Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions

An unusual way to validate regional chemistry-transport models

Laurent MENUT[1], Sylvain MAILLER[1], Bertrand BESSAGNET[2], Guillaume SIOUR[3], Augustin COLETTE[2], Florian COUVIDAT[2], and Frédérik MELEUX[2]

[1]Laboratoire de Météorologie Dynamique, Ecole Polytechnique, IPSL Research University, Ecole Normale Supérieure, Université Paris-Saclay, Sorbonne Universités, UPMC Univ Paris 06, CNRS, Route de Saclay, 91128 Palaiseau, France
[2]INERIS, National Institute for Industrial Environment and Risks, Parc Technologique ALATA, F-60550 Verneuil-en-10 Halatte, France
[3]LISA, Créteil, France

Correspondence to: Laurent Menut, menut@lmd.polytechnique.fr

Abstract. A simple and exhaustive model evaluation technique for regional chemistry-transport is discussed. It is based on the concept that we can learn more on models performances by comparing the results to in situ measurements available for other time periods than the period originally targeted in the simulation. First, the usual scores (spatial and temporal correlation) are computed for a reference period, using the actual temporal synchronization and spatial location of measurements . Second, the same scores are calculated for several other years by conserving only the actual spatial locations and days of the year. The difference between the two score provides complementary insights to the following questions: (i) is the model performing well only because the situation is persistent? (ii) is the model representative enough of the measurements for all variables? (iii) if the pollutants concentrations are not well modelled, is it due to meteorology or chemistry? In order to synthetize the large amount of results, a new score is proposed: the "multi-year variability", designed to compare the several indicators between all the years of validation and to quantify if the studied period was well modelled and, if yes, for the good reasons.

1 Introduction

Chemistry transport models (CTM) aim at simulating the air pollutants concentrations in the lowest layers of the atmosphere where humans and the environment can be affected when air quality is poor. Air pollution results from the presence of chemical components emitted into the air due to anthropogenic activities and by natural sources (biogenic emissions from vegetation, soil erosion, sea salts, volcanic activity, and wild-land fires). CTMs are used to represent the dynamic and chemical processes that drive spatial and temporal features of the atmospheric composition.

To estimate the quality of CTMs, output results are usually compared with available observations. In areas where the monitoring network are dense enough, such as in Europe, comparisons can be made with observations from surface stations that provide hourly concentrations of O_3, NO_2 for gas and $PM_{2.5}$ and PM_{10} for aerosols. In order to quantify transport of aerosols in dense plumes aloft, observations from lidar or the AERONET network (to have the optical depth) are increasingly used with regional models.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

But there can me multiple reasons for a model simulation to agree or disagree with observations. That is because the result of a simulation is the integrated budget of several processes, and it is challenging to easily identify why a model would exhibit an inappropriate behavior.

A fundamental difference between models and observations is the spatial representativeness (Valari and Menut, 2008; Solazzo and Galmarini, 2015). To isolate problems intrinsic to the models, several methodologies have been developed to extract the relevant information in the simulations, particularly to identify what could be the processes most responsible for model discrepancies. These methods are effective but often with huge computation time. Among these approaches, ensemble modeling is used in analysis of case studies and forecasting, Kioutsioukis and Galmarini (2014); Marécal et al. (2015); Lemaire et al. (2016). By performing several perturbed simulations, one can identify if there is a general tendency on the error. But if the case study consists of a complex real situation, the analysis can be challenging. Adjoint modeling allows tracking the behavior of chemical species with respect to model input parameters. But it requires tedious model developments and the result is generally valid for an infinitesimal perturbation since the problem to solve was linearized, (Menut, 2003; Pison et al., 2007). In practice, the validity of this approach is limited to chemical species with a long lifetime as presented in Kopacz et al. (2010); Mao et al. (2015).

In chemistry transport modeling, emissions are well known to constitute one of the most uncertain forcings. There are therefore studies devoted to scenario simulations in order to quantify the relative weight of each pollutant emitted in the final calculated concentration budget, (Rea et al., 2015).

More recently, Solazzo and Galmarini (2016) proposed to decompose statistical scores to better understand the errors in surface ozone modeling. Finally, other studies also use observations to adjust the result by implementing methods to unbias simulation without changing the model, as in Porter et al. (2015) for ozone on the United States.

In the present study, we try to provide a simple method to improve the validation of a simulation and to identify the processes responsible for the differences between the model and the available observations. For this, we compute several correlation scores, to identify the model accuracy compared to observations. Afterwards, we apply these scores to a model simulation and several different observations datasets. The originality of the approach presented here is that we do not compare the simulation of a case study only to the corresponding observational dataset (in time and space) but we use all available data of the other years. The new dataset of scores will highlight the differences between specific and systematic errors.

