Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: an ensemble approach applied to ozone modeling

Vivien Mallet and Bruno Sportisse

Cerea: Teaching and Research Center in Atmospheric Environment
Joint laboratory École Nationale des Ponts et Chaussées / Électricité de France R&D
http://www.enpc.fr/cerea/
Clime
Joint team Inria / École Nationale des Ponts et Chaussées
http://www-rocq.inria.fr/clime/

Abstract. This paper estimates the uncertainty in the outputs of a chemistry-transport model due to physical parameterizations and numerical approximations. An ensemble of twenty simulations is generated from a reference simulation in which one key parameterization (chemical mechanism, dry deposition parameterization, turbulent closure, etc.) or one numerical approximation (grid size, splitting method, etc.) is changed at a time. Intercomparisons of the simulations and comparisons with observations allow us to assess the impact of each parameterization and numerical approximation, and the robustness of the model. An ensemble of sixteen simulations is also generated with multiple changes in the reference simulation in order to estimate the overall uncertainty. The case study is a four-month simulation of ozone concentrations over Europe in 2001 performed using the modeling system Polyphemus. It is shown that there is a high uncertainty due to the physical parameterizations (notably the turbulence closure and the chemical mechanism). The low robustness suggests that ensemble approaches are necessary in most applications.

1. Introduction

Chemistry-transport models are now widely used in air-quality applications ranging from impact studies to daily forecasts. To date, they perform satisfactory simulations, in both basic cases such as passive tracer tracking [e.g., *Nodop*, 1997] and in complex cases involving photochemical mechanisms. The reliability of the models is partially assessed through comparisons with measurements and numerous statistical-measures [as those defined in *US EPA*, 1991]. These comparisons are performed with intensive observation periods from specific campaigns or with daily measurements from regular monitoring sites. A large set of comprehensive and reliable 3D Eulerian chemistry-transport models has been "validated" this way, such as Chimere [*Schmidt et al.*, 2001], CMAQ [Community Multiscale Air Quality, *Byun and Ching*, 1999], DEHM [Danish Eulerian Hemispheric Model, *Christensen*, 1997], EMEP [European Monitoring and Evaluation Programme, *Simpson et al.*, 2003], Eurad [European Air Pollution Dispersion, *Hass*, 1991], Lotos [Long Term Ozone Simulation, *Builtjes*, 1992], Polair3D [*Boutahar et al.*, 2004].

These models have usually been "tuned" in order to deliver satisfactory model-to-observation statistics. Also while the "validations" give the error of the simulations, they do not give information on the uncertainty associated with these simulations. The origin of the uncertainty is three-fold: the underlying physical parameterizations (biogenic emissions, deposition velocities, turbulent closure, chemical mechanism, etc.), the input data (land use data, emission inventories, raw meteorological fields, chemical data, etc.) and the numerical approximations (mesh sizes, time step and number of chemical species). The best characterization of the uncertainty would be the probability density functions of the simulation errors. Computing a probability density function (PDF) for given model outputs (such as forecast error statistics) is in practice a difficult task primarily because of the computational costs.

There are specific techniques to assess uncertainties. The first-order derivatives of model outputs with respect to model inputs can give "local" estimates of uncertainties [e.g., *Schmidt*, 2002]. Monte Carlo simulations based on different values for given input parameters or fields can provide an approximation to the probability density functions if the number of simulations is large enough [*Hanna et al.*, 2001]. An alternative approach, which is now widely used in meteorology [*Toth and Kalnay*, 1993; *Houtemaker et al.*, 1996; *Buizza et al.*, 1999] and which is a promising method in air quality modeling (e.g., *Delle Monache and Stull* [2003] for photochemistry or *Galmarini et al.* [2004] for radionuclides), is the so-called ensemble approach based on a set of models supposed to account for the range of uncertainties.

This paper uses an ensemble approach to provide estimates of the uncertainty in photochemical forecasts due to the parameterizations and some data associated with them. It also deals with numerical issues such as mesh size.

The study is performed with a four-month European-scale simulation, from May to August 2001. A comparison between the reference simulation and a similar simulation but for one change in a parameterization enables us to estimate the impact of this parameterization. For each modified parameterization, the reliability of the simulation is checked with comparisons to measurements, which allows us to assess the robustness of the whole modeling system. The same experiment is finally performed with a set of simulations in which several parameterizations may be changed (at the same time, in the same simulation). It allows us to study the robustness of the system with respect to cumulated uncertainties.

1

This paper is organized as follows. Section 2 briefly summarizes the relevant methods to estimate uncertainties, details the specific aims of this paper and describes the methodology. Section 3 details the model, the reference simulation and the involved parameterizations. In the last section, the results are analyzed with intercomparisons of the simulations and comparisons to observations.

2. Methodology

2.1. Definitions

We define:

• The error. It is the discrepancy between model outputs and field observations.

• The uncertainty. It is the range of values in which the model outputs may lie with a high degree of confidence. In this paper, we only deal with a priori uncertainties, i.e. uncertainties estimated without taking into account observations.

• The spread. Hereafter we refer to the variability of an ensemble as its spread. The spread is a measure of the uncertainty and it can be quantified by a standard deviation.

• The variability. Herein the variability solely refers to the spatial or/and temporal variabilities of a concentration field. For the sake of clarity, the variability of an ensemble is called a spread.

2.2. Motivation

Assessing the uncertainties in model outputs is a field of growing interest in environmental forecasting, especially in meteorology. In meteorology, the dynamics of models have a "chaotic" behavior. The uncertainties in initial conditions have therefore a strong impact and the issue is to propagate these uncertainties through "ensemble forecasts" [*Toth and Kalnay*, 1993; *Houtemaker et al.*, 1996; *Buizza et al.*, 1999]. In air quality applications, there is not such a strong dependence on initial conditions. The impact of uncertainties in the input data (e.g., emissions, meteorological fields), in the parameterizations (e.g., deposition velocities, turbulence closure) and in the numerical algorithms is much stronger.

The actual errors of a model, given by comparisons to observational data, may be low with high uncertainties in the results. A model may be tuned to fit the observations (and all models are improved this way), which leads to low errors. Nevertheless, if this model is used with different parameterizations (assumed to be valid physical parameterizations), other data or alternative numerical schemes, then it could lead to very different results, including those far from the measurements, with the magnitude of spread depending on the actual uncertainty. This is, of course, a strong limitation of the models, and the uncertainty has to be estimated in order to assess the "robustness" of the models. One may refer to *Russell and Dennis* [2000] for an overview of the strengths and limitations of photochemical models.

It is impossible to compute the error in all meteorological conditions, at every point in a given simulation domain (even at ground level), for all chemical species, and at every time. In the absence of observations, an estimation of the uncertainty is essentially the only means to assess the quality of the results. In an operational context, the models may be used for risk assessment. The reliability of the results is then a crucial issue and, if available, the full PDFs associated with these results would be highly valuable. For instance, in prospective or screening studies (e.g., impact studies related to different emission scenarios), the models may be used with uncommon input data (e.g., strongly corrected emissions) and without any available observations with which to tune the models. From the research point of view, an estimate of the uncertainty is necessary for other communities to assess the feasibility and the relevance of given applications. For instance, the effect of pollution on health may or

may not be effectively estimated, depending on the accuracy of the underlying air-quality models. For each model, the development is also oriented to improve the description in the parameterizations responsible for the main uncertainty.

2.3. A review of existing methods

There are several methods to estimate the uncertainty and to identify its sources. As for the uncertainty due to the input data, one can compute first-order derivatives of the model outputs with respect to the model inputs [e.g., *Schmidt*, 2002]. This provides "local" sensitivities from which the uncertainty in the outputs can be derived, taking into account the uncertainty in the input data.

Ideally one would want to compute the full PDF associated with the results. It would mean solving the Fokker-Planck equation [the equation satisfied by the output PDF, *Gardiner*, 1996] which is unfeasible. Instead, the Monte Carlo methods can generate approximations of the PDF. The idea is to generate a set of N input fields that roughly describe the PDF associated with the input data. The model is then run N times, which provides an approximation of the output PDF. These methods may be well suited but they are restricted to the uncertainty due to input data or parameters in parameterizations, that is, due to continuous variables. A related method, which could be viewed as a Monte Carlo method too, is the use of a set of N input fields generated by another model. In practice, the ensemble forecasts from the meteorological centers may be used as input to the air quality models. It leads to promising applications but it is restricted to the meteorological fields [*Warner et al.*, 2002].

Another method is the use of different air quality models. This technique has already been used but with a fairly low number of models [e.g., four models in *Delle Monache and Stull*, 2003]. It is hard to assemble enough models to claim a reliable estimate of the uncertainty. Moreover intercomparisons are difficult because the models may not be operated under the same conditions (e.g. with the same meteorological fields). Note that this technique involves the uncertainties of several models and is not suited to assess the uncertainty of a given model. Moreover the models have usually been tuned in comparisons to measured data; hence they do not embrace the whole uncertainty in the physics and the chemistry.

