificially broadened (middle left and lower left panel in Fig. 8).

Besides such off-line sampling artefacts resulting from day and night mismatches for photochemically active species and photolysis rates, similar sampling artefacts can arise for all quantities with fast varying gradients and/or smaller-scale features, such as streamers, tropopause folds, emission plumes, clouds, etc. With the classical off-line sampling approach such phenomena can either be overlooked, if they occur in between the model output time steps, or overstated, if they are present in the model output time step, but

```
&CPL
!# SYNTAX: NAME, TRACK-DATA FILE BASE, UPDATE-SWITCH, COLUMN OUTPUT ?,
!#         OUTPUT ALL MODEL TIME STEPS ALONG TRACK, FILL VALUE,
!#         LIST OF CHANNEL OBJECTS
!#
!# NOTES:
!#   - NAME <= 8 CHARACTERS
!#   - UPDATE SWITCH: -1: NEVER (SWITCHED OFF)
!#                     0: DAILY
!#                     1: MONTHLY
!#   - TRACK-DATA FILE NAMES
!#              <path>/<prefix><YYYY><MM><DD>.pos  (daily)   -> 0 !
!#              <path>/<prefix><YYYY><MM>.pos      (monthly) -> 1 !
!#   - SYNTAX FOR CHANNEL OBJECT LIST:
!#        "channel:object,object,object;channel:object;"
!#        (in object-names wildcards (*,?) can be used)
!#
!#   - THE TRACK-DATA FILES MUST CONTAIN:
!#      year month day hour minute second longitude latitude pressure [hPa]
!#
TRACK(1) = 'TEST_D', '$INPUTDIR_MESSY/s4d/misc/test_', 0, T, T, -1.E+34,
           "tracer_gp:Rn,14CO,H2O;tropop:PV;PHYS:press;g3b:aps;",
TRACK(2) = 'TEST_M', '$INPUTDIR_MESSY/s4d/misc/test_', 1, F, T, -1.E+34,
           "tracer_gp:Rn,14CO,H2O;tropop:tp*;PHYS:pressi,press;g3b:aps;",
/
```

Fig. 7. Example namelist of the submodel S4D: For `TRACK(1)` S4D looks at every first model time step of a new day (0) for the presence of a file named `test_YYYYMMDD.pos`, in the path `$INPUTDIR_MESSY/s4d/misc`, where YYYY, MM and DD are replaced by the current model year, time and day, respectively. For `TRACK(2)`, the position files are split into monthly files (1), and S4D searches likewise for the position file `test_YYYYMM.pos` at the first time step every new month in the same path. If the respective files are not present, a warning is output and no data is sampled. S4D creates the new *channels* `S4D_TEST_D` (`TRACK(1)`) and `S4D_TEST_M` (`TRACK(2)`), respectively. Along `TRACK(1)` the complete model column is sampled every model time step (T,T), whereas along `TRACK(2)` the requested 3-D channel objects are vertically interpolated to the actual pressure altitude position in each model time step (F,T). Since every model time step is sampled on both tracks, the missing value (–1.E+34) is meaningless in the given examples. Along `TRACK(2)` all objects which name starts with `tp` from channel `tropop` are sampled (`tropop:tp*;`), whereas along `TRACK(1)` only the object `PV` of channel `tropop` is sampled. The latter results in the channel object `tropop_PV` in channel S4D_TEST_D.

not anymore a few steps later. An a-posteriori correction of such artefacts is impossible, since the required information is lost. The solutions are either to increase the model output frequency (which drastically increases the storage demands), or to perform the sampling on-line, as done with S4D.

5.4 SORBIT: sampling along sun-synchronous satellite orbits

A fast growing source of geoscientific data for model evaluation is emerging from remote sensing instruments on satellites. A specific class of satellites is defined by the sun-synchronous orbiters. Their orbit is nearly polar and their orbit inclination (cf. Fig. 9) and altitude is chosen such that the gravitational force gradient resulting from the Earth's oblateness causes a precession rate of the orbital plane (with respect to the celestial sphere) of one full circle per year. As a result, any given point of the Earth's surface is passed by the satellite at the same local mean solar time. This implies constant light conditions, which is favourable for remote sensing in-

strumentation relying on sunlight. The local (index L) time $T_{L,O}$ (hour of day) of the orbiter's (index O) flyover at a given latitude θ is, as derived from the spherical rectangular triangle (see Fig. 9 and Eq. 3.203g in Bronstein et al., 2005):

$$T_{L,O}(\theta) = \left(T_{L,O}(0) \pm \arcsin\left(\frac{\tan\theta}{\tan\delta}\right)\frac{12}{\pi} + 48 \right) \bmod 24. \quad (3)$$

$T_{L,O}(0)$ is the equator crossing local time, δ is the inclination of the orbital plane, and the sign is positive for the ascending and negative for the descending parts of the orbit, respectively (see also Leroy, 2001). This simple relation for the orbit geometry of sun-synchronous orbiters nicely allows the tailor-made on-line sampling of data from a model for direct comparison with the retrieved satellite observations without the requirement of knowing the actual position of the satellite at a given time. For a given scalar variable X in grid-point representation, a second variable X_O is defined as
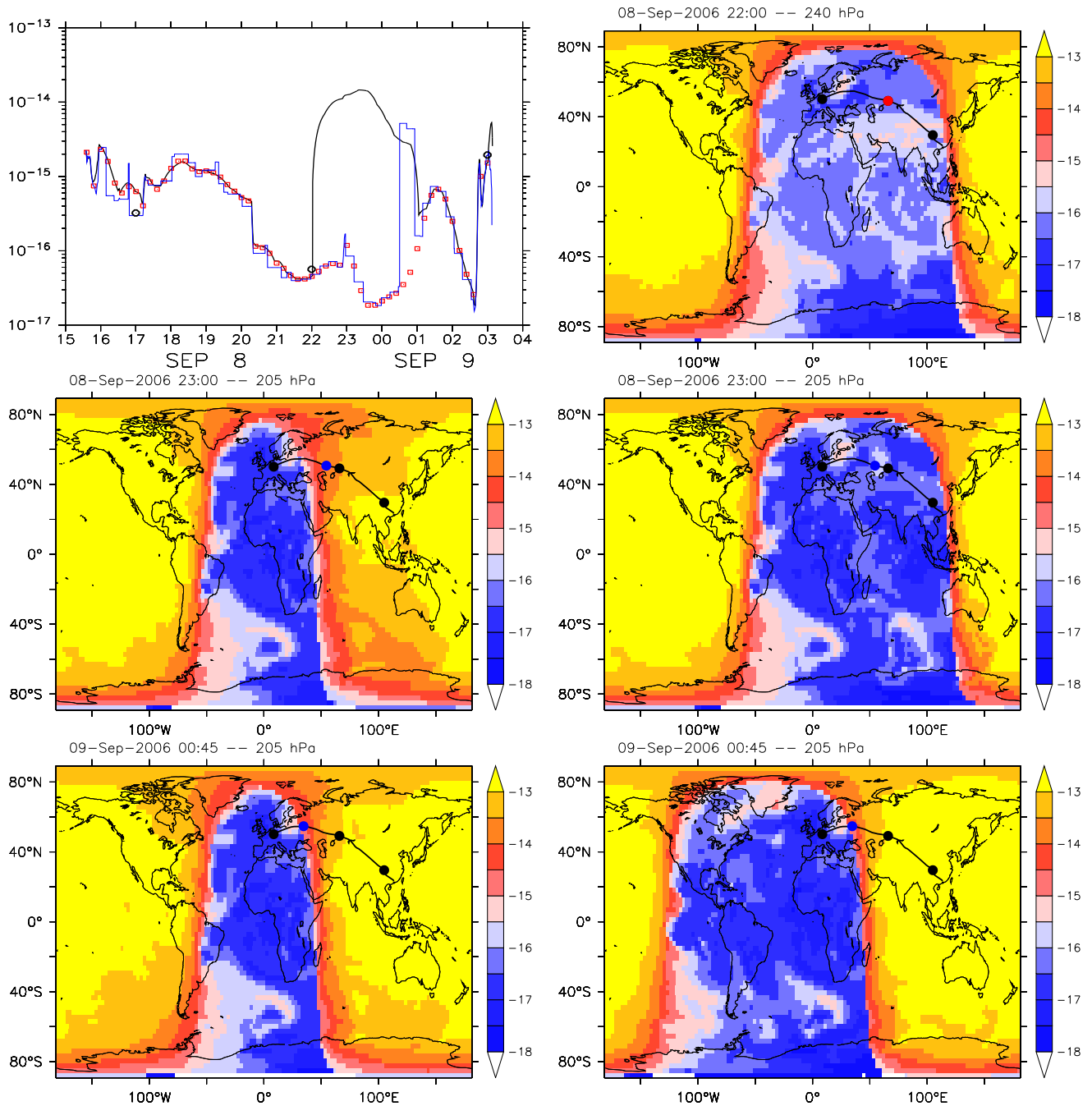
Fig. 8. The upper left panel shows the OH mixing ratio (in mol/mol, simulated with EMAC2) along a flight path of the CARIBIC aircraft on 8 and 9 September 2006 from China to Germany. The red boxes denote the results of the S4D on-line sampling (with a model time step of 12 min), the black line results from the spatio-temporal off-line interpolation based on the 5-hourly standard 3-D model output, and the blue line is the result of the off-line sampling with neither spatial nor temporal interpolation (i.e., the nearest neighbour is used). The black circles indicate the 5-hourly standard model output time steps at 17:00, 22:00, and 03:00 UTC, respectively. The other panels show the global log(OH/(mol/mol)) at the time and corresponding flight level (in hPa) of the aircraft, as labeled at the top of each panel. The black lines depict the flight path of the aircraft, filled circles highlight the aircraft position at the 5-hourly model output time steps (black) and at the time labeled at the top of the panel (blue). Filled circles in red indicate that time of aircraft position and 5-hourly model output time step coincide. Panels on the left side show the off-line interpolated results, those on the right side the nearest neighbours of the 5-hourly output time steps, i.e., without interpolation in space and time. The corresponding output times are (right column) 8 September, 22:00 UTC (exact match, upper), 8 September 22:00 UTC (middle) and 9 September, 03:00 UTC (lower).
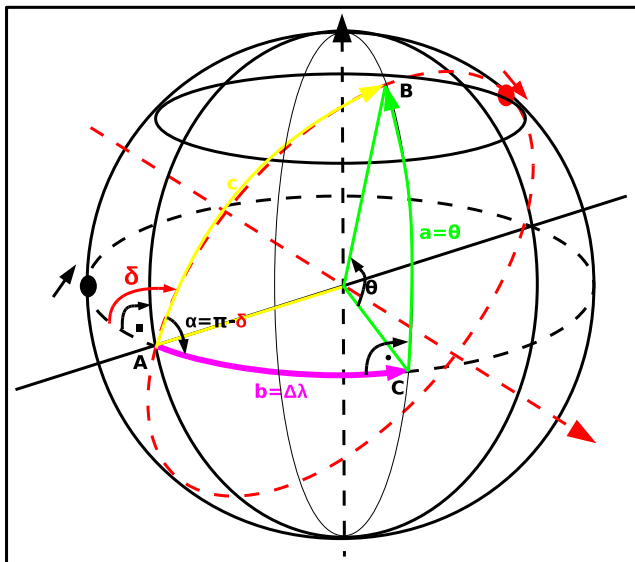
Fig. 9. Orbit geometry of a sun-synchronously orbiting satellite: the black and the red dot indicate arbitrary positions of an equator-following and a sun-synchronous orbiter, respectively. The corresponding arrows indicate the flight direction. The orbital plane of the sun-synchronous orbiter (dashed red ellipse) is rotated from the equatorial plane (dashed black line) by the inclination δ. The points A, B and C form a rectangular spherical triangle with $a = \theta$ being the latitudinal distance (green), $b = \Delta\lambda$ the longitudinal distance (pink) and c the orbital distance (yellow) of an arbitrary satellite position from the equatorial intersection (A).

$$X_O(i,j,k,l) =$$
$$\begin{cases} X(i,j,k,l) & \text{if } |T_L(i,j,l) - T_{L,O}(\theta(j))| \leq \Delta T \\ X_U & \text{otherwise,} \end{cases} \quad (4)$$

where i, j and k are the grid-box indices in longitudinal, latitudinal and vertical direction, respectively, l is the time step of the model, and T_L is the local solar time (hours of day) in the corresponding grid-box (with indices i and j) at model time step l:

$$T_L(i,j,l) = T_{UTC}(l) + \frac{\lambda(i,j)}{360} 24. \quad (5)$$

λ is the geographical longitude (in degrees) and $T_{UTC}(l)$ the model time (UTC) at time step l. Furthermore, X_U in Eq. (4) indicates an undefined value, and due to the discrete grid, a time interval ΔT is required. A natural choice is half the model time step length, i.e., $\Delta T = \Delta t/2$, or the width of the grid-box expressed in units of time, e.g., $\Delta T = (86\,400/N_x)/2$ (in seconds), where N_x is the number of longitude points of the grid.

During the model integration X_O is successively filled at each model time step l. After 24 h, the (latitude dependent) local solar time of the orbiter ($T_{L,O}$) was reached in every

grid-box of the model. The resulting global field X_O is output and re-initialised by X_U for the next 24 h.

The implementation of this algorithm in SORBIT is, as the other diagnostic submodel presented here, utilising the capabilities of the channel interface (Sect. 2) and entirely controllable by the CPL-namelist (see Fig. 10). Two entries are used to control the output interval (`lout_auto`) and the undefined value X_U (`r_init`). For `lout_auto = T` the output is triggered every 24 h (i.e., if the X_O field is full). The option `lout_auto = F` is only useful for testing. In that case, the output frequency is specified in the CPL-namelist of the channel interface (Sect. 2). The orbit definitions consist of

- the keyword `ORB` followed by an arbitrary, but unique number (between 1 and 50) in parentheses,

- the name of the orbit (up to 8 characters),

- a flag for calculating the orbiter's overflight local solar time $T_{L,O}(\theta)$ according to Eq. (3), i.e., latitude dependent (`T`), or for using the equator crossing local time $T_{L,O}(0)$ at all latitudes (`F`),

- the inclination δ of the orbital plane,

- a flag to select either the ascending (`T`) or descending (`F`) part of the orbit,

- the hour of the equator crossing local solar time,

- the minute of the equator crossing local solar time,

- a switch for selecting $\Delta T = \Delta t/2$ (`F`) or, more restrictive (`T`), $\Delta T = \min(\Delta t, 86\,400/N_x)/2$ (see Eq. 4), and

- a list of *channel objects* to be sampled along the orbit; the list needs to be specified as a semicolon-separated list of channel names, each followed by a colon and a comma-separated list of channel object names; channel objects can be summarised by using wildcards (`*` and `?`).

For each defined orbit, SORBIT produces an additional output channel named `sorbit_` followed by the orbit name specified in the namelist (second item in the list above). The sampled *channel objects* are, depending on the object to be sampled, in 2- or 3-D grid-point representation. The *channel objects* in the new SORBIT *channel* are named as the original *channel*, followed by an underscore and the name of the original *channel object*.