Therefore, we want to elaborate scores that provides answer to the subsequent questions: *Are the performances of the model satisfactory because the model is accurate or just because the model is able to reproduce a situation which is persistent from year to year? does the model have a good spatial representativeness compared to available observations for a given variable?*, and *Are the biases introduced by meteorological or emissions variability or by the processes parameterized in the chemistry-transport model itself?*

The issue to solve and the tools developed are presented in section 2. The new methodology with the presentation of the score developed for this study are presented in section 3. The results and discussions to point out the drivers of model errors are presented in section 4.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions
Open Access
EGU

2 The problem to solve

The problem to solve is presented in a general way by presenting the principle of chemistry-transport modeling. Then, the studied case and the models used are presented.

2.1 Regional chemistry-transport modeling

Figure 1 presents the several forcing and processes involved in a typical chemistry-transport model (CTM). The objective of this simple figure is to remind the dependencies between each "geophysical compartment" involved in such modeling tools.

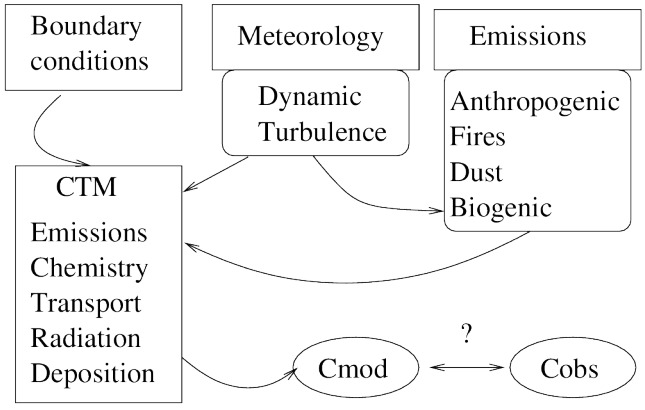

Figure 1. *The several processes taken into account in a regional chemistry-transport model.*

These processes may be divided in four categories: (i) boundary conditions, (ii) dynamics, (iii) emissions and (iv) chemistry and transport.

The boundary conditions prescribe the concentration in chemical species which may enter the modeled area during the simulation. Usually for large domains, they are issued from global models as monthly climatologies. They correspond to averaged values suitable to characterize the background concentrations of long-lived species such as ozone, carbon monoxide, mineral dust.

The meteorological variables influence transport and mixing processes, with a direct effect on gas and aerosol plumes locations and their vertical distribution. Cloudiness and temperature impact the photolysis efficiency, the boundary layer height impact the surface mixing of pollutants, rainfall impact the wet deposition. Moreover, meteorology impact emissions: wind variability is the prevalent driver for dust emissions, and it has a strong impact on wildfires emissions. Both temperature and solar irradiance influence the magnitude of biogenic emissions. The spatial variability of landuse has also a strong impact on all these natural emissions.

Anthropogenic emissions are prescribed from databases and the influence of meteorology is limited in the model. On the other hand, biogenic, fires and mineral dust emissions also depend both on landuse and meteorology. These emissions are difficult to measure; this is not possible to quantify their realism.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

The chemistry-transport model is a numerical integration tool of all the forcings and processes. The chemistry mechanism prescribes the amount of the chemical species (production and loss) when the deposition is the only sink of species. With the model, the spatial (horizontal and vertical) and temporal resolutions are also defined, directly impacting the simulation representativeness and thus the realism of the modeled air pollutant concentrations when they are compared to the available observations.

2.2 The studied case and the models

We focus on a case study for the summer of 2013 (1st May to 31 August) in the Euro-Mediterranean region, this period is called "reference period" in this paper. This case has already been modeled (using WRF and CHIMERE) and the results were discussed in Menut et al. (2015). The same simulation is used in this study, all parameters are identical. The observational data come from different sources depending on the variable and they are presented in Table 1. Originally provided hourly or three-hourly, they are used as daily averaged in the present study.

Variable	Network	Spatial coverage	Vertical coverage	Temporal frequency	Unit
O_3, NO_2	EMEP	Europe	Surface	Hourly	ppb
$PM_{2.5}$, PM_{10}	EMEP	Europe	Surface	Hourly	μg m^{-3}
AOD, Angström	AERONET	Global	Column	Hourly	ad.
T_{2m}	BADC	Global	Surface	Tri-hourly	^{o}C
U_{10m}	BADC	Global	Surface	Tri-hourly	m s^{-1}
Precipitation	BADC	Global	Surface	Tri-hourly	mm day^{-1}

Table 1. *List of measurements data used for the statistical comparison with the model results. All data used are issued from surface stations, representative of their own environment.*

3 The proposed methodology

The proposed methodology follows three steps: (i) compute the correlation scores (spatial and temporal) between the measurements and the model and during the whole reference period, (ii) recalculate these scores between the modeled reference period and the observed data for the similar period in 2008, 2009, 2010, 2011 and 2012, (iii) build and use a synthetic score to quantify if the model had high scores for good reasons or not. This is summarized in Figure 2.