The method applied in this paper mainly takes advantage of the multiple parameterizations that should be available in a well designed modeling system [*Mallet et al.*, 2005]. The model is run in many configurations with respect to the available state-of-the-art parameterizations, but also with respect to changes in the parameters and the base input-data needed for these parameterizations. The impact of the numerical approximations is studied as well. This method allows fair comparisons since the framework is exactly the same for all simulations. It gives an accurate view of the uncertainty due to the parameterizations of a given model. Notice that the method introduces discrete changes, which is the only means to assess the impact of the parameterizations. There is no continuous transition between two parameterizations or between their base input-data sets. Details about the method are provided below.

2.4. The multi-configurations approach

The air quality system with which the experiments have been performed relies on many parameterizations (see Section 3.1). There are often several valid parameterizations to compute the same field. Furthermore most parameterizations depend on input-data sets (including scalar parameters). For instance, the deposition velocities depend on the

land use coverage which may be given by USGS[1] data or by GLCF[2] data (see below). The alternatives between the parameterizations themselves and their input-data sets introduce a finite number of choices. Hence the method deals with discrete dependencies.

The impact of numerical options are also assessed through discrete changes, e.g. by changing a numerical scheme. Nonetheless a few values that belong to a continuous interval are studied as well. They are modified as if they were discrete variables, i.e. only a few values are allowed for them. For example, the time step is a continuous variable but it can be restricted to a set of three values (a reference time step, a small one and a large one).

For the sake of clarity, the changes in the input data to the parameterizations will be viewed as changes in the parameterizations themselves. Since the numerical issues are treated in the same way as the parameterizations (they are associated with a finite number of choices), they are also viewed as parameterizations hereafter.

Assume that the model is written in the form:

$$y = f(p_1, p_2, ..., p_N) = f(p) \qquad (1)$$

Every input parameter $p_i \in \{0, ..., n_i - 1\}$ is associated with a given parameterization that has n_i possible values. f is the model itself. The output y may be the pollutant concentrations, deposition fields, evaluation statistics, etc. Notice that f is already a discretized model.

The reference simulation is associated with a reference vector assumed to be zero: $p_{\text{ref}} = 0$. The idea is to estimate the uncertainty and the impact of every parameterization by changing one parameterization at a time, i.e. computing all $f(p)$ where $p_i = 0$ for all i except for one component. There are $\sum_{i=1}^{N}(n_i - 1)$ such simulations. This is only a small subset of the $\Pi_{i=1}^{N} n_i$ possible combinations, but the computational cost makes it impossible to run all simulations.

This method allows us to estimate the impact of each parameterization. The impact is estimated with the resulting changes in the output concentrations. It is analyzed with the concentration distributions and their spatial and temporal variabilities. In addition, for each change, an evaluation of the output can be performed. It shows whether the modified parameterization leads to an improved agreement with the measurements and, therefore, maybe to a better description of the physics. The fact that not all combinations $(p_i)_i$ are available restricts the study: it is hard to decide whether a parameterization should be discarded because its drawbacks may be canceled by changes in other parameterizations. There are still useful conclusions to draw: for instance, it may be shown that a given parameterization limits the variability in the results.

Furthermore, the results are enhanced by combined changes, but only with a few selected parameterizations to reduce the computational cost of the study. Four parameterizations are selected mainly due to their significant impact (even if other parameterizations have a similar importance). The model is then put in the form $y = f(\tilde{p})$ where the vector $\tilde{p}$ has four components. Each component can take two values (0 or 1); therefore there are 16 possible combinations. It provides a rough estimate of the overall uncertainty.

3. The Experiments Setup

3.1. The Modeling System

This study is based on the modeling system Polyphemus (available under the GNU General Public License at `http://www.enpc.fr/cerea/polyphemus/`). This system is divided into four parts:

1. the databases: they incorporate the data needed in the parameterizations (one may also include the meteorological fields here).

2. the libraries: they provide (1) facilities to manage the multidimensional data involved in atmospheric chemistry, (2) useful functions associated with the physical and chemical fields (e.g. coordinate transformations) and (3) the parameterizations.

3. a set of programs: these programs make the calls to the libraries to generate the input data needed by the chemistry-transport model. Their flexibility is made possible by the input configuration files that they read.

4. the chemistry-transport model: it is responsible for the time integration of the chemistry-transport equation. It therefore computes the output concentrations.

The databases contain the raw data: the land use coverage, the anthropogenic emission inventories, chemical constants, etc. The meteorological fields may also be included even if they strongly depend on the application.

The libraries play a major role in this study since they provide the basis of the flexibility of the parameterizations. They first provide the data structures and functions needed for data processing. They then provide a set of parameterizations. Most of the changes to the simulations are made with a different call to the libraries, specifically to the library dedicated to physical parameterizations, the C++ library AtmoData [*Mallet and Sportisse*, 2005].

The programs of Polyphemus make calls to the libraries in order to process the raw data. They format the raw data for the chemistry-transport model, but the primary function of the programs is to use the parameterizations from the library AtmoData to compute the needed fields. These programs read configuration files in which many options are specified, including which parameterizations are to be used and with which input data and parameters. Roughly speaking, there exists a set of configuration files for every vector p (vector defined in Section 2.4). This study therefore relies heavily on the flexibility characteristic of the programs.

Finally, the Eulerian chemistry-transport model Polair3D computes the output concentrations through the numerical integration of the transport-chemistry equation. With respect to this study, a strong advantage of Polair3D is its ability to deal with multiple chemical mechanisms. Details about Polair3D may be found in *Boutahar et al.* [2004].
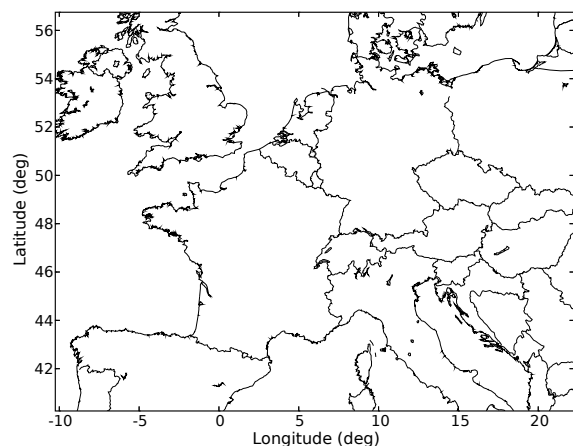

Figure 1. Domain $[40.25°N, 10.25°W] \times [56.75°N, 22.25°E]$ of the reference simulation.

Further details about the architecture of the whole system may be found in *Mallet et al.* [2005]. A complete description is not relevant here because not all features of the system are used in this study. The system is able to handle many applications (many chemical mechanisms, data assimilation, Monte Carlo simulations, etc.) and its flexibility enables the multiple experiments presented in this paper. The next subsection describes the base application.

3.2. The Reference Simulation

The impact of the parameterizations is evaluated by the changes they introduce with respect to the reference simulation. The reference simulation takes place at European scale during summer 2001 (22 April 2001 to 31 August 2001). A validation, over the same domain and the same period, similar to the reference simulation, may be found in *Mallet and Sportisse* [2004].

The domain is $[40.25°N, 10.25°W] \times [56.75°N, 22.25°E]$ and is shown in Figure 1. The first layer is located between 0 m and 50 m; the concentrations are thus computed at 25 m. The thickness of the other layers is about 600 m with the top of the last layer at 3000 m. RACM is the photochemical mechanism used in this simulation [*Stockwell et al.*, 1997]. Since the best results are obtained for ozone and the number of ozone measurements is significantly higher than for other species, this study focuses on ozone. We are notably concerned with the ozone peaks since they are often of high interest in forecasts (due to the regulations that mostly limit the peaks).

Here is a review of the main components of the reference simulation:

1. meteorological data: the best ECMWF data available for the period (i.e. $0.36° \times 0.36°$, the TL511 spectral resolution in the horizontal, 60 levels, time step of 3 hours, 12 hours forecast-cycles starting from analyzed fields);

2. land use coverage: USGS finest land cover map (24 categories, 1 km Lambert);

3. emissions: the EMEP[3] inventory, converted according to *Middleton et al.* [1990];

4. biogenic emissions: computed as advocated in *Simpson et al.* [1999];

5. deposition velocities: the revised parameterization proposed in *Zhang et al.* [2003];

6. vertical diffusion: within the boundary layer, the Troen and Mahrt parameterization as described in *Troen and Mahrt* [1986], with the boundary-layer height provided by the ECMWF; above the boundary layer, the Louis parameterization [*Louis*, 1979];

7. boundary conditions: output of the global chemistry-transport model Mozart 2 [*Horowitz et al.*, 2003] run over a typical year;

8. numerical schemes: a first-order operator splitting, the sequence being advection–diffusion–chemistry; a direct space-time third-order advection scheme with a Koren flux-limiter [*Verwer et al.*, 1998]; a second-order order Rosenbrock method for diffusion and chemistry.

The performance of the reference simulation has been evaluated through a comparison of the forecasted ozone peaks with the observations from 242 stations distributed over Europe (in a network with mixed stations: urban, peri-urban and rural stations). With the first five days excluded (because of the rough initial conditions), the root mean square (with all observations put together) is $23.5 \mu g \cdot m^{-3}$, the correlation is 71.4% and the bias $-4.5 \mu g \cdot m^{-3}$ (the mean of observed values being $94.7 \mu g \cdot m^{-3}$) – the statistical measures are defined in the appendix. The results therefore show a reasonable agreement with observations [*Hass et al.*, 1997; *Schmidt et al.*, 2001].

