HAL
open science

What Can We Expect from Data Assimilation for Air Quality Forecast ? Part II : Analysis with a Semi-Real Case

Bertrand Bessagnet, Laurent Menut, Florian Couvidat, Frédérik Meleux, Guillaume Siour, Sylvain Mailler

To cite this version:

Bertrand Bessagnet, Laurent Menut, Florian Couvidat, Frédérik Meleux, Guillaume Siour, et al.. What Can We Expect from Data Assimilation for Air Quality Forecast ? Part II : Analysis with a Semi-Real Case. Journal of Atmospheric and Oceanic Technology, 2019, 36 (7), pp.1433-1448. 10.1175/JTECH-D-18-0117.1 . ineris-03319065

HAL Id: ineris-03319065

https://ineris.hal.science/ineris-03319065

Submitted on 11 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

What Can We Expect from Data Assimilation for Air Quality Forecast? Part II: Analysis with a Semi-Real Case

BERTRAND BESSAGNET[a]

Institut National de l'Environnement Industriel et des Risques, Verneuil en Halatte, France

LAURENT MENUT

Laboratoire de Météorologie Dynamique, Ecole Polytechnique, IPSL Research University, Ecole Normale Supérieure, Université Paris-Saclay, Sorbonne Universités, UPMC Université Paris 06, CNRS, Palaiseau, France

FLORIAN COUVIDAT AND FRÉDÉRIK MELEUX

Institut National de l'Environnement Industriel et des Risques, Verneuil en Halatte, France

GUILLAUME SIOUR

Laboratoire Inter-Universitaire des Systèmes Atmosphériques, UMR CNRS 7583, Université Paris Est Créteil et Université Paris Diderot, Institut Pierre Simon Laplace, Créteil, France

SYLVAIN MAILLER

Laboratoire de Météorologie Dynamique, Ecole Polytechnique, IPSL Research University, Ecole Normale Supérieure, Université Paris-Saclay, Sorbonne Universités, UPMC Université Paris 06, CNRS, Palaiseau, France

(Manuscript received 12 July 2018, in final form 6 May 2019)

ABSTRACT

Assimilation of observational data from ground stations and satellites has been identified as a technique to improve air quality model results. This study is an evaluation of the maximum benefit expected from data assimilation in chemical transport models. Various tests are performed under real meteorological conditions; the injection of various subsets of "simulated observational data" at the initial state of a forecasting period is analyzed in terms of benefit on selected criteria. This observation dataset is generated by a simulation with perturbed input data. Several criteria are defined to analyze the simulations leading to the definition of a "tipping time" to compare the behavior of simulations. Assimilating three-dimensional data instead of ground observations clearly adds value to the forecast. For the studied period and considering the expected best favorable data assimilation experiment, the maximum benefit is higher for particulate matter (PM) with tipping times exceeding 80 h; for ozone (O_3) the gain is on average around 30 h. Assimilating O_3 concentrations with a delta calculated on the first level and propagated over the vertical direction provides better results on O_3 mean concentrations when compared with the expected best experiment corresponding to the injection of the O_3 "observations" 3D dataset, but for maximum O_3 concentrations the opposite behavior is observed. If data assimilation of secondary pollutant concentrations provides an improvement, assimilation of primary pollutant emissions can have beneficial impacts when compared with an assimilation of concentrations, after several days on secondary pollutants like O_3 or nitrate concentrations and more quickly for the emitted primary pollutants. An assimilation of ammonia concentrations has slightly better performances on nitrate, ammonium, and PM concentrations relative to the assimilation of nitrogen or sulfur dioxides.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JTECH-D-18-0117.s1.

[a] Current affiliation: Hangzhou Futuris Environmental Technology Co., Ltd., Hangzhou, Zhejiang, China.

Corresponding author: Bertrand Bessagnet, bertrand.bessagnet@futuris-environment.com

DOI: 10.1175/JTECH-D-18-0117.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

1. Introduction

The deterioration of air quality is one of the major environmental threats throughout the world. The impact of pollutants on health, ecosystems, and climate are clearly highlighted by many studies (Lim et al. 2012; Faustini et al. 2014).

Modeling tools, the so-called chemical transport models such as "CHIMERE," are useful to evaluate, analyze, and forecast air quality from urban to global scales (Mailler et al. 2017). In France, the national forecast platform "PREV'AIR" (http://www.prevair.org/) provides a 4-day air quality forecast now at approximately 4-km resolution over France and at approximately 10 km over Europe (Rouïl et al. 2009).

Assimilation of observational data issued from ground stations and satellite data has been identified as a technique to improve the air pollution mapping but also the air quality forecast by injecting available observational data at the initial state (Park et al. 2016; Pagowski et al. 2010). In these works, a gain for O_3 and fine particles concentrations was observed up to 24 h but some of these studies reported an increase of errors attributed to model parameterization of the nocturnal chemistry. Moreover, the modification of the initial state is expected to be a source of instabilities by introducing a change of chemical regimes in the case of species involved in nonlinear chemical schemes. Sandu and Chai (2011) provides an overview of several methods used in chemical data assimilation (DA); they insist on the influence of the initial conditions that fade in time and the concentration fields becoming largely driven by emission and removal processes (and by lateral boundary conditions in regional simulations). For air quality forecasting, Menut and Bessagnet (2019) analyzed the maximum potential benefit of DA techniques by modifying the initial state of the simulation. In the present study, which constitutes a follow-up of this latter study, similar principles are applied to a semi-real case focusing on a particulate matter (PM) pollution outbreak that occurred in March of 2014 (Vieno et al. 2016). This pollution event is an opportunity to investigate the role of DA over Europe with a chemical regime leading to high ammonium nitrate concentrations in the fine fraction of PM in some regions of Europe. Ozone and nitrogen dioxides are also studied even if these species are of minor concern during these episodes, but they contribute to the PM formation.

With academic cases, Menut and Bessagnet (2019) investigated frontier cases including a more realistic case using covariance techniques to propagate the correction in two or three dimensions. They find a maximum benefit of DA estimated at 60 h for ozone concentrations and the use of vertical error covariance to propagate the correction provided only a slight benefit. The objective of this study as a follow-up of Menut and Bessagnet (2019) is not to evaluate the model; this has been performed in past studies over many countries and time scales (Bessagnet et al. 2016; Vivanco et al. 2017; Rea et al. 2015; Menut et al. 2015, 2013; Bessagnet et al. 2008; Vivanco et al. 2009). Here our goal is to evaluate various ideal data assimilation experiments (DAEs) in a forecast perspective, based on a real meteorological case and focusing on particulate matter. Briefly, the principles, as detailed in depth in the next section, are 1) to create a control 5-day simulation considered as a reference forecast, 2) to create a coherent (in terms of chemical regime) "observation" 3D dataset through the use of perturbed input data, and 3) to modify the initial state of various forecast test cases with an analysis of the evolution of relevant species along the 5-day forecast by the use of appropriate criteria.

2. Method–principle of the analysis

a. Reference simulations

In this study, we investigate the maximum potential of DA for air quality forecasting in a semi-real case. In Menut and Bessagnet (2019), the potential of DA was studied for controlled ideal cases; here, the technique is applied over a large European domain to cover various chemical regimes and for meteorological conditions that favor high PM concentrations. A high pressure system developed from the north of France to Scandinavia leading to favorable conditions for the development of a PM pollution event over western Europe. This kind of conditions are frequent in early spring and has been described in previous works (Petit et al. 2017; Bessagnet et al. 2005). A large amount of ammonium nitrate is produced by the reaction of ammonia (mainly emitted by the agricultural sector) and nitric acid issued from the oxidation of nitrogen oxides (involving O_3 chemistry). Although this case is a typical PM episode, the behavior of O_3 and NO_2 concentrations will also be investigated since the photochemistry is already active in early spring.

Prior the description of the DAEs, both reference simulations are presented. The first one [control run (CTRL)] is the forecast simulation with the input data (emissions, boundary conditions, and meteorological conditions) as described in section 2c. The second reference simulation is performed to mimic coherent observational data by perturbing in a random way two input datasets: anthropogenic emissions and meteorological conditions. In doing so, an hourly 3D dataset for all CHIMERE species is created as a representation of a pseudo "real world," which will be our "observations" dataset called SOBS.

This dataset is chemically coherent and includes concentrations of all chemical species. To build this dataset, a random perturbation factor is applied both on emissions and meteorological conditions. For the meteorological conditions, the horizontal wind speed and specific humidity are the selected parameters. For the emissions, the perturbation factors are applied in the range [0.1, 1.9]. For the meteorological parameters, the interval is reduced to [0.7, 1.3] to remain realistic. To increase the efficiency of perturbations, these perturbations are applied over small subdomains of 30×30 and 40×40 grid points, respectively, for the meteorological conditions and the emissions. Perturbing emissions and meteorological conditions leads to a coherent perturbations of concentrations accounting for all chemical processes. The model and the chemical mechanism remain the same for all experiments. CHIMERE diagnoses at each time step the vertical wind speed to ensure mass conservation.

These parameters and the size of subdomains have been selected to obtain a minimum of 10% difference between SOBS and CTRL on average for the main pollutants. This criterion is arbitrary and is mainly used to obtain readable differences. To be less dependent on this criterion, an adequate error is described in the next section to evaluate the return to the CTRL simulation after injection of the SOBS dataset.

b. *The DAEs*

The CTRL and SOBS reference simulations are run over 15 days from 1 to 15 March 2014, breaking down into three blocks of 5 days with a restart at the beginning of each elementary 5-day simulation block. Only the last 5-day block is analyzed, and it constitutes the reference "forecast" period of interest; the first 10 days are considered to be a spinup period to reach a chemical equilibrium for both simulations. Even if the meteorological conditions are a reanalysis, the last 5-day block is considered to be a forecast then starting with an initialization on 0000 UTC 11 March 2014. Therefore, SOBS starting on 0000 UTC 11 March is considered to be the "observations" dataset with hourly available data for all levels and all species. Maps of differences between SOBS and CTRL simulations are reported in Fig. S1 in the online supplementary material. Despite a different meteorological and emissions dataset, on average over the target period, the spatial patterns are in agreement but the magnitude is different in several areas—for instance, lower NO_2 concentrations over Poland, north of Italy, and over "Benelux" and lower PM2.5 (PM $<$ 2.5 μm) concentrations over the north of France in the SOBS dataset. On average, O_3 concentrations are slightly higher in the SOBS simulation.