Figure 11 illustrates the increase of the valuable (i.e., useful for direct comparison) model output data density in comparison to the data density of satellite observations. The left column of Fig. 11 shows a standard snapshot output of a model simulation. For an intended point-by-point comparison with observations from a satellite instrument, only very few measurements from only one orbit correspond to the

```
&CPL
! T: automatic daily output (highly recommended)
! F: output according to CPL in channel.nml (for testing)
lout_auto = T,
! value for re-initialisation at first time step of every day
r_init    = -1.0E+34,
!
!# SYNTAX: - NAME,
!#          - LATITUDE DEPENDENT LOCAL TIME (T,F), ORBIT INCLINATION (deg),
!#            ASCENDING(+1) (T) OR DESCENDING(-1) (F)
!#          - LOCAL TIME HOUR, LOCAL TIME (MINUTE) [EQUATOR CROSSING TIME],
!#            LIMIT DT TO LOCAL TIME DISTANCE ? (T,F)
!#          - LIST OF CHANNEL OBJECTS
!#
!# NOTES:
!#    - NAME <= 8 CHARACTERS
!#    - SYNTAX FOR CHANNEL OBJECT LIST:
!#        "channel:object,object,object;channel:object;"
!#        (in object-names wildcards (*,?) can be used)
!#
ORB(1)='ENVI-AN',  T, 98.5451, T, 22,00, F, 'g3b:aps,albedo;PHYS:geopot,qm1,tm1;tracer_gp:*',
ORB(2)='ENVI-DN',  T, 98.5451, F, 10,00, F, 'g3b:aps,albedo;PHYS:geopot,qm1,tm1;tracer_gp:*',
/
```

Fig. 10. Example namelist of the submodel SORBIT: the SORBIT *channels* (sorbit_ENVI-AN and sorbit_ENVI-DN) are output every 24 h simulation time (i.e., when the sampled fields are full), since `lout_auto = T`. After the output the SORBIT *channel objects* are re-initialised with `r_init = -1.0E+34`. Two sun synchronous orbits are defined, both with latitude dependent local time calculation (T): 'ENVI-AN' for the ascending part (T) of ENVISAT with an equator crossing local time of 22:00 and 'ENVI-DN' for the descending part (F) of ENVISAT with an equator crossing local time of 10:00. The orbit inclination in both cases is 98.5451°, the *channel objects* aps and albedo from *channel* g3b, the *channel objects* geopot, qm1, tm1 from *channel* PHYS, and all chemical species from *channel* tracer_gp are sampled at the orbit local times, i.e., if the corresponding grid-box local time is within $T_O \pm \Delta t/2$ (F), where T_O is the orbit local time and Δt the model time step.

output simulation time. The exact number depends on the time difference between the orbit position, the model simulation output time, and a user-defined concurrence threshold. This is indicated by the red symbols (strict definition allowing only the nearest neighbours of the exact time) and the blue symbols (weaker definition allowing an interval of $\pm \Delta t$ around the exact time, with Δt being the model time step). The SORBIT output (middle column), however, is constructed to output each grid-box at the time of the day, which corresponds to a potential overflight of the orbiter, simply by selecting the correct, latitude dependent, local solar time. As a consequence, for all orbit positions of all orbits, the SORBIT output contains the corresponding model value with the strict criterion for the allowed time deviation applied (red symbols in the right panels of Fig. 11). In other words, for each satellite observation, a corresponding model simulated value is directly available in the output. Erroneous interpolation in time between snapshots is not required anymore. This is the highest valuable data density that can be reached within the limitations due to the coarse model grid, and the fact that the derived geometry is – strictly speaking – only valid for measurements in NADIR mode.

Figure 12 illustrates the systematic error arising, if climatological averages derived from sun-synchronously orbiting instruments are wrongly compared to climatological averages derived from standard (i.e., snapshot) model output. The examples show (arbitrarily selected) monthly (January 2006) averages of NO at 50 hPa (left) and the total ozone column density (right), once calculated from 5-hourly standard snapshot output (upper) and once calculated from SORBIT output (mid). The latter represents here the climatology derived from satellite observations. The lower panels show the corresponding absolute differences (in the respective units). The systematic biases are apparent and non-negligible. It is straightforward to argue that the systematic error is the larger, the shorter the photochemical lifetime, or likewise, the larger the diurnal amplitude of the selected species is.

6 New process submodels

A well suited method for the evaluation of Earth System models is the simulation and subsequent comparison to observations of so-called tracers of opportunity. The fate and distribution of such tracers are characteristically determined
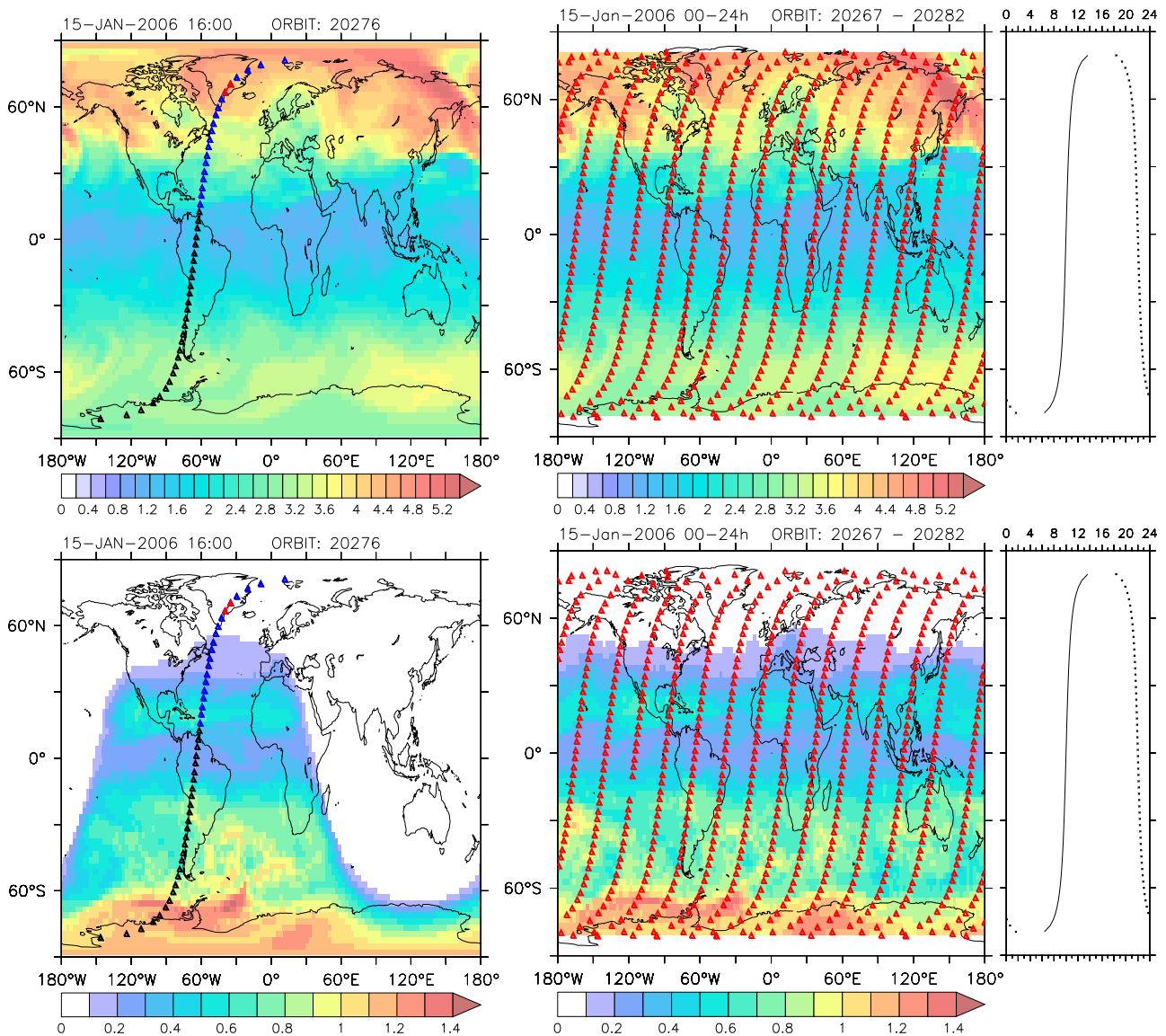
Fig. 11. Output of one model time step ("snapshot", left) for 15 January 2006, 16:00 UTC compared to SORBIT output for the same day (mid). The upper row shows O_3 (in μmol/mol) and the lower row NO (in nmol/mol), both at 50 hPa. The triangles denote the footprints of the descending parts of the ENVISAT orbits (orbit numbers are indicated at the top of the panels). The right panels show the corresponding latitude dependent local time (hour of day) of the descending (line) and ascending (symbols) parts of the orbits. In the left and middle column, red symbols denote those orbit positions, which correspond directly to the time of the underlying model output, blue symbols indicate positions which are at maximum $\pm\Delta t$ away from the given model output time (Δt is the model time step length), and black symbols show orbit positions outside this time interval. The example shows results from a T42L90MA simulation with a time step $\Delta t = 12$ min. In the SORBIT output (mid) all orbit positions correspond – per construction – to the time of the underlying output, as the SORBIT submodel samples the model data at all possible satellite positions each time step.

by a few processes only. Here, we introduce two new MESSy submodels for the simulation of ^{222}Rn and ^{210}Pb (DRADON, Sect. 6.1) and ^{14}CO (D14CO, Sect. 6.2) and one for artificial tracers (TREXP, Sect. 6.3).

6.1 DRADON: radon as diagnostic tracer

^{222}Rn is widely applied for the evaluation of the atmospheric transport characteristics (mainly in the vertical direction) of models of the atmosphere (e.g., Mahowald et al., 1997; Dentener et al., 1999; Allen et al., 1996; Zhang et al., 2008). The tracer is emitted from soil as radioactive decay product of
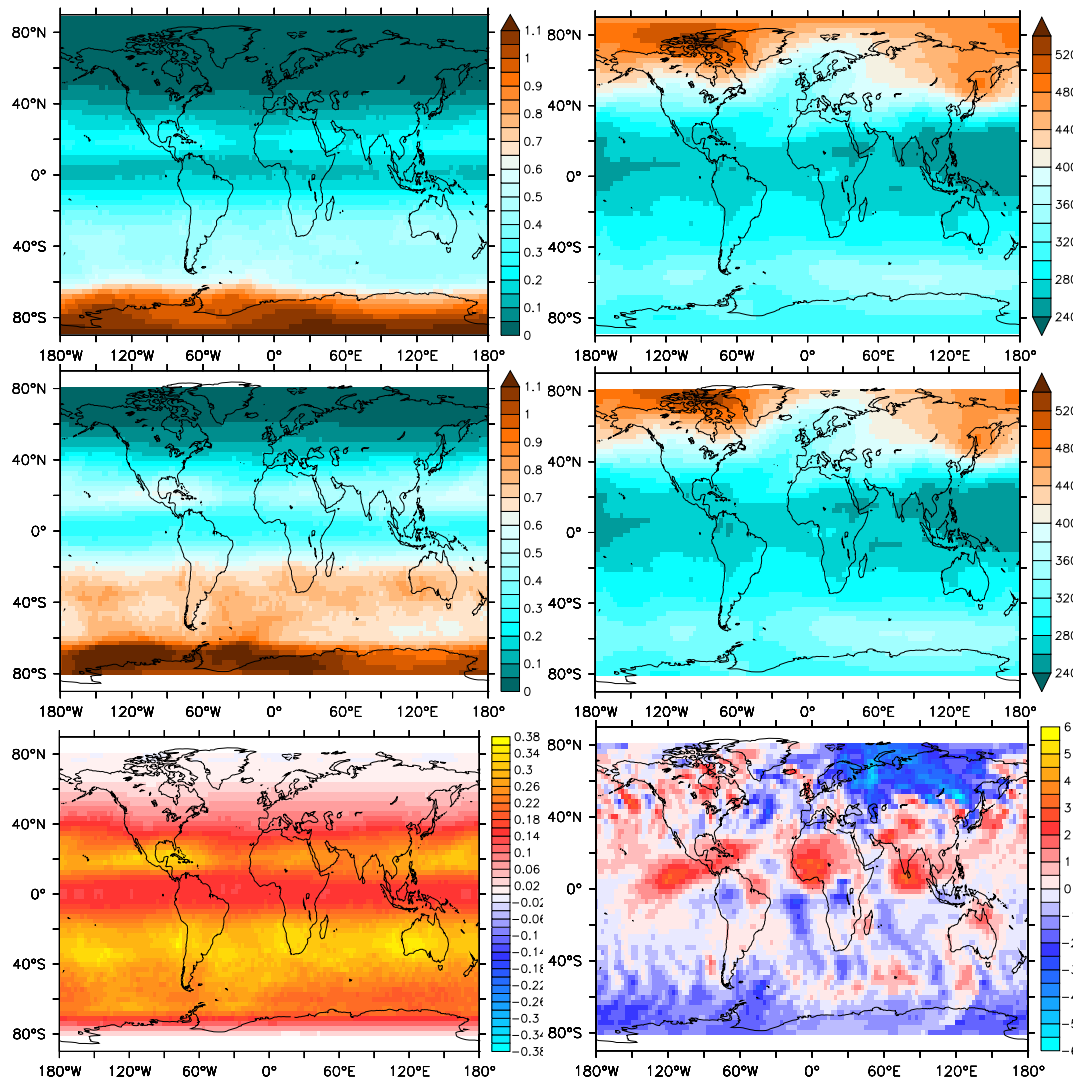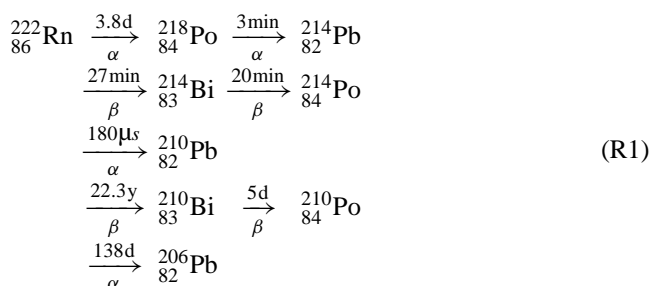
Fig. 12. Simulated monthly (January 2006) average NO in nmol/mol at 50 hPa (left) and total ozone in DU (right). The upper row shows the averages calculated from the 5-hourly standard model output (snapshots), averages calculated from the SORBIT output (ENVISAT, descending) are shown in the middle row. The lower row shows the differences (in the respective units) between both averages (SORBIT – 5-hourly).

^{226}Ra and itself decays further with a half-life of 3.8 days:

$$
\begin{aligned}
^{222}_{86}\text{Rn} &\xrightarrow[\alpha]{3.8\,\text{d}} {}^{218}_{84}\text{Po} \xrightarrow[\alpha]{3\,\text{min}} {}^{214}_{82}\text{Pb} \\
&\xrightarrow[\beta]{27\,\text{min}} {}^{214}_{83}\text{Bi} \xrightarrow[\beta]{20\,\text{min}} {}^{214}_{84}\text{Po} \\
&\xrightarrow[\alpha]{180\,\mu s} {}^{210}_{82}\text{Pb} \\
&\xrightarrow[\beta]{22.3\,\text{y}} {}^{210}_{83}\text{Bi} \xrightarrow[\beta]{5\,\text{d}} {}^{210}_{84}\text{Po} \\
&\xrightarrow[\alpha]{138\,\text{d}} {}^{206}_{82}\text{Pb}
\end{aligned}
\tag{R1}
$$

A common assumption is a ^{222}Rn emission rate of 1 atom/(cm^2 s) over (ice and snow free) soil and zero elsewhere. A more detailed source estimation has been derived by Schery and Wasiolek (1998).

Within the decay chain of ^{222}Rn, one daughter provides an additional, well suited tracer of opportunity for model evaluation: ^{210}Pb (with a half-life of 22.3 years). After its production ^{210}Pb immediately sticks to aerosol, due to its high adhesiveness. It is therefore ideally suited to evaluate model simulated aerosol dry and wet deposition rates (e.g., Rehfeld and Heimann, 1995; Preiss and Genthon, 1997; Liu et al., 2001; Koch et al., 1996). Due to these removal processes from the atmosphere, the atmospheric life-time of ^{210}Pb is

```
&CTRL
I_Rn_flux_method = 1    ! 0 = const; 1 = offline
! ### only for I_Rn_flux_method = 0:
R_Rn_cflux_land  = 10000.0 ! atoms m^(-2) s^(-1)
R_Rn_cflux_ocean = 0.0     ! atoms m^(-2) s^(-1)
/

&CPL
I_GP_emis_method  = 2  ! emission method for GP (1,2)
L_GP_chain        = T  ! 222Rn -> ... -> 210Pb
! ### only for L_GP_chain = T:
C_GP_210Pb_aermod = 'ptrac',
I_GP_210Pb_mode   = 4,
/

! ### only for I_Rn_flux_method = 1:
&RGTEVENTS
RG_TRIG(1) = 1,'months','first',0, 'Rn_flux', 1, 1, 12, $START_MONTH, 'Rn_flux',
/

&REGRID
!outfile  = "ncrg_${HRES}_Rn_flux_1x1_schery1998.nc",
infile    = "$INPUTDIR_MESSY/dradon/Rn_flux_1x1_schery1998.nc",
i_latm    = "LAT",
i_latr    = -90.0,90.0,
i_lonm    = "LON",
i_lonr    = -180.,180.,
i_timem   = "MONTH_REG",
!            convert mBq m^(-2) s^(-1) to atoms m^(-2) s^(-1)
var       = "Rn_flux=RN_FLUX:INT,0.0473665E+04"
/
```

Fig. 13. Example namelist file of the submodel DRADON: in the CTRL namelist, the choice is made, whether the ^{222}Rn source is constant (I_Rn_flux_method = 0) with values (in atoms/(m^2 s)) of R_Rn_cflux_land over (snow free) land and R_Rn_cflux_ocean over the oceans, respectively, or whether the ^{222}Rn source is provided off-line (I_Rn_flux_method = 1). Switched in the CPL namelist, the ^{222}Rn emission is either applied as tracer tendency to the lowest model grid box (I_GP_emis_method = 1), or as lower boundary condition of the vertical diffusive flux (I_GP_emis_method = 2), i.e., in the same way as tracer surface emissions are handled in the submodel OFFLEM (see Kerkweg et al., 2006b). In case not only the ^{222}Rn decay (L_GP_chain = F) is calculated, but the whole decay chain up to ^{210}Pb (L_GP_chain = T), C_GP_210Pb_aermod and I_GP_210Pb_mode specify the aerosol model and the corresponding mode for the ^{210}Pb tracer, respectively. The corresponding aerosol mean radius and aerosol radius standard deviation are retrieved from the selected aerosol submodel. In the example setup above, the submodel PTRAC (Jöckel et al., 2008) is used as simple aerosol model. The namelists RGTEVENTS and REGRID are used to import a time varying, prescribed (off-line) global ^{222}Rn source distribution via the MESSy NCREGRID interface (see Jöckel, 2006), in case I_Rn_flux_method = 1.

considerably shorter (days to weeks) than its decay half-life of 22.3 years. Measurements of both, the atmospheric abundance and the rain-out of ^{210}Pb, exist for a direct comparison with model simulations (e.g., Preiss et al., 1996; Preiss and Genthon, 1997).