3.3. The Parameterizations

The modified parameterizations were chosen according to the relevance and the availability of alternative parameterizations. Only state-of-the-art parameterizations or, at least,

widely used parameterizations are involved. The list of the parameterizations (and the data associated with them) used in this study is shown in Table 1.

The changes first include prominent processes such as the chemistry (RADM 2). Several chemical mechanisms are available in Polair3D but reliable emission inventories were available only for RACM and for RADM 2. The same speciation [for volatile organic compounds, *Passant*, 2002] was used for the two mechanisms. A drawback is that both mechanisms are too close to embrace the diversity of the chemical mechanisms available in air quality modeling. Nevertheless, as shown hereafter, there is a substantial difference between the two mechanisms.

The sensitivity to the turbulence closure is assessed with the comparison between the Troen & Mahrt parameterization (well suited for models with a low vertical resolution – which is the case with only five layers) and the Louis parameterization. The Louis closure is used above the boundary layer (for all simulations), in the boundary layer in stable conditions (simulation 3) and in any condition (simulation 2). One should note that the leading contribution to ground concentrations comes from the vertical diffusion coefficient at the top of the first layer. It determines the transfer between this and the above layer (up to 600m) which roughly corresponds to the residual layer in the night. The Troen & Mahrt parameterization and the Louis parameterization are designed in two different ways: the first one is independent of the vertical discretization while the second one relies on finite differences. There is a clear difference in the coefficients computed by the two parameterizations: the averages at the top of the first layer are $7.6 m^2 \cdot s^{-1}$ (Troen & Mahrt) and $5.7 m^2 \cdot s^{-1}$ (Louis). The correlation of 60% also shows the gap between the two parameterizations for coarse vertical discretizations (the differences decrease as the vertical mesh is refined).

Another known important process with multiple parameterizations is dry deposition. An alternative to the reference velocities computed as proposed in *Zhang et al.* [2003] is based on the method by *Wesely* [1989] (simulation 4) which includes a reasonable parameterization and is widely used. The two parameterizations rely on the same fundamentals and the differences in the computed deposition velocities come as much from the input data (resistances, land use descriptions) as from the parameterization itself. As for ozone deposition velocities, the relative bias between the two parameterizations is only 3%, the correlation is 96% but the ratio of the standard deviation of the difference and the mean velocity is high: 0.31. In addition, the surface flux used to compute the aerodynamic resistance can be the heat flux or the momentum flux, although the heat flux is usually assumed to be more suitable for a scalar variable such as the concentration of a pollutant.

The two last physical-parameterizations (simulations 6 and 7) deal with the attenuation coefficients. The reference option is based on the optical depth, as described in *Chang et al.* [1987] and *Madronich* [1987], estimated with the cloud liquid water content. The liquid water content is integrated within the clouds and the cloud fraction is calculated based on the relative humidity q and its critical value q_c:

$$\text{cloud fraction} = \frac{1 - q}{1 - q_c} \tag{2}$$

$$q_c = 1 - \alpha \sigma^a (1 - \sigma)^b \left(1 + \beta \left(\sigma - \frac{1}{2}\right)\right) \tag{3}$$

Table 1. Parameterizations, raw input data and numerical choices for the reference simulation and their alternatives. The impact of the parameterizations are assessed in this study through the use of the alternatives shown in this table.

#	Parameterization	Reference	Alternative(s)	Comment
			Physical parameterizations	
1.[a]	Chemistry	RACM	RADM 2 [*Stockwell et al.*, 1990]	
2.	Vertical diffusion	Troen & Mahrt	Louis [*Louis*, 1979]	
3.			Louis in stable conditions	Troen & Mahrt kept in unstable conditions
4.	Deposition velocities	Zhang [*Zhang et al.*, 2003]	Wesely [*Wesely*, 1989]	
5.	Surface flux	Heat flux[b]	Momentum flux[b]	For the aerodynamic resistance (in deposition velocities)
6.	Cloud attenuation	RADM method [*Chang et al.*, 1987; *Madronich*, 1987]	Esquif (*ESQUIF* [2001])	
7.	Critical relative humidity	Depends on σ	Two layers	Used in the RADM method to compute cloud attenuation
			Input data	
8.	Emissions vertical distribution	All in the first cell	All in the two first cells	
9.	Land use coverage	USGS	GLCF	For deposition velocities[c]
10.	Land use coverage	USGS	GLCF	For biogenic emissions[c]
11.	Exponent p in Troen & Mahrt	2	3	
12.	Photolytic constants	JPROC	Depends on the zenith angle (only)	
			Numerical issues	
13.	Time Step	600 s	100 s	
14.			1800 s[d,e]	
15.	Splitting method	First order	Strang splitting	
16.	Horizontal resolution	0.5°	0.1° [e]	
17.			1.0°	
18.	Vertical resolution	5 layers	9 layers	The first layer height remains 50 m
19.	First layer height	50 m	40 m	The top of every other layer does not change

[a] The reference simulation will be referred to as simulation #0.

[b] Computed using the Louis formulae.

[c] The consistency between the land use coverage used for the deposition velocities and for the biogenic emissions is not required. Indeed a large part of the uncertainty lies in the data associated with the land use categories (e.g. resistances for deposition and emission factors for the biogenic emissions). Moreover a given description may be more suited only for the emissions (vegetation) or only for the deposition (roughness, etc.).

[d] The advection is integrated over submultiples of 1800 s so as to satisfy the CFL (Courant-Friedrichs-Lewy) condition.

[e] The numerical scheme is also slightly modified in this simulation: it uses source splitting. It is used to enforce the stability but has only slight consequences in the results.

where $\sigma = \frac{P}{P_s}$, P is the pressure, P_s is the surface pressure, $\alpha = 1.1$, $\beta = \sqrt{1.3}$, $a = 0$ and $b = 1.1$. In an alternative simulation (#7), the critical relative humidity is simply constant over two distinct layers: $q_c = 0.75$ below $700 hPa$ and $q_c = 0.95$ above.

Another set of simulations is derived from changes in the input data. The land use coverage is described by the USGS data (24 categories, 1 km Lambert) or by the GLCF data (14 categories, $0.0083°$). The GLCF data contain less categories worldwide but they involve more categories over Europe than the USGS data do. The impact of the land use description is assessed through the deposition velocities (simulation 9) and the biogenic emissions (simulation 10) independently.

In the Troen & Mahrt parameterization, the vertical diffusion coefficients depend on several parameters, particularly an exponent p [see *Troen and Mahrt*, 1986] which determines the shape of the vertical profile. Since it is a free parameter (with $p = 2$ or $p = 3$ recommended), the exponent is set to 2 in the reference simulation and also set to 3 as an alternative (which increases the diffusion coefficients – simulation 11).

The emission inventories are a concern of most modelers, especially the time and spatial distributions associated with them. The time distribution is known to have a slight impact at continental scale [*Tao et al.*, 2004]. The horizontal distribution is given with the EMEP inventory. Meanwhile the vertical distribution is not well known and is chosen by the modeler. In the reference simulation, all emissions are released in the first layer (therefore below 50 m). In an alternative simulation (#8), the emissions from industrial combustion (sectors 1 and 3 in the EMEP inventory) and from the waste treatment (sector 9 in the EMEP inventory) are released in the second layer. The emissions due to the combustion of non-industrial plants (sector 2 in the EMEP inventory) are released in both layers (one half in each).

Finally the impact of the database for the photolytic constants is estimated. The reference simulation takes advantage of the photolytic constants computed by JPROC [part of CMAQ, *Byun and Ching*, 1999] that are a function of the latitude, the altitude, the day in the year and the hour angle. The alternative simulation uses a coarser description with a single dependence on the zenith angle derived from the values given in *Stockwell et al.* [1997].

The last set of simulations involves changes in the numerical approximations. The time step is set to 100 s (simulation 13) and 1800 s (simulation 14) instead of 600 s (reference simulation). In the reference simulation, the splitting method is a first-order method (advection–diffusion–chemistry). An alternative simulation (#15) takes advantage of the Strang splitting method (advection–chemistry–diffusion over $\frac{\Delta t}{2}$ and then diffusion–chemistry–advection still over $\frac{\Delta t}{2}$), *Sportisse* [2000]. Finally the spatial discretization is changed horizontally (simulations 16 and 17) and vertically (simulation 18 and 19). When the spatial discretization changes, the raw meteorological fields (ECMWF fields) are interpolated on the new grid.

The nineteen alternative simulations address a reasonable range of the choices that can be made in a forecasting system, in the parameterizations, their input data and the numerical options.

4. Results and discussion

Due to the coarse initial conditions, the five first days of the simulations are excluded in the following comparisons. The comparison is limited to hourly ozone concentrations in the first layer. Moreover the comparisons between the computed fields (not with the observations) are not performed in all cells to avoid the influence of the boundary conditions. A three-cell band at the domain borders is excluded from the comparisons in Subsections 4.1, 4.2 and 7.

The first subsection compares all simulations in order to estimate the spread due to the parameterizations (and numerical choices). The second subsection focuses on the impact of each parameterization. The comparisons are relative to the reference simulation. In the third subsection, a comparison with the observations evaluates the impact of the parameterizations on the forecasts. In the last section, the impact of combined changes in the parameterizations is performed to give an estimate of the overall uncertainty.