Injecting a subset of these data in the CTRL simulation forecast at the initialization time allows us to analyze the effect of DA at initial state on the performance of the model to forecast air quality. Injecting the whole 3D dataset SOBS in the CTRL forecast simulation is expected to give the best performances of such a DA technique; it constitutes one of the DAEs discussed later. Many data subsets are possible: by species, by level, or a combination of both; only some of them were selected and analyzed. The different cases are run over the 11–15 March period with the new initialization on 0000 UTC 11 March; they can be compared with the CTRL or the SOBS reference simulations. The expected best case for DA is to inject at the initial state all of the observations from the SOBS run at all levels into the CTRL run; this simulation is called SALLA. Different cases will be studied, they are defined and reported in Table 1. The last two characters of DAE "L1" or "LA" respectively refer to the "first" and "all" levels are impacted by the DAE definition. For all simulations, the way they return toward the CTRL simulation forecast will be investigated, or in other words, how far this simulation diverges from the pseudo-observations represented by SOBS and returns toward the CTRL simulation. Two additional cases are also considered: 1) the use of DA to improve the emissions instead of the initial state concentrations (DAE called EMILA) and 2) the use of the same O_3 concentration delta (SOBS–CTRL) calculated at the first level and propagated to all levels (called SORLA). All simulations in this study share the same chemical boundary conditions. The use of a large European domain allows us to be less dependent on boundary conditions when focusing on the center of the domain.

c. *Model setup*

The model grid covers a large part of Europe using a regular longitude/latitude grid with a resolution of 0.15° (east–west) $\times$ 0.1° (south–north). For the boundary conditions, global chemical fields are issued from the Copernicus program (http://www.copernicus.eu/). The vertical discretization has nine layers extending from approximately 997 to 500 hPa. The period of 1–15 March 2014 is simulated by three successive blocks of 5-day elementary simulations; the analysis will be performed over the last block: 11–15 March. The model configuration is summarized here, but the reader can refer to the reference CHIMERE publication (Mailler et al. 2017) for details on the corresponding model components and references as well as non-user-specific model characteristics. The gas-phase chemical mechanism is "MELCHIOR2," which consists of a simplified version (40 species and 120 reactions) of the full chemical mechanism based on the concept of chemical operators.

TABLE 1. List of data assimilation experiments. For the case of DAE affecting concentrations, the initialization state is modified at 0000 UTC 11 Mar 2014. For the case EMILA, the use of the SOBS emission dataset is used all along the period of interest 11–15 Mar 2014.

DAE name	Changes	Impacted species	Impacted levels	Comments
CTRL	—	—	—	Control run
SOBS	Random changes (meteorological conditions and emissions)	—	—	"Observations" dataset
SALLA	Initialization with SOBS	All	All	All CHIMERE species
SALL1	"	All	1	All CHIMERE species
SONLA	"	O_3 and NO_2	All	—
SMALA	"	Main species	All	All particulate species, NO_2, NH_3, O_3, and SO_2
SMAL1	"	Main species	1	All particulate species, NO_2, NH_3, O_3, and SO_2
SNXLA	"	NO_2	All	—
SNXL1	"	NO_2	1	—
SOZLA	"	O_3	All	—
SSDLA	"	SO_2	All	—
SSDL1	"	SO_2	1	—
SSIL1	"	SIA	1	Only the particulate phase of SIA
SSUL1	"	Sulfate	1	Only sulfate
SNHL1	"	Total ammonia	1	Ammonia + ammonium
SNHLA	"	Total ammonia	All	Ammonia + ammonium
SORLA	"	O_3	All	Application of the first-level delta to the other levels
EMILA	Emissions	All emitted species	All	Use of emission dataset from SOBS

Modeled particulate matter includes primary particulate matter, and secondary inorganic (nitrate, sulfate, and ammonium based on the "ISORROPIA" thermodynamic equilibrium model) and organic aerosol. The secondary production of organic aerosols is represented according to a single-step oxidation of the relevant anthropogenic and biogenic precursors and gas-particle partitioning of the condensable oxidation products. Biogenic emissions are computed with the Model of Emissions of Gases and Aerosols from Nature (MEGAN), version 2.04; sea-salt and mineral dust emissions are also considered. The particle size ranges from 10 nm to 40 μm over 10 bins according to these ranges from bin 1 to bin 10, which are respectively [10.00–22.01 nm], [22.01–48.43 nm], [48.43–106.7 nm], [106.7–234.7 nm], [234.7–516.2 nm], [516.2 nm–1.14 μm], [1.14–2.50 μm], [2.5–5.0 μm], [5.0–10.0 μm], [10–40 μm]. This paper uses the following abbreviations: secondary inorganic aerosol (SIA; SIA is the sum of nitrate, sulfate, and ammonium), secondary organic aerosols (SOA; anthropogenic and biogenic in origin), natural mineral dust (DUST), sea salt (SALT), and total primary particulate matter (TPPM; TPPM is total primary anthropogenic carbonaceous and noncarbonaceous species). The attribute "-10," "-25," or "coarse" respectively refers to the PM10, PM2.5, or coarse PM (difference between PM10 and PM2.5) matrices.

Horizontal transport is solved with the second-order van Leer scheme (van Leer 1979). The CHIMERE model is forced by external and precalculated meteorological fields. For this study, we are using ECMWF Integrated Forecasting System reanalyzes (with a resolution of 0.125°). About the wind field, CHIMERE reads the horizontal wind components u and v only and rediagnoses the vertical component w to ensure mass conservation. In urban areas, the horizontal wind component is reduced to represent shallower mixing close to the surface (Terrenoire et al. 2015). Although the study is performed in a forecasting perspective, meteorological data are issued from reanalyses. Emissions data for the main pollutants are issued by the European Monitoring and Evaluation Programme (EMEP; EMEP is a cooperative program for monitoring and evaluation of long-range transmission of air pollutants in Europe) and are available online (http://www.emep.int).

3. Results and discussion

a. Analysis criteria

Simulation SALLA as defined in Table 1 represents the most favorable DA case since all CHIMERE species concentrations are initialized with the SOBS dataset. For the analysis, averaged concentrations values or criteria defined latter in this paper will be aggregated over selected countries coded with three letters as France (FRA), Great Britain (GBR), Germany (DEU), Poland (POL), Romania (ROU), Italy (ITA), the Netherlands (NLD), Sweden (SWE), Switzerland (CHE), Spain (ESP), and a group of the previous countries (GRP). The analysis will be performed for averaged concentrations when aggregated over a country. Relative criteria are adopted to be less dependent on absolute concentrations.

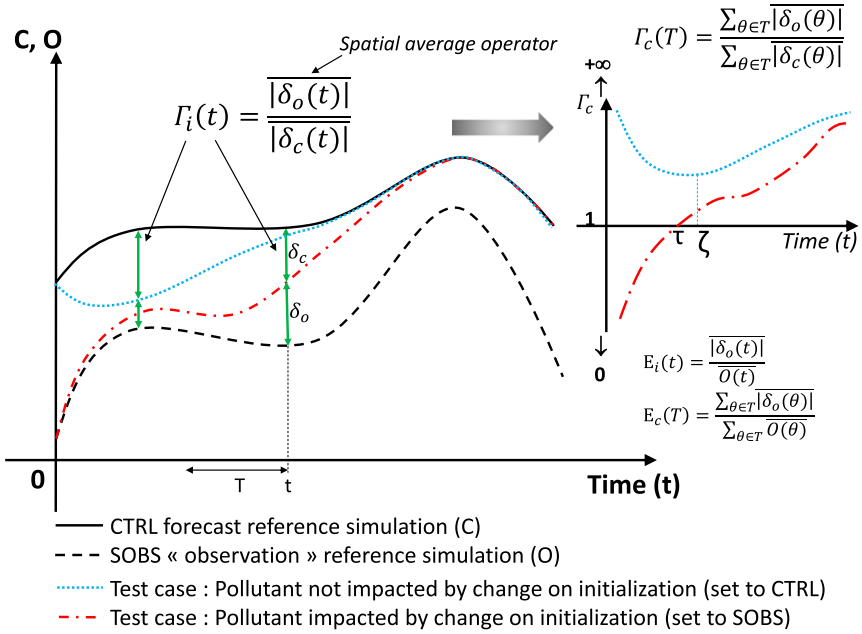

FIG. 1. Schematic representation of the criteria Γ and E (instantaneous i; averaged/cumulated c over a period T) and tipping times τ and ζ for a given DAE.