Of course it seems apparently awkward to evaluate day by day a model with observational data from another year. For a given station at a given day of the reference year air concentrations will be affected by a different local meteorology, emissions and also long range transport of chemical species. But we can consider that to take the same day for another is strictly the same

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

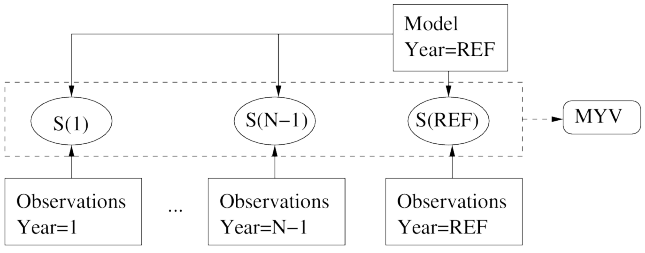

Figure 2. *Principle of the multi-year variability score's calculation, using one modelled year and several observations years.*

that to choose randomly a day in the same season. This trivial method can emphasize how a model is affected by large scale patterns and long term temporal cycles. The correlation is the more appropriate statistical metric for such analysis.

3.1 Calculation of usual correlation scores

To compute the correlation coefficient, it is important that, for all years of validation, the same list of stations with valid measurements is used. Each correlation provides specific information on the quality of the simulation:

The spatial correlation, noted R_s, is calculated from the temporal mean averaged values of observations and model for each location where observations are available. A good correlation shows that the model correctly locates known sources and plumes during long range transport. For processes leading to large plumes (dust, fires, volcanoes), this indicator indicates that the model is using realistic emissions and is able to reproduce a correct transport. For all the studied parameters, it is also an indicator that the resolution of the model is adapted to the variable considered.

The temporal correlation, noted R_t, is estimated station by station. This indicator is directly related to the variability from day to day, for each station. The longer the atmospheric lifetime of the species, the lower the relevance of temporal correlation .

For the correlations, obviously better scores are expected for the reference year compared to the other, particularly for the temporal correlation. This would confirm that during the transport of pollutants, the model is able to correctly model the day to day variability.

3.2 The multi-year variability 'MYV' score

We aim to develop a simple indicator that would increase with correlation but would be moderated if the differences with other years are low. We thus first estimate the differences, D, between all years as:

$$D = \frac{1}{N-1}\left(\sum_{i=1}^{N-1} |s_i - s_N|\right) \quad (1)$$

with s_N the score for the actual year being modelled and s_i the score computed using observations corresponding to other meteorological years (from 1 to $N-1$ if there is $N-1$ other available years for the observations).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions
Open Access
EGU

We can thus estimate a "Multi Year Variability", noted MYV as:

$$MYV = s_N \times \left(1 - exp(-D_s)^{\delta}\right) \quad (2)$$

The value for δ is arbitrary but it should be larger than unity. This tuning parameter enables to adapt the relative weight we want to attribute to the absolute value of the scores for the selected year and the differences between all years. In general, we want that a good score for the studied year have a largest weight that the differences: in the present case , we choose to select δ=4.

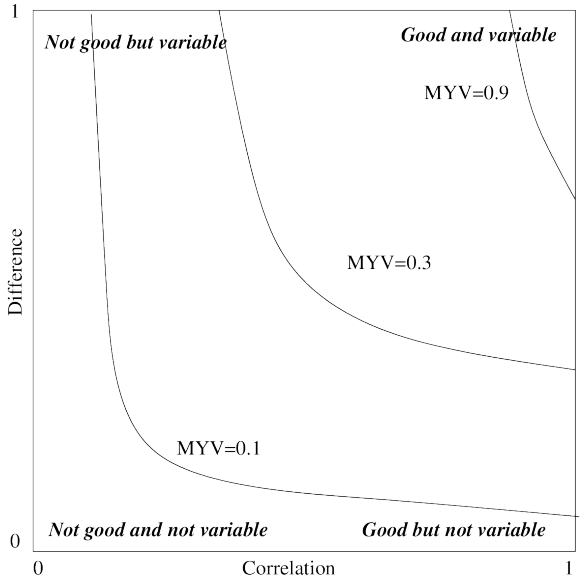

Figure 3. *Scheme of the MY score as a function of the studied year correlation and the multi-years differences.*

The behavior of MYV is plotted on Figure 3 for academic values of the scores and the differences. Ideally we would hope that the model performs well for the absolute scores but also variable i.e. able to reproduce very different situations. When MYV tends to 1 this means that the correlation value is close to 1 and the differences of the modeled studied year compared to the other years are also close to 1. In reality, this ideal situation is rarely obtained since we are modeling a very complex atmospheric system, based on processes with different variabilities and uncertainties. On the other hand, if the correlation is close to zero, the model is definitely poor. And if the difference is also close to zero, we can conclude that the low performance of the model is independent from the selected year: in that case, MYV is also close to 0.