The submodel DRADON (diagnostic Radon) is controlled by Fortran95 namelists (see Fig. 13) with various options:

– The source can either be specified as constant emission rates over (ice and snow free) land and ocean, respectively, or an external source distribution (e.g., Schery and Wasiolek, 1998)[10] can be prescribed.

– The source can be either applied as tendency of the tracer in the lowest model layer, or as lower boundary condition of the vertical diffusive flux. This is the same way as tracer surface emissions are handled in the submodel OFFLEM (see Kerkweg et al., 2006b).

– Only the decay of ^{222}Rn can be simulated, or the decay chain up to ^{210}Pb (including its decay). In the latter case, the corresponding system of coupled ordinary differential equations (ODE) is integrated with the Bareman solution described by Pressyanov (2002).

– ^{210}Pb, in case it is simulated (see above), is associated to an aerosol submodel and mode. These informations

---

[10]http://infohost.nmt.edu/~schery/mapdata.html

determine the aerosol characteristics (such as the mean radius and the radius standard deviation) required for the calculation of the aerosol relevant processes (sedimentation (submodel SEDI), scavenging (submodel SCAV), dry deposition (submodel DRYDEP)).

Figure 14 shows the summarised results of an EMAC2 simulation in T42L90MA resolution with the DRADON submodel. We follow the analysis of Zhang et al. (2008, for observations see references therein), and present the atmospheric abundance of ^{222}Rn in comparison to observations at several ground based stations. 8 simulated years (2000–2007, nudged towards ECMWF operational analysis data) have been averaged for comparison.

The simulation results are of comparable quality as derived by Zhang et al. (2008), who used also ECHAM5 as basemodel. The individual time series (monthly climatologies) are provided in the Supplement (Fig. 24 in MESSy2_evaluation.pdf).

Figure 15 shows the results of the same EMAC2 simulation for the atmospheric abundance and deposition of ^{210}Pb in comparison to the climatology of observations provided by Preiss and Genthon (1997). The scatter of both, the ^{210}Pb abundance and the rain-out flux, is comparable between the observations and the simulation results (Fig. 15 lower row), however, the correlation between the simulation results and the climatology is only moderate due to the large scatter. The regression analysis yet indicates a good agreement of the simulated abundance with the observations (a slope of 0.87), whereas the deposition flux is by a factor of 4 underestimated by the model (slope of 0.24). This is particularly surprising, since both, the near-surface abundance (Fig. 15, left; see also Fig. 20 in MESSy2_evaluation.pdf in the Supplement) and the vertical profiles (compiled by Emmons et al., 2000, see Fig. 16 in MESSy2_evaluation.pdf in the Supplement) are quite well represented by the model. Moreover, the global mass flux of ^{210}Pb is well balanced, ruling out a model error: the (8 year) average annual global source flux of ^{222}Rn decaying into ^{210}Pb is calculated to 13 ± 0.01 kg/year, the deposition fluxes by scavenging, dry deposition and sedimentation are 10.75 ± 0.02, 0.51 ± 0.01, and 1.38 ± 0.02 kg/year, respectively. The uncertainty ranges are the multi-annual (2000–2007) standard deviations. The annual change in burden is less than ± 0.002 kg/year. This implies a small mass deficit (deposition – emission) of -0.35 kg/year or -3% of the annual emission, which is explained by the radioactive decay of ^{210}Pb.

6.2 D14CO: ^{14}CO as diagnostic tracer

^{14}CO is a well suited tracer for the evaluation of global atmospheric chemistry models w.r.t. the simulated global distribution of the OH radical and the capability of the model to accurately represent the stratosphere-to-troposphere exchange (Jöckel et al., 2002; Jöckel, 2000, and references therein).
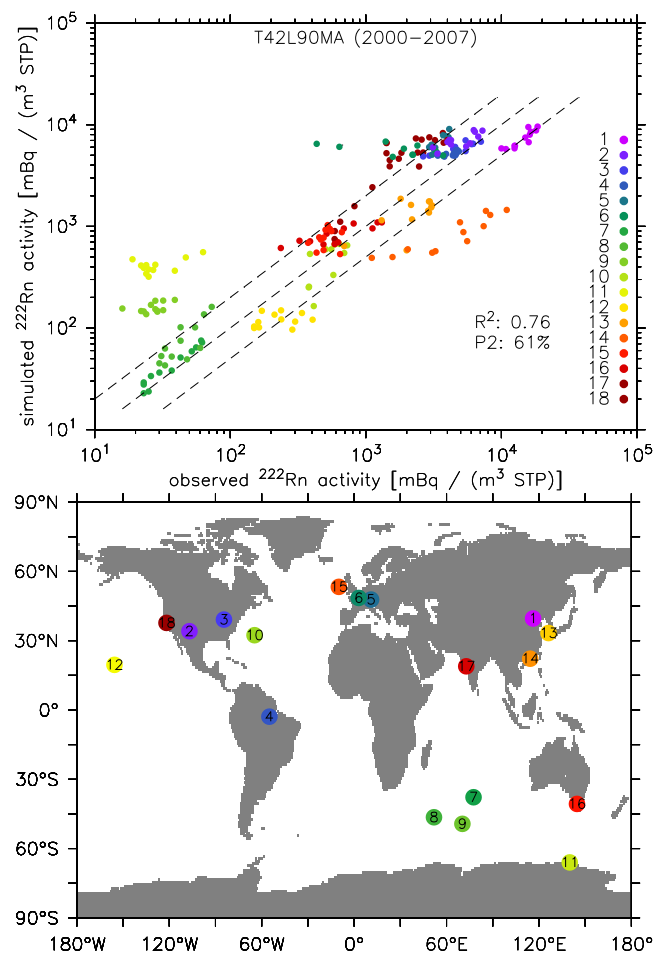

Fig. 14. Observed versus simulated monthly average ^{222}Rn activity in mBq/(m^3 STP) at ground level (upper panel). The model results are climatological monthly averages of the years 2000–2007 from an EMAC2-DRADON simulation in T42L90MA resolution. The figure corresponds to Fig. 9 (upper left panel) by Zhang et al. (2008, for observations see references therein), i.e., the dashed lines indicate the range within a factor of 2 of the measurements, P2 is the percentage of samples within this range, and R^2 is the correlation coefficient between the simulated and observed data. The colors and numbers indicate the positions on the globe (lower panel).

The primary origin of atmospheric ^{14}CO is production by cosmic radiation. High energy cosmic rays (mainly protons) induce large nucleonic particle cascades in the atmosphere and produce atmospheric neutrons. Most of them diffuse and thermalize before they are captured by nitrogen nuclei forming ^{14}C (^{14}N$(n, p)^{14}$C). The recoil ^{14}C atom rapidly oxidizes to ^{14}CO, with a yield that has been determined to be approximately 95% (Pandow et al., 1960; MacKay et al., 1963). In this way, a natural tracer is produced throughout the atmosphere, almost equally partitioned between the stratosphere and the troposphere, however with its maximum in polar regions caused by the influence of the geomagnetic field on the primary cosmic ray particles. The average source strength
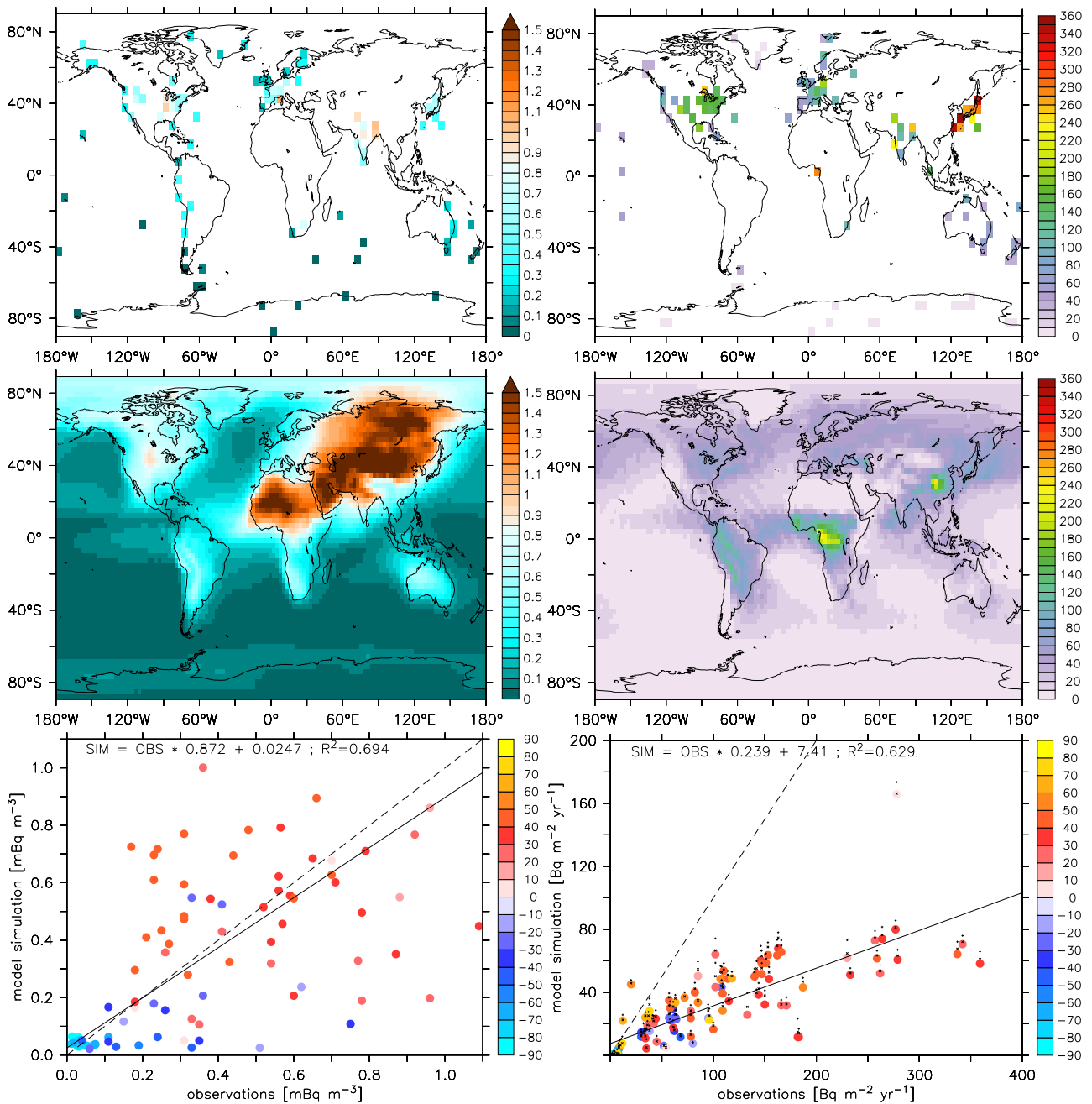
Fig. 15. ^{210}Pb activity (mBq/(m^3 STP), left) and ^{210}Pb deposition flux (Bq/(m^2 yr), right): the upper row shows the climatology compiled by Preiss and Genthon (1997), the middle row the averaged (2000–2007) results of an EMAC2-DRADON simulation in T42L90MA resolution with a constant ^{222}Rn emission of 1 atom/(cm^2 s) over ice-free land. The lower row shows the respective correlations of the model results (SIM) with the observed climatologies (OBS), color coded by latitude. For the regression analysis, the model results have been transformed to the 5° × 5° grid of the observed climatology. The black lines indicate the results of the linear regression analysis, the dashed line the line of perfect correspondence. R^2 is Pearson's correlation coefficient. Colored symbols in the lower right panel indicate the simulated wet deposition fluxes of ^{210}Pb, the tiny symbols indicate the corresponding accumulated fluxes of wet and dry deposition, and the total flux including also deposition through sedimentation, respectively.

```
&CPL
! -------------------------------------------------------
!# SWITCHES: - 1: INTEGRATION ON (T) / OFF (F)
!#           - 2: TENDENCY ADJUSTMENT ON (T) / OFF (F)
S_SWITCH(1) = T, T,
! -------------------------------------------------------
!### SOURCE DISTRIBUTION:
!#   - channel, object
!#   - '+', event name (must begin with 'CS' to identify source)
S_14CO(1)   = '+','CS_MA_N',
! -------------------------------------------------------
!### TROPOPAUSE FOR STE (14COs + 14COt)
!#            | diagnosed   |climatological |const. press.|
!#      switch:|     1       |       2       |     3       |
S_TP_STE(1) = 1, 'tropop', 'tp', 30000., 21500., 10000.,
! -------------------------------------------------------
!### TROPOPAUSE FOR OH (OHs + OHt)
!#            | diagnosed   |climatological |const. press.|
!#      switch:|     1       |       2       |     3       |
S_TP_OH(1)  = 2, 'tropop', 'tp', 30000., 21500., 10000.,
! -------------------------------------------------------
!### STRATOSPHERIC OH
!#   - channel, object
!#   - '+', event name (must begin with 'OH' to identify OH)
S_OHs(1)    = 'tracer_gp', 'OH',
! -------------------------------------------------------
!### TROPSPHERIC OH
!#   - channel, object
!#   - '+', event name (must begin with 'OH' to identify OH)
S_OHt(1)    = 'tracer_gp', 'OH',
! -------------------------------------------------------
/
&RGTEVENTS
RG_TRIG(1) = 1, 'years' ,'first',0, 'CS_MA_N', 13, 0, 13, 13,
          'VAR=P14CO_MA_N;UNIT=molec/(g s);ZSCALE=1.0',
/
&REGRID
infile    = "$INPUTDIR_MESSY/d14co/nP_14C.nc",
i_latm    = "LAT",
i_latr    = -90.0,90.0,
i_lonm    = "LON",
i_timem   = "PHI",
i_hyam    = "PRESS",
i_p0      = "100.0",
var       = "P14CO_MA_N=NP14C_MA",
/
```

Fig. 16. Example namelist file of the submodel D14CO: the example shows a definition of one setup. Three tracers, CO_14C_01, CO_14Cs_01, and CO_14Ct_01 will be defined by the submodel for the total ^{14}CO, the ^{14}CO produced in the stratosphere and for the ^{14}CO produced in the troposphere, respectively. The tropopause to distinguish between the production domains is set by S_TP_STE(1), here the diagnosed tropopause (*channel object* 'tp' from *channel* 'tropop') is used. The alternatives are a climatological tropopause (e.g., at 300–200×cos^2(latitude) hPa) or a constant pressure level (e.g., at 100 hPa). For the OH distribution, the corresponding tracer (*channel object* 'OH' from *channel* 'tracer_gp') is used for both, the troposphere (S_OHt(1)) and the stratosphere (S_OHs(1)). Alternatively, off-line provided OH fields can be chosen (in the same way as the source below), separately for stratosphere and troposphere, combined at the tropopause selected with S_TP_OH(1) (similar as S_TP_STE(1)). The 3-D source distribution is set with S_14CO(1), here the '+' indicates that the event (see TIMER, Sect. 3) with name 'CS_MA_N' must be evaluated. The corresponding event RG_TRIG(1) triggers the import of the variable 'P14CO_MA_N' in units of molec/(g s) via NCREGRID (Jöckel, 2006) using the corresponding REGRID-namelist. The vertical axis of this field requires no scaling (ZSCALE = 1.0), because it is already in pressure units. The first switch in S_SWITCH(1) can be used to deactivate this specific setup (set to F), the second switch, if T, forces a re-adjustment (linearisation) of the tracer tendencies in every model time step, such that always CO_14C_01 = CO_14Cs_01 + CO_14Ct_01. Up to 10 setups can be defined in this way, independent of each other. This allows sensitivity studies w.r.t. the source distribution, the OH distribution etc. in only one model simulation.

is 1.6–2 molecules per second and per square-centimeter of the Earth's surface, corresponding to a total production of approximately 13–16 kg ^{14}CO per year. Since the cosmic ray flux reaching the atmosphere is modulated by the solar wind intensity, the cosmogenic ^{14}CO production rate oscillates with a phase of 11 years (solar cycle) with higher production rates during times of low solar activity. The secondary ("biogenic") contribution, comprising 20–25% of the total source, consists of recycled ^{14}CO from the biosphere, entering or evolving in the atmosphere by oxidation of natural methane and higher hydrocarbons, and by biomass burning. The use of fossil fuel does not contribute to atmospheric ^{14}CO as geological production times vastly exceed the ^{14}C half life of about 5730 years.