Table 2. Means and standard deviations of the hourly ozone concentrations ($\mu g \cdot m^{-3}$) of the twenty simulations. The reference simulation is indexed by 0 and the other simulations are indexed as in Table 1. On the left, the simulations are sorted by their mean; on the right, they are sorted by their standard deviation. The relative standard deviation of the means is 5.5%, and the relative standard deviation of the standard deviations is 10.4%.

| | *Sorted by mean* | | | *Sorted by standard deviation* | |
#	Mean	Standard deviation	#	Mean	Standard deviation
15	90.30	25.73	2	68.94	34.23
18	89.15	24.27	3	79.15	28.83
14	87.49	26.14	11	84.14	26.22
7	85.92	25.16	16	83.27	26.20
6	85.92	24.75	4	85.18	26.19
4	85.18	26.19	14	87.49	26.14
17	85.15	24.74	15	90.30	25.73
0	84.92	25.11	19	82.62	25.22
13	84.73	24.99	7	85.92	25.16
8	84.23	21.81	0	84.92	25.11
11	84.14	26.22	12	84.00	25.06
12	84.00	25.06	13	84.73	24.99
9	83.96	24.39	5	81.38	24.87
16	83.27	26.20	6	85.92	24.75
19	82.62	25.22	17	85.15	24.74
5	81.38	24.87	9	83.96	24.39
10	81.30	22.88	18	89.15	24.27
3	79.15	28.83	10	81.30	22.88
1	77.11	21.14	8	84.23	21.81
2	68.94	34.23	1	77.11	21.14

Table 3. Percentiles of ozone concentrations ($\mu g \cdot m^{-3}$), sorted by the mean of the percentiles.

#	10th	20th	30th	40th	50th	60th	70th	80th	90th
15	58	69	77	84	91	97	104	111	122
18	60	69	76	83	88	94	101	109	120
14	55	66	74	81	87	94	101	109	121
6	55	66	74	80	86	92	98	106	117
7	55	66	73	80	86	92	98	106	117
4	53	64	72	79	85	91	98	107	118
17	55	65	72	79	85	91	97	105	117
0	54	65	72	79	85	91	97	105	116
13	54	65	72	79	84	90	97	105	116
8	57	66	73	79	84	89	95	102	111
9	54	65	72	78	84	90	96	104	114
11	52	63	71	77	84	90	97	105	117
12	53	64	71	78	84	90	96	104	115
16	52	63	71	77	83	89	96	104	115
19	51	63	70	77	83	89	95	103	114
10	52	63	70	76	82	88	94	100	110
5	50	61	69	75	81	87	94	101	112
3	41	55	65	73	80	87	95	103	115
1	50	61	68	73	78	84	89	95	102
2	28	40	50	58	67	75	85	96	111

4.1. Intercomparison between the Simulations

The distribution of the spatio-temporal means and standard deviations of the fields is shown in Table 2. The means and the standard deviations are well spread considering that the simulations differ only at most in two parameterizations. The mean is particularly affected by the splitting method, the number of layers, the time step (1800 s), the chemical mechanism and the vertical diffusion. The standard deviation increases due to the turbulent closure and decreases with RADM 2, the vertically distributed emissions and the land use coverage used to compute the biogenic emissions. From these comparisons, it appears that the turbulent closure and the chemical mechanism have a strong impact on the output ozone concentrations. Even the use of the Louis closure only in stable conditions modifies both the mean and the standard deviation. The numerical issues also have a clear impact on the ozone mean. Finally the emissions can modify the standard deviation of the output concentrations. These conclusions may already be known issues, but this study shows the prominent impacts.

Notice that the standard deviation has a greater spread than the mean. The relative standard deviation[4] of the means and of the standard deviations shown in Table 2 are 5.5% and 10.4% respectively.

More details are provided by the distribution of each ozone field. The percentiles associated with the simulations are shown in Table 3. In addition, Figure 2 shows the relative frequency distributions of:

1. simulation 15 (splitting order) which provides the highest concentrations (highest mean and highest percentiles) with a standard deviation close to the reference simulation;

2. simulation 2 (Louis turbulence closure) which has the lowest concentrations and the highest standard deviation;

3. simulation 1 (chemical mechanism RADM 2) which is associated with the lowest standard deviation.

These three simulations exhibit the most extreme behavior and therefore give a good idea of the uncertainty due to the changes in the parameterizations.

The comparisons deal with the concentrations computed over the whole (restricted) domain and at all time steps. A finer analysis of the variability deals with the spatial and temporal variabilities.

The spatial variability is estimated from the time average of the spatial standard deviations (the standard deviations
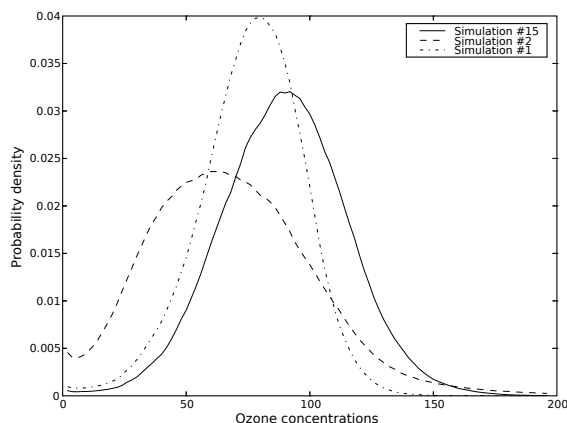
Table 4. Means and standard deviations of the ozone daily maxima ($\mu g \cdot m^{-3}$) of the twenty simulations. On the left, the simulations are sorted by their mean; on the right, they are sorted by their standard deviation. The relative standard deviation of the means is 3.6%, and the relative standard deviation of the standard deviations is 13.5%.

	Sorted by mean			Sorted by standard deviation	
#	Mean	Standard deviation	#	Mean	Standard deviation
15	108.79	22.62	2	95.71	33.50
14	104.40	23.87	16	101.39	24.35
18	104.11	22.87	11	101.04	23.98
4	102.39	23.38	14	104.40	23.87
7	102.03	22.58	4	102.39	23.38
6	101.71	22.05	3	99.73	23.17
16	101.39	24.35	12	99.59	22.94
11	101.04	23.98	18	104.11	22.87
0	100.92	22.62	0	100.92	22.62
17	100.62	22.49	15	108.79	22.62
13	100.52	22.57	7	102.03	22.58
9	99.77	21.68	13	100.52	22.57
3	99.73	23.17	17	100.62	22.49
12	99.59	22.94	19	98.96	22.25
19	98.96	22.25	6	101.71	22.05
5	98.34	21.61	9	99.77	21.68
8	97.90	20.10	5	98.34	21.61
10	96.31	19.32	8	97.90	20.10
2	95.71	33.50	10	96.31	19.32
1	91.49	16.75	1	91.49	16.75

computed with the concentrations in all cells and at a given time step). It appears that the spatial variability and the standard deviation on the whole field at once lead to the same conclusions. The correlation between the spatial variability and the global standard deviation of all simulations can be as high as 98.7%.

The temporal variability is estimated from the spatial average of the temporal standard deviations (computed with all concentrations in a given cell). The correlation of this
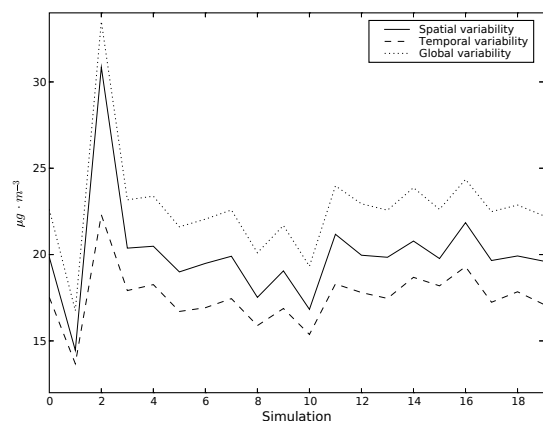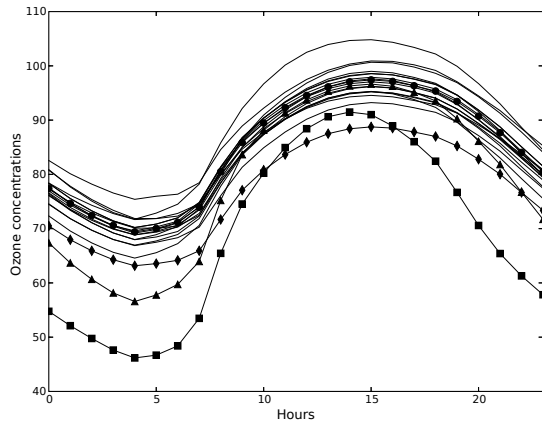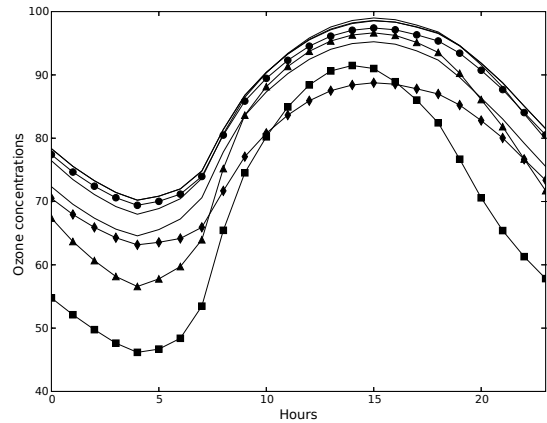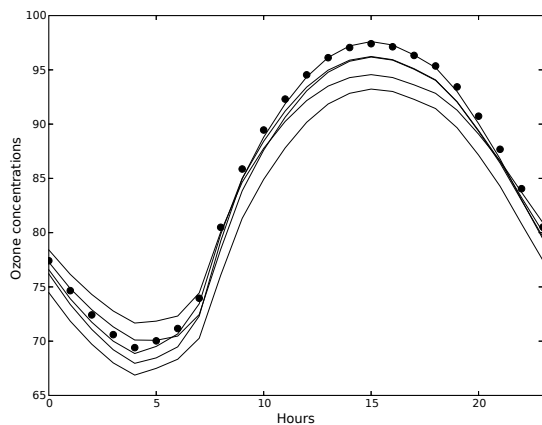