At a given time t or for a given period T, the average relative (r) or absolute (a) distance (instantaneous i or cumulated c over time) between SOBS and CTRL can be calculated according to Eq. (1). These criteria are useful to complement the analysis and are defined as

$$\varepsilon_{i,r}(t) = \frac{\overline{|C_{\mathrm{CTRL}}(t) - C_{\mathrm{SOBS}}(t)|}}{\overline{C_{\mathrm{SOBS}}(t)}},$$

$$\varepsilon_{c,r}(T) = \frac{\sum_{\theta\in T}\overline{|C_{\mathrm{CTRL}}(\theta) - C_{\mathrm{SOBS}}(\theta)|}}{\sum_{\theta\in T}\overline{C_{\mathrm{SOBS}}(\theta)}},$$

$$\varepsilon_{i,a}(t) = \overline{|C_{\mathrm{CTRL}}(t) - C_{\mathrm{SOBS}}(t)|}, \quad \text{and}$$

$$\varepsilon_{c,a}(T) = \frac{1}{N_T} \times \sum_{\theta\in T}\overline{|C_{\mathrm{CTRL}}(\theta) - C_{\mathrm{SOBS}}(\theta)|}. \qquad (1)$$

Similarly, a relative error E is defined as Eq. (2) for each experiments described in Table 1. This criterion evaluates the ability to remain close to the observations SOBS simulation; however, this simulation can also strongly diverge from both SOBS and CTRL simulation. Then a second criterion Γ inspired from Menut and Bessagnet (2019) is defined for a given time t to calculate the relative difference of distances δ of each simulated concentration C from SOBS ($\delta_o = C_{\mathrm{DAE}} - C_{\mathrm{SOBS}}$) and CTRL ($\delta_c = C_{\mathrm{DAE}} - C_{\mathrm{CTRL}}$) reference simulations [Eq. (3) and, graphically, Fig. 1]. As previously mentioned, a usual root-mean-square error (even normalized) would have been sensitive to the initial difference between SOBS and CTRL. In this study as well as in Menut and Bessagnet (2019) the goal is to evaluate the behavior of each experiment between the SOBS and CTRL simulations. This L1-based error statistic is more robust and stable when the denominator is close to 0, that is, C_{DAE} is close to C_{CTRL}. For a given country the mean values of $|\delta_o|$ and $|\delta_c|$ are considered. For both criteria E and Γ, an instantaneous criterion can be calculated for a given time t (or for a maximum value), or cumulated over a given period T (on a daily basis or from the initial state) to smooth the time evolution of criteria. These criteria are respectively indexed with i and c subscript letters, and x indicates maximum. This smoothing procedure is used because Γ is not a monotonic function of time; sometimes the convergence toward the CTRL or SOBS can pass through several local extrema values. In the case of pollutants concentrations initialized with observations issued from the SOBS simulation, $\Gamma_i = 0$; in all other cases, $\Gamma_i = +\infty$ (default initialization to CTRL):

$$E_i(t) = \frac{\overline{|C_{\mathrm{DAE}}(t) - C_{\mathrm{SOBS}}(t)|}}{\overline{C_{\mathrm{SOBS}}(t)}},$$

$$E_c(T) = \frac{\sum_{\theta\in T}\overline{|\delta_o(\theta)|}}{\sum_{\theta\in T}\overline{C_{\mathrm{SOBS}}(\theta)}}, \qquad (2)$$

$$\Gamma_i(t) = \frac{\overbrace{\overline{|C_{\mathrm{DAE}}(t) - C_{\mathrm{SOBS}}(t)|}}^{\overline{\delta_c(t)}}}{\underbrace{\overline{|C_{\mathrm{DAE}}(t) - C_{\mathrm{CTRL}}(t)|}}_{\overline{\delta_c(t)}}}, \quad \text{and}$$

$$\Gamma_c(T) = \frac{\sum_{\theta \in T} \overline{|\delta_o(\theta)|}}{\sum_{\theta \in T} \overline{|\delta_c(\theta)|}}. \tag{3}$$

For pollutants initialized with SOBS (or partially initialized for the case of PM, which is an aggregation of individual species), a "tipping time" τ is defined as the first time when the cumulated criterion Γ becomes greater than 1 [Eq. (4) and Fig. 1], that is, when the concentrations become closer to the CTRL forecast. For all other pollutants that are not affected (or partially affected) by an initialization to SOBS, a tipping time ζ [Eq. (5) and Fig. 1] is defined as the maximum of gain toward the SOBS simulation based on the cumulated criterion Γ from the initial state. For example, considering the SOZLA experiment, only O_3 concentrations are initialized to SOBS, and nitrate concentrations that will be initialized with CTRL should benefit from the O_3 concentrations initialization before returning toward the CTRL forecast. Therefore, for a given DAE, a pollutant is characterized by either a tipping time τ or ζ; the comparison between two scenarios can be possible only through the tipping time τ, whereas the comparison of simulations with ζ should be carefully interpreted and must be accompanied by a discussion with the distance to the observations (E criterion). A DA simulation will exhibit good performances when τ will be high; that is, its time to return to the CTRL simulation will be long. A species initialized at CTRL being indirectly affected by DA will display good performance when the concentration remains close to the minimum at time ζ and is associated with a low distance to SOBS. In our cases, if τ or $\zeta = 120$ h, it means that these tipping times are at least equal to 120 h. The tipping times are indicated as follow:

$$\tau = \{[\Gamma_c(t) - 1](_{-}\nearrow^{+})\}^{-1} \quad \text{and} \tag{4}$$

$$\zeta = \left\{\min_t[\Gamma_c(t)]\right\}^{-1}. \tag{5}$$

The ratio Π [Eq. (6)] is very convenient to analyze two DAEs as a way to compare their proximity to the SOBS dataset; this ratio can be calculated on the basis of either the Γ or E criterio. The Π ranges from 0 to $+\infty$, with 1 being the boundary at which DAE A and DAE B will be respectively closer to the SOBS dataset for $\Pi_\Gamma < 1$ and $\Pi_\Gamma > 1$. Then a third characteristic time ω [Eq. (7)] can be defined to compare two DAEs; $t = \omega$ when $\Pi_\Gamma(t)$ passes through the unit value 1. This characteristic time determines, between DAE A and DAE B, the simulation A or B that benefits the most from DA. However, as in other defined criteria, the Π ratio as a function of time is not necessary monotonic and can cross the unit value 1 several times; $t = \omega$ (like τ) is considered as the last time this value is crossed. The ratio and time are respectively given as

$$\Pi_{\Psi_{A,B}} = \Psi_{\mathrm{DAE}_A}/\Psi_{\mathrm{DAE}_B}; \quad \Psi = E, \Gamma \tag{6}$$

$$\omega = \{[\Pi_{\Gamma_c}(t) - 1](_{-}\nearrow^{+} \text{ or } {}^{+}\searrow_{-})\}^{-1}. \tag{7}$$

The online supplementary material contains additional figures to support the analysis. Figure S1 reports maps of concentrations for the reference simulations CTRL and SOBS averaged over the forecast period 11–15 March 2014 for some key species; Figs. S2–S17 report the time evolution of PM10, PM2.5, O_3, NO_2, SO_2, particulate nitrate, sulfate, ammonium, and nitric acid ground concentrations (50th and 90th percentile) aggregated over the group of countries GRP for the DAE in comparison with CTRL and SOBS simulations; Figs. S18–S26 report maps of Π_{Γ_c} ratio values for some 2 by 2 comparisons for two DAEs averaged over the forecast period 11–15 March 2014; Figs. S27–S34 report the maps of Γ_c ratio values for two DAEs averaged over the forecast period 11–15 March 2014; Figs. S35–S53 report the maps of averaged latitudinal cross sections of Π_{Γ_c} ratio values for key pollutants for the first three days of the forecast from 11 to 13 March 2014.

b. *Analysis of the best expected DAE*

As previously described, SALLA is the experiment that is expected to provide the best performances of DA (Fig. 2). In this simulation, all CHIMERE species from the SOBS are injected at all levels on 0000 UTC 11 March in the CTRL forecast simulation. In Fig. 2 and Table 2, on average for GRP, NO_2, SO_2, O_3 ground concentrations quickly return toward the CTRL simulation in 17, 20, and 30 h, respectively. PM2.5, PM10, and the constituting species nitrate and sulfate remain influenced by the initial state during a longer period: respectively 81, 85, 58, and at least 120 h. If the Γ function is monotonic on average over the GRP area, Romania exhibits more unstable behavior (Fig. 3) for some species like O_3 and PM2.5 that return toward observations before rising again toward the CTRL simulation. However, for O_3 Romania exhibits the best performance and the worst performance is observed in Spain, with a gain lasting only 6 h on average. In Italy and France, good performances for both sulfate and SO_2

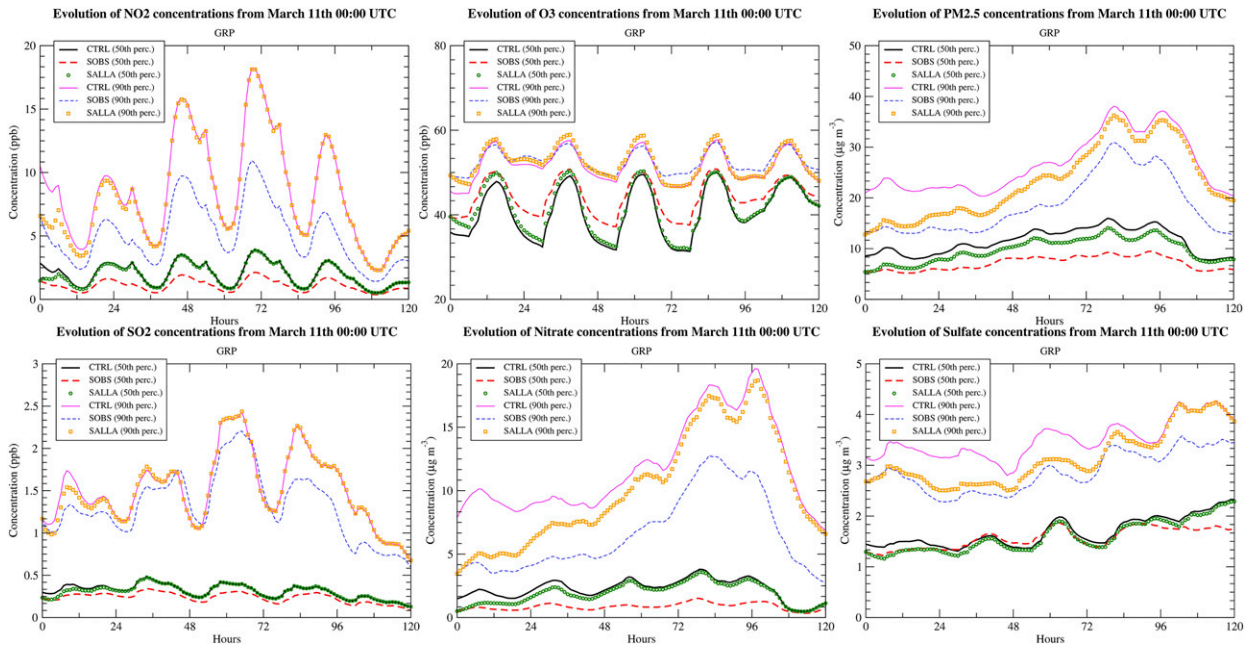

FIG. 2. Time evolution of NO_2, O_3, PM2.5, SO_2, particulate nitrate, and sulfate ground concentrations (50th and 90th percentile) aggregated over the group of countries GRP for the simulation SALLA in comparison with the CTRL and SOBS simulations.

concentrations are observed; for this latter compound τ reaches 38 h on average. For PM10 and PM2.5, the Netherlands and Spain show the best performances, with a high tipping time τ, and the United Kingdom and Sweden show the lowest performances. For O_3, clearly Spain and Switzerland show the lowest performance, with a fast return to the CTRL in less than 9 h; these mountainous countries could be more influenced by the top boundary conditions. For nitrates, Spain, France, and the Netherlands show the best performances; in France the tipping time τ is 67 h, and it reaches 120 h over the Netherlands. For sulfate, low τ values are noteworthy in Romania and Poland; indeed, these regions emit large amounts of sulfur dioxides driving the fast return to the CTRL simulation.