The significance of ^{14}CO is that it constitutes a natural tracer that can be used to assess the hydroxyl radical (OH) abundance, because ^{14}CO + OH is its main sink reaction, with an average tropospheric lifetime of ^{14}CO of about 2–3 months and a stratospheric lifetime of about 4–7 months (Jöckel et al., 2000, see their Table 1). Jöckel and Brenninkmeijer (2002) compiled a climatology of cosmogenic ^{14}CO, i.e., a zonally averaged seasonal cycle at the surface comprising 1088 ^{14}CO observations from 4 institutes and Jöckel et al. (2002) used this climatology to evaluate two different chemistry transport models (CTMs).

The submodel D14CO implements the methodology as a MESSy submodel for further investigations and for the evaluation of EMAC2. Up to 10 independent setups are possible, each comprising three tracers, one for total ^{14}CO, one for ^{14}CO produced in the stratosphere, and one for ^{14}CO produced in the troposphere, respectively. Different concurrent setups allow sensitivity studies within one model simulation. The definition of one exemplary setup is shown in Fig. 16, which explains the D14CO namelists.

Figure 17 shows the result of an EMAC2-D14CO simulation in T42L90MA resolution from 1998 to 2008 (June), nudged towards the ECMWF operational analysis data.

The three ^{14}CO tracers have been analysed (i.e., compared to the ^{14}CO climatology) according to the procedure described by Jöckel et al. (2002), however, the results have been aggregated monthly. Thus, the upper and middle panel of Fig. 17 can be directly compared to Fig. 2 in Jöckel et al. (2002). It is important to note that for the simulation presented here, no re-scaling of the stratospheric contribution is applied. Furthermore, a different source distribution with a weaker vertical gradient has been applied here (Masarik and Beer, 1999 instead of Lingenfelter, 1963) tending to lower ^{14}CO mixing ratios at the surface. Note further that the peak values are smoothed, since the results have been monthly averaged.

EMAC2-D14CO, as the models used by Jöckel et al. (2002), shows a distinct asymmetry of wintertime surface ^{14}CO between the Northern (NH) and Southern Hemisphere (SH). The fraction of ^{14}CO originating from the stratosphere at the surface is larger in the NH compared to the SH.
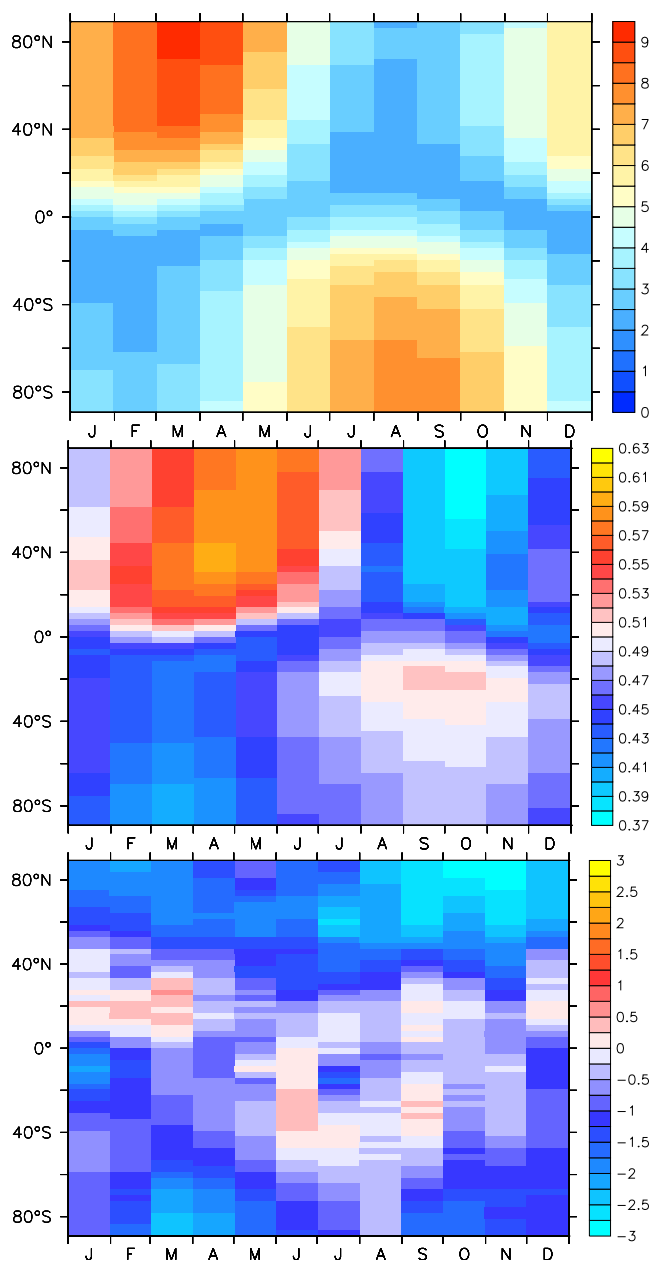

Fig. 17. Simulated multi-annual (1998–June 2008) monthly average near-surface cosmogenic ^{14}CO mixing ratio (in molec/(cm^3 STP), upper panel) and corresponding stratospheric fraction (middle panel). The lower panel shows the deviation (model – climatology) from the climatology based on observations compiled by Jöckel and Brenninkmeijer (2002). For the simulation the cosmogenic ^{14}C source distribution from Masarik and Beer (1999) was applied, normalised to a global average production rate of 1 molec/(cm^2 s). The climatology based on observations has been normalised to the same global average production rate by dividing it by the 1955–1988 (i.e., three solar cycles) average production rate of 1.91 molec/(cm^2 s). This number is based on the relationship between the global average ^{14}CO production rate and heliospheric potential (Masarik and Beer, 1999) and the reconstruction of the heliospheric potential (Usoskin et al., 2005).

```
&CPL
! ALLOW POINT EMISSIONS ALSO FOR NON-TREXP-TRACERS ?
l_force_emis = T,
!
! ### LIST OF TRACERS
! ### O : ORDER OF REACTION
!         = 0: dx/dt = -ka * x + p ,        [ka] = 1/s , Ta = 0
!         = 1: dx/dt = -ka * x * [Y] + p , [ka] = cm3/s
!         (p: production, x: mixing ratio, Y: reaction partner
! ### ka: O=0: ARRHENIUS A FACTOR          [ka] = cm3/s
!         O=1: DECAY CONSTANT              [ka] = 1/s
! ### Ta: O=0: Ta=0
!         O=1: ACTIVATION TEMPERATURE [K]
! SYNTAX:
!                                    ONLY FOR ORDER=1
!                                    |=====================|
!    'name', 'subname', ORDER, ka, Ta, 'channel', 'object'
TR( 1) = 'T01', , 1, 0.5625E-13,  1.5550E+03,'tracer_gp','OH',
TR( 2) = 'R01', , 0, 1.981875e-10,            ,           ,   ,
!
! ### LIST OF RELEASE POINTS AND TIME
! ### LEV  [hPa], MASS [kg], LON [-180 ... 180], LAT [-90 ... 90]
! SYNTAX:
!   LON, LAT, LEV, MASS [kg],
!   YYYY, MM, DD, HH, MI, SE, YYYY, MM, DD, HH, MI, SE, tracer_list
!   |===================|   |===================|
!             START                    STOP
POINT( 1) =   11.88,  78.90, 1000.0, 0.0978E+03,
    1978, 1, 1,  0, 0, 0,     1978, 1, 2, 0, 0, 0, 'T01;R01;',
/
```

Fig. 18. Example namelist file of the submodel TREXP: with `l_force_emis = T` the point sources (defined further down with `POINT`) can also comprise emissions of tracers, which have been defined by other submodels. Per default (`l_force_emis = F`) point sources are only allowed for tracers defined by TREXP (defined with `TR`). Two tracers are defined here (both without sub-name): the first (`TR(1)`) is T01, which reacts with OH from channel tracer_gp with a rate coefficient of $k = 0.5625 \times 10^{-13} \exp(-1555/T)\, \mathrm{cm}^3/\mathrm{s}$, where T is the temperature in K. The second (`TR(2)`) is R01, a radioactively decaying species with a decay constant of $1.981875 \times 10^{-10}\, 1/\mathrm{s}$. The mass of 97.8 kg of both tracers is released continuously on 1 January 1978 at 11.88° E, 78.90° N at 1000 hPa pressure altitude.

Due to its finer vertical resolution, EMAC2 represents the stratosphere-to-troposphere exchange (amount and phase) much better than the two CTMs used by Jöckel et al. (2002). Common to all three models is the still unexplained significant underestimation of near surface [14]CO in NH autumn. A further, more detailed analysis including a comparison to more recent observations is clearly required, however, far beyond the scope of this model documentation.

6.3 TREXP: tracer release experiments from point sources

The submodel TREXP was initially implemented to provide a submodel for artificial tracer studies as discussed previously by Jöckel et al. (2003). As all MESSy submodels, it is completely controlled via its namelist, more specifically its CPL namelist (for an example see Fig. 18). TREXP is used to define new tracers (including one degradation reaction) and point sources of tracers. A new tracer and degradation reaction is defined by

- the keyword `TR` followed by an arbitrary, but unique number (between 1 and 50) in parentheses,

- the unique name of the tracer,

- an optional sub-name of the tracer,

- the order of its degradation reaction, which is either 0 for radioactively decaying species or 1 for species reacting with one further educt,

- the decay constant (in 1/s for 0-order decay) or the Arrhenius A-factor (k_a in cm^3/s for first order reactions),

- and, in addition only for first order reactions, the activation temperature (T_a in K) and,

- the *channel* and *channel object* names (see Sect. 2) of the reaction partner (educt).

Note that the reaction partner in first order reactions is not altered by the reaction. Sources for the tracers defined this

way by TREXP can be specified – as for any other tracer – via the submodels OFFLEM or TNUDGE. As an alternative TREXP allows also to specify point sources. Such a point source is defined by

- the keyword POINT followed by an arbitrary, but unique number (between 1 and 100),

- the longitude (in ° E),

- the latitude (in ° N),

- the pressure level (in hPa),

- the mass (in kg) to be released,

- the start date and time (year, month, day, hour, minute, second) of the tracer release,

- the stop date and time (year, month, day, hour, minute, second) of the tracer release,

- the semicolon-separated list of tracers to be released.

Note that point sources defined in TREXP can also be used to emit tracers, which are not defined by TREXP, but by other submodels. For this l_force_emis needs to be T. The specified mass is continuously released within the time interval determined by the start and stop dates and times.

Besides the originally intended application to simulate artificial tracer release experiments, TREXP can also be used to simulate strong point sources, (e.g., volcano eruptions, megacities, etc.). Figure 19 shows the propagation of a volcanic plume as an example.

7 Improvements of the chemistry setup

For the development cycle 2 of MESSy introduced here, we implemented several updates and improvements for simulating the atmospheric gas phase and heterogeneous chemistry. These changes mainly comprise

- plugging in an additional update of the Module Efficiently Calculating the Chemistry of the Atmosphere (MECCA, Sander et al., 2005),

- plugging in the Multiphase Stratospheric Box Model (MSBM) for consistently calculating the heterogeneous reaction rates on polar stratospheric cloud (PSC) particles and on stratospheric background aerosol,

- an improvement of the submodel LNOX for the calculation of the tropospheric NO_x produced by lightning activity.

Further details of these changes/updates are described in this section.
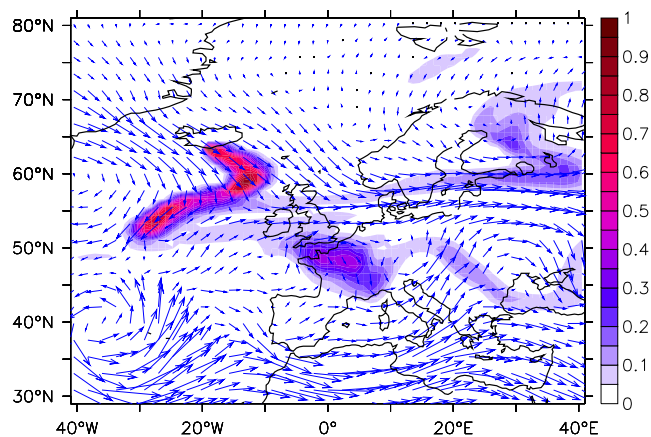

Fig. 19. Simulated ash plume of the Eyjafjallajökull over Europe at 18 April 2010, 16:00 UTC. The color scale indicates the column density (normalised to the maximum in the selected geographical region) in relative units. The blue arrows indicate the wind vector at 500 hPa. The EMAC2 simulation was performed in T106L90MA resolution nudged to the operational ECMWF analysis data. The submodel TREXP was used to define a continuous point-source at 63.629° N, 19.630° W in 500 hPa, starting 14 April 2010 at 00:00 UTC.

7.1 MECCA

The Module Efficiently Calculating the Chemistry of the Atmosphere (MECCA, Sander et al., 2005) has been renamed to MECCA1. In addition we plugged in an updated version (now called MECCA), which is also applied in the photochemical box model CAABA (Sander et al., 2011). MECCA (as MECCA1) is used to create (and solve) tailor-made chemical mechanisms.

7.1.1 The Kinetic Pre-Processor Post Processor (KP[4])

An atmospheric chemistry mechanism is mathematically described by a system of coupled ordinary differential equations (ODEs). For the integration in MECCA1 and MECCA we use the Kinetic Pre-Processor (KPP) to generate the Fortran95 code for the specified kinetic system and to combine it with a suited (depending on the stiffness of the corresponding ODE system) numerical solver. Whereas MECCA1 is based on version 1.1 of KPP (Damian et al., 2002), MECCA is based on version 2.2 of KPP (Sandu and Sander, 2006). Note that the syntax of the KPP input files (in particular the equation and the species files) and the generated Fortran95 code are different between both versions.

A major drawback of the KPP generated Fortran95 code is that it is – at least without further modifications – not running with satisfactory run-time performance. This is not an issue for photochemical (0-D) box models (which require a few CPU-seconds on a standard PC for a typical integration), but a huge problem for three dimensional models, where the integration of the kinetic systems requires up to 80% or more

of the used CPU time. The main reason for the bad run-time performance is a special form of nested loops (e.g., in the subroutine `KppDecomp`), which are used to calculate the LU-decomposition (lower – upper) of the (sparse) Jacobian:

```
DO k=1,NVAR
    ...
  DO kk = LU_CROW(k), LU_DIAG(k)-1
      j = LU_ICOL(kk)
      a = -W(j) / JVS( LU_DIAG(j) )
      W(j) = -a
      DO jj = LU_DIAG(j)+1, LU_CROW(j+1)-1
          W( LU_ICOL(jj) ) = W( LU_ICOL(jj) ) &
             + a*JVS(jj)
      END DO
      ...
  END DO
  ...
END DO
```

The operations in the nested loops contain integer arrays for the indirect addressing of arrays, which prevent the compiler from an efficient optimisation (e.g., because no memory prefetching is possible). Since the indices are known a-priori and do not change during the integration, but only depend on the chosen kinetic mechanism, the loops can be easily replaced by a sequence of operations without the need of index arrays and indirect addressing:

```
...
a = - W(32)/ JVS(93)
W(32)= -a
W(106)= W(106)+ a *JVS(94)
W(109)= W(109)+ a *JVS(95)
...
```

A further problem occurs on vector architectures, because the layout of the KPP generated code is for solving a kinetic system in one single box. This implies that for the application in models with a spatial resolution (usually 2- or 3-D), an outer loop over the spatial positions (e.g., model grid boxes) is required. On vector architectures (and potentially also on scalar architectures depending on the cache structure) it is more favourable, however, to have this loop over the spatial position as innermost loop.