Figure 3. Spatial, temporal and global variabilities ($\mu g \cdot m^{-3}$) of ozone daily maxima for the twenty simulations. The spatial variability is estimated with the time average of the spatial standard deviations (the standard deviations computed with the peaks in all cells, for a given day). The temporal variability is estimated with the spatial average of the temporal standard deviations (computed with all daily maxima in a given cell). The global variability is measured by the standard deviation of all daily maxima.

Figure 2. Relative frequency distributions of ozone concentrations ($\mu g \cdot m^{-3}$) for the simulations 15, 2 and 1 which show the most extreme behavior (in terms of mean and standard deviation).
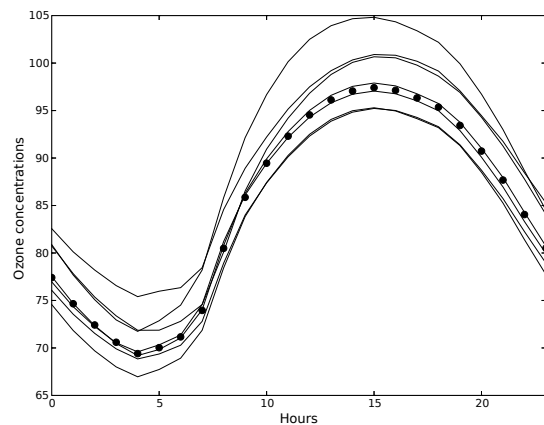
(a) All simulations.

(b) Physical parameterizations.

(c) Input data. The lowest concentrations are reached by simulation 10 (GLCF land use coverage for biogenic emissions).

(d) Numerical approximations. The highest concentrations are reached by simulation 15 (Strang splitting).

Figure 4. Ozone daily profile for the twenty simulations and for the three groups related to (b) the physical parameterizations, (c) the input data and (d) the numerical approximations. The dots represent the reference simulation (#0), the squares represent simulation 2 (Louis closure), the triangles represent simulation 3 (Louis closure in stable conditions) and the diamonds represent simulation 1 (RADM 2).

temporal variability with the global standard deviations is still above 98%. The variability can also be estimated daily and then averaged over the days. In this case, the correlation with the global standard deviation is 94%, which is still high. Hence the parameterizations that introduce some variability increase both the temporal and the spatial variabilities. In Section 7, the most strongly impacted regions are identified.

A key point in ozone forecasts is the daily maximum. Table 4 shows the distribution of the means and the standard deviations of the ozone daily peaks. The behavior of the daily maxima differs from the field averages previously analyzed.

While the means of the maxima are less widely spread, the standard deviations are strongly spread and range from $16.75\mu g \cdot m^{-3}$ (simulation 1, RADM 2) to $33.50\mu g \cdot m^{-3}$ (simulation 2, Louis parameterization). On one hand, the highest standard deviation is reached with the Louis closure. Notice that if the Louis closure is only used in stable conditions (simulation 3), the impact on the daily maxima is much lower, which means that the nighttime concentrations have a small influence on the peaks. On the other hand, the low-

est standard deviation comes from the chemical mechanism RADM 2. It is also associated with the lowest concentrations, which is consistent with *Gross and Stockwell* [2003].

Contrary to the concentration averages, the daily maxima are more variable in space than in time, as shown in Figure 3. The simulation with RADM 2 is essentially the only simulation for which the temporal variability is similar to the spatial variability. The impact of the parameterizations is also greater on the spatial variability than on the temporal variability: their relative standard deviations are 15.1% and 9.5% respectively.

4.2. Comparisons with the Reference Simulation

Now that the global variability has been analyzed, comparisons with the reference simulation allow us to give details about the impact of each change in the parameterizations or in the numerical choices.

First examined are the bias, the standard deviation of the distance to the reference simulation (namely the difference with the reference simulation) and the correlation with the reference simulation. These values are shown in Table 5.

The changes in the turbulence closure (simulation 2 and 3) lead to the largest differences. The chemical mechanism

RADM 2 (simulation 1) also has an impact but not as strong as what was seen in the previous subsection with respect to the variability. On the contrary, the fine resolution (0.1°, simulation 16) leads to strong differences with the reference simulation, even if this was not obvious from the previous analyses. It is noteworthy that simulation 17, with a 1.0° resolution, has a lower but still significant impact.

The other main changes are due to the splitting method (simulation 15), the vertical resolution (nine levels, simulation 18), the emission vertical distribution (simulation 8) and the land use coverage used for the biogenic emissions (simulation 10).

For each change in the model, there is an explanation for its low or high impact on the output concentrations. We do not provide such explanations due to the number of simulations and because the purpose of the paper is to describe the global uncertainty due to the parameterizations and the numerical choices. What should be emphasized instead is that the results are sensitive to the physical parameterizations, the input data sets *and* the numerical issues.

The relative frequency distribution of all concentrations of the reference simulation is shown in Figure 5 with its uncertainty due to the parameterizations. The concentration distribution is sensitive to the parameterizations. The uncertainty (estimated with the relative standard deviation) in the relative frequency distribution is about 25%–30% for concentrations in $[30\mu g \cdot m^{-3}, 170\mu g \cdot m^{-3}]$. This high un-

Table 5. Biases and standard deviations of the distance to the reference simulation ($\mu g \cdot m^{-3}$) for the nineteen simulations. On the left, the simulations are sorted by their bias; on the right, they are sorted by their standard deviation.

	Sorted by bias			Sorted by standard deviation		
#	Bias	Standard deviation	#	Bias	Standard deviation	Correlation
15	5.39	6.65	2	-15.97	18.60	0.85
18	4.23	5.10	3	-5.77	10.40	0.93
14	2.57	3.60	16	-1.64	8.95	0.94
7	1.00	1.32	1	-7.81	7.55	0.96
6	1.00	2.63	15	5.39	6.65	0.97
4	0.26	2.69	8	-0.69	5.62	0.98
17	0.24	5.07	18	4.23	5.10	0.98
13	-0.18	0.98	17	0.24	5.07	0.98
8	-0.69	5.62	10	-3.62	5.04	0.98
11	-0.78	2.81	5	-3.54	3.69	0.99
12	-0.92	2.48	14	2.57	3.60	0.99
9	-0.96	2.30	11	-0.78	2.81	0.99
16	-1.64	8.95	4	0.26	2.69	1.00
19	-2.30	2.34	6	1.00	2.63	0.99
5	-3.54	3.69	12	-0.92	2.48	1.00
10	-3.62	5.04	19	-2.30	2.34	1.00
3	-5.77	10.40	9	-0.96	2.30	1.00
1	-7.81	7.55	7	1.00	1.32	1.00
2	-15.97	18.60	13	-0.18	0.98	1.00

Table 6. The amount of negative biases among the daily biases, the daytime biases, the nocturnal biases and the spatial biases. The simulations are sorted by their standard deviation.

#	Standard deviation ($\mu g \cdot m^{-3}$)	Daily negative bias (%)	Daytime negative bias (%)	Nocturnal negative bias (%)	Spatial negative bias (%)
2	18.60	100	100	100	98
3	10.40	100	100	100	100
16	8.95	100	100	100	79
1	7.55	100	100	100	100
15	6.65	0	0	6	1
8	5.62	70	84	44	64
18	5.10	0	0	0	0
17	5.07	27	25	33	44
10	5.04	100	100	100	100
5	3.69	100	100	100	100
14	3.60	0	13	0	0
11	2.81	88	84	90	76
4	2.69	35	30	60	33
6	2.63	15	17	15	1
12	2.48	81	90	71	91
19	2.34	100	100	100	100
9	2.30	85	85	83	79
7	1.32	0	0	0	0
13	0.98	61	42	95	82

Table 7. Percentages of stations that meet the most-restrictive EPA recommendations on hourly concentrations (cutoff of $80\mu g \cdot m^{-3}$) for the mean bias, the mean gross error and the unpaired (in time, paired in space) peak prediction; the root mean square ($\mu g \cdot m^{-3}$), the correlation and the overall bias ($\mu g \cdot m^{-3}$) for the ozone daily peaks (all stations put together).