The evolution of the daily averaged error E (Fig. 3) is very dependent on pollutants and areas. Generally, the error increases with time but for O_3 the evolution is very flat and low along the 5-day forecast with a clear diurnal cycle of instantaneous errors. The error is generally the highest at night in line with higher differences between SOBS and CTRL simulations.

TABLE 2. Tipping times τ (h) for several species, for the SALLA and the corresponding cumulated errors from 11 Mar E_c (%) at this time τ over some areas in Europe. The average cumulated absolute error $\varepsilon_{c,a}$ between CTRL and SOBS at τ or ζ is given in parentheses (parts per billion for gases O_3, NO_2, and SO_2 and micrograms per meter cubed for particulate matter and constitutive species nitrate, sulfate and ammonium).

	PM10		PM2.5		O_3		NO_2		SO_2		Nitrate		Sulfate		Ammonium	
Country	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)	τ	E_c ($\varepsilon_{c,a}$)
GRP	81	23 (5.0)	85	25 (4.5)	30	6 (4.8)	17	38 (1.2)	20	8 (0.1)	58	58 (1.9)	120	20 (0.4)	61	27 (0.6)
FRA	79	31 (7.8)	79	34 (7.6)	26	7 (5.8)	19	60 (2.2)	32	21 (0.2)	67	71 (4.4)	109	5 (0.2)	54	39 (1.3)
GBR	99	8 (1.5)	105	8 (1.2)	28	6 (4.9)	8	21 (0.8)	9	5 (0.0)	27	20 (0.7)	25	10 (0.3)	61	4 (0.1)
DEU	94	22 (5.3)	94	24 (5.2)	30	10 (7.7)	27	17 (1.3)	14	17 (0.2)	77	51 (3.0)	80	6 (0.1)	101	20 (0.8)
POL	56	36 (4.9)	58	42 (4.1)	26	11 (9.2)	19	64 (2.6)	7	10 (0.3)	54	276 (1.8)	8	7 (0.0)	55	78 (0.6)
ROU	78	14 (1.5)	75	14 (1.3)	60	4 (3.7)	20	7 (0.1)	15	21 (0.2)	49	49 (0.4)	61	1 (0.0)	77	12 (0.1)
ITA	78	25 (6.9)	76	27 (6.7)	25	4 (3.8)	13	50 (2.1)	38	16 (0.1)	33	42 (1.6)	120	21 (0.5)	36	31 (0.9)
NLD	120	18 (2.1)	120	13 (1.0)	20	6 (4.8)	7	12 (0.8)	15	18 (0.1)	120	36 (0.3)	75	9 (0.1)	83	4 (0.2)
SWE	32	23 (3.1)	49	17 (1.3)	53	3 (2.9)	17	18 (0.1)	17	16 (0.0)	25	59 (0.1)	47	4 (0.0)	19	43 (0.0)
CHE	92	13 (3.5)	79	19 (3.8)	19	7 (5.5)	12	66 (2.1)	6	7 (0.0)	51	49 (2.4)	119	6 (0.1)	58	31 (0.8)
ESP	120	27 (4.0)	120	32 (4.0)	6	4 (3.1)	8	45 (1.5)	11	26 (0.2)	85	60 (1.4)	120	12 (0.3)	98	30 (0.6)

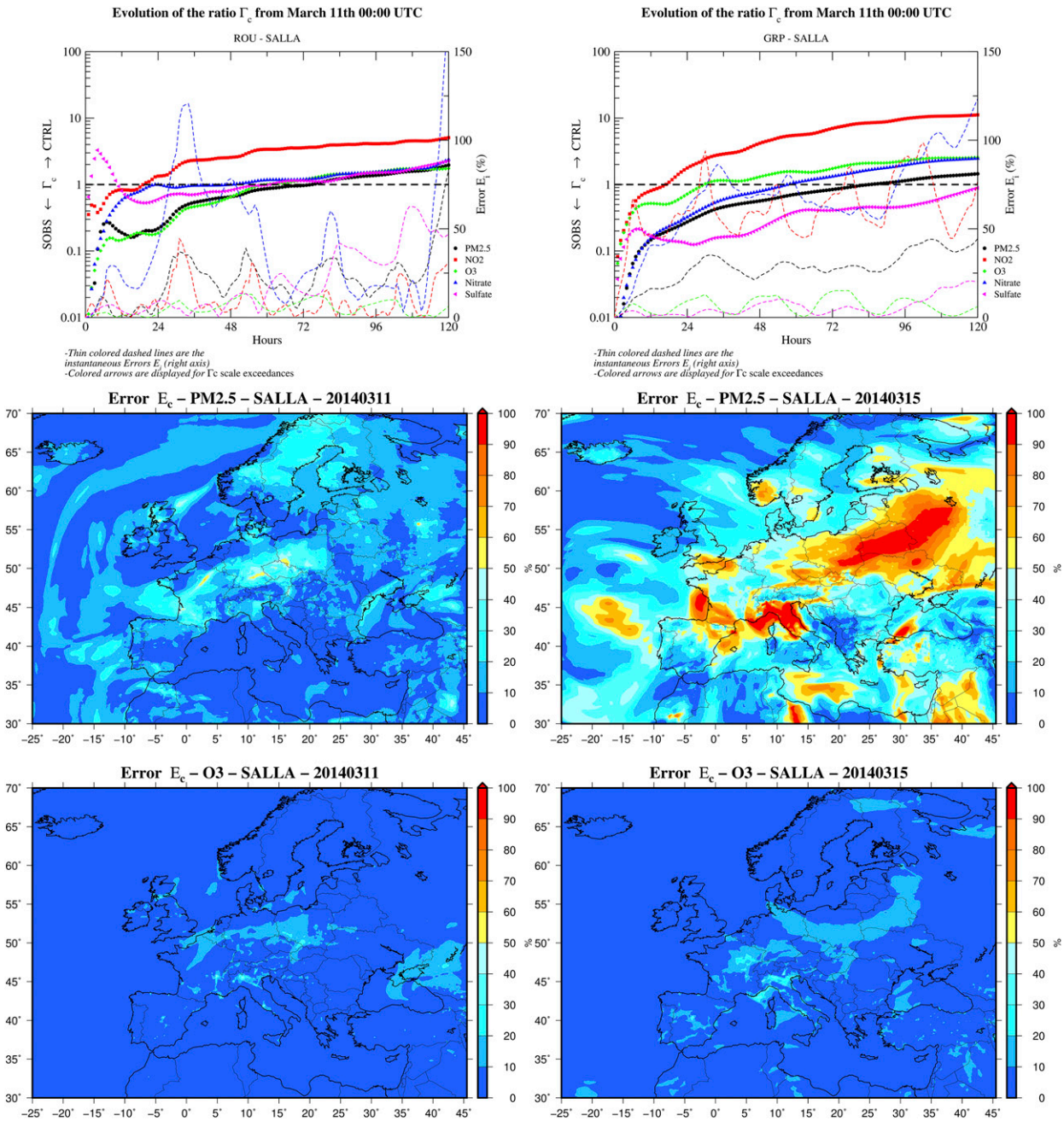

FIG. 3. (top) Evolution of the Γ and E criteria for some species issued from the SALLA simulation for Romania (ROU) and the GRP group. Also shown is the evolution of daily average E for (middle) PM2.5 and (bottom) O_3 surface concentrations for 11 and 15 Mar 2014.