In MECCA1 the loop structures with indirect addressing (first issue above) are replaced by a sequence of operations with the help of a shell-script (containing mainly `sed` commands), which automatically generates a Fortran95 program from the KPP generated code, compiles and runs it, and thereby explicitly executes the loops and writes the alternative Fortran95 code. To address the second issue (vector blocking) in MECCA1, we have to maintain a specific ("vectorised") SMIL module for MECCA1 and additional scripts to modify the SMCL modules. Unfortunately the scripts are not very robust w.r.t. minor modifications of the KPP input files (in particular the equation file), and the need to maintain two different Fortran95 modules for the same purpose but

only for different architectures is error-prone and therefore unsatisfactory.

Consequently, for MECCA (based on KPP 2.2), we use the KPP post processor (KP[4]) to modify the KPP generated Fortran95 code (including the modification of the loop structure and optionally the vector blocking) to achieve an improved run-time performance. The core of KP[4] consists of a Fortran95 parser and is implemented in C++. This core is controlled by a shell-script (kp4.sh), which in turn is called from the script (xmecca), which guides the user through the complete process of generating a tailor-made chemical mechanism.

If selected (required for the SMIL of the 3-D model, optional for the CAABA box model), KP[4] combines all KPP output files (Fortran95) to one Fortran95 module. Within this module, the nested `DO` loops with access to the index arrays are replaced by a sequence of operations (as shown above). If an additional vector blocking is desired (the user needs to specify the vector length), all variables and operations are modified accordingly. The KPP variables, which need to be expanded by one rank for the vector blocking are listed in the wrapper script (kp4.sh), additional user specified variables (e.g., the pressure) in the KPP equation file need to be encapsulated in an additional directive of the form (no line break at the \):

```
!KPPPP_DIRECTIVE vector variable \
                 definition start
  REAL(dp) :: press   ! pressure [Pa]
    ...
!KPPPP_DIRECTIVE vector variable \
                 definition end
```

The dots indicate further variables. In order to allow both, either scalar code or the internal vector blocking with the same SMIL module, KP[4] in addition automatically generates a "fill-subroutine" (named `fill_var`) for each variable (`var`) subject to vector blocking. By means of these routines, the SMIL simply "fills" the corresponding variables at all spatial positions (i.e., grid boxes) at once (at the beginning of the time step) into corresponding SMCL arrays. The vector loop is thus part of the automatically generated SMCL, and not anymore (as in MECCA1) part of the (not automatically generated) SMIL. If in vector mode a solver with automatic time stepping is applied, a special Fortran95 module (`messy_main_tools_kp4_compress.f90`) of KP[4] takes care of shuffling the grid-boxes along the vector, depending on whether the final stage is reached or further sub-time-steps are required.

The run-time performance improvement, after these measures are applied, depends on the stiffness of the kinetic system, but mainly on the architecture. For a typical setup (EMAC2 in T42L90MA resolution) comprising tropospheric and stratospheric chemistry (similar as described by Jöckel et al., 2006), we achieved, after the replacement of the nested

loops (no vector blocking), a factor of 10 speedup (compared to the original KPP output) on an IBM Power6.
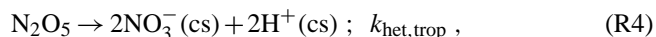
7.1.2 MECCA_KHET

An important improvement of the chemical setup is the new differentiation of tropospheric from stratospheric heterogeneous reactions on aerosol surfaces. In the previous setup (see Jöckel et al., 2006) the submodels PSC (Buchholz, 2005; Kirner, 2008; Kirner et al., 2010) and HETCHEM were used to calculate the heterogeneous reaction coefficients. In a typical model setup, PSC was used to calculate the polar stratospheric cloud (PSC) properties in a defined PSC region (polar lower stratosphere). HETCHEM then calculated in (every grid box of) the stratosphere the corresponding reaction coefficients (on PSCs) within this PSC region and on a climatological background (sulphate) aerosol outside the PSC region. For the troposphere the reaction coefficients (on aerosol surface) were either calculated on an external tropospheric aerosol surface density climatology, or by coupling a dynamical aerosol model. The resulting reaction coefficients in all regions were combined within HETCHEM and delivered to MECCA1 as pseudo first-order reaction coefficients. Due to the implementation, no distinction between typical stratospheric and tropospheric reaction pathways is possible. In particular, the reaction

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \; ; \quad k_{het} \qquad (R2)$$

is calculated within MECCA1 with the reaction coefficient (from HETCHEM) throughout the atmosphere, i.e., on PSCs, stratospheric sulphate aerosol and on tropospheric aerosol. However, in the troposphere N_2O_5 reacting heterogeneously on aerosol surfaces is converted into aqueous-phase nitrate, whereas in the dry stratosphere, gaseous HNO_3 is produced. Therefore, in MECCA, Reaction (R2) is split into two alternative pathways,

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \; ; \quad k_{het,strat} \qquad (R3)$$

and

$$N_2O_5 \rightarrow 2NO_3^-(cs) + 2H^+(cs) \; ; \quad k_{het,trop} \; , \qquad (R4)$$

which take into account that in the troposphere the reaction product is not released as HNO_3, but mostly remains in the aerosol phase (coarse mode, indicated by cs) in a dissociated state. This approach requires a distinction between stratospheric and tropospheric reaction coefficients, which are zero in the respective other domain. These are provided by the new sub-submodel MECCA_KHET. For the stratosphere, MECCA_KHET is coupled (via its CPL-namelist, see Fig. 20) to a stratospheric aerosol model, e.g., MSBM (see Sect. 7.2), which calculates the $k_{het,strat}$ and provides in addition a region flag (*channel object* STRAT_region), which is 0 in the stratosphere and 1 in the troposphere.

In the troposphere (defined by the STRAT_region flag), MECCA_KHET calculates the $k_{het,trop}$ either based on an

aerosol surface density climatology (aerosurf_clim in the CPL_KHET namelist, see Fig. 20), or based on the information provided by a dynamic tropospheric aerosol submodel (asm(.), see namelist). As a nice feature, the tropospheric aerosol surface area densities and corresponding reaction coefficients can be calculated (and output via the channel interface, Sect. 2) for several aerosol models running concurrently for diagnostics, but of course only one set of reaction coefficients can be used in the MECCA kinetic calculation. The latter is selected by the namelist switch asm_cpl.

The tropospheric heterogeneous reaction rate coefficients are calculated in MECCA_KHET as mass transfer rates (in 1/s)

$$k_{het,trop} = \frac{1}{4} v \gamma A \qquad (6)$$

(see Jacobson, 2005, Eq. 11.150), where γ is the dimensionless uptake coefficient, A the aerosol surface area density (in m^2/m^3) and v the mean molecular velocity (in m/s) of the Maxwell-Boltzmann distribution

$$v = \left(\frac{8 R_{gas} T}{M_m \pi} \right)^{\frac{1}{2}} , \qquad (7)$$

where T is temperature (in K), $R_{gas} = 8.314409\, J/(K\,mol)$ and M_m is the molar mass of the reactant (in kg/mol). The area surface density is calculated as

$$A = \sum_{m=1}^{M} 4\pi r_m^2 n_m \exp\left(2 \frac{(\log_{10}(\sigma_m))^2}{(\log_{10}(e))^2} \right) , \qquad (8)$$

where the sum is over the M aerosol modes (listed in asm(.)), r_m is the ambient radius of aerosol mode m, σ_m is the corresponding radius standard deviation, and n_m the particle number density of aerosol mode m.

7.2 MSBM

The Multiphase Stratospheric Box Model (MSBM) is basically a combination of the PSC submodel (Buchholz, 2005; Kirner, 2008) and the calculation of the stratospheric heterogeneous reaction rate coefficients of HETCHEM (Jöckel et al., 2006, and references therein). Indeed, the code of MSBM is to a large degree identical to the code of PSC. A detailed description of the current status of the polar stratospheric clouds (PSCs) parameterisations is beyond the scope of this overall model documentation and is published elsewhere (Kirner et al., 2010). Furthermore, since the MSBM namelists are essentially identical to those of PSC, we will not repeat them here, but refer to Kirner et al. (2010).

The reasons for combining the submodels PSC and parts of the submodel HETCHEM into MSBM are straightforward: first of all, it is favourable to have only one consistent submodel for the calculation of the heterogeneous reaction

```
&CTRL_KHET
l_troposphere  = T
l_stratosphere = T
/

&CPL_KHET
! channel and object with aerosol surface climatology:
aerosurf_clim = 'offlem', 'aerosurf_clim'
!
! aerosol submodel and modes:
asm(3) = 'm7',    '2,3,4'
!
! aerosol chemistry coupling; submodel to calculate rate
! coefficients (0 = aerosol surface climatology):
asm_cpl = 0
!
! stratosphere
strat_channel = 'msbm'
/
```

Fig. 20. Example namelists of the sub-submodel MECCA_KHET: The logical switches `l_troposphere` and `l_stratosphere` in `CTRL_KHET` determine if MECCA_KHET is performing calculations (`T`=true) for the troposphere and/or stratosphere, respectively, or not (`F`=false). In the `CPL_KHET` namelist, `aerosurf_clim` (optionally) names the *channel* and the *channel object* (Sect. 2) of an aerosol surface density climatology for which the heterogeneous reaction coefficients $k_{\mathrm{het,trop}}$ shall be calculated. The array `asm` lists the dynamic (tropospheric) aerosol submodel(s) and corresponding aerosol mode number(s), for which the aerosol surfaces and the heterogeneous reaction coefficients $k_{\mathrm{het,trop}}$ shall be calculated. Any listed aerosol submodel (`m7` in the example here) needs to deliver the ambient radius in m (*channel object* `wetradius` with rank 4, where the 3rd rank determines the aerosol mode) and the corresponding radius standard deviation (*channel object* `sigma` with rank 1 specifying the aerosol mode). In addition, for each aerosol mode the information on the particle number density must be available as TRACER (Jöckel et al., 2008). The parameter `ams_cpl` is used to select the set of $k_{\mathrm{het,trop}}$, which is used for the kinetic calculations in MECCA; 0 is used for the aerosol climatology defined by `aerosurf_clim`, any other number x for the corresponding aerosol submodel listed as `asm(x)`. Finally, `strat_channel` names the *channel* of a stratospheric aerosol model, which delivers the stratospheric heterogeneous coefficients $k_{\mathrm{het,strat}}$ and the *channel object* `STRAT_region` for flagging the stratosphere and troposphere (see text).

rate coefficients on stratospheric aerosol. Indeed, the distinction between "PSC region" and "rest of the stratosphere" must be consistent, if more than one submodel (like PSC and HETCHEM) are involved. This unavoidably leads to a close coupling, which renders the separation rather artificial. Moreover, large parts of the code (SMCL) of PSC and HETCHEM, in particular the functions and subroutines for the calculation of various heterogeneous pseudo-first order reaction coefficients are identical, because the same parameterisation for reactions on stratospheric background (sulphate) aerosol surfaces (HETCHEM) and PSC surfaces are applied. Maintaining different modules with the same code, however, is error-prone and the risk is always present that the different parts diverge after updates and/or bug-fixes, for instance that different (inconsistent) parameter settings occur.

A second process, which is implemented in both, PSC and HETCHEM, is the re-partitioning of H_2O into the gas-liquid- and ice-phase in the stratosphere. Since this re-partitioning affects directly the hydrological cycle, and therefore exerts a feedback on the dynamical part of the ESM (or GCM), it is also desirable to calculate this process consistently throughout the stratosphere.

In short summary, reasonable setups for stratospheric calculations always had to involve the submodels PSC and HETCHEM, now in MSBM we put together what reasonably belongs together. It is important to note, however, that the obsolete submodels PSC and HETCHEM remain available for compatibility with MESSy1. As a result, the user can now select between two different setups for stratospheric chemistry: one involving MECCA1, PSC and HETCHEM and another involving MECCA (with its sub-submodel MECCA_KHET, see Sect. 7.1.2) and MSBM.

7.3 LNOX

An important source of atmospheric nitrous oxide is contributed by lightning activity (Schumann and Huntrieser, 2007). Since single flashes cannot be resolved by global models of the atmosphere, this contribution has to be parameterised. The MESSy submodel LNOX (Jöckel et al., 2006; Tost et al., 2007) provides several parameterisations based on convective mass-flux (Grewe et al., 2001), based on cloud-top-height (Price and Rind, 1992, 1994), based on the updraft strength at a given altitude (Allen and Pickering, 2002), or based on the amount of convective precipitation at the

```
&CTRL
!
i_ffcalc      = 1, ! calculation of flash frequency according to
                   ! 1: Price & Rind
                   ! 2: Grewe
                   ! 3: Allen & Pickering (Massflux)
                   ! 4: Allen & Pickering (Conv. precipitation)
                   ! 9: diagnostically all parameterisations
!
i_iccg        = 2, ! 1: CG only, 2: CG + IC
i_shape       = 2, ! 1: const. mix. ratio,
                   ! 2: C-shape according to Pickering
i_scal        = 1, ! 1: Price & Rind, 2: Grewe
!
!
r_scal_ff     = 1.0  ! scaling factor for flash-frequency
!
r_noxpf       = 15.6, ! average NOx production per CG flash
                     ! [kgN/flash]
r_eff         = 0.1,  ! NOx production efficiency ratio (IC/CG)
/

&CPL
i_ff_cpl     = 1   ! if i_ffcalc == 9 => select parameterisation
                   ! (see list) for emission into NO tracer
!
c_updr       = 'convect','cu_uvelo',
c_top        = 'convect','cu_top',
c_bot        = 'convect','cu_bot',
c_freeze     = 'convect','cu_freeze',
!
l_midlevel   = F,  ! consider also mid-level convection
                     ! for lightning NOx production ?
c_top_mid    = 'convect','cu_top_mid',
c_bot_mid    = 'convect','cu_bot_mid',
c_freeze_mid = 'convect','cu_freeze_mid',
!
/
```

Fig. 21. Example CTRL and CPL namelists of the LNOX submodel. With i_ffcalc the parameterisation (for references see text) is selected, in case i_ffcalc = 9, all parameterisations are concurrently calculated diagnostically. With i_iccg either only cloud-to-ground (CG) flashes are considered (1) , or intra-cloud flashes as well (2). The vertical distribution of the lightning NO_x production within the convective column is determined with i_shape: 1: uniform, 2: C-shape according to Pickering et al. (1998). The resolution dependent scaling method is selected with i_scal as indicated, an additional global scaling factor for the flash-frequency is r_scal_ff, r_noxpf denotes the average NO_x production per flash (in kg(N)/flash), and r_eff the NO_x production efficiency ratio of IC versus CG flashes. With i_ff_cpl the parameterisation for the NO production is selected, this is only required for i_ffcalc = 9. The required input parameters (i.e., the *channel objects*) from the convection scheme are: c_updr for the updraft velocity (in m/s), c_top, c_bot and c_freeze for the top, bottom and freezing level (indices) of deep convection, respectively. In case mid-level convection (depending on the convection parameterisation) should also be considered for lightning NO_x production (l_midlevel = T), also the corresponding top, bottom and freezing level (indices) of mid-level convection are required: c_top_mid, c_bot_mid, and c_freeze_mid, respectively.

surface (Allen and Pickering, 2002). Each parameterisation can be selected via the Fortran95 namelist (Fig. 21) of LNOX and further be combined with the C-shape like vertical distribution of the resulting NO_x (Pickering et al., 1998). All parameterisations, except for the one by Grewe et al. (2001), distinguish between flashes over land and flashes over the ocean for the calculation of the flash frequency. In previous versions of LNOX, the flash frequency in each model grid box was either calculated for land or for ocean, depending on an integer land-sea mask. This has been changed and now both calculations, for land and ocean, are performed. The resulting flash frequency in each grid box is then determined as the sum of both contributions, weighted with the fractional land-sea mask. The impact on the global flash frequency

distribution is shown in Fig. 22 exemplarily for the parameterisation by Price and Rind (1994) as has been applied by Jöckel et al. (2006).

As expected, the largest effect (Fig. 22, lower panel) is a reduction of the flash frequency over Indonesia and Central America, where (in the applied resolution) most grid-boxes were previously "oceanic", whereas with the revised LNOX the land fraction (with a reduced flash frequency) is properly taken into account. Since the normalised flash frequency is shown and both simulation results have been normalised with their respective average (in time) integrated (in space) flash frequency, Fig. 22 (lower panel) indicates a relative redistribution of flashes into the continental, lightning-active regions. For equal scaling factors, this in turn implies that in grid-boxes over land and over sea no change in the absolute lightning-NO$_x$ production occurs, whereas in coastal (i.e., mixed land-sea) grid boxes the lightning production is either increased or decreased, depending on the integer landsea mask used before. The algorithmic modification is a clear improvement, since we have shown earlier that the flash frequency over Indonesia and Central America was considerably overestimated by LNOX (Tost et al., 2007).