#	MNBE (%)	MNGE (%)	UPA (%)	RMS (peaks)	Correlation (peaks)	Bias (peaks)
0	43	100	61	23.54	0.71	-4.47
1	7	97	21	29.22	0.65	-14.19
2	13	85	39	24.86	0.73	-8.26
3	32	98	59	23.54	0.72	-5.27
4	52	100	69	22.73	0.73	-2.33
5	24	99	48	24.47	0.71	-7.25
6	45	100	61	24.03	0.69	-3.36
7	48	100	64	23.67	0.70	-3.18
8	35	100	52	25.03	0.69	-6.29
9	36	99	56	24.02	0.71	-5.50
10	21	99	36	26.40	0.67	-9.65
11	43	99	62	23.02	0.73	-4.38
12	29	99	45	24.59	0.70	-6.75
13	41	100	55	23.85	0.71	-4.99
14	57	100	76	22.26	0.74	0.01
15	47	100	64	25.11	0.67	5.39
16	62	100	75	22.83	0.72	-0.39
17	33	99	51	24.12	0.71	-6.60
18	38	98	52	24.48	0.68	-2.70
19	46	100	58	23.72	0.71	-4.67

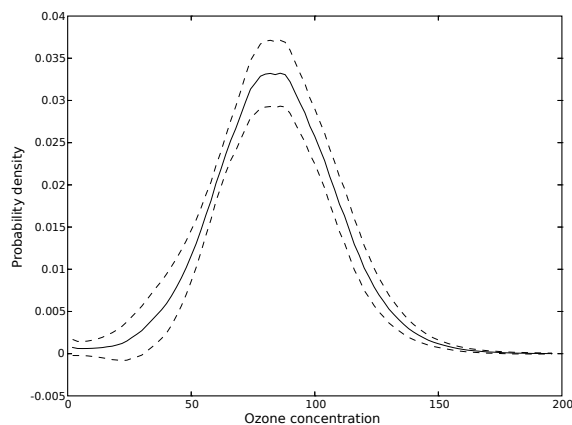**Figure 5.** Relative frequency distribution $f([O_3])$ of the ozone concentrations ($\mu g \cdot m^{-3}$) of the reference simulation and the functions $f([O_3]) \pm \sigma_f([O_3])$ where $\sigma_f([O_3])$ is the standard deviation of $f([O_3])$ computed from all simulations.

certainty is consistent with the uncertainty roughly shown in Figure 2.

In Figure 4, the mean of the daily evolution for ozone over the whole domain and all days is shown for all simulations. This daily evolution is also shown independently for the ensembles generated with changes (1) in the physical parameterizations, (2) in the input data and (3) in the numerical approximations. From these figures, the prominent changes are the turbulence closure and the chemical mechanism. The profile is less sensitive to the input data, but this is an average profile that can hide spatial or temporal variabilities (which are analyzed below). The spread (estimated with the relative standard deviation) is 4% on the peak and 6% for the whole profile. It reaches 9% at 0400 UT, which is high since the impacts are not cumulative (see Section 7 for the cumulative effects).

One question about the nature of the variability introduced by each change lies in the bias and its nocturnal and diurnal evolutions. The standard deviation of the difference with the reference simulation does not provide this information since it may be low even with systematic biases in the night or in the daytime. In the same way, systematic biases may appear in given regions. In Table 6, the amount of negative biases (concentrations less that the reference concentrations) are reported for:

1. the daily biases: the biases (mean over all grid cells of the difference) computed for each day;

2. the daytime biases: the biases computed for each day but only during the daytime (from 0400 UT to 1800 UT);

3. the nocturnal biases: the biases computed for each day but only during the night;

4. the spatial biases: the biases (mean over all time steps of the difference) computed for each grid cell.

It first indicates that the simulations with the largest biases are characterized by a clear trend: they are either above or below the reference simulation at nearly all hours. It also demonstrates that the biases at night and during the daytime can strongly differ. This is true for simulation 8 (emission vertical distribution) due to the fact that the pollutants emitted at night and in the second layer barely influence the ground concentrations. During the daytime, the emissions are mixed in the boundary layer which decreases the impact of the vertical location of the pollutants at release time. In simulation 4 (Wesely's parameterization), the nocturnal concentrations are often below the reference concentrations but only slightly below since the total amount of negative biases is very close to the amount associated with the daytime. The last simulation that shows such differences is simulation 13 (100 s as a time step) for unclear reasons. The active chemical reactions are not the same at night as in the daytime, which may explain why the numerical time step has a different impact.

Another point lies in the spatial and temporal differences. The day-by-day bias may hide spatial inhomogeneities of the bias. Simulations 16 and 17 (0.1° and 1.0° horizontal resolution respectively) are good examples. Loosely speaking, choices in the simulation setup may impact the spatial distributions independently from their temporal effects.

4.3. Comparisons with Observations

The results of the simulations are compared to ozone peaks which are usually a major concern of forecasting systems. The comparisons are performed with 242 stations over Europe. Each selected station has a reasonable amount of measurements (at least 30 peak measurements during the 126 days of the comparison period). There are 27,000 peak observations and 620,000 hourly observations.

Following *US EPA* [1991], we first evaluate the results with the normalized bias (MNBE), the mean normalized gross error (MNGE) and the daily unpaired (in time, paired in space) peak prediction (UPA). A cutoff level of $80\mu g \cdot m^{-3}$ is used and the errors are evaluated as "computed minus observed" (a positive bias represents overestimation). A simulation is assessed through the amount of stations that match the most-restrictive EPA suggested performances: ±5% for the normalized bias, ±30% for the normalized gross error and ±15% for the unpaired peak prediction accuracy. The root mean square (RMS), the correlation and the overall bias of all daily ozone peaks are also reported in Table 7.

The mean normalized gross error is within the EPA limits at almost all stations for all simulations except for the simulation 2 with the Louis closure: for this simulation, the underestimation is indeed too high ($-8.26\mu g \cdot m^{-3}$ on the peaks). Simulations 1 and 10 also have a strong bias on the daily peaks, but they do not underestimate all concentrations above $80\mu g \cdot m^{-3}$ as much as simulation 2. In this case, the MNGE does not distinguish the simulations, even if other indicators give a wide spread. One might consider that the uncertainty due to the parameterizations is below the error that the EPA limit on the MNGE can detect. This

Table 8. The 16 simulations set up with the four alternative parameterizations.

#	Emissions	Deposition	Turbulence	Chemistry
a	Ground	Zhang	Troen & Mahrt	RACM
b	Ground	Zhang	Troen & Mahrt	RADM 2
c	Ground	Zhang	Louis	RACM
d	Ground	Zhang	Louis	RADM 2
e	Ground	Wesely	Troen & Mahrt	RACM
f	Ground	Wesely	Troen & Mahrt	RADM 2
g	Ground	Wesely	Louis	RACM
h	Ground	Wesely	Louis	RADM 2
i	Two layers	Zhang	Troen & Mahrt	RACM
j	Two layers	Zhang	Troen & Mahrt	RADM 2
k	Two layers	Zhang	Louis	RACM
l	Two layers	Zhang	Louis	RADM 2
m	Two layers	Wesely	Troen & Mahrt	RACM
n	Two layers	Wesely	Troen & Mahrt	RADM 2
o	Two layers	Wesely	Louis	RACM
p	Two layers	Wesely	Louis	RADM 2

Table 9. Means and standard deviations of the ozone concentrations ($\mu g \cdot m^{-3}$) and their daily peaks for the sixteen simulations. As for the whole concentrations, the relative standard deviation of the means is 15%, and the relative standard deviation of the standard deviations is 20%. As for the daily maxima, the relative standard deviations are 8% and 28% respectively.

	All concentrations			*Daily maxima*	
#	Mean	Standard deviation	#	Mean	Standard deviation
g	69.26	35.71	g	97.58	35.03
c	68.94	34.23	c	95.71	33.50
o	70.14	31.94	o	95.32	31.39
k	69.73	30.52	k	93.29	30.17
h	56.17	27.08	e	102.39	23.38
e	85.18	26.19	a	100.92	22.62
d	56.26	26.07	h	80.91	21.75
a	84.92	25.11	m	99.36	20.70
p	58.47	24.49	d	79.80	20.67
l	58.41	23.46	i	97.90	20.10
m	84.49	22.84	p	80.67	19.82
f	77.27	21.96	l	79.29	18.96
i	84.23	21.81	f	92.72	17.12
b	77.11	21.14	b	91.49	16.75
n	77.51	19.30	n	90.85	15.73
j	77.34	18.50	j	89.61	15.50

is rather speculative since only single changes were introduced in the simulations and, in the next subsection, it is shown that the amount of stations below the MNGE limit can increase if several changes are introduced at the same time.

Meanwhile, there is a high uncertainty in the concentrations above $80 \mu g \cdot m^{-3}$ according to the normalized bias statistics. The amount of stations whose bias is acceptable ($\pm 5\%$) ranges from 7% to 62%. In the opposite way as for the MNGE, the MNBE test may be questionable precisely because of its variability. A conclusion may be that this test is too severe to be relevant: a good percentage of acceptable stations would come mainly from a favorable configuration of the model. The uncertainty in the model is too high to grant a reasonable validity for this test.