Figure 4 shows the maps of τ values for the forecasting period. For PM2.5 the τ values are very high from the very north of France to the Netherlands as previously mentioned, mainly driven by the performance of DA on nitrate in this area with low errors E. A similar behavior is observed over the Mediterranean basin. These maps show stiff τ gradients for all pollutants, certainly the effect of chemistry and the definition of this criterion based on the calculation of a threshold exceedance. Moreover, for some pollutants the return to the CTRL simulation can pass through several extrema and we remind here that τ is defined as the last cross of the unit value 1. Over the south part of the Atlantic Ocean τ values are generally high because of the weak effect of boundary conditions because of northeasterly fluxes in this region and low emissions. Over the north, boundary

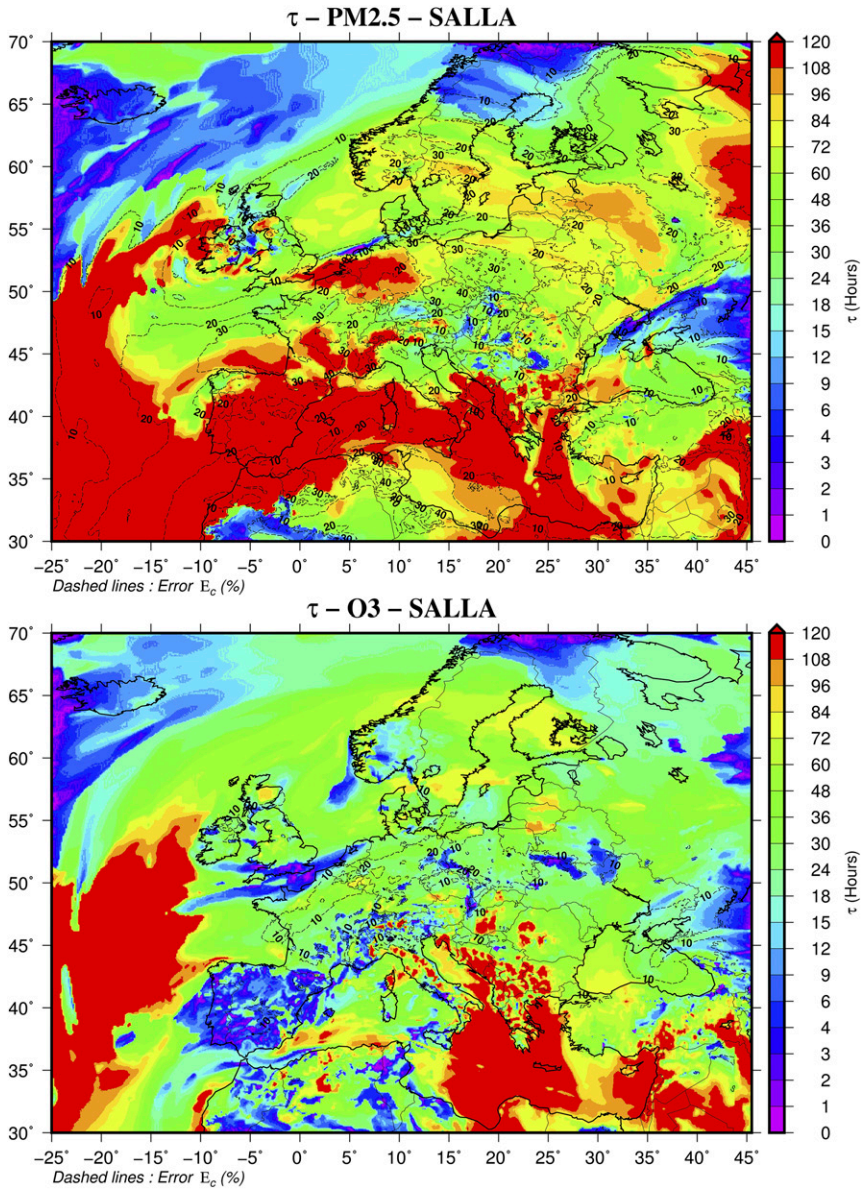

FIG. 4. Map of tipping time τ for (top) PM2.5 and (bottom) O_3 and corresponding cumulated errors from 11 Mar (dashed contour lines) for the DAE SALLA.

conditions associated with westerly winds quickly annihilate the benefit of a better initialization with very low τ values.

c. *Analysis of the other DAE*

1) IMPACT OF VERTICAL LEVELS

Clearly, on average for all experiments and most species the tipping times are the lowest when the DA concerns only the first level; the information injected at the initial state is rapidly lost; the influence of the CTRL simulation from the upper levels annihilates the benefit of DA (Tables 3 and 4). For primary and secondary pollutants ζ and τ tipping times are in favor of an injection of SOBS data at all levels. For PM10, the difference of tipping times τ between SALLA and SALL1 is very large: respectively 81 and 2 h. The difference is less impressive for NO_2 concentrations, with 17 and 3 h, respectively, for the tipping times. For the case of DAE on primary pollutants, comparing, for instance, SNXLA and SNXL1 leads to smaller differences in ζ tipping times on secondary pollutants, with, for instance, ζ values on nitrate concentrations that are 15 and 10 h, respectively, with this DA of NO_2 concentrations. The maximum

TABLE 3. Tipping times τ (boldface type) and ζ (italic type) for the pollutants, for the group of countries GRP, and the corresponding cumulated errors E_c (%) at this time for all assimilation cases.

	PM10		PM2.5		O_3		NO_2		SO_2		Nitrate		Sulfate		Ammonium	
Cases	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c	τ, ζ	E_c
SALLA	**81**	**23**	**85**	**25**	**30**	**6**	**17**	**38**	**20**	**8**	**58**	**58**	**120**	**20**	**61**	**27**
SOZLA	*10*	*62*	*10*	*68*	**44**	**6**	*5*	*76*	*17*	*17*	*10*	*141*	*11*	*10*	*10*	*57*
SALL1	**2**	**29**	**2**	**32**	**2**	**8**	**3**	**41**	**1**	**4**	**2**	**70**	**5**	**5**	**2**	**27**
SMALA	**73**	**24**	**76**	**26**	**49**	**5**	**16**	**39**	**19**	**8**	**47**	**62**	**120**	**20**	**54**	**28**
SONLA	*18*	*58*	*18*	*62*	**47**	**5**	**16**	**40**	*17*	*16*	*19*	*122*	*14*	*12*	*20*	*53*
SNXLA	*14*	*58*	*14*	*62*	*67*	*9*	**15**	**40**	*17*	*16*	*15*	*120*	*13*	*12*	*21*	*52*
SNXL1	*10*	*60*	*10*	*65*	*4*	*15*	**3**	**42**	*17*	*16*	*10*	*131*	*15*	*11*	*10*	*54*
SORLA	*59*	*52*	*58*	*58*	**120**	**5**	*29*	*67*	*77*	*11*	*11*	*143*	*86*	*6*	*81*	*48*
SNHL1	**0**	**47**	**0**	**53**	*66*	*10*	*0*	*68*	*1*	*7*	*3*	*127*	*6*	*9*	**2**	**30**
SSIL1	**1**	**37**	**1**	**39**	*44*	*10*	*5*	*81*	*3*	*9*	**2**	**71**	**5**	**5**	**2**	**28**
SSUL1	**0**	**51**	**0**	**57**	*0*	*9*	*0*	*68*	*2*	*8*	*4*	*128*	**5**	**5**	*3*	*53*
SNHLA	**0**	**47**	**0**	**53**	*43*	*10*	*29*	*72*	*2*	*9*	*16*	*108*	*86*	*6*	**12**	**28**
SSDLA	*23*	*57*	*23*	*63*	*21*	*10*	*12*	*81*	**18**	**9**	*24*	*130*	*16*	*10*	*24*	*56*
SSDL1	*0*	*43*	*0*	*48*	*0*	*9*	*0*	*68*	**1**	**4**	*6*	*127*	*85*	*6*	*5*	*53*
EMILA	*115*	*38*	*116*	*41*	*101*	*8*	*120*	*29*	*6*	*10*	*120*	*62*	*93*	*11*	*67*	*43*

benefit for nitrate concentrations is reached earlier in SNXL1 with a higher cumulated error (131% for SNXL1 against 120% for SNXLA).

2) IMPACT OF SPECIES

On average, assimilating all of the species instead of a group of species leads to better results; values of tipping times τ have been provided in the previous section devoted to the SALLA DAE. The case of O_3 will be discussed later in a dedicated subsection; the behavior of this secondary species is peculiar. In comparing the DAE involving all species (SALLA) and the main species (SMALA), it is seen that the added values of assimilating all volatile organic compounds (VOCs) and intermediary species is low, with, for instance, a tipping time τ ranging from 73 h for SMALA to 81 h for SALLA on the PM10 concentrations (Table 2).

The differences between DAE over all levels of total ammonia (SNHLA), SO_2 (SSDLA), and NO_2 (SNXLA) on particulate matter is first analyzed. Total ammonia and NO_2 concentrations have an influence on nitrate concentrations and therefore on PM and indirectly on O_3 chemistry. The differences of tipping time ζ between SNHLA and SNXLA for nitrates are low on average, but an assimilation of ammonia concentrations is performing slightly better on nitrate and ammonium concentrations and then on PM. Assimilation of NO_2 concentrations shows the best performances over the Netherlands and the lowest over the United Kingdom on nitrate concentrations. The vertical cross section of Π ratio in Fig. S49 of the supplementary material shows a benefit of SNXLA at altitudes up to 500 m on average over emission zones latitude for nitric acid and up to 60°N for the first forecast day 11 March 2014. Elsewhere over remote places ammonia concentrations variation play an important role and in higher altitudes a change of concentrations modifies the chemical regimes with a clear advantage on assimilating ammonia versus NO_2 on nitrate and nitric acid.

An assimilation of SO_2 is expected to have an influence on PM through the formation of ammonium sulfate. Except for SO_2 and sulfate concentrations SNXLA displays the best performances relative to SSDLA data experiments. This is observed in Fig. S51 of the supplementary material on the map of Π ratios; the spatial footprint for PM2.5 concentrations clearly shows a benefit of DA of NO_2 concentrations over lands largely influenced by nitrate concentrations even in eastern countries where sulfur dioxide emissions are high. The vertical cross sections of Π ratio shows an advantage of SSDLA on nitrate concentrations at altitude (visible bubble at about 2 km) around 50°N as a result of industrial emissions over Poland emitted at altitude that have a more beneficial impact on the chemistry through thermodynamic equilibrium as compared with an update of NO_2 concentrations (particular chemical regime in this region). A comparison of DA of SO_2 versus total ammonia concentrations shows better performances on PM, NO_2, nitrate, and ammonium concentrations in favor of the SNHLA simulation while SSDLA shows better performances for O_3, sulfate and SO_2 concentrations. Sulfur dioxide has an impact on the OH radical consumption, leading to a slight effect on O_3 chemistry.