8 Other changes

In addition to the major extensions described in the previous sections, we here list minor (but yet important) updates in the development cycle 2 of MESSy for the sake of completeness. These updates comprise:

- Complementing the generic submodels described in Sects. 2–4 and the generic submodel TOOLS, which is also part of MESSy1, MESSy2 provides the new generic submodels

    - BLATHER providing subroutines for the standardised output to the standard output and standard error units, i.e., to the log-files.
    - RND provides uniformly distributed random numbers between 0 and 1, calculated with either the standard Fortran90 function RANDOM_NUMBER or the Mersenne Twister algorithm (Matsumoto and Nishimura, 1998). Based on these, RND can also produce normally-distributed random numbers centered around zero, using the Marsaglia polar method[11].
    - GRIDTRAFO providing algorithms for various grid-transformations. The details will be described elsewhere (Pozzer et al., 2011).
    - TRANSFORM providing high-level routines for the transformation and transposition of decomposed fields between the different representations/parallel decompositions.

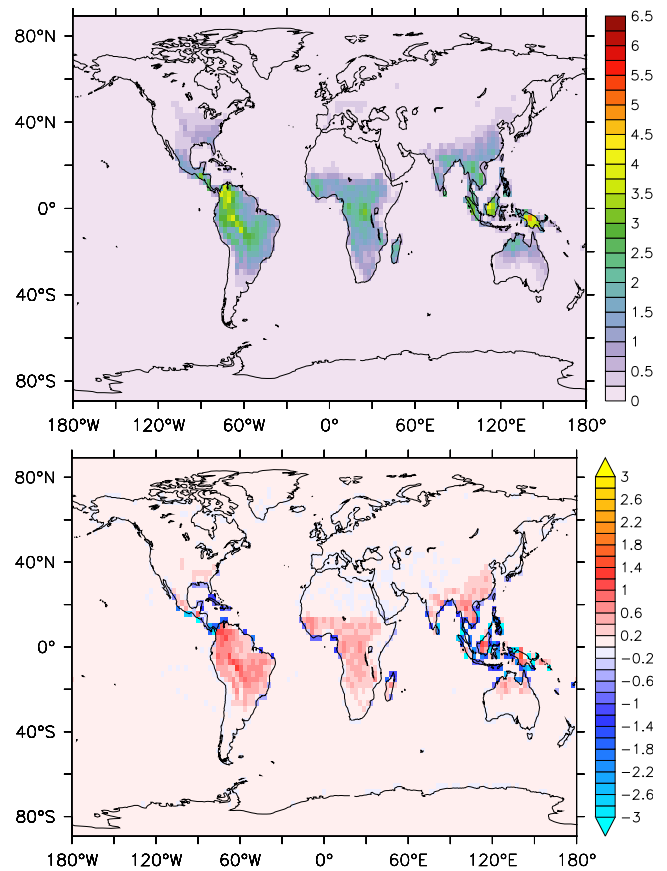[11] http://en.wikipedia.org/wiki/Marsaglia_polar_method

Fig. 22. Simulated two year (2003–2004) average flash frequency (upper panel) normalised to the total flash frequency (horizontally integrated, averaged in time) in 10^{-14} m^{-2} s^{-1} calculated with the parameterisation proposed by Price and Rind (1994) and taking into account the fractional land-sea mask. The lower panel shows the deviation of this flash frequency distribution from that calculated (data from Jöckel et al., 2006) with a previous version of LNOX based on the same parameterisation, but without accounting for the fractional land-sea mask. Both simulations were performed in T42L90MA resolution.

    - MPI containing high level routines and variables as interface to the message passing interface (MPI) library, which is used for the distributed memory parallelisation.

- The directory structure of the distribution has been changed to better reflect the 4 layer structure (subdirectories smcl, smil and bmil, respectively) and to allow different basemodels in the same distribution. As a prototype case, two different versions of ECHAM5 (5.3.01 and 5.3.02) are now included as basemodels. The distribution contains the file DIRSTRUCT with further information.

- The basemodel interface layer (BMIL) Fortran95 modules have now the suffix _bi.f90.

– The submodel interface layer (SMIL) Fortran95 modules have now the suffix _si.f90, if they are used for different basemodels (not only different versions of the same basemodel).

– The configure/gmake process has been revised.

– The universal run-script has been revised.

– The automatic restart facility has been revised. It is based on the generic submodel CHANNEL (see Sect. 2). Restart files are enumerated with the restart cycle number and therefore not overwritten, as previously. A new script (init_restart) is provided to prepare a model setup starting from restart files.

9 A re-evaluation simulation

With the modifications presented in Sect. 7 and the additional diagnostics, of which some results have been presented in the sections above, we performed a re-evaluation simulation with EMAC2 similar to the S2 simulation with EMAC1 presented by Jöckel et al. (2006) and further evaluated by Pozzer et al. (2007). The applied submodels are marked with an asterisk in Table 1.

9.1 Model configuration and setup

The applied spectral resolution of the ECHAM5 basemodel (version 5.3.02) is again T42, corresponding to a horizontal quadratic Gaussian grid of approximately $2.8° \times 2.8°$ in latitude and longitude. The vertical discretisation comprises 90 layers reaching from the ground into the mesosphere, more precisely up to 0.01 hPa (mid of uppermost layer). This T42L90MA (MA = middle atmosphere) setup is "nudged" (by Newtonian relaxation) towards the operational analysis data of the ECMWF. More specifically, the prognostic variables (logarithm of) surface pressure, vorticity, divergence and temperature are nudged, the latter three only between the 4th model layer above the ground to approximately 200 hPa. The simulation time covers the years 1998 (January) to 2008 (June).

The main differences compared to the S2 simulation of Jöckel et al. (2006) are:

– The submodels MECCA (with sub-submodel MECCA_KHET) and MSBM are used, instead of MECCA1, PSC and HETCHEM (see Sects. 7.1, 7.1.2 and 7.2).

– The submodel LNOX takes into account the fractional land-sea mask (see Sect. 7.3). Yet, the total emission is rescaled to achieve approximately 5 Tg/year NO emission by lightning.

– The biomass burning emissions of the Global Fire Emission Database (GFED version 2.1, Randerson et al.,

2007; van der Werf et al., 2006) for the years 1998 to 2006 are used (monthly averages); for the years 2007 and 2008 we applied climatological monthly averages of the previous years (because at the time of our simulation the data for these years were not available). Note that in Jöckel et al. (2006) we applied the monthly data of the year 2000 (GFED version 1) repeatedly for all years. A table with biomass burning emission totals for all years is contained in MESSy2_evaluation.pdf in the Supplement.

– The submodel AIRSEA (Pozzer et al., 2006) is applied to calculate the emissions of C_5H_8, DMS and CH_3OH from the ocean.

– We take into account the emissions of the bromocarbons $CHCl_2Br$, $CHClBr_2$, CH_2ClBr, CH_2Br_2 and $CHBr_3$ (as described by Kerkweg et al., 2008b) with the submodel OFFLEM (Kerkweg et al., 2006b), as well as the release of Br from sea-salt (Kerkweg et al., 2008b) with the submodel ONLEM (Kerkweg et al., 2008b). The Br-flux from sea-salt is calculated by scaling the mass flux of sea-salt by the ratio of the bromine to chlorine abundance in sea-salt (1.5×10^{-3}) and by assuming a bromine depletion within the sea salt of 50% (leading to an additional factor 0.5).

– In S2 described by Jöckel et al. (2006) we applied a precompiled (monthly climatological) aerosol surface area density in the troposphere for the calculation of heterogeneous reactions on aerosol surface. In the new simulation, the submodel M7 (Vignati et al., 2004; Kerkweg et al., 2008a) is applied to include tropospheric aerosol; and the sub-submodel MECCA_KHET is coupled to M7, i.e., the tropospheric heterogeneous reaction rate coefficients are calculated by MECCA_KHET (and applied in MECCA as pseudo-first order reactions) for the aerosol surface density of the M7 aerosol. For the N_2O_5 reaction on aerosol surface (Reaction R4), we used an uptake coefficient (see Eq. 6) of $\gamma = 0.02$ as suggested by Evans and Jacob (2005).

– The chemistry of mercury was implemented using the reaction mechanism by Xie et al. (2008). Mercury is a highly toxic element that has both natural and anthropogenic sources. Gaseous elemental mercury (GEM, Hg0) is quite inert and can be transported in the atmosphere over long distances. Once oxidised to reactive gaseous mercury (RGM), it is deposited to aerosols and to the surface quickly. Global mercury emission fields were taken from the GEOS-Chem model for the year 2004 (Selin et al. (2008) and Bess Sturges Corbitt, personal communication, 2008). Loss by scavenging was accounted for assuming uptake coefficients of $\gamma = 0.001$ for GEM and $\gamma = 0.1$ for RGM. Uptake of
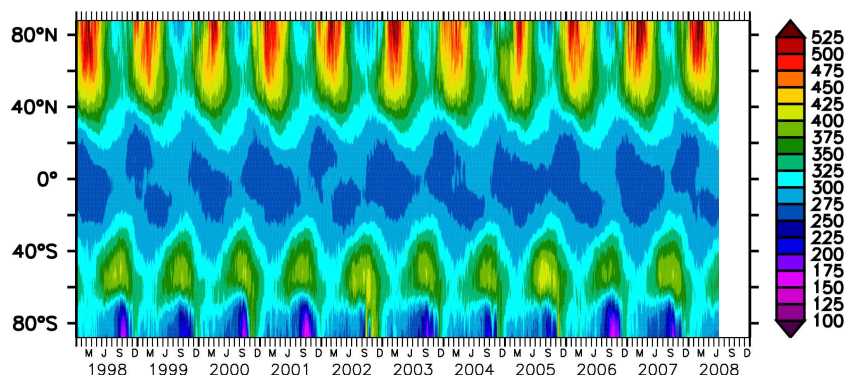
Fig. 23. Time series of total ozone (zonally averaged, in DU) simulated with EMAC2.

RGM on aerosols was calculated using an uptake co-efficient $\gamma = 0.1$. Due to its high volatility, uptake of GEM onto aerosols was not considered.

– The reaction coefficients in MECCA have been updated to JPL2006 (Sander et al., 2006). The chemical mechanism as applied in MECCA is listed in detail in the Supplement.

– We applied a more comprehensive aqueous phase mechanism in the submodel SCAV (Tost et al., 2006a), which is listed in detail in the Supplement.

– We included the heterogeneous reaction

$$\text{HNO}_3 \rightarrow \text{NO}_3^-\,(\text{cs}) + \text{H}^+(\text{cs}) \qquad \text{(R5)}$$
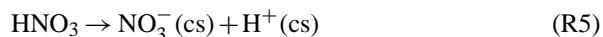
on tropospheric aerosol (reaction number HET202, see MESSy2_mecca.pdf in the Supplement). The reaction rate coefficient is calculated by MECCA_KHET according to Eqs. (6), (7) and (8). According to Sander et al. (2006), the accommodation coefficient for uptake on liquid water is ≥ 0.05. For our study, we have used $\gamma = 0.1$.

9.2 Results

Since the overall chemical setup has already been evaluated in detail (Jöckel et al., 2006; Pozzer et al., 2007), the same effort is not repeated here. We rather summarise some results to show that the presented modifications either improve the overall results, or at least do not deteriorate them. Moreover, we highlight some new findings and provide, in the Supplement (MESSy2_evaluation.pdf), a detailed inter-comparison of previous (EMAC1, simulation S1, Jöckel et al., 2006) and new (EMAC2, this study) model results with various observations. This Supplement is meant to serve as a reference and benchmark for future model simulations. Additional, supporting results on ^{222}Rn and ^{210}Pb have been mentioned

already in Sect. 6.1. The Supplement further contains figures showing the simulated QBO, the total (zonally averaged) ozone column (see also Fig. 23), and a snapshot of total ozone during the antarctic vortex split in September 2002, all in comparison to earlier model results and to observational data. As Fig. 23 shows, the issue of the strongly underestimated Antarctic ozone hole in spring remains unresolved as in many other models (Austin et al., 2010).

Due to the application of transient biomass-burning emissions (GFED 2.1, see above), the inter-annual variability of near-surface CO is – as to be expected – generally larger compared to the S1 simulation. Fig. 24 shows the time series of the observed and simulated near-surface CO and the corresponding anomalies (multi-annual average seasonal cycle subtracted) at a selected high northern latitude station (Point Barrow, Alaska) of the National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL, Novelli et al., 1998). The corresponding results for all other NOAA/ESRL stations are provided in Fig. 18 of MESSy2_evaluation.pdf in the Supplement. Figure 24 indicates that the simulated high northern latitude spring maximum of CO is strongly underestimated in comparison to the observations. In the new EMAC2 simulation (red line) this is even more severe than in the previous EMAC1 S1 simulation (blue line), due to the fact that in the EMAC2 simulation (GFED 2.1, transient) the multi-annual average CO emission from biomass burning is about 6%[12] lower compared to the EMAC1 S1 simulation (GFED 1, repeated 2000), whereas all other CO emissions are virtually the same in both simulations. Röckmann et al. (2002) showed through stable isotope analysis of CO that the strong biomass burning emissions in 1998 had a significant impact on the remote CO mixing ratio at high northern latitudes. This is confirmed by our model simulations, and also for the years 2002 and 2003, since the EMAC2 simulated CO anomaly (lower panel in Fig. 24) correlates much better with

---

[12]Detailed numbers of the biomass burning emissions are provided in Table 1 of the Supplement MESSy2_evaluation.pdf.

ALT (−62.52°; 82.45°)
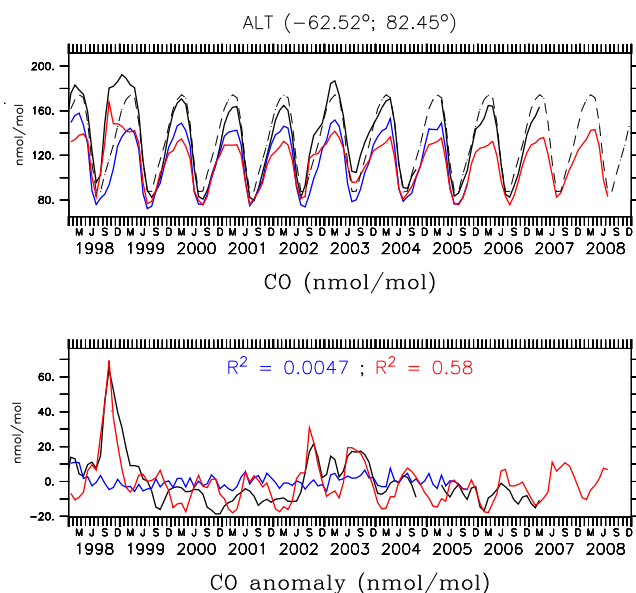
CO (nmol/mol)

CO anomaly (nmol/mol)

Fig. 24. Comparison of simulated monthly average CO mixing ratio with observations (at Point Barrow, Alaska) provided by the National Oceanic and Atmospheric Administration Earth System Research Laboratory (NOAA/ESRL; as originally presented by Novelli et al., 1998). The geographical position is indicated at the top of the upper panel. The upper panel shows the absolute values: the black line denotes the observations, the dashed black line the corresponding climatological monthly average, the blue line the results of the simulation S1 (Jöckel et al., 2006) and the red line the results of the new EMAC2 simulation. The lower panel shows the deviations (anomalies) from the corresponding climatological monthly average in absolute units: black: observations, blue: simulation S1, red: new EMAC2 simulation. The R^2 are the corresponding Pearson's correlation coefficients between the simulated and observed anomalies.

the observed anomaly than the EMAC1 simulated anomaly (the latter has no inter-annual variation). This is clearly visible also at other high northern latitude stations (see Supplement). The comparison of the simulated CO anomalies without (here EMAC1 S1, GFED 1, 2000 repeated) and with (here EMAC2 with transient GFED 2.1) inter-annual variations in biomass burning emissions provides therefore a direct measure of the role of biomass burning emissions at a given location. Overall, the correlation of simulated and observed CO anomalies is rather poor, indicating that other processes are important drivers of the inter-annual variability of CO, e.g., probably also anthropogenic emissions.