It is also possible to have a good score but a low variability: in that case, MYV is also close to zero. This is because the MYV score has to be viewed as complementary to the score (and not replacing the score). Thus, if we already know that the score is good, we can rely on MYV to ensure that the score was also good for other years.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions
Open Access
EGU

From a subjective point of view, and from the Figure 3, we can consider the model is accurate and has an acceptable variability for MYV > 0.3. Of course, this directly depends of the δ parameter value.

3.3 Detailed examples of MYV calculation

To better understand the relevance of the score MYV, two examples are detailed in this section. The scores are done for 2m temperature, T_{2m}, and for the surface concentration of nitrogen dioxide, NO_2. Results are presented in Table 2.

These two variables are presented here because they represent very different variables in a CTM simulation:

- T_{2m} is a meteorological variable, constraining processes both for meteorology and chemistry. Its diurnal cycle is well marked as its latitudinal variability (for large model domains), ensuring a good spatial correlation. In general, it is the less uncertain of modelled meteorological variables.
- NO_2 is both a primary and secondary species. Mostly emitted in urbanized areas, the diurnal cycle of this species is well constrained. Depending on meteorological conditions, its lifetime may vary a lot, from hours to days. Modelling this species with CTMs is challenging several uncertainties are acting at the same time, including the spatial representativity of the model cell.

T_{2m}			NO_2		
Year	R_s	R_t	Year	R_s	R_t
2008	0.58	0.36	2008	0.44	0.00
2009	0.57	0.38	2009	0.42	-0.04
2010	0.60	0.30	2010	0.66	-0.04
2011	0.62	0.26	2011	0.79	-0.03
2012	0.61	0.40	2012	0.76	0.04
2013	0.61	0.94	2013	0.88	0.22
D	0.02	0.60	D	0.27	0.23
MYV	0.04	0.85	MYV	0.58	0.13

Table 2. *Scores for T_{2m} and NO_2. The reference year is 2013.*

3.3.1 Analysis of T_{2m} scores

The spatial correlation is good for all years, ranging from 0.57 (2009) to 0.62 (2011). For the studied year (2013), the score is 0.61, thus lower than for 2011. Even if the correlation for the selected year is good, it is not significantly better than for the other year, with D=0.02, and this yields to MYV(R_s)=0.04. This means that the model reproduces fairly well a spatial pattern that is observed every year. Indeed, the simulation domain is large and the temperature has a latitudinal variability larger than between each measurements stations. This temporal correlation ranges from 0.26 to 0.94. And the best score is for 2013 leading

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

to a good score of MYV(R_t)=0.85. The model is thus performing well in capturing the day to day variability for T2m and for the good reasons.

3.3.2 Analysis of NO_2 scores

The second example is related to the surface concentrations of NO_2. This species is a secondary species quickly produced by oxidation of NO and the scores show at the same time if the sources are properly placed, if the photochemistry and transport processes have been well simulated. In general, at low model resolution, the scores for this species are less good than for ozone, its spatial extent of its representativeness being more limited (emissions from traffic in urban environments etc.), even if the measurements stations considered in this study are all background sites.

We can see that the spatial correlation gives a score of R_s=0.88 for 2013. Being the best comparison, we obtain MYV(R_s)=0.58. This shows the importance of NO_x source location that is the main driver of spatial performances. The temporal correlation is low for 2013, R_t=0.22, but is close to 0 for other years. In the end, we have a low score with MYV(R_t)=0.13 even if the simulated year is better. These two scores show that the model certainly captures the right location of emission sources (low variability of R_s). For temporal variability, the model is not able to reproduce the day to day variability, but it remains significantly better for the reference year compare to the others.

4 Results and discussion

The scores are calculated for all variables described in Table 1 and for the years 2008 to 2013, it is reminded that only the May to August 2013 period was modeled. Results are presented as time series in Figure 4. Using all scores values, a MYV is estimated for each score type and each variable. Results are presented in Table 3. These results are discussed in the following sections.