More and more instruments allow one to provide hourly SIA concentrations. Assimilating SIA at the initial state (SSIL1 experiment) has a positive impact in terms of error E, which is a low 39% as compared with the ε error of 61% between SOBS and CTRL, but the

TABLE 4. Two-by-two comparisons of data assimilation experiments for the group of countries GRP. Labels A or B alone mean that DAE A or DAE B, respectively, shows the best performances, and A $\overset{\omega}{\rightarrow}$ B means that DAE A exhibits first the best performances until $\omega - 1$ hours, with DAE B displaying a better benefit afterward. Tables S1–S20 in the online supplementary material report ω per country.

DAE A vs DAE B	PM10	PM2.5	O_3	NO_2	SO_2	Nitrate	Sulfate	Ammonium
SALLA vs EMILA	A	A	A	A $\overset{22}{\rightarrow}$ B	A $\overset{44}{\rightarrow}$ B	A $\overset{99}{\rightarrow}$ B	A	A
SALL1 vs SALLA	B	B	B	B	B	B	B	B
SONLA vs SOZLA	B $\overset{7}{\rightarrow}$ A	B $\overset{7}{\rightarrow}$ A	A	A	A	B $\overset{5}{\rightarrow}$ A	A	B $\overset{7}{\rightarrow}$ A
SONLA vs SNXLA	B	B	A	A	A $\overset{5}{\rightarrow}$ B	B	A $\overset{5}{\rightarrow}$ B	B
SONLA vs EMILA	B	B	A	A $\overset{22}{\rightarrow}$ B	B	A $\overset{23}{\rightarrow}$ B	A $\overset{9}{\rightarrow}$ B	B
SOZLA vs SORLA	A $\overset{7}{\rightarrow}$ B	A $\overset{7}{\rightarrow}$ B	A $\overset{3}{\rightarrow}$ B	A $\overset{6}{\rightarrow}$ B	A $\overset{6}{\rightarrow}$ B	A $\overset{7}{\rightarrow}$ B	A $\overset{5}{\rightarrow}$ B	A $\overset{7}{\rightarrow}$ B
SOZLA vs SNXLA	B	B	A	B	A $\overset{2}{\rightarrow}$ B	B	A $\overset{2}{\rightarrow}$ B	B
SALLA vs SORLA	A	A	A $\overset{4}{\rightarrow}$ B	A	A	A	A	A
SSIL1 vs SNXL1	A	A	B	B	A $\overset{12}{\rightarrow}$ B	A	A	A
SNHL1 vs SALL1	B	B	B	B	B	B	B	B
SNHL1 vs SNHLA	B	B	B	B	B	B	B	B
SNHL1 vs SNXL1	A	A	B	B	A	A $\overset{8}{\rightarrow}$ B	A $\overset{3}{\rightarrow}$ B	A
SNXL1 vs SNXLA	B	B	B	B	B	B	B	B
SNXL1 vs EMILA	B	B	B	A $\overset{8}{\rightarrow}$ B	B	A $\overset{10}{\rightarrow}$ B	B	B
SNHLA vs SNXLA	A	A	B	B	A	A	A $\overset{6}{\rightarrow}$ B	A
SSDLA vs SNHLA	B	B	B	B	A	B	B $\overset{4}{\rightarrow}$ A	B
SSDLA vs SNXLA	B	B	B	B	A	B	A	B
SSDL1 vs SSDLA	B	B	B	B	B	B	B	B

information is quickly lost in 1 h, on average. The same behavior is observed for nitrate and ammonium for which the tipping time τ does not exceed on average 5 h. Acting on NO_2 for DA at the first level is more beneficial. A gain is observed during 14 h for the PM10 and PM2.5, but the error E remains high, and SSIL1 is more beneficial in terms of DA relative to SNXL1 for the PM species because of the partial update of concentrations at the initial state. The Π ratio in Fig. S53 of the online supplementary material between SSUL1 and SSDL1 allows a comparison of the benefit between respectively assimilating sulfate and sulfur dioxide concentrations at the first level. Except for SO_2 there are more benefits to assimilate the sulfate concentrations for all compounds close to the ground.

3) THE CASE OF O_3

In this section, the case of O_3 is analyzed mainly over GRP (Table 3). Comparing the expected best case SALLA with all species from SOBS injected at the initial state to the DAE with only O_3 injected at the initial state (SOZLA) the influence on τ is very low with surprisingly, a higher tipping time τ for SOZLA case: 44 h as compared with 30 h for SALLA. This means that an initial update of other concentration species has little influences over this period on the O_3 chemistry. If NO_2 is added to O_3 in the DA process (simulation SONLA), performances are better with a tipping time τ reaching 47 h with a lower error on average; this time reaches 49 h if other species are added (SMALA experiment). This could indicate that CHIMERE does not benefit from the update of all remaining species (radicals, VOCs) for the O_3 concentrations; this feature is confirmed by the simulation without those species (SMALA) and leading to a tipping time τ for O_3 reaching 49 h. Updating only NO_2 (SNXLA) concentrations leads to a tipping time ζ of 67 h on O_3 concentrations but with a higher error (9%) relative to the other cases (about 5%); therefore, an initial update of NO_2 has a positive impact but performances are less beneficial when compared with the previous cases with DA of O_3 concentrations (SOZLA, SALLA, and SONLA).

In line with the previous surprising outcomes, other counterintuitive findings arise from the analysis of the SORLA simulation. This SORLA experiment only updates O_3 at the first level with the SOBS dataset and propagates the corresponding delta of O_3 concentrations along the column. This is a first DA that is very feasible for modelers. On average, this simulation provides the best performances with a high tipping time exceeding 120 h and a low cumulated error along this period. However, looking at maximum values and the time series (Fig. 5) the best performances are observed for nighttime concentrations values for the SORLA simulation. However, the daily maximum concentrations seem impaired with an overshoot during the three first days if we compare with the full update with SOBS data (SOZLA simulation). This feature supports the assumption that other species (VOCs, radicals, intermediary species) could have a more positive influence during daytime with a more beneficial influence by starting at the equilibrium for our chemical mechanism. The vertical cross-section Π ratio in Fig. S40 of the supplementary material shows this quick advantage of

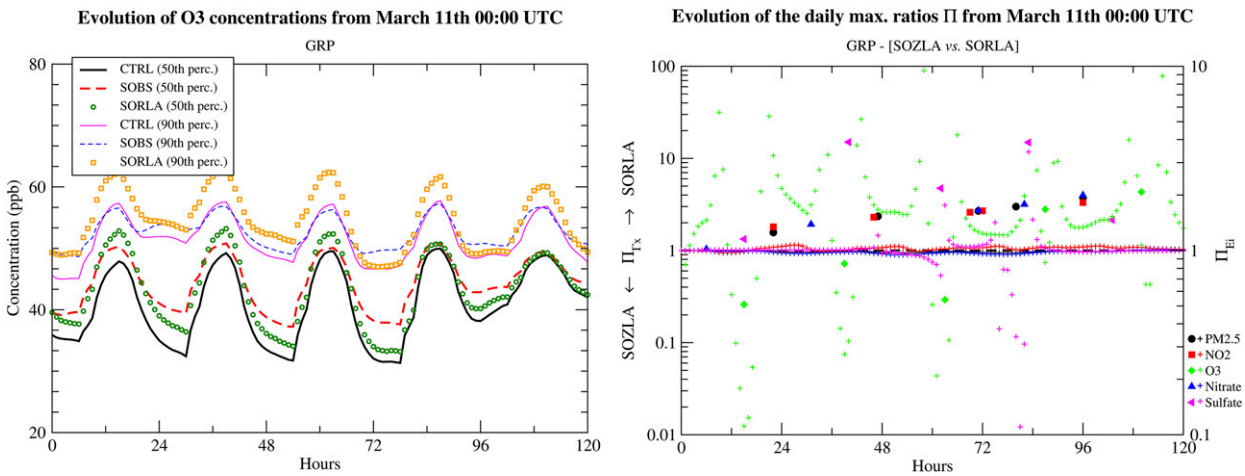

FIG. 5. (left) Time series of O_3 concentrations for the SORLA experiment. (right) Evolution of the Π ratio based on daily maximum Γ_x in comparison with the SOZLA experiment; for this panel, the filled colored symbols correspond to each given species (left *y* axis), and the plus sign is the corresponding Π ratios of instantaneous errors from 11 Mar E_c (right *y* axis).

SORLA compared to SOZLA at the surface and the higher latitudes. A closer look at the differences between SORLA and SOZLA experiments on NO_2 vertical cross section shows an advantage of the SORLA simulation at higher altitudes close to the top boundary conditions the first day; the opposite is observed for O_3. Two explanation are proposed. A first possible explanation could be that forcing with a constant O_3 concentration delta along the column involves first a positive influence on species involved in the O_3 cycle formation and therefore on O_3 concentrations through the chemical process. This delta could compensate the differences of chemical regime at boundary conditions issued from a global model and the CHIMERE chemical regime. A more plausible and complementary explanation could be that this homogeneous delta injected along the column at 0000 UTC will be a recorded information for the next hours up to the ground through the vertical mixing that counterbalance the time fading of low concentrations, and therefore the ground level benefits from this initial injection along the whole column. These two explanations are supported by Fig. 6, which displays this quick advantage of SORLA experiment at the first level; in upper altitudes the benefits appear later because of the chemical processes inertia.

4) IMPACT OF DA OF EMISSIONS INSTEAD OF CONCENTRATIONS

A debate in the community concerns the use of DA techniques applied on emissions instead of concentrations, the question being, Could we use observational data to adjust emissions (spatial patterns and/or magnitude of emissions) instead of concentrations, hoping for a better benefit? The EMILA data experiment assimilates the anthropogenic emissions that were used to create the SOBS dataset; therefore, the remaining difference with the CTRL reference simulation is due to the meteorological conditions. In our case during the five days, on average the assimilation of concentrations is better for PM10, PM2.5, sulfate, ammonium, and O_3. At time $\omega = 22$ h for NO_2, 99 h for nitrates, and 44 h for SO_2 the simulation EMILA provides better performances (Table 5). If we consider only the DA of NO_2 at the first level this time falls to 8 h on NO_2 concentrations. On average, the tendency of the Π ratio is in favor of better results with a DA of emissions, but it can take time to reach this ω tipping time. A closer look over some countries in Poland, Switzerland, the Netherlands, and Italy shows that performances of EMILA turn better on O_3 concentrations after respectively 91, 87, 84, and 69 h. For all simulations, the performance turns better for the two primary compounds SO_2 and NO_2 that are the most influenced by their emissions; the switch is particularly fast over the Netherlands (8 h for NO_2 concentrations); this time is the highest over Germany for NO_2 concentrations (after 70 h).