The UPA statistics are also well spread but they show a lower variability: their relative standard deviation is 9% against 37% for the MNBE. In comparison, the RMS and the correlations vary slightly with 6% and 3% respectively of relative standard deviation among the simulations. Therefore the correlations do not provide substantial information. Moreover the correlation between RMS and 1 - UPA reaches 90%, which means that it is essentially useless to compute both.

Finally there is a rather high uncertainty in the peak levels: almost $20 \mu g \cdot m^{-3}$ of bias between simulations 1 and 15, with mean observed-peaks at $103 \mu g \cdot m^{-3}$. Nevertheless the standard deviation of the biases is only $4 \mu g \cdot m^{-3}$. This means that an overall bias is not detailed enough to show the uncertainty.

Even if these results can only barely be generalized, defining indicators adequately related to the uncertainty in the models is not an easy task. They are supposed to distinguish simulations and, at the same time, to be robust enough to changes (within the uncertainty range) in the models. None of the previous indicators seem to be balanced enough for this purpose.

4.4. Combined Changes

In this section, we try to estimate the impact of combined changes: several parameterizations are changed at the same time. All combinations of the parameterizations and the numerical choices introduced in Table 1 cannot be applied because of the computational costs (there would be 184,320 simulations). Hence only four alternatives are kept: the Louis closure, the RADM 2 mechanism, the deposition velocities as computed in *Wesely* [1989] and the vertically distributed emissions. The first two parameterizations are included due to their strong impact, the third one due its improvements in the results as compared to the observations and the fourth one because of its low variability. Refer to Table 8 for the list of the simulations.

Table 9 shows the spread on ozone peaks for the new set of simulations. It should be compared to Tables 2 and 4. The spread is clearly higher in the new set of simulations. As a consequence, the whole uncertainty due to the parameterizations cannot be easily assessed on the basis of single changes in the parameterizations. The results from the previous sections cannot claim more than an estimate of a lower bound on the uncertainty.

The same is true about the error statistics. Even the MNGE limitation ($\pm 30\%$) is not satisfied by more than 90% of the stations for 6 simulations of the 16. Five simulations have a root mean square above $30 \mu g \cdot m^{-3}$ whereas none of the previous twenty simulations reaches such a RMS (see Table 7). It is obviously due to the combination of changes that individually contribute to decrease the concentrations: the underestimation is then worsened.

The "cumulative underestimations" can be seen in Figure 6 (to be compared to Figure 4). For instance, the simulation that combines the Louis closure and the chemical mechanism RADM 2 shows low concentrations. The non-linearity even increases this effect. In Figure 7 we compare
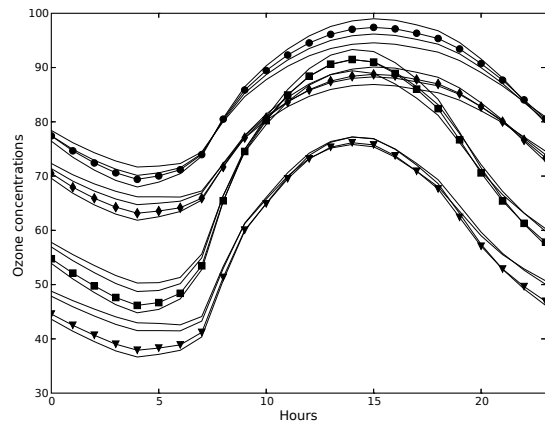

Figure 6. Ozone daily profile for the sixteen simulations. The dots represent the reference simulation (#a), the diamonds represent simulation b (RADM 2), the squares represent simulation c (Louis closure) and the triangles represent simulation d that combines the Louis closure and RADM 2.
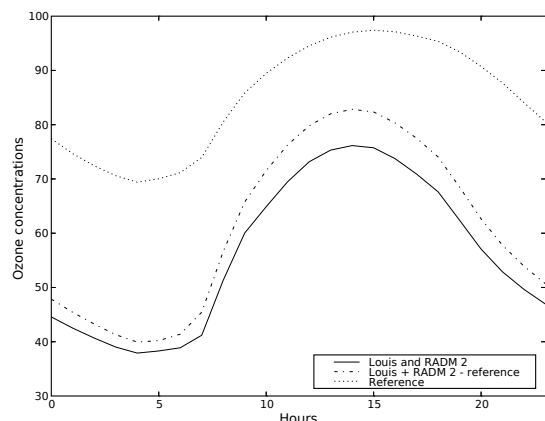

Figure 7. Ozone daily profile of the reference simulation a, simulation d (Louis closure and RADM 2) and the linear combination "#c + #b - #a" that adds linearly the effects of simulations b (RADM 2) and c (Louis closure). The non-linearity increases the impact of the parameterizations and the concentrations of simulation d are closer to the reference concentrations than the concentrations of the linear combination. It is especially true for the peak.

the simulation d (Louis closure and RADM 2) and the linear combination "#c + #b - #a". Both should be equal if the dependences were linear. The concentrations of the simulation d are even lower than the concentrations of the linear combination. This means that the uncertainty is not additive.

The mean spread (relative standard deviation of the ensemble) of the daily-profile concentrations reaches 16% (against 6% with single changes). The highest spread is reached at 0400 UT with 23%. The spread on the peak is 9%. Notice that this spread is a measure of the uncertainty.

Another measure is the relative standard deviation of all concentrations, not of the mean profile. The relative standard deviation is computed for concentrations in $[40\mu g \cdot m^{-3}, 130\mu g \cdot m^{-3}]$ to include only the main concentrations (refer to the relative frequency distribution shown in Figure 5). Figure 8 shows the relative standard deviation versus ozone concentrations. The average of the relative standard deviation on this interval is 17%. The lowest concentrations have the largest uncertainty. It means that the processes at night are sensitive to the available parameterizations. The turbulence closure plays an important role at night when the values of the vertical diffusion coefficients are hard to estimate in stable conditions and at the top of the first layer. As for the daily peaks, the relative standard deviation, computed over the whole domain and with all days, reaches 11%. It is difficult to determine the reason why the peaks have a low uncertainty as compared to the other concentrations (Figure 8): it may be due to less uncertainty in all parameterizations, or the peaks may be sensitive only to a few leading processes. For instance, the turbulence may be less determinant in well mixed conditions whereas the photochemical activity is strong at the same time.

The ensemble is not equally spread everywhere in the domain as shown in Figure 9. The uncertainty measured by the
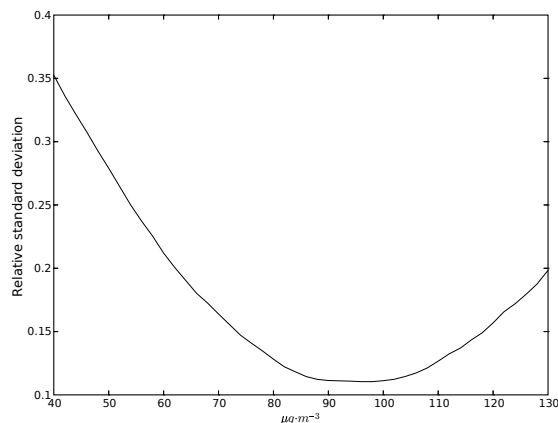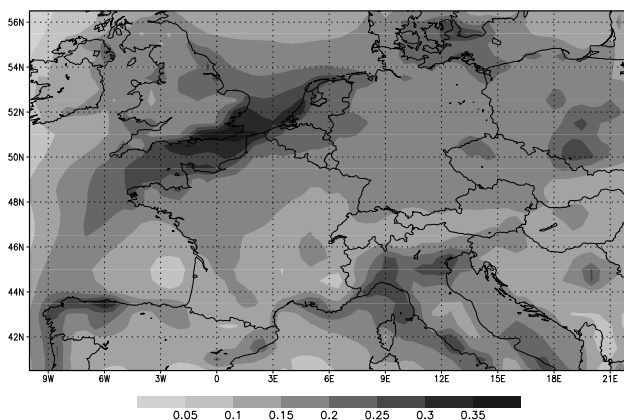
standard deviation (for concentrations in $[40\mu g \cdot m^{-3}, 130\mu g \cdot m^{-3}]$) is high around the coasts and it tends to be high in polluted regions (Southern regions and, due to the emissions, in Great Britain and Poland). In Northern Italy, the Alps are also associated with a high uncertainty. The uncertainty on the peaks has the same spatial distribution. The turbulence closure may notably explain these uncertainties, and the chemical mechanism has probably a strong impact close to the emission locations. However a more detailed study of each process is necessary to properly analyze the spatial inhomogeneities due to each process.

5. Conclusion

It has been shown that a chemistry-transport model is sensitive to its physical parameterizations, to the associated input data and to the numerical approximations. The turbulent closure and the chemical mechanism introduce the highest uncertainty. The overall uncertainty, measured with the relative standard deviation of an ensemble of sixteen simulations, is estimated at 17% for the common concentration levels and at 11% for the daily peaks. It has been shown that this uncertainty was notably high along the coasts. The uncertainty is too high to let any configuration of the chemistry-transport model fully satisfy the common requirements in comparisons with observations. This low robustness suggests that ensemble approaches are necessary in most applications.

A remaining question is whether these conclusions are limited to the Polyphemus system, even if this system includes commonly used parameterizations. Moreover, this work should be extended to aerosol modeling for which many physical parameterizations and numerical algorithms are also available (hybrid models, nucleation laws, etc.). Another extension deals with the uncertainty due to the input fields to the model such as the meteorological fields or the emissions.