4.1 Meteorological variables

Scores for T_{2m} were discussed in the previous section. The calculation of u_{10m} also gives satisfactory results but for the temporal correlation is only R_t=0.60 and MYV=0.54. The spatial correlation is not correct and very variable for one year to another, leading to MYV=0.03. As for T_{2m}, we also have an effect of the model resolution and the representativity of the variable. Scores for the precipitation are correct, with a very good spatial correlation leading to MYV(R_s)=0.54. For the day to day variability, the score is less good with MYV(R_t)=0.21 but significantly higher for 2013. These scores showed that the meteorological forcing is well retrieved, and better for the year being considered compared to other years.

4.2 Optical properties

The optical properties are directly linked to the atmospheric composition of aerosol and may be quantified using the Aerosol Optical Depth (AOD) and the Angström exponent (ANG).

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

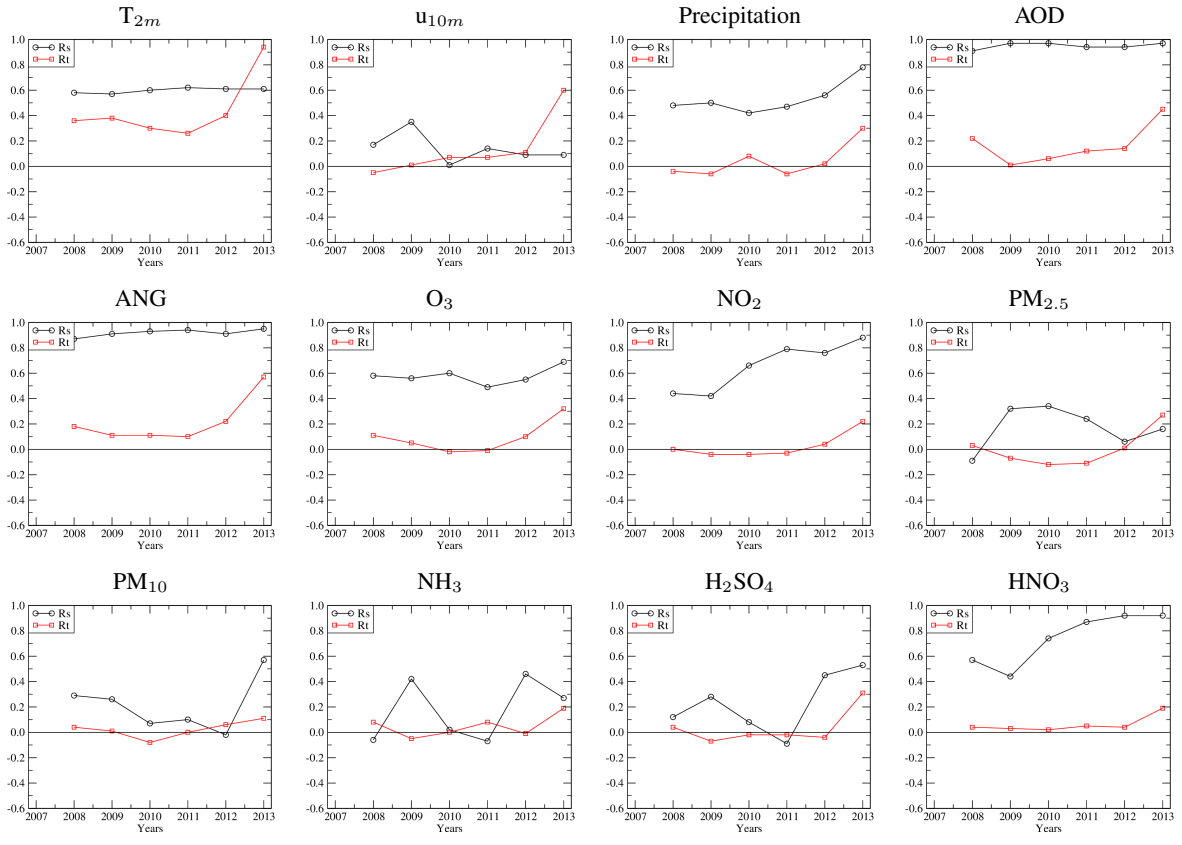

Figure 4. *Multi years scores for the 2m temperature, the 10m wind speed, AOD and the Angström coefficient. The reference year is 2013.*

For the AOD, the spatial correlation is very good for 2013, R_s=0.97 but it is as good or better for other years. This means that we model a rather recurring phenomenon: every year the same stations are exposed to aerosol plumes: MYV(R_s)=0.09. The temporal correlation is lower with R_t=0.45 but much better than for other years: MYV(R_t)=0.33. This means that the model reproduced partly the observed temporal variability but the events are changing from one year to another and the model captures well these changes. The AOD are sensitive to desert dust in summer in that region, this means that large scale systems are driving the aerosol plumes, they are spatially recurrent and temporally better estimated for the year being considered than for other years.