On average, over the forecast period 11–15 March 2014, the map of Γ ratios in Fig. 7 for O_3 concentrations shows better performances for DA of observations relative to emissions over remote areas while close to emission source areas it is the opposite, particularly during this period over shipping areas (the English Channel and Riviera shores). For the PM2.5 concentrations the spatial patterns are similar except over the Channel. If we consider the "more realistic" DAE SONLA, which corresponds to a DA of NO_2 and O_3 over all levels, although performances are better within the first two hours for the SONLA simulation on PM2.5 concentrations, after

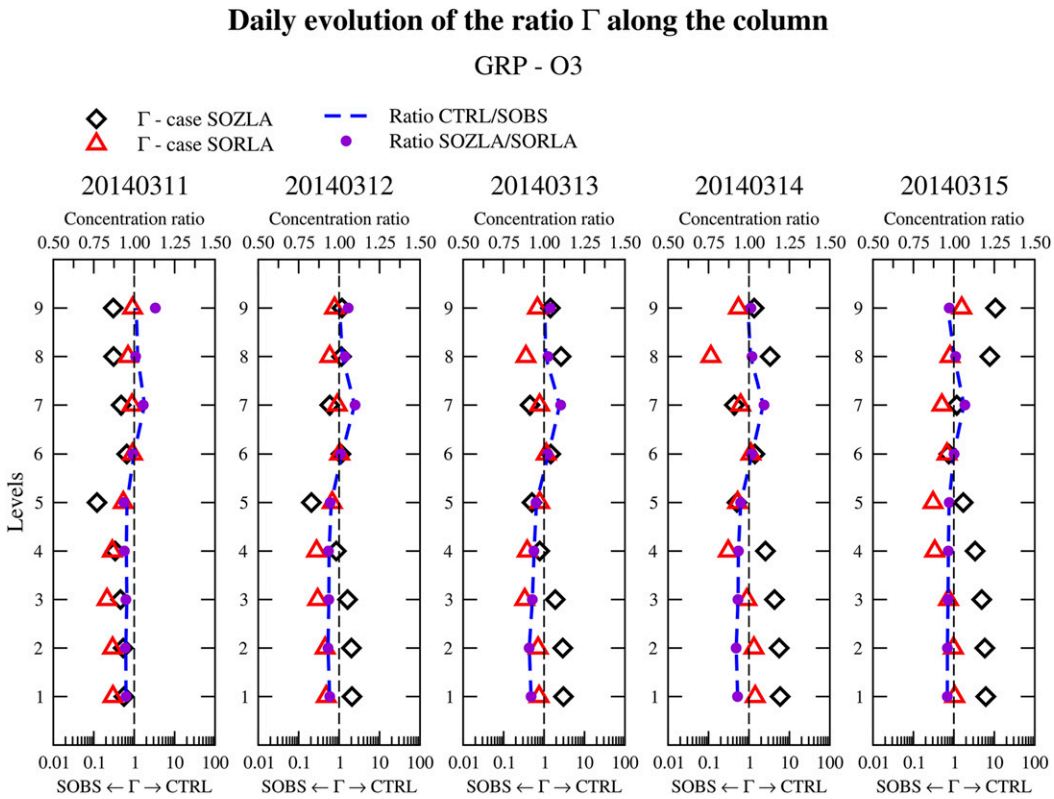

FIG. 6. Daily mean Γ ratio vertical profiles for O_3 concentrations over the group of countries GRP.

about 20 h the DA on emission provides better benefits. Over the Netherlands, it is noteworthy to observe a very fast impact of emissions on ammonium concentrations (certainly due to the update of NO_2 and NH_3 emissions) that occurs in less than 1 h, and 13 h later the update of total ammonia concentrations provides better results. In this region a positive impact of emission assimilation is clearly highlighted on average over the forecast period for the nitrates and therefore the PM2.5 concentrations while over Italy and Poland an opposite behavior is observed. The difference of thermodynamic chemical regimes and meteorological conditions is certainly the reason for these differences.

4. Conclusions

This work proposes an evaluation framework providing an order of magnitude for the potential data assimilation

TABLE 5. Comparisons of data assimilation experiments SALLA (A) vs EMILA (B) for several countries in Europe. Labels A or B alone mean that DAE A or DAE B, respectively, shows the best performances, and A $\underset{\rightarrow}{\omega}$ B means that DAE A exhibits first the best performances until $\omega - 1$ hours, with DAE B displaying a better benefit afterward.

Country	PM10	PM2.5	O_3	NO_2	SO_2	Nitrate	Sulfate	Ammonium
GRP	A	A	A	A $\underset{\rightarrow}{22}$ B	A $\underset{\rightarrow}{44}$ B	A $\underset{\rightarrow}{99}$ B	A	A
FRA	A	A	A	A $\underset{\rightarrow}{22}$ B	A $\underset{\rightarrow}{68}$ B	$\underset{\rightarrow}{A}$	A	A
GBR	A	A	A	A $\underset{\rightarrow}{10}$ B	A $\underset{\rightarrow}{20}$ B	A	A $\underset{\rightarrow}{37}$ B	A $\underset{\rightarrow}{62}$ B
DEU	A	A	A	A $\underset{\rightarrow}{70}$ B	A $\underset{\rightarrow}{50}$ B	A	A $\underset{\rightarrow}{105}$ B	$\underset{\rightarrow}{A}$
POL	A $\underset{\rightarrow}{85}$ B	A $\underset{\rightarrow}{84}$ B	A $\underset{\rightarrow}{91}$ B	A $\underset{\rightarrow}{22}$ B	A $\underset{\rightarrow}{10}$ B	A $\underset{\rightarrow}{57}$ B	A $\underset{\rightarrow}{13}$ B	A $\underset{\rightarrow}{58}$ B
ROU	$\underset{\rightarrow}{A}$	$\underset{\rightarrow}{A}$	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{47}$ B	A $\underset{\rightarrow}{36}$ B	$\underset{\rightarrow}{A}$	$\underset{\rightarrow}{A}$	$\underset{\rightarrow}{A}$
ITA	A $\underset{\rightarrow}{113}$ B	A $\underset{\rightarrow}{109}$ B	A $\underset{\rightarrow}{69}$ B	A $\underset{\rightarrow}{19}$ B	A $\underset{\rightarrow}{115}$ B	A $\underset{\rightarrow}{47}$ B	A	A $\underset{\rightarrow}{57}$ B
NLD	$\underset{\rightarrow}{A}$	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{84}$ B	A $\underset{\rightarrow}{8}$ B	A $\underset{\rightarrow}{17}$ B	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{76}$ B	$\underset{\rightarrow}{B}$
SWE	A	A	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{43}$ B	A $\underset{\rightarrow}{27}$ B	A	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{71}$ B
CHE	A	A	A $\underset{\rightarrow}{87}$ B	A $\underset{\rightarrow}{17}$ B	A $\underset{\rightarrow}{14}$ B	A $\underset{\rightarrow}{79}$ B	A	A $\underset{\rightarrow}{87}$ B
ESP	A	A	$\underset{\rightarrow}{A}$	A $\underset{\rightarrow}{11}$ B	A $\underset{\rightarrow}{17}$ B	$\underset{\rightarrow}{A}$	A	$\underset{\rightarrow}{A}$

