ECMWF COPERNICUS REPORT

Copernicus Atmosphere Monitoring Service

Annual report on the verification of interim re-analyses

IRA2017

Issued by: METEO-FRANCE/ G. Collin

Date: 15/03/2019

Ref:

CAMS50_2015SC3_D50.5.3.1-2017_201903_IRA2017_verification_report_v1

IMPLEMENTED BY ECMWF

Contributors

INERIS
F. Meleux
L. Rouïl

METEO-FRANCE
G. Collin
N. Assar

Table of Contents

Table of Figures

Executive summary

The present report provides a performance analysis of the Regional interim air quality re-analyses throughout Europe, produced by CAMS for Year 2017.

The CAMS Regional services include the provision of ENSEMBLE air quality re-analyses, resulting from the combination of seven well-validated and documented chemistry-transport models' results. So-called "interim" re-analyses are data assimilated fields of air pollutant concentrations, based on up-to-date observation data. Since October 1st, 2015, according to EU Decision 2011/850/EU *on reciprocal exchange of information and reporting on ambient air quality*, EU Member States must report to the European Environment Agency (EEA) observation data as soon as it is produced, even if the necessary validation process is not completed. Such data is thus flagged as "non-validated" or "non-verified" data. Up-to-Date (UTD) data should be considered as provisional or "interim" data, until they are flagged as "validated" by the Member States, which can formally happen more than one year after their production[1].

Nevertheless, it is interesting to elaborate interim re-analyses as first guess of air pollution patterns and levels that developed in Europe in 2017. Such information can be used to support Member States for the regulatory reporting duty on air quality (according to Directive 2008/50/EC). This is the reason why it is important to carefully evaluate the simulations against observations that are not used for the re-analyses production.
INERIS performed this evaluation process and computed several performance indicators and scores for ozone, nitrogen dioxide, PM_{10} and $PM_{2.5}$ concentrations. They are presented in this report. Globally the models performed as expected and the ENSEMBLE median model generally gives good results, but not always the best ones (see PM simulations). Consistency with previous validated interim re-analyses results is ensured. However, as with the previous exercise in 2016, this evaluation demonstrates the difficulties encountered by the ensemble approach based on the median to simulate properly exceedances of threshold (limit) values. This should be clearly highlighted as a need for the CAMS_50 service evolution.
The interim re-analyses maps can be considered relevant for policy support, even if some care should be taken, as usual with provisional results.

We can highlight the following points:

- Too few up-to-date observation data was available to perform an extensive evaluation of interim re-analyses over the whole of Europe. Eastern and Southern European regions are not correctly covered, which is a pity since they correspond to areas where there are more uncertainties (especially because of emissions).
- In Western and Central Europe, where there are more stations for the evaluation of the models' performances, results are generally more representative and correct despite an overall

[1] Validated observations related to year Y-1 are reported by the 30th September of year Y by the Member States

degradation compared to the previous years. The reasons for this discrepancy have not been demonstrated, but lack of robustness of the process consisting in gathering interim observations can partially explain this situation. The European Environment Agency should strengthen quality assurance procedures in the coming years and more countries, especially in Southern Europe, are supposed to deliver up-to-date data, which will impact positively the interim re-analyses production process.