For ANG, the spatial correlation is very good, R_s=0.91 but also persistent leading to a low score of MYV(R_s) = 0.14. The temporal correlation is much better for 2013 than other years with MYV(R_t) = 0.49. This is probably due to a size distribution that is not necessarily well simulated from one day to another (showed by AOD) but correct relative contributions of fine and coarse aerosol atmospheric load. This feature highlights the high sensitivity of the AOD calculation depending on the modeled aerosol size distribution, although the overall mass emitted and transported could be realistic.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions
Open Access
EGU

Variable	R_s			R_t		
	Value	D	MYV	Value	D	MYV
T_{2m}	0.61	0.02	0.04	0.94	0.60	**0.85**
u_{10m}	0.09	0.09	0.03	0.60	0.56	**0.54**
precip	0.78	0.29	**0.54**	0.30	0.31	0.21
AOD	0.97	0.02	0.09	0.45	0.34	**0.33**
ANG	0.91	0.04	0.14	0.59	0.44	**0.49**
O_3	0.69	0.13	0.29	0.32	0.27	0.21
NO_2	0.88	0.27	**0.58**	0.22	0.23	0.13
$PM_{2.5}$	0.16	0.15	0.07	0.27	0.32	0.20
PM_{10}	0.57	0.43	**0.47**	0.11	0.10	0.04
NH_3	0.20	0.13	0.08	0.21	0.20	0.12
H_2SO_4	0.51	0.21	0.29	0.31	0.34	0.23
HNO_3	0.15	0.51	0.13	0.09	0.08	0.03

Table 3. *The MYV scores for all variables: the meteorology with T_{2m}, u_{10m} and precipitation rate, the vertically integrated column of aerosols with the Aerosol Optical Depth (AOD) and the Angström exponent (ANG), the surface concentrations of all aerosols in term of size distribution with $PM_{2.5}$ and PM_{10} and for the inorganic species with $D_p < 10$ µm. Values of MYV up to 0.3 are bolded. Units of the variables are detailed in Table 1.*

Globally, AOD and ANG reflect the model's ability to retrieve the long range transport of long-lived aerosols. This mixes a lot of processes (emissions, transport, and deposition). With these scores, we can conclude that the model is able to retrieve these yearly recurrent plumes but that the mass distributed into the model size distribution needs improvements.

4.3 Surface concentrations

The scores for the surface concentrations of gaseous species (O_3, NO_2) and aerosol ($PM_{2.5}$ and PM_{10}) are very sensitive to the species. The spatial correlation is good for O_3, NO_2 and PM_{10}, with R_s=0.69, 0.88 and 0.57 respectively. For $PM_{2.5}$ this correlation is low with R_s=0.16. The PM_{10} tends to show that the largest particles are well modelled over the whole domain, and this was also the conclusion seeing at the scores for AOD and ANG. The low score of $PM_{2.5}$ show that, in the aerosol distribution, the fine mode is less well modeled than the coarse mode. This is confirmed by the scores of the aerosol inorganic species, NH_3, H_2SO_4 and HNO_3. Except for H_2SO_4 (with R_s=0.51), the spatial correlations are 0.15 for HNO_3 and 0.20 for NH_3. Thus, the fine part of the aerosol is not well modeled mainly due to a deficiency in the modeling of nitrates species.

The temporal correlations have a completely different behavior that the spatial correlations. The values are generally low, from R_t=0.09 for HNO_3 to R_t=0.32 for O_3. Surprisingly, the PM_{10} have a good spatial correlation but a poor temporal correlation. This is due to the long lifetime in the atmosphere of non-reactive species such as mineral dust: large plumes are correctly modeled over regions but the day to day variability needs improvements. Another point is the good spatial correlation for NO_2

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Geoscientific Model Development Discussions
Open Access
EGU

(and for the good reasons with MYV=0.58) but its low temporal correlation with R_t=0.22 and a low MYV=0.13. In this case, this means we have a correctly localized anthropogenic emissions inventory (main source of NO_2) but difficulties to model the day to day chemistry.

In conclusion for the surface concentrations, we can conclude that O_3, NO_2 and PM_{10} concentrations are spatially well modelled and this is not due to a persistent behavior, the MYV scores having high values. For the aerosol, the problem is more related to the fine mode, where $PM_{2.5}$ concentrations are not well located. This modeling problem is also obvious with the low correlations and MYV values for the inorganic species. For the temporal correlations, the scores are always lower that for the spatial correlation but also always higher for the year being considered than for the other years.