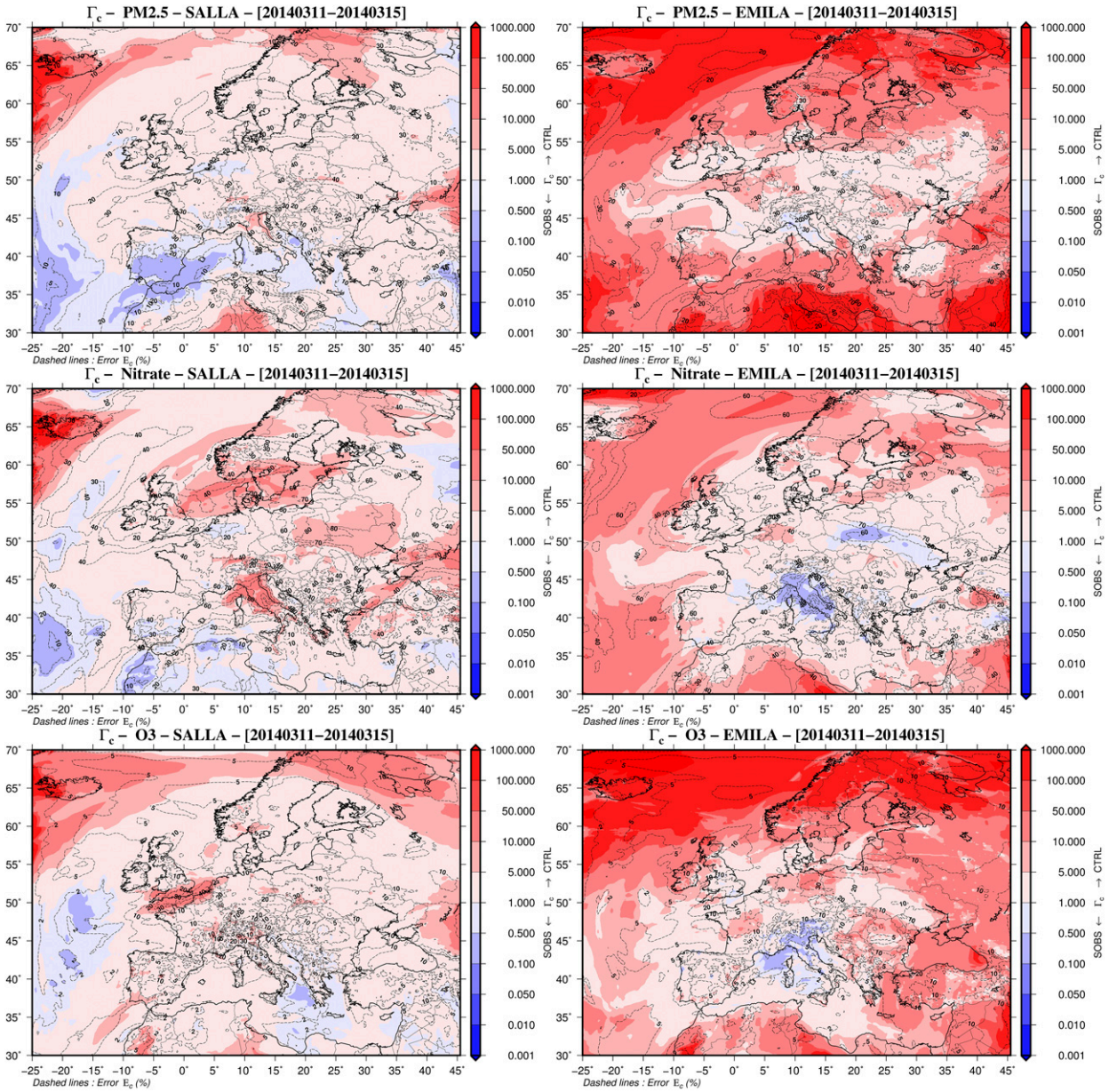

FIG. 7. The Γ_c ratios over the forecast period for the (left) SALLA and (right) EMILA DAEs for (top) PM2.5, (middle) nitrate, and (bottom) O_3. The dashed lines represent E_c, the average relative distance of DAE to SOBS.

benefit in a forecast context of air quality; however, the outcomes are insightful for the general use of DA over more extended applications. For the studied period and considering the expected best favorable DAEs, the maximum benefit is higher for PM than for O_3 concentrations. The tipping time τ is generally low for primary species like NO_2 and SO_2 largely influenced by emissions. A DA at the first level only exhibits low performances with a quick return toward the control simulation (less than 10 h). Indeed, the first level is quickly ''contaminated'' by upper levels that remain at the CTRL concentrations. Applying DA for given compounds can have a positive influence on other species; for instance, assimilating only O_3 and NO_2 has a positive impact on nitrate concentration during 19 h on average over the group of selected countries (GRP). Ozone has an atypical behavior; the expected best case assimilating all species at all levels is not the best on average except for the maximum concentrations. On average, the experiment that consists of propagating the first-level delta between CTRL and SOBS for O_3 to all levels provides the best performance because of the mixing processes

from upper levels that artificially counterbalance the bias, particularly for nocturnal concentrations. However, the maximum concentrations are negatively impacted; the benefit is mainly obtained for nighttime concentrations. Therefore, if one is interested in O_3 mean concentrations, a propagation of the delta over the whole column could be sufficient as a first DA technique. Despite the update of initial state of inorganic compounds, the assimilation of SIA injected at the first level only is quickly lost with a tipping time of 2 h on average for nitrate concentrations; however, it seems better than injecting primary species as NO_2 at the first level. For primary species, assimilating their emissions has a positive impact on their concentrations relative to the injection of the concentrations at the initial state.

This study only focused on an episode but over a large area in Europe. The results cannot be fully generalized, and additional work could follow this framework to enrich the analysis. However, it confirms—in a forecast perspective—the benefit of using DA in 3D (from lidar or satellites). Of course, the SALLA experiment is not realistic; however, for O_3 a simple propagation of the ground delta along the vertical is enough to increase the model performances. For primary sources, injecting better emissions improves the performances but for secondary pollutants like O_3 or secondary PM, an update of initial concentrations shows better performances than acting on precursors emissions. The DA of ammonia concentrations seems to have a better effect on average on PM concentrations when compared with a DA of NO_2 concentrations; this findings confirms the relevance of current works of ammonia DA from satellite data.

The impact of the hour of DA could be another issue to investigate; it could have an impact on the expected benefits through the chemical and meteorological processes. As another perspective of work, this framework could be applied to an analysis of the impact of boundary conditions through the advection and chemical processes involved in models used over limited areas.

Acknowledgments. This work is partly funded by the French Ministry of Ecological and Solidarity Transition (MTES).

REFERENCES

Bessagnet, B., A. Hodzic, O. Blanchard, M. Lattuati, O. L. Bihan, H. Marfaing, and L. Rouïl, 2005: Origin of particulate matter pollution episodes in wintertime over the Paris basin. *Atmos. Environ.*, **39**, 6159–6174, https://doi.org/10.1016/j.atmosenv.2005.06.053.

——, and Coauthors, 2008: Regional modeling of carbonaceous aerosols over Europe—Focus on secondary organic aerosols. *J. Atmos. Chem.*, **61**, 175–202, https://doi.org/10.1007/s10874-009-9129-2.

——, and Coauthors, 2016: Presentation of the EURODELTA III intercomparison exercise—Evaluation of the chemistry transport models' performance on criteria pollutants and joint analysis with meteorology. *Atmos. Chem. Phys.*, **16**, 12 667–12 701, https://doi.org/10.5194/acp-16-12667-2016.

Faustini, A., R. Rapp, and F. Forastiere, 2014: Nitrogen dioxide and mortality: Review and meta-analysis of long-term studies. *Eur. Respir. J.*, **44**, 744–753, https://doi.org/10.1183/09031936.00114713.

Lim, S. S., and Coauthors, 2012: A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the global burden of disease study 2010. *Lancet*, **380**, 2224–2260, https://doi.org/10.1016/S0140-6736(12)61766-8.

Mailler, S., and Coauthors, 2017: CHIMERE-2017: From urban to hemispheric chemistry-transport modeling. *Geosci. Model Dev.*, **10**, 2397–2423, https://doi.org/10.5194/gmd-10-2397-2017.

Menut, L., and B. Bessagnet, 2019: What can we expect from data assimilation for air quality forecast? Part I: Quantification with academic test cases. *J. Atmos. Oceanic Technol.*, **36**, 269–279, https://doi.org/10.1175/JTECH-D-18-0002.1.

——, and Coauthors, 2013: CHIMERE 2013: A model for regional atmospheric composition modelling. *Geosci. Model Dev.*, **6**, 981–1028, https://doi.org/10.5194/gmd-6-981-2013.

——, G. Rea, S. Mailler, D. Khvorostyanov, and S. Turquety, 2015: Aerosol forecast over the Mediterranean area during July 2013 (ADRIMED/CHARMEX). *Atmos. Chem. Phys.*, **15**, 7897–7911, https://doi.org/10.5194/acp-15-7897-2015.

Pagowski, M., G. A. Grell, S. A. McKeen, S. E. Peckham, and D. Devenyi, 2010: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: Some results using the Weather Research And Forecasting–Chemistry model and grid-point statistical interpolation. *Quart. J. Roy. Meteor. Soc.*, **136**, 2013–2024, https://doi.org/10.1002/qj.700.

Park, S.-Y., D.-H. Kim, S.-H. Lee, and H. W. Lee, 2016: Variational data assimilation for the optimized ozone initial state and the short-time forecasting. *Atmos. Chem. Phys.*, **16**, 3631–3649, https://doi.org/10.5194/acp-16-3631-2016.

Petit, J.-E., and Coauthors, 2017: Characterising an intense PM pollution episode in March 2015 in France from multi-site approach and near real time data: Climatology, variabilities, geographical origins and model evaluation. *Atmos. Environ.*, **155**, 68–84, https://doi.org/10.1016/j.atmosenv.2017.02.012.

Rea, G., S. Turquety, L. Menut, R. Briant, S. Mailler, and G. Siour, 2015: Source contributions to 2012 summertime aerosols in the Euro-Mediterranean region. *Atmos. Chem. Phys.*, **15**, 8013–8036, https://doi.org/10.5194/acp-15-8013-2015.

Rouïl, L., and Coauthors, 2009: PREV'AIR: An operational forecasting and mapping system for air quality in Europe. *Bull. Amer. Meteor. Soc.*, **90**, 73–83, https://doi.org/10.1175/2008BAMS2390.1.

Sandu, A., and T. Chai, 2011: Chemical data assimilation—An overview. *Atmosphere*, **2**, 426–463, https://doi.org/10.3390/atmos2030426.

Terrenoire, E., and Coauthors, 2015: High-resolution air quality simulation over Europe with the chemistry transport model CHIMERE. *Geosci. Model Dev.*, **8**, 21–42, https://doi.org/10.5194/gmd-8-21-2015.

van Leer, B., 1979: Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method.

J. Comput. Phys., **32**, 101–136, https://doi.org/10.1016/0021-9991(79)90145-1.

Vieno, M., and Coauthors, 2016: The UK particulate matter air pollution episode of March–April 2014: More than Saharan dust. *Environ. Res. Lett.*, **11**, 044004, https://doi.org/10.1088/1748-9326/11/4/044004.

Vivanco, M. G., I. Palomino, R. Vautard, B. Bessagnet, F. Martin, L. Menut, and S. Jimenez, 2009: Multi-year assessment of photochemical air quality simulation over Spain. *Environ. Modell. Software*, **24**, 63–73, https://doi.org/10.1016/j.envsoft.2008.05.004.

——, and Coauthors, 2017: Joint analysis of deposition fluxes and atmospheric concentrations of inorganic nitrogen and sulphur compounds predicted by six chemistry transport models in the frame of the EURODELTAIII project. *Atmos. Environ.*, **151**, 152–175, https://doi.org/10.1016/j.atmosenv.2016.11.042.