In a next step, one may want to take advantage of the ensemble to provide improved forecasts. The point is to find (actually to forecast) the best combination of the models, as an improvement of the ensemble-mean or ensemble-median approaches.

Figure 8. Relative standard deviation of the ensemble versus ozone concentrations. The average standard deviation is 17%.

Notes

1. U.S. Geological survey.
2. Global Land Cover Facility.
3. Co-operative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe.
4. The relative standard deviation is the standard deviation divided by the mean.

Figure 9. Relative standard deviation of the ensemble for concentrations in $[40\mu g \cdot m^{-3}, 130\mu g \cdot m^{-3}]$. The average standard deviation is 17%. The uncertainty is notably high along the coasts.

References

Boutahar, J., S. Lacour, V. Mallet, D. Quélo, Y. Roustan, and B. Sportisse, Development and validation of a fully modular platform for numerical modelling of air pollution: POLAIR, *Int. J. Env. and Pollution, 22*(1/2), 17–28, 2004.

Builtjes, P., The LOTOS – LOng Term Ozone Simulation – project, summary report, *Tech. Rep. R92/240*, TNO, Delft, the Netherlands, 1992.

Buizza, R., M. Miller, and T. N. Palmer, Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, *Quart. J. Roy. Meteor. Soc., 125*, 2,887–2,908, 1999.

Byun, D. W., and J. K. S. Ching (Eds.), *Science algorithms of the EPA models-3 community multiscale air quality (CMAQ) modeling system*, EPA, 1999.

Chang, J., R. Brost, I. Isaken, S. Madronich, P. Middleton, W. Stockwell, and C. Walcek, A three-dimensional Eulerian acid deposition model: physical concepts and formulation, *J. Geophys. Res.*, *92*(D12), 14,681–14,700, 1987.

Christensen, J. H., The Danish Eulerian hemispheric model – a three-dimensional air pollution model used for the arctic, *Atmos. Env.*, *31*, 4,169–4,191, 1997.

Delle Monache, L., and R. B. Stull, An ensemble air-quality forecast over western Europe during an ozone episode, *Atmos. Env.*, *37*, 3,469–3,474, 2003.

ESQUIF, Étude et simulation de la qualité de l'air en île de France – rapport final, 2001.

Galmarini, S., et al., Ensemble dispersion forecasting – part I: concept, approach and indicators, *Atmos. Env.*, *38*(28), 4,607–4,617, 2004.

Gardiner, C. W., *Handbook of Stochastic Methods – For Physics, Chemistry and the Natural Sciences*, Springer, 1996.

Gross, A., and W. R. Stockwell, Comparison of the EMEP, RADM2 and RACM mechanisms, *Journal of Atmospheric Chemistry*, *44*, 151–170, 2003.

Hanna, S. R., Z. Lu, H. C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D. A. Hansen, Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain, *Atmos. Env.*, *35*(5), 891–903, 2001.

Hass, H., Description of the EURAD chemistry-transport-model version 2 (CTM2), *Tech. Rep. 83*, Institute of Geophysics and Meteorology, University of Cologne, 1991.

Hass, H., P. J. H. Builtjes, D. Simpson, and R. Stern, Comparison of model results obtained with several European regional air quality models, *Atmos. Env.*, *31*(19), 3,259–3,279, 1997.

Horowitz, L. W., et al., A global simulation of tropospheric ozone and related tracers: description and evaluation of MOZART, version 2, *J. Geophys. Res.*, *108*(D24), 2003.

Houtemaker, P. L., L. Lefaivre, J. Derome, H. Ritchie, and H. L. Mitchell, A system simulation approach to ensemble prediction, *Mon. Wea. Rev.*, *124*(6), 1,225–1,242, 1996.

Louis, J.-F., A parametric model of vertical eddy fluxes in the atmosphere, *Boundary-Layer Meteor.*, *17*, 187–202, 1979.

Madronich, S., Photodissociation in the atmosphere: 1. actinic flux and the effects of ground reflections and clouds, *J. Geophys. Res.*, *92*(D8), 9,740–9,752, 1987.

Mallet, V., and B. Sportisse, 3-D chemistry-transport model Polair: numerical issues, validation and automatic-differentiation strategy, *Atmos. Chem. Phys. Discuss.*, *4*, 1,371–1,392, 2004.

Mallet, V., and B. Sportisse, Data processing and parameterizations in atmospheric chemistry and physics: the AtmoData library, *Tech. Rep. 12*, CEREA, 2005.

Mallet, V., D. Quélo, and B. Sportisse, Software architecture of an ideal modeling platform in air quality – A first step: Polyphemus, *Tech. Rep. 11*, CEREA, 2005.

Middleton, P., W. R. Stockwell, and W. P. L. Carter, Aggregation and analysis of volatile organic compound emissions for regional modeling, *Atmos. Env.*, *24A*(5), 1,107–1,133, 1990.

Nodop, K. (Ed.), *ETEX symposium on long-range atmospheric transport, model verification and emergency response*, 1997.

Passant, N. R., Speciation of UK emissions of NMVOC, *Tech. Rep. AEAT/ENV/0545*, AEA Technology, 2002.

Russell, A., and R. Dennis, NARSTO critical review of photochemical models and modeling, *Atmos. Env.*, *34*, 2,283–2,234, 2000.

Schmidt, H., Sensitivity studies with the adjoint of a chemistry transport model for the boundary layer, in *Air pollution modelling and simulation*, edited by B. Sportisse, pp. 400–410, Springer, 2002.

Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann, A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe, *Atmos. Env.*, *35*, 6,277–6,297, 2001.

Simpson, D., H. Fagerli, J. E. Jonson, S. Tsyro, P. Wind, and J.-P. Tuovinen, Transboundary acidification, eutrophication and ground level ozone in Europe – part I: unified EMEP model description, *Tech. rep.*, EMEP, 2003.

Simpson, D., et al., Inventorying emissions from nature in Europe, *J. Geophys. Res.*, *104*(D7), 8,113–8,152, 1999.

Sportisse, B., An analysis of operator splitting techniques in the stiff case, *J. Comp. Phys.*, *161*(1), 140–168, 2000.

Stockwell, W. R., P. Middleton, J. S. Chang, and X. Tang, The second generation regional acid deposition model chemical mechanism for regional air quality modeling, *J. Geophys. Res.*, *95*(D10), 16,343–16,367, 1990.

Stockwell, W. R., F. Kirchner, M. Kuhn, and S. Seefeld, A new mechanism for regional atmospheric chemistry modeling, *J. Geophys. Res.*, *102*(D22), 25,847–25,879, 1997.

Tao, Z., S. M. Larson, A. Williams, M. Caughey, and D. J. Wuebbles, Sensitivity of regional ozone concentrations to temporal distribution of emissions, *Atmos. Env.*, *38*(37), 6,279–6,285, 2004.

Toth, Z., and E. Kalnay, Ensemble forecasting at NMC: the generation of perturbations, *Bull. Amer. Meteor. Soc.*, *74*, 2,317–2,330, 1993.

Troen, I., and L. Mahrt, A simple model of the atmospheric boundary layer; sensitivity to surface evaporation, *Boundary-Layer Meteor.*, *37*, 129–148, 1986.

US EPA, Guideline for regulatory application of the urban airshed model, *Tech. Rep. EPA-450/4-91-013*, US EPA, 1991.

Verwer, J. G., W. Hundsdorfer, and J. G. Blom, Numerical time integration for air pollution models, *Tech. rep.*, CWI, 1998.

Warner, T. T., R.-S. Sheu, J. F. Bowers, R. I. Sykes, G. C. Dodd, and D. S. Henn, Ensemble simulations with coupled atmospheric dynamic and dispersion models: illustrating uncertainties in dosage simulations, *J. Applied Meteor.*, *41*, 488–504, 2002.

Wesely, M. L., Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, *Atmos. Env.*, *23*, 1,293–1,304, 1989.

Zhang, L., J. R. Brook, and R. Vet, A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, *3*, 2,067–2,082, 2003.

Vivien.Mallet@cerea.enpc.fr, Bruno.Sportisse@cerea.enpc.fr

Statistical measures

Notations: Let y be the vector of model outputs and let o be the vector of the corresponding observations. These vectors both have n components. Their means are $\bar{y}$ and $\bar{o}$.

Relative standard deviation:

$$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}{\bar{y}} \tag{A1}$$

Bias:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right) \tag{A2}$$

Root mean square error (RMS):

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right)^2} \tag{A3}$$

Correlation:

$$\text{correlation} = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(o_i - \bar{o}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\sum_{i=1}^{n}\left(o_i - \bar{o}\right)^2}} \tag{A4}$$

Mean normalized bias error (MNBE):

$$\text{MNBE} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - o_i}{o_i} \tag{A5}$$

Mean normalized gross error (MNGE):

$$\text{MNGE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - o_i|}{o_i} \tag{A6}$$

Unpaired peak prediction accuracy (UPA): For one day:

$$\text{UPA}_{\text{day}} = \frac{y_{\text{max}} - o_{\text{max}}}{o_{\text{max}}} \tag{A7}$$

The UPA is then averaged over all days.