4.4 All results on a single plot

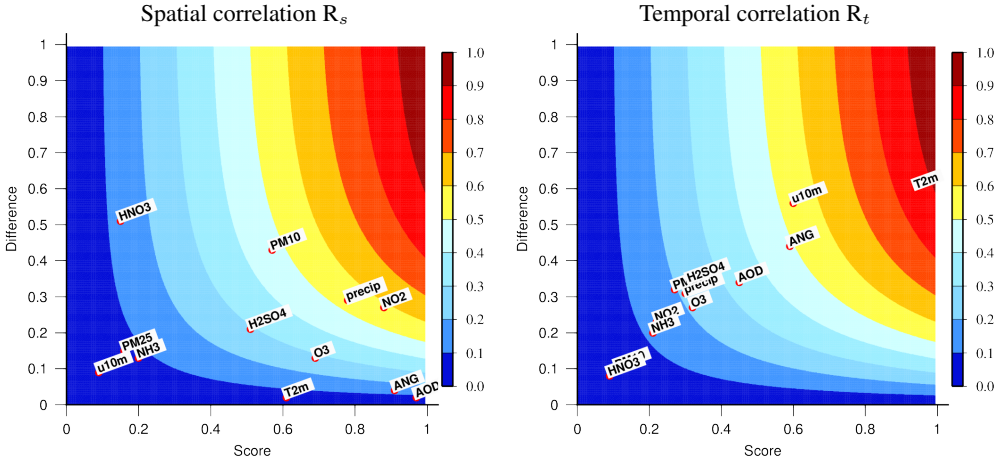

Figure 5. *Results of the MYV scores for the spatial and temporal correlations. For each model variable its value is represented using the correlation on the x-axis and the difference between the studied year and the others on the y-axis. The colors represent the MYV values.*

Complementary to the Table 3, Figure 5 reports the results on a simple plot. The x-axis represents the correlation (spatial or temporal), the y-axis represents the differences between all years, D. For each studied variables, their values are reported on the Figure where the colors represent the value of the score MYV. The interpretation of these results follows the quality criteria presented in the academic scheme in Figure 3.

This presentation shows an important spread for the spatial correlation results. If the relative differences D range from 0 to 0.6, the correlations range from 0.09 (for the 10m wind speed) to 0.97 (for AOD). The common point is that there is no variable with differences up to 0.5. This means that, spatially, the studied problem shows systematic patterns from year to year. The low values of correlations thus showed that some variables are systematically on specific parts of the domain and not over the whole region.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

The representation of all temporal correlations shows a specific linear pattern . The largest the correlation values, the most important the differences. This temporal correlation represents the day to day variability for each location. This means that the studied problem is based on high day to day variability without similar consecutive days (in this case, one would have high correlations but low differences). This illustrates the fact that the studied problem is primarily a problem of sporadic peaks and that the model is able to correctly find this variability from one day to another.

5 Conclusions

At first glance, using a different year than the simulated one for the day to day evaluation seems awkward. However, we can learn more about the performances of chemistry transport models than when using a single statistical indicator. Of course, this approach will never replace a strict evaluation of a pollution case analysis using time series, vertical profiles and usual error statistics. But it can afford to have a very fast and integrated vision of the strengths and weaknesses of a model with very little calculation. This may also allow, for example, to compare different models during inter-comparison exercises.

To answer the questions presented in the introduction, and for this particular model and simulated period: the model always better simulates the studied year than any other meteorological year and it is able to reproduce the day to day variability for high concentrations of pollutants.

The spatial correlation is good for 2m temperature and precipitation rate, but not for wind speed: this highlights the fact that the modeled domain is large and the resolution not optimized for small scale processes. The spatial correlation is also very good for the long-range transport of particles as demonstrated with R_s=0.96 and 0.90 for AOD and ANG. But, since this feature is recurring every year, this leads to low MYV scores. This means that for a large domain, the main spatial patterns of particle concentrations are recurrent and well modeled. The chemical species that are best modelled are either species with a long atmospheric lifetime (PM_{10}) or species well spatially constrained on the domain (such as NO_2 mainly due to anthropogenic emissions). For aerosol, the results depend on the size distribution: the largest particles are better models than the finest ones.

The conclusions are different for the temporal correlation. The scores are calculated using daily observations and modeled outputs. Thus, these scores reflect the ability of the model to retrieve the day to day variability. As for the spatial correlation, scores are good for the meteorological variables. For the aerosol, and mainly for the long-lived species (such as mineral dust), the temporal correlation is also correct as the MYV scores: MYV=0.33 and 0.49 for AOD and ANG respectively. But for the short-live species the temporal correlation and the MYV scores are low. This means that improvements have to be done in priority for the day to day variability compared to the locations of emissions. This may probably be dued to the atmospheric transport, the spatial variability of 10m wind speed being poorly simulated. But overall, the temporal correlation is better for the studied year than for the others, showing that the problem is highly variable from year to year, but the model is significantly able to catch the evolution of atmospheric composition.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

6 Code and/or data availability

This study presenting a methodology using existing data and models, all required informations are already included in this article.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

References

Kioutsioukis, I. and Galmarini, S.: *De praeceptis ferendis*: good practice in multi-model ensembles, Atmospheric Chemistry and Physics, 14, 11 791–11 815, doi:10.5194/acp-14-11791-2014, http://www.atmos-chem-phys.net/14/11791/2014/, 2014.