A comparison (not shown)[13] of the simulated mixing ratios from this study with observed vertical profiles (com-

_____

[13]The Supplement (MESSy2_evaluation.pdf) contains a series of figures with comparisons of the simulation results (both, from the S1 simulation presented by Jöckel et al. (2006) and the new simulation presented here) with observations compiled by Emmons et al. (2000).

piled by Emmons et al., 2000, from various field campaigns) confirms what has been obtained using the model results from simulation S1 and described by Pozzer et al. (2007). More specifically, for the C_2–C_3 alkanes and alkenes included in the chemical mechanism (ethane, propane, ethene and propene), the largest differences (up to 50%) between the simulated mixing ratios from the new simulation and from simulation S1 to the observations occur in the lower troposphere and in the planetary boundary layer (PBL). Differences in the biomass burning emissions are clearly indicated by the differently simulated mixing ratios of these compounds at the west-African coast. However, as shown previously for simulation S1, also here alkenes are largely underestimated, while a good agreement with observations is achieved for alkanes. Large discrepancies between results from simulation S1 and results from the new simulation occur for some oxygenated organic compounds (methanol and acetone). For acetone, the new quantum yield proposed by Blitz et al. (2004) has been used, while for methanol an increased natural emission (151 Tg/yr instead of 62 Tg/yr) as suggested by Jacob et al. (2005) has been applied. Although these modifications reduce the discrepancy between the simulated and observed mixing ratios in comparison to the original S1 simulation, the simulated mixing ratios are still too low. For CO and H_2O_2, the main differences between results from simulation S1 and the new simulation are located in regions influenced by biomass burning emissions. The similarity of the results confirms that HO_x ($OH + HO_2$) has been reproduced in analogy to simulation S1. Finally, no significant differences of the simulated mixing ratios are apparent between the new simulation and simulation S1 for the other compounds (e.g., HCHO and O_3).

For mercury, several model scenarios were tested (not shown). In agreement with Holmes et al. (2006) and Seigneur and Lohman (2008), we also found that oxidation of mercury is dominated by Br. A balanced mercury budget could best be reproduced if the oxidations by O_3 and OH were switched off. Note that it is very difficult to measure the reactions $Hg + OH$ and $Hg + O_3$ in the laboratory. Therefore, the uncertainties of the rate coefficients are very large and testing with a model, if these reactions are needed at all is a viable approach. Although it is normally assumed that O_3 and OH are the main sinks for most atmospheric pollutants, our results are in line with the study of Holmes et al. (2010).

The reaction of HNO_3 on aerosol results in a fast conversion of tropospheric HNO_3 into NO_3^- (cs) (Reaction R5), but the sum of both species is similar to the HNO_3 mixing ratio of simulation S1 (see Fig. 25; more examples are provided in Figs. 13 and 14 in MESSy2_evaluation.pdf of the Supplement). Keene et al. (1998) discuss the difficulties to measure gas-phase HNO_3 without interference from aerosol, and the implied potential that HNO_3 reported from observations might be biased high to the expense of NO_3^- (cs). Nevertheless, the rapid conversion of HNO_3 into NO_3^- (cs)
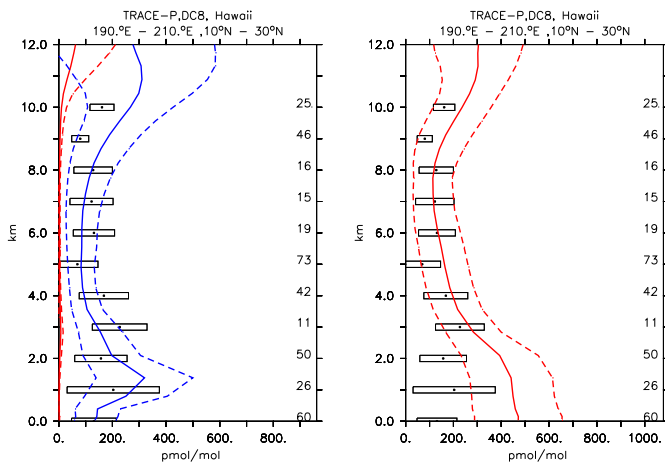
Fig. 25. Simulated vertical profiles (year 2000) of HNO$_3$ compared to selected campaign data (TRACE-P, DC8, near Hawaii) from the compilation of observational data (mainly from aircraft campaigns) by Emmons et al. (2000). The blue lines are the results of the S1 simulation (Jöckel et al., 2006), the red lines those of the new EMAC2 simulation. The left panel shows HNO$_3$, the right panel HNO$_3$ plus NO$_3^-$ (cs) (EMAC2 only). The left vertical axes show the altitude (in km), the horizontal axes show the mixing ratio (in pmol/mol). Numbers close to the right vertical axes list the number of measurements, the asterisks show the mean values of the measurements, the boxes indicate the corresponding standard deviation. The lines denote the average within the geographical area (as listed above the panels) and during the time of the year of the respective campaign, the dashed lines the corresponding (plus/minus) spatio-temporal standard deviations.

seems to be overestimated. Equation (6) shows that the reaction coefficient linearly depends on the uptake coefficient (γ) and the aerosol surface (A). Since there is no indication that the simulated aerosol surface is orders of magnitudes too large, a much smaller uptake coefficient (γ) compared to laboratory studies might be valid for the atmosphere. An alternative explanation is that the reaction is probably oversimplified, since it does not take into account any saturation effect of HNO$_3$ being resorbed on the aerosol surface. This is, however, subject of further investigations which go beyond the scope of this discussion. Luckily, the effect of the conversions hardly affects other species, most probably because HNO$_3$ and NO$_3^-$ (cs) share the same efficient removal processes, namely mainly scavenging and dry deposition.

10 Summary and outlook

The second development cycle of the Modular Earth Submodel System (MESSy2) expands this modelling system by an improved, basemodel-independent model infrastructure. The new developments comprise (1) a completely new designed memory and meta-data management, data trans-

fer and export interface (including restart facility), which is the generic submodel CHANNEL, (2) the generic submodel TIMER for the overall time control, and (3) the generic submodel QTIMER for the efficient application within job-scheduling systems. The Supplement contains detailed documentations of CHANNEL and TIMER. The new infrastructure submodels, mainly CHANNEL and TIMER, are required for a further development towards a comprehensive Earth System Model (ESM). More specifically, our current activities of coupling a dynamical and bio-geochemical ocean model, the on-line nesting of a regional model and the extension of the model domain into the thermosphere by coupling of an upper atmosphere model rely on this infrastructure.

For the regular evaluation of model results in comparison to observations, new diagnostic (data sampling) submodels are provided: (4) VISO for diagnosing vertically layered, 2-D iso-surfaces in 3-D scalar fields in Eulerian representation and for mapping 3-D scalar fields in Eulerian representation onto these surfaces, (5) SCOUT for sampling vertical model columns with higher output frequency at stationary observational sites, (6) S4D for sampling model data along the tracks of moving platforms with the maximum possible frequency (i.e., every model time step), and (7) SORBIT for sampling model data along orbits of sun-synchronously polar orbiting satellites. These diagnostic submodels are fully controllable by the MESSy user interface i.e., Fortran95 namelists, which are explained in detail. The submodels are designed with minimum overhead w.r.t. additional computational and memory demands, and – if applied – considerably increase the information contents (less valuable information is lost) of model output without blowing up the file space demands. For all diagnostic submodels introduced here, it is shown in detail that alternative approaches based on the post-processing of standard model output either introduce arbitrarily large interpolation errors, or significantly decrease the statistical sample sizes, if only perfect matches in space and time are considered.

In addition to the sampling submodels, also three new diagnostic process submodels exploiting so-called "tracers of opportunity" for the model evaluation are introduced: (8) DRADON for the simulation of ^{222}Rn and (optionally) ^{210}Pb, (9) D14CO for the simulation of cosmogenic ^{14}CO, and (10) TREXP for simulating point-sources of virtual and real species and for simulating simple zero order (i.e., decay) and first order reactions. Like the data sampling submodels, these submodels are also controllable by Fortran95 namelists, which are documented here. DRADON is evaluated for both, ^{222}Rn and ^{210}Pb, with available observations, D14CO with a previously published climatology, which is also based on observations. The simulation results are in good agreement with the observations.

Furthermore, the chemistry setup has been improved: (11) a new version of MECCA is implemented in addition to the previous version, which is now called MECCA1. For

the efficient application of MECCA in 3-D models, we apply the kinetic pre-processor (KPP) post processor (KP[4]) for automatic code optimisation on vector and scalar architectures. MECCA comprises the sub-submodel (12) MECCA_KHET for the calculation of heterogeneous reaction rate coefficients on tropospheric aerosol and for combining them with corresponding stratospheric heterogeneous rate coefficients. The latter are now provided by (13) the new submodel MSBM, which is a combination of the PSC submodel and the HETCHEM submodel, thus constituting a multi-phase stratospheric box model. It is important to note that, according to the MESSy philosophy, both setups, i.e., either MECCA1 – PSC – HETCHEM, or MECCA – MECCA_KHET – MSBM are available in the same model environment. The (14) submodel LNOX for the calculation of lightning NO production now takes into account the fractional land-sea mask for those parameterisations, which distinguish flashes over land from those over sea.

MESSy2 (as MESSy1) with the improved chemistry setup, i.e., using MECCA – MECCA_KHET – MSBM and other modifications/extensions (e.g., improved emissions, a more complex aqueous phase mechanism, a mercury mechanism included, tropospheric heterogeneous chemistry on prognostic M7-aerosol surfaces) has been successfully coupled to the Atmospheric General Circulation Model (GCM) ECHAM5, together constituting the ECHAM/MESSy Atmospheric Chemistry (EMAC) model. The new version is evaluated in comparison to previous results and in comparison to observations. Overall, the results are satisfactory and confirm the very high standard of the model system.

Supplementary material related to this article is available online at: http://www.geosci-model-dev.net/3/717/2010/ gmd-3-717-2010-supplement.zip.

References

Allen, D. J. and Pickering, K. E.: Evaluation of lightning flash rate parameterizations for use in a global chemical transport model, J. Geophys. Res., 107, 4711, doi:10.1029/2002JD002066, 2002.

Allen, D. J., Rood, R. B., Thompson, A. M., and Hudson, R. D.: Three-dimensional radon 222 calculations using assimilated meteorological data and a convective mixing algorithm, J. Geophys. Res., 101, 6871–6881, 1996.

Austin, J., Struthers, H., Scinocca, J., Plummer, D. A., Akiyoshi, H., Baumgaertner, A. J. G., Bekki, S., Bodeker, G. E., Braesicke, P., Brühl, C., Butchart, N., Chipperfield, M. P., Cugnet, D., Dameris, M., Dhomse, S., Frith, S., Garny, H., Gettelman, A., Hardiman, S. C., Jöckel, P., Kinnison, D., Kubin, A., Lamarque, J. F., Langematz, U., Mancini, E., Marchand, M., Michou, M., Morgenstern, O., Nakamura, T., Nielsen, J. E., Pitari, G., Pyle, J., Rozanov, E., Shepherd, T. G., Shibata, K., Smale, D., Teyssèdre, H., , and Yamashita, Y.: Chemistry-climate model simulations of spring Antarctic ozone, J. Geophys. Res., 115, D00M11, doi:10.1029/ 2009JD013577, http://dx.doi.org/10.1029/2009JD013577, 2010.

Baumgaertner, A. J. G., Jöckel, P., and Brühl, C.: Energetic particle precipitation in ECHAM5/MESSy1 – Part 1: Downward transport of upper atmospheric NO$_x$ produced by low energy electrons, Atmos. Chem. Phys., 9, 2729–2740, doi:10.5194/acp-9-2729-2009, 2009.

Baumgaertner, A. J. G., Jöckel, P., Riede, H., Stiller, G., and Funke, B.: Energetic particle precipitation in ECHAM5/MESSy – Part 2: Solar proton events, Atmos. Chem. Phys., 10, 7285–7302, doi:10.5194/acp-10-7285-2010, 2010a.

Baumgaertner, A. J. G., Jöckel, P., Steil, B., Tost, H., and Sander, R.: A fast stratospheric chemistry solver: the E4CHEM submodel for the atmospheric chemistry global circulation model

EMAC, Geosci. Model Dev., 3, 321–328, doi:10.5194/gmd-3-321-2010, 2010b.

Blitz, M., Heard, D., Pilling, M., Arnold, S. R., and Chipperfield, M. P.: Pressure and temperature-dependent quantum yields for the photodissocution od acetone between 279 and 327.5 nm, Geophys. Res. Lett., 31, L06111, doi:10.1029/2003GL018793, 2004.

Brenninkmeijer, C. A. M., Crutzen, P., Boumard, F., Dauer, T., Dix, B., Ebinghaus, R., Filippi, D., Fischer, H., Franke, H., Frieß, U., Heintzenberg, J., Helleis, F., Hermann, M., Kock, H. H., Koeppel, C., Lelieveld, J., Leuenberger, M., Martinsson, B. G., Miemczyk, S., Moret, H. P., Nguyen, H. N., Nyfeler, P., Oram, D., O'Sullivan, D., Penkett, S., Platt, U., Pupek, M., Ramonet, M., Randa, B., Reichelt, M., Rhee, T. S., Rohwer, J., Rosenfeld, K., Scharffe, D., Schlager, H., Schumann, U., Slemr, F., Sprung, D., Stock, P., Thaler, R., Valentino, F., van Velthoven, P., Waibel, A., Wandel, A., Waschitschek, K., Wiedensohler, A., Xueref-Remy, I., Zahn, A., Zech, U., and Ziereis, H.: Civil Aircraft for the regular investigation of the atmosphere based on an instrumented container: The new CARIBIC system, Atmos. Chem. Phys., 7, 4953–4976, doi:10.5194/acp-7-4953-2007, 2007.

Bronstein, I. N., Semendjajew, K. A., Musiol, G., and Mühlig, H.: Taschenbuch der Mathematik, Verlag Harri Deutsch, 2005.

Buchholz, J.: Simulations of physics and chemistry of polar stratospheric clouds with a general circulation model, Ph.D. thesis, University of Mainz, Germany, http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:hebis:77-8187, 2005.

Damian, V., Sandu, A., Damian, M., Potra, F., and Carmichael, G. R.: The kinetic preprocessor KPP - a software environment for solving chemical kinetics, Comput. Chem. Eng., 26, 1567–1579, 2002.

Dentener, F., Feichter, J., and Jeuken, A.: Simulation of the transport of Radon^{222} using on-line and off-line global models at different horizontal resolutions: A detailed comparison with measurements, Tellus, 51B, 573–602, 1999.

Emmons, L. K., Hauglustaine, D. A., Müller, J.-F., Carroll, M. A., Brasseur, G. P., Brunner, D., Staehelin, J., Thouret, V., and Marenco, A.: Data composites of airborne observation of tropospheric ozone and its precursor, J. Geophys. Res., 105, 20497–20538, 2000.

Evans, M. J. and Jacob, D. J.: Impact of new laboratory studies of N_2O_5 hydrolysis on global model budgets of tropospheric nitrogen oxides, ozone and OH, Geophys. Res. Lett., 32, L09813, doi:10.1029/2005GL022469, 2005.

Giorgetta, M. A. and Bengtsson, L.: The potential role of the quasi-biennial oscillation in the stratosphere-troposphere exchange as found in water vapour in general circulation model experiments, J. Geophys. Res., 104, 6003–6019, 1999.

Grewe, V., Brunner, D., Dameris, M., Grenfell, J. L., Hein, R., Shindell, D., and Staehelin, J.: Origin and variability of upper tropospheric nitrogen oxides and ozone at northern mid-latitudes, Atmos. Environ., 35, 3421–3433, 2001.

Gromov, S., Jöckel, P., Sander, R., and Brenninkmeijer, C. A. M.: A kinetic chemistry tagging technique and its application to modelling the stable isotopic composition of atmospheric trace gases, Geosci. Model Dev., 3, 337–364, doi:10.5194/gmd-3-337-2010, 2010.

Holmes, C. D., Jacob, D. J., and Yang, X.: Global lifetime of elemental mercury against oxidation by atomic bromine in the free troposphere, Geophys. Res. Lett., 33, L20808, doi:10.1029/2006GL027176, 2006.

Holmes, C. D., Jacob, D. J., Corbitt, E. S., Mao, J., Yang, X., Talbot, R., and Slemr, F.: Global atmospheric model for mercury including oxidation by bromine atoms, Atmos. Chem. Phys., 10, 12037–12057, doi:10.5194/acp-10-12037-2010, 2010.

Jacob, D. J., Field, B. D., Li, Q., Blake, D. R., de Gouw, J., Warneke, C., Hansel, A., Wisthaler, A., Singh, H. B., and Guenther, A.: Global budget of methanol: Constraints from atmospheric observations, J. Geophys. Res., 110, D08303, doi:10.1029/2004JD005172, 2005.