- Whatever the pollutant, the performances are always of lower quality than what can be achieved with the validated re-analyses process for which more stations are available and observation datasets are validated.
- For ozone, the ENSEMBLE model gives the best results. Ozone daily maxima are generally underestimated. The correlation coefficient ranges between 0.8 and 0.9 and the RMSE between 15 and 17 $\mu g/m^3$ at rural and suburban locations. At urban locations, these scores decrease significantly, the correlation coefficient ranging between 0.5 and 0.6 and the RMSE between 20 and 25 $\mu g/m^3$. This difference with respect to station typologies is more pronounced than in the previous assessments.
- Good model scores for simulating ozone in Western Europe are hampered by inferior performances at few stations in Eastern and Southern Europe.
- The situation of nitrogen dioxide re-analyses is much more worrying with performances significantly degraded compared to the previous years. Since such poor results are observed for all individual models, the limited number of available stations and perhaps the lack of quality of observations can explain this degradation of the performances for NO_2.
- For PM_{10}, even if the results are quite satisfactory considering the state of the art, the statistical scores remain lower than what is usually achieved with validated re-analyses, and what was observed in the previous years. Up-to-date PM_{10} observation datasets need to be more consolidated in the future.
- PM_{10} is the pollutant for which model responses range in the largest interval: the correlation coefficient from 0.4 to 0.9 and the RMSE from 16 to 9 $\mu g/m^3$, depending on the model and the station typology. The results are the best for suburban stations in Western and central Europe.
- Moreover, the evaluation demonstrates how the ensemble approach, based on a median average of involved models is not appropriate to simulate exceedances of threshold values. Only 17% (2 points less than for 2016) of the exceedances of the PM_{10} daily limit values were correctly caught by the ENSEMBLE (against about 80% for some individual models).
- Finally, although only little $PM_{2.5}$ measurement data was available for the evaluation, the results obtained for this pollutant are promising. The individual models' responses are quite consistent and the Ensemble median generally gives the best results. The correlation coefficient ranges from 0.4 to 0.9 according to the location and the station typology and the RMSE from 3 to 15 $\mu g/m^3$, which is very reasonable. Once again, the conclusions are limited by the low number of stations available in some geographical areas and should be consolidated and improved in future interim assessments, when the up-to-date data gathering process at the EEA is strengthened.

Introduction

This report gives an overview of the performances of the European air quality **interim re-analysis** process developed by the CAMS Regional services and implemented to simulate air quality in Europe during Year 2017.

Air quality interim re-analyses result from a combination of chemistry-transport models results that simulate the spatio-temporal evolution of regulatory air pollutant concentrations (according to the ambient Air quality Directive 2008/50/EC), and observations assimilated in each model to correct and improve its results. Each team providing air quality re-analyses developed appropriate and validated data assimilation chains to provide best estimates of air pollution patterns according to available observation data.

The models implemented to calculate these interim re-analyses are the set of seven models run in other CAMS Regional near-real time services. The models' set-up is described in a series of reports published in 2018[2]. The models are CHIMERE (INERIS, France), EMEP (MET Norway, Norway), EURAD-IM (RIU-UK, Germany), LOTOS-EUROS (KNMI-TNO, The Netherlands), MATCH (SMHI, Sweden), MOCAGE (METEO-FRANCE, France), and SILAM (FMI, Finland). The results presented in the following paragraphs highlight the difficulties encountered by the MATCH model. The results delivered by SMHI showed huge discrepancies whatever the pollutant and the indicator. The reasons for such unexpected behaviour remained unexplained at the time of delivery of the interim re-analyses and it was not possible for the modelling team to provide new results that could be included in the ENSEMBLE calculations. Therefore, MATCH results have been excluded from all ENSEMBLE results presented and commented in the present report.

Observations are issued from the regulatory air quality monitoring networks that report to the European Environment Agency (EEA), according to Air quality Directive 2008/50/EC and Decision 2011/850/EU on reciprocal exchange and reporting on ambient air quality. "Interim re-analyses" are so called because the observation data used are not formally validated yet. The 2011 decision stipulates that Member States must report monitoring data as soon as they are produced, in near-real time, with an appropriate flag indicating that they are not verified or validated yet. This set of data is named "Up-To-Date (UTD) data". The data is gathered in the commonly named AQ e-reporting database. "Interim data" are UTD data collected on the EEA website within a certain delay, to leave enough time to have a chance to get verified data[3]. We estimate that 20 days is an appropriate time-lag to get the data and run the re-analyses for a given day.

The set of observation sites reported to the EEA is split into two subsets, one for data assimilation in the interim re-analyses and the other for verification. Those datasets do not overlap, and verification cannot be biased by use of data for both assimilation and verification processes. It should be noted that not all Members States reported UTD data. Consequently, data assimilation and evaluation cannot be performed in some geographical areas, and it will not be possible to draw some clear conclusions about the model capacities in those regions. Southern Europe (included Italy) is more particularly concerned.

[2] Reports are referenced as "CAMS50_2015SC3_D50.3.3.2.MODEL-2017_201806_Annual_IRA_Report".

[3] Member states can check, verify and validate their data when they want and resubmit with the appropriate flag as many times as they wish. Formal validation is expected only in September the year after.

The evaluation focuses on the seven individual models and the ENSEMBLE as well. The ENSEMBLE is the result of the median of the seven models and is considered as the best estimate of air pollution patterns and levels, since it combines the strengths of the other models. This is what will be checked in the present report.

Statistical indicators (bias, root mean square error, correlation coefficient) are presented to compare the models' results against observations. Maps, histograms and Taylor diagrams are proposed for a better understanding and analysis of the performances. They are computed for the four regulatory pollutants targeted by the service: ozone (O_3), nitrogen dioxide (NO_2), particulate matter (PM_{10} and $PM_{2.5}$). Metrics relevant for policy purposes (regarding the content of the air quality directives) and for health impacts are considered for the evaluation.

All results are presented below, after a short introduction on the computed performance indicators.

1. Performance indicators

The model performances are evaluated based on classical statistical indicators that measure objectively the gap between the model results and the observations at the available stations: bias, root mean square error (RMSE) and correlation coefficient are the most classical. Comparison of observed and modelled averages is generally considered as well.
Obviously, the behaviour of performance indicators depends on the station typology and on the considered pollutant: the models used in the CAMS Regional service run at the European scale and their spatial resolution is about 20 to 10 km in the best case. Consequently, for pollutants which are largely influenced by local sources (NO_2, PM in some situations), these regional models are not able to reproduce hot spots monitored by traffic or industrial stations, and performance indicators will not be assessed. Difficulties can even be encountered at urban stations.

Conversely for pollutants characterized by long residence time in the atmosphere and large impacted areas (typically ozone and PM in some cases), performance indicators evaluated at all types of stations (except traffic and industrial sites) make sense.

The definitions of the various performance indicators used in the report are given below. They are very usual[4] in evaluation processes:

- Bias indicates, on average, if the simulations under or over-predict the actual measured concentrations. In our case, negative values indicate under-prediction, whereas positive values indicate over-prediction; values close to 0 are the best ones:

  $$\frac{1}{N}\cdot\sum_{i=1}^{N}(P_i - O_i)$$

  Where N is the number of observations, P_i refers to the predictions and O_i to the observations. It is expressed in µg/m^3.
- Root Mean Square Error (RMSE) gives information about the skill of the model in predicting the overall magnitude of the observations. It should be as weak as possible:

  $$\sqrt{\frac{1}{N}\cdot\sum_{i=1}^{N}(P_i - O_i)^2}$$

  Where N is the number of observations, P_i refers to the predictions and O_i to the observations. It is expressed in µg/m^3.
- Correlation is a measure of whether predictions and observations change together in the same way (i.e. at the same time and/or place). The closer the correlation is to one, the better is the correspondence of extreme values of the two data sets.

  $$r = \frac{\mathrm{cov}(P_i, O_i)}{\sqrt{\mathrm{var}(P_i)}\cdot\sqrt{\mathrm{var}(O_i)}}$$

  Where N is the number of observations, P_i refers to the predictions and O_i to the observations. This is a non-dimensional number.

[4] Chang J.C. et Hanna S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys. 87, 167–196.

Taylor diagrams synthesize on a unique quadrant, various statistical indicators for different models: the radii correspond to the correlation coefficient values, the x-axis and the y-axis delimits arcs with bias values and the internal semi-circles correspond to the RMSE values. Therefore, this is a pedagogical way to present an overview of the relative performances of a set of models, often used in model intercomparison exercises.

For indicators related to threshold values, for instance the number of days, hours when a certain concentration level is exceeded, some “contingency tables” giving the percentages of correct predictions (GP), false alarms (FA), or missing events (ME) are estimated. These concepts come from the weather or air quality forecasting world. Although they are very severe and not objectively representative of the intrinsic model performance (because of the threshold cut-effect, a result close to the threshold can fall arbitrary in one or the other category), they can give useful information to compare various models’ behavior in different geographical regions. GP, FA and ME are expressed in percentage (%).

Several representations of the models’ skills are proposed:

- Maps with colored patches at the location of the stations selected for the evaluation process. The color scale indicates how the model performs.
- Taylor diagrams provide a wider overview of the model performances.
- Histograms with model performances sorted by station typology and by European sub-region (Western, Northern, Southern, Central, Eastern) are proposed as well.

2. Performance indicators for ozone

In this evaluation, we focused on the ability of the model to correctly predict the ozone daily maximum (hourly average), which is the most relevant considering regulatory indicators like the number of exceedances of information and alert thresholds. The evaluation is performed over the "summer" period when ozone increases, reaching levels that may impact human health and ecosystems.

Erreur ! Source du renvoi introuvable. shows the Taylor diagram that synthesizes performances of individual CAMS models and the ENSEMBLE to simulate hourly daily maximum of ozone in the summer period. The graph highlights huge problems encountered by the MATCH model (from SMHI) and quite good performances of the other models. In particular, ENSEMBLE results (in light blue) are very satisfactory with correlation coefficients of about 0.85 and RMSE lower than 15 μg/m^3 for rural and suburban areas. Performances decrease for all models at urban locations (correlation coefficients about 0.6 and RMSE about 23-24 μg/m^3).

In-depth analysis of the ENSEMBLE interim re-analyses can be elaborated considering the spatial distribution of the statistical indicators over Europe. 0 presents maps of bias, correlation coefficient and RMSE related to the ENSEMBLE, for daily maxima from April 1st to September 30th, 2017. Bias ranges in most parts of Europe between -5 and 5 μg/m^3. Higher bias values (underestimation) can be found in some specific locations in the Southern part of Europe, rather at rural locations. However, it should be noted that evaluation cannot be conducted in several Southern countries (Italy, Greece, Slovenia, Croatia) and Eastern countries (Romania, Bulgaria), because of a lack of reported interim observation data. It is interesting to note local discrepancies of the quality of the indicator in mountainous areas as well.
Correlation coefficient is excellent with values higher than 0.95 in most cases. The same can be seen for RMSE, although for a major number of stations, RMSE ranges between 15 and 20 μg/m^3, which is quite good for interim results, but higher than what is usually achieved with validated assimilated results (rather between 5 and 15 μg/m^3). This can be a consequence of using partial and non-validated observation data, and results should improve when the validated re-analyses are performed. However, results remain acceptable compared to the state of the art.
Performances decrease for stations around the Mediterranean area (Portugal, Spain) and in Eastern Europe, with values that may be higher than 25 μg/m^3.

To help in the interpretation of those maps, one can consider the same performance indicators for each individual model and the ENSEMBLE and various station typologies. **Erreur ! Source du renvoi introuvable.** and **Erreur ! Source du renvoi introuvable.** present bias, correlation coefficient and RMSE scores for all models at rural and suburban stations respectively. The indicators are sorted per geographical region: Western Europe (EUW), Central Europe (EUC), Southern Europe (EUS), Northern Europe (EUN), Eastern Europe (EUE). The interpretation of results is hampered by the low (sometimes null) number of stations available for verification is some areas: in Southern Europe, no station was available and the verification process has not been performed since very few countries in that area report UTD observation data to the European Environment Agency. The situation is expected to improve in the coming years. In Eastern and Northern Europe, the evaluation has been performed against a very little number of stations, which may be a problem regarding the representativeness of the obtained results.

As mentioned previously, performances from MATCH are very poor whatever the situation and will not be further commented.
Where observation data is available, it is interesting to note very few differences in the performances for rural and suburban sites. At rural and suburban sites, all models except LOTOS-EUROS underestimate ozone peak values, which is consistent with what was observed for the previous years. Correlation coefficient is quite high ranging from 0.8 to 0.9 in the best case which is the ENSEMBLE model.
Regarding RMSE, we can note once again good consistency between results they are even slightly better at rural stations. Better results are obtained for the ENSEMBLE model (from 13 to 17 µg/m^3) at rural and suburban sites.
Obviously, there are more uncertainties in the models in Southern, Eastern and Northern regions due to uncertainties in emissions, and the complexity of the photochemical processes and meteorology. However, there are also much fewer stations than in other regions, making the scores very sensitive to the weak performance of one or two stations. For this reason, conclusions should be established with care and refined when validated re-analyses for 2017 are available.
Nevertheless, overall performances of the models to simulate ozone daily maxima are satisfactory and consistent with previous results obtained in the past and with the state of the art.

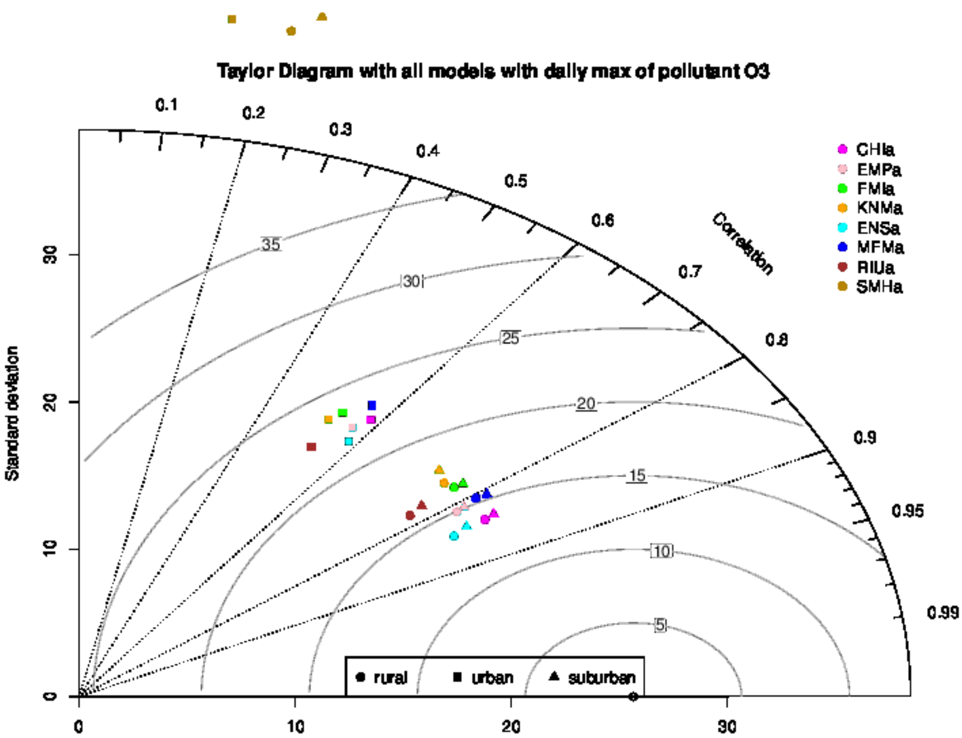
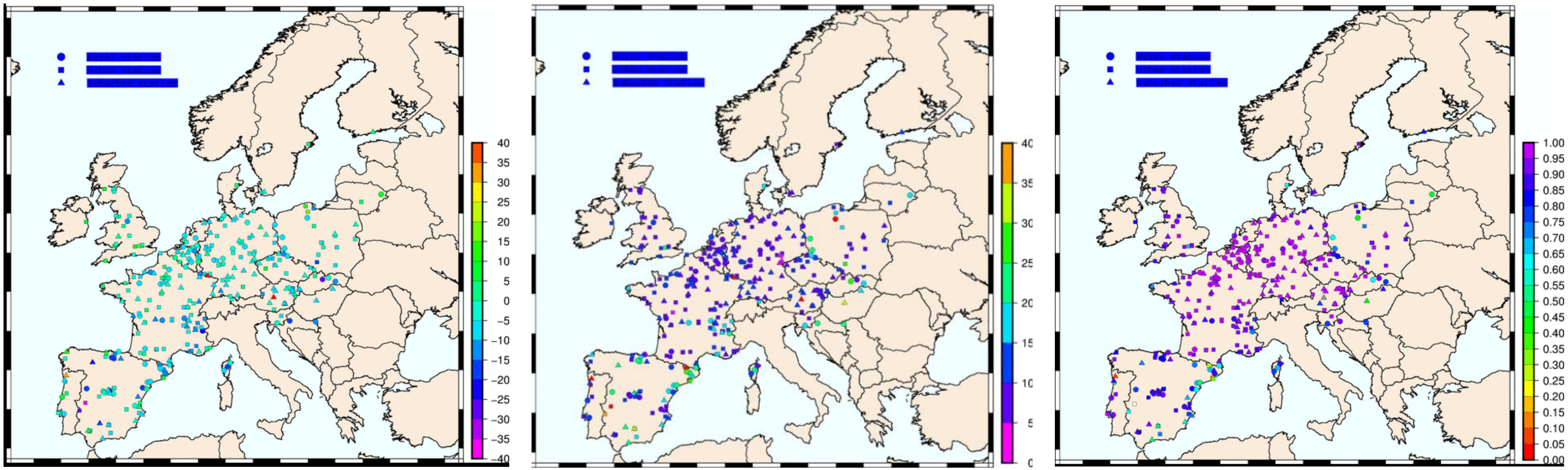