Kopacz, M., Jacob, D. J., Fisher, J. A., Logan, J. A., Zhang, L., Megretskaia, I. A., Yantosca, R. M., Singh, K., Henze, D. K., Burrows, J. P., Buchwitz, M., Khlystova, I., McMillan, W. W., Gille, J. C., Edwards, D. P., Eldering, A., Thouret, V., and Nedelec, P.: Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmospheric Chemistry and Physics, 10, 855–876, doi:10.5194/acp-10-855-2010, http://www.atmos-chem-phys.net/10/855/2010/, 2010.

Lemaire, V. E. P., Colette, A., and Menut, L.: Using statistical models to explore ensemble uncertainty in climate impact studies: the example of air pollution in Europe, Atmospheric Chemistry and Physics, 16, 2559–2574, doi:10.5194/acp-16-2559-2016, http://www.atmos-chem-phys.net/16/2559/2016/, 2016.

Mao, Y. H., Li, Q. B., Henze, D. K., Jiang, Z., Jones, D. B. A., Kopacz, M., He, C., Qi, L., Gao, M., Hao, W.-M., and Liou, K.-N.: Estimates of black carbon emissions in the western United States using the GEOS-Chem adjoint model, Atmospheric Chemistry and Physics, 15, 7685–7702, doi:10.5194/acp-15-7685-2015, http://www.atmos-chem-phys.net/15/7685/2015/, 2015.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., Denier van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, A., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geoscientific Model Development, 8, 2777–2813, doi:10.5194/gmd-8-2777-2015, http://www.geosci-model-dev.net/8/2777/2015/, 2015.

Menut, L.: Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2, J. Geophys. Res., 108, 8562, doi:10.1029/2002JD002549, 2003.

Menut, L., Mailler, S., Siour, G., Bessagnet, B., Turquety, S., Rea, G., Briant, R., Mallet, M., Sciare, J., Formenti, P., and Meleux, F.: Ozone and aerosol tropospheric concentrations variability analyzed using the ADRIMED measurements and the WRF and CHIMERE models, Atmospheric Chemistry and Physics, 15, 6159–6182, doi:10.5194/acp-15-6159-2015, http://www.atmos-chem-phys.net/15/6159/2015/, 2015.

Pison, I., L.Menut, and G.Bergametti: Inverse modeling of surface NOx anthropogenic emissions fluxes in the Paris area during the ESQUIF campaign, Journal of Geophysical Research, Atmospheres, 112, D24 302, doi:10.1029/2007JD008871, 2007.

Porter, P. S., Rao, S. T., Hogrefe, C., Gego, E., and Mathur, R.: Methods for reducing biases and errors in regional photochemical model outputs for use in emission reduction and exposure assessments, Atmospheric Environment, 112, 178 – 188, doi:http://dx.doi.org/10.1016/j.atmosenv.2015.04.039, 2015.

Rea, G., Turquety, S., Menut, L., Briant, R., Mailler, S., and Siour, G.: Source contributions to 2012 summertime aerosols in the Euro-Mediterranean region, Atmospheric Chemistry and Physics, 15, 8013–8036, doi:10.5194/acp-15-8013-2015, http://www.atmos-chem-phys.net/15/8013/2015/, 2015.

Solazzo, E. and Galmarini, S.: Comparing apples with apples: Using spatially distributed time series of monitoring data for model evaluation, Atmospheric Environment, 112, 234 – 245, doi:http://dx.doi.org/10.1016/j.atmosenv.2015.04.037, 2015.

Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-153, 2016
Manuscript under review for journal Geosci. Model Dev.
Published: 24 June 2016

Solazzo, E. and Galmarini, S.: Error apportionment for atmospheric chemistry-transport models Ž013 a new approach to model evaluation, Atmospheric Chemistry and Physics, 16, 6263–6283, doi:10.5194/acp-16-6263-2016, http://www.atmos-chem-phys.net/16/6263/2016/, 2016.

Valari, M. and Menut, L.: Does increase in air quality models resolution bring surface ozone concentrations closer to reality?, Journal of Atmospheric and Oceanic Technology, doi:10.1175/2008JTECHA1123.1, 2008.