Jacobson, M. Z.: Fundamentals of atmospheric modeling, Cambridge University Press, 2nd edn., 2005.

Jöckel, P.: Cosmogenic ^{14}CO as tracer for atmosperic chemistry and transport, Ph.D. thesis, Combined Faculties for the Natural Sciences and for Mathematics of the Rupertus Carola University of Heidelberg, Germany, http://www.ub.uni-heidelberg.de/archiv/1426, 2000.

Jöckel, P.: Technical note: Recursive rediscretisation of geoscientific data in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 3557–3562, doi:10.5194/acp-6-3557-2006, 2006.

Jöckel, P. and Brenninkmeijer, C. A. M.: The seasonal cycle of cosmogenic ^{14}CO at the surface level: A solar cycle adjusted, zonal average climatology based on observations, J. Geophys. Res., 107, 4656, doi:10.1029/2001JD001104, 2002.

Jöckel, P., Brenninkmeijer, C. A. M., and Lawrence, M. G.: The atmospheric response time of cosmogenic ^{14}CO to changes in solar activity, J. Geophys. Res., 105, 6737–6744, 2000.

Jöckel, P., Brenninkmeijer, C. A. M., Lawrence, M. G., Jeuken, A. B. M., and van Velthoven, P. F.: Evaluation of stratosphere - troposphere exchange and the hydroxyl radical distribution in 3-dimensional global atmospheric models using observations of cosmogenic ^{14}CO, J. Geophys. Res., 107, 4446, doi:10.1029/2001JD001324, 2002.

Jöckel, P., Brenninkmeijer, C. A. M., and Crutzen, P. J.: A discussion on the determination of atmospheric OH and its trends, Atmos. Chem. Phys., 3, 107–118, doi:10.5194/acp-3-107-2003, 2003.

Jöckel, P., Sander, R., Kerkweg, A., Tost, H., and Lelieveld, J.: Technical Note: The Modular Earth Submodel System (MESSy) – a new approach towards Earth System Modeling, Atmos. Chem. Phys., 5, 433–444, doi:10.5194/acp-5-433-2005, 2005.

Jöckel, P., Tost, H., Pozzer, A., Brühl, C., Buchholz, J., Ganzeveld, L., Hoor, P., Kerkweg, A., Lawrence, M. G., Sander, R., Steil, B., Stiller, G., Tanarhte, M., Taraborrelli, D., van Aardenne, J., and Lelieveld, J.: The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere, Atmos. Chem. Phys., 6, 5067–5104, doi:10.5194/acp-6-5067-2006, 2006.

Jöckel, P., Kerkweg, A., Buchholz-Dietsch, J., Tost, H., Sander, R., and Pozzer, A.: Technical Note: Coupling of chemical processes with the Modular Earth Submodel System (MESSy) submodel TRACER, Atmos. Chem. Phys., 8, 1677–1687, doi:10.5194/acp-8-1677-2008, 2008.

Keene, W. C., Sander, R., Pszenny, A. A. P., Vogt, R., Crutzen, P. J., and Galloway, J. N.: Aerosol pH in the marine boundary layer: A review and model evaluation, J. Aerosol Sci., 29, 339–356, doi:10.1016/S0021-8502(97)10011-8, 1998.

Kerkweg, A., Buchholz, J., Ganzeveld, L., Pozzer, A., Tost, H., and Jöckel, P.: Technical Note: An implementation of the dry removal processes DRY DEPosition and SEDImentation in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 4617–4632, doi:10.5194/acp-6-4617-2006, 2006a.

Kerkweg, A., Sander, R., Tost, H., and Jöckel, P.: Technical note: Implementation of prescribed (OFFLEM), calculated (ON-LEM), and pseudo-emissions (TNUDGE) of chemical species in the Modular Earth Submodel System (MESSy), Atmos. Chem. Phys., 6, 3603–3609, doi:10.5194/acp-6-3603-2006, 2006b.

Kerkweg, A., Sander, R., Tost, H., Jöckel, P., and Lelieveld, J.: Technical Note: Simulation of detailed aerosol chemistry on the global scale using MECCA-AERO, Atmos. Chem. Phys., 7, 2973–2985, doi:10.5194/acp-7-2973-2007, 2007.

Kerkweg, A., Jöckel, P., Pozzer, A., Tost, H., Sander, R., Schulz, M., Stier, P., Vignati, E., Wilson, J., and Lelieveld, J.: Consistent simulation of bromine chemistry from the marine boundary layer to the stratosphere – Part 1: Model description, sea salt aerosols and pH, Atmos. Chem. Phys., 8, 5899–5917, doi:10.5194/acp-8-5899-2008, 2008a.

Kerkweg, A., Jöckel, P., Warwick, N., Gebhardt, S., Brenninkmeijer, C. A. M., and Lelieveld, J.: Consistent simulation of bromine chemistry from the marine boundary layer to the stratosphere – Part 2: Bromocarbons, Atmos. Chem. Phys., 8, 5919–5939, doi:10.5194/acp-8-5919-2008, 2008b.

Kirner, O.: Prozessstudien der stratosphaerischen Chemie und Dynamik mit Hilfe des Chemie-Klima-Modells ECHAM5/MESSy1, Ph.D. thesis, Universitaet Karsruhe, http://digbib.ubka.uni-karlsruhe.de/volltexte/1000010199, 2008.

Kirner, O., Ruhnke, R., Buchholz-Dietsch, J., Jöckel, P., Brühl, C., and Steil, B.: Simulation of polar stratospheric clouds in the chemistry-climate-model EMAC via the submodel PSC, Geosci. Model Dev. Discuss., 3, 2071–2108, doi:10.5194/gmdd-3-2071-2010, 2010.

Koch, D. M., Jacob, D. J., and Graustein, W. C.: Vertical transport of tropospheric aerosols as indicated by and in a chemical tracer model, J. Geophys. Res., 101, 18651–18666, 1996.

Landgraf, J. and Crutzen, P. J.: An efficient method for online calculations of photolysis and heating rates, J. Atmos. Sci., 55, 863–878, 1998.

Lauer, A., Eyring, V., Hendricks, J., Jöckel, P., and Lohmann, U.: Global model simulations of the impact of ocean-going ships on aerosols, clouds, and the radiation budget, Atmos. Chem. Phys., 7, 5061–5079, doi:10.5194/acp-7-5061-2007, 2007.

Leroy, S. S.: The Effects of Orbital Precession on Remote Climate Monitoring, J. Climate, 14, 4330–4337, 2001.

Lingenfelter, R. E.: Production of carbon 14 by cosmic-ray neutrons, Rev. Geophys., 1, 35–55, 1963.

Liu, H., Jacob, D. J., Bey, I., and Yantosca, R. M.: Constraints from ^{210}Pb and ^{7}Be on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields, J. Geophys. Res., 106, 12109–12128, 2001.

MacKay, C., Pandow, M., and Wolfgang, R.: On the chemistry of natural radiocarbon, J. Geophys. Res., 68, 3929–3931, 1963.

Mahowald, N. M., Rasch, P. J., Eaton, B. E., Whittlestone, S., and Prinn, R. G.: Transport of 222Radon to the remote troposphere using the Model of Atmospheric Transport and Chemistry and assimilated winds from ECMWF and the National Center for Environmental Prediction/NCAR, J. Geophys. Res., 102, 28139–28151, 1997.

Marenco, A., Thouret, V., Nédélec, P., Smit, H., Helten, M., Kley, D., Karcher, F., Simon, P., Law, K., Pyle, J., Poschmann, G., Von Wrede, R., Hume, C., and Cook, T.: Measurement of ozone and water vapor by Airbus in-service aircraft: The MOZAIC airborne program, An overview, J. Geophys. Res., 103, 25631–25642, 1998.

Masarik, J. and Beer, J.: Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere, J. Geophys. Res., 104, 12099–12111, 1999.

Matsumoto, M. and Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Trans. Model. Comput. Simul., 8, 3–30, 1998.

Nissen, K. M., Matthes, K., Langematz, U., and Mayer, B.: Towards a better representation of the solar cycle in general circulation models, Atmos. Chem. Phys., 7, 5391–5400, doi:10.5194/acp-7-5391-2007, 2007.

Novelli, P. C., Masarie, K. A., and Lang, P. M.: Distributions and recent changes in carbon monoxide in the lower troposphere, J. Geophys. Res., 103, 19015–19033, 1998.

Pandow, M., MacKay, C., and Wolfgang, R.: The reaction of atomic carbon with oxygen: Significance for the natural radio-carbon cycle, J. Inorg. Nucl. Chem., 14, 153–158, 1960.

Pickering, K. E., Wang, Y., Tao, W.-K., Price, C., and Müller, J.-F.: Vertical distribution of lightning NO_X for use in regional and chemical transport models, J. Geophys. Res., 103, 31203–31216, 1998.

Pozzer, A., Jöckel, P., Sander, R., Williams, J., Ganzeveld, L., and Lelieveld, J.: Technical Note: The MESSy-submodel AIRSEA calculating the air-sea exchange of chemical species, Atmos. Chem. Phys., 6, 5435–5444, doi:10.5194/acp-6-5435-2006, 2006.

Pozzer, A., Jöckel, P., Tost, H., Sander, R., Ganzeveld, L., Kerkweg, A., and Lelieveld, J.: Simulating organic species with the global atmospheric chemistry general circulation model ECHAM5/MESSy1: a comparison of model results with observations, Atmos. Chem. Phys., 7, 2527–2550, doi:10.5194/acp-7-2527-2007, 2007.

Pozzer, A., Jöckel, P., Kern, B., and Haak, H.: The Atmosphere-Ocean General Circulation Model EMAC-MPIOM, to be submitted to Geosci. Model Dev. Discuss., 2011.

Preiss, N. and Genthon, C.: Use of a new database of lead 210 for global aerosol model validation, J. Geophys. Res., 102, 25347–25357, 1997.

Preiss, N., Mélières, M.-A., and Pourchet, M.: A compilation of data on lead 210 concentration in surface air and fluxes at the air-surface and water-sediment interfaces, J. Geophys. Res., 101, 28847–28862, 1996.

Pressyanov, D. S.: Short solution of the radioactive decay chain equations, Am. J. Phys., 70, 444–445, 2002.

Price, C. and Rind, D.: A simple Lightning Parametrization for Calculating Global Lightning Distributions, J. Geophys. Res., 97, 9919–9933, 1992.

Price, C. and Rind, D.: Modeling Global Lightning Distributions in a General Circulation Model, Mon. Weather Rev., 122, 1930–1939, 1994.

Pringle, K. J., Tost, H., Message, S., Steil, B., Giannadaki, D., Nenes, A., Fountoukis, C., Stier, P., Vignati, E., and Lelieveld, J.: Description and evaluation of GMXe: a new aerosol sub-

model for global simulations (v1), Geosci. Model Dev., 3, 391–412, doi:10.5194/gmd-3-391-2010, 2010.

Randerson, J. T., van der Werf, G. R., Giglio, L., Collatz, G. J., and Kasibhatla, P. S.: Global Fire Emissions Database, Version 2 (GFEDv2.1), Tech. rep., Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi:10.3334/ORNLDAAC/849, http://daac.ornl.gov/, 2007.

Rehfeld, S. and Heimann, M.: Three dimensional atmospheric transport simulation of the radioactive tracers ^{210}Pb, ^{7}Be, ^{10}Be, and ^{90}Sr, J. Geophys. Res., 100, 26141–26161, 1995.

Röckmann, T., Jöckel, P., Gros, V., Bräunlich, M., Possnert, G., and Brenninkmeijer, C. A. M.: Using ^{14}C, ^{13}C, ^{18}O and ^{17}O isotopic variations to provide insights into the high northern latitude surface CO inventory, Atmos. Chem. Phys., 2, 147–159, doi:10.5194/acp-2-147-2002, 2002.

Roeckner, E., Brokopf, R., Esch, M., Giorgetta, M., Hagemann, S., Kornblueh, L., Manzini, E., Schlese, U., and Schulzweida, U.: Sensitivity of simulated climate to horizontal and vertical resolution in the ECHAM5 atmosphere model, J. Climate, 19, 3771–3791, 2006.

Sander, R., Kerkweg, A., Jöckel, P., and Lelieveld, J.: Technical note: The new comprehensive atmospheric chemistry module MECCA, Atmos. Chem. Phys., 5, 445–450, doi:10.5194/acp-5-445-2005, 2005.

Sander, S. P., Friedl, R. R., Golden, D. M., Kurylo, M. J., Moortgat, G. K., Keller-Rudek, H., Wine, P. H., Ravishankara, A. R., Kolb, C. E., Molina, M. J., Finlayson-Pitts, B. J., Huie, R. E., and Orkin, V. L.: Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation Number 15, JPL Publication 06-2, Jet Propulsion Laboratory, Pasadena, CA, 2006.

Sander, R., Baumgaertner, A., Gromov, S., Harder, H., Jöckel, P., Kerkweg, A., Kubistin, D., Regelin, E., Riede, H., Sandu, A., Taraborrelli, D., Tost, H., and Xie, Z.-Q.: The atmospheric chemistry box model CAABA/MECCA-3.0, to be submitted to Geosci. Model Dev. Discuss., 2011.

Sandu, A. and Sander, R.: Technical note: Simulating chemical systems in Fortran90 and Matlab with the Kinetic PreProcessor KPP-2.1, Atmos. Chem. Phys., 6, 187–195, doi:10.5194/acp-6-187-2006, 2006.

Schery, S. D. and Wasiolek, M. A.: Modeling radon flux from the earth's surface, in: Radon and Thoron in the Human Environment, Proceedings of the 7 th Tohwa University International Symposium, edited by: Katase, A. and Shimo, M., World Scientific, Singapore, 1998.

Schumann, U. and Huntrieser, H.: The global lightning-induced nitrogen oxides source, Atmos. Chem. Phys., 7, 3823–3907, doi:10.5194/acp-7-3823-2007, 2007.

Seigneur, C. and Lohman, K.: Effect of bromine chemistry on the atmospheric mercury cycle, J. Geophys. Res., 113D, D23309, doi:10.1029/2008JD010262, 2008.

Selin, N. E., Jacob, D. J., Yantosca, R. M., Strode, S., Jaeglé, L., and Sunderland, E. M.: Global 3-D land-ocean-atmosphere model for mercury: Present-day versus preindustrial cycles and anthropogenic enrichment factors for deposition, Glob. Biogeochem. Cy., 22, GB2011, doi:10.1029/2007GB003040, 2008.

Tost, H.: Global Modelling of Cloud, Convection and Precipitation Influences on Trace Gases and Aerosols, Ph.D. thesis, University of Bonn, Germany, http://hss.ulb.uni-bonn.de/diss_online/math_nat_fak/2006/tost_holger), 2006.

Tost, H., Jöckel, P., Kerkweg, A., Sander, R., and Lelieveld, J.: Technical note: A new comprehensive SCAVenging submodel for global atmospheric chemistry modelling, Atmos. Chem. Phys., 6, 565–574, doi:10.5194/acp-6-565-2006, 2006a.

Tost, H., Jöckel, P., and Lelieveld, J.: Influence of different convection parameterisations in a GCM, Atmos. Chem. Phys., 6, 5475–5493, doi:10.5194/acp-6-5475-2006, 2006b.

Tost, H., Jöckel, P., and Lelieveld, J.: Lightning and convection parameterisations – uncertainties in global modelling, Atmos. Chem. Phys., 7, 4553–4568, doi:10.5194/acp-7-4553-2007, 2007.

Usoskin, I. G., Alanko-Huotari, K., Kovaltsov, G. A., and Mursula, K.: Heliospheric modulation of cosmic rays: Monthly reconstruction for 1951–2004, J. Geophys. Res., 110, A12108, doi:10.1029/2005JA011250, 2005.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.

Vignati, E., Wilson, J., and Stier, P.: M7: An efficient size-resolved aerosol microphysics module for large-scale aerosol transport models, J. Geophys. Res., 109, D22202, doi:10.1029/2003JD004485, 2004.

Xie, Z.-Q., Sander, R., Pöschl, U., and Slemr, F.: Simulation of atmospheric mercury depletion events (AMDEs) during polar springtime using the MECCA box model, Atmos. Chem. Phys., 8, 7165–7180, doi:10.5194/acp-8-7165-2008, 2008.

Zhang, K., Wan, H., Zhang, M., and Wang, B.: Evaluation of the atmospheric transport in a GCM using radon measurements: sensitivity to cumulus convection parameterization, Atmos. Chem. Phys., 8, 2811–2832, doi:10.5194/acp-8-2811-2008, 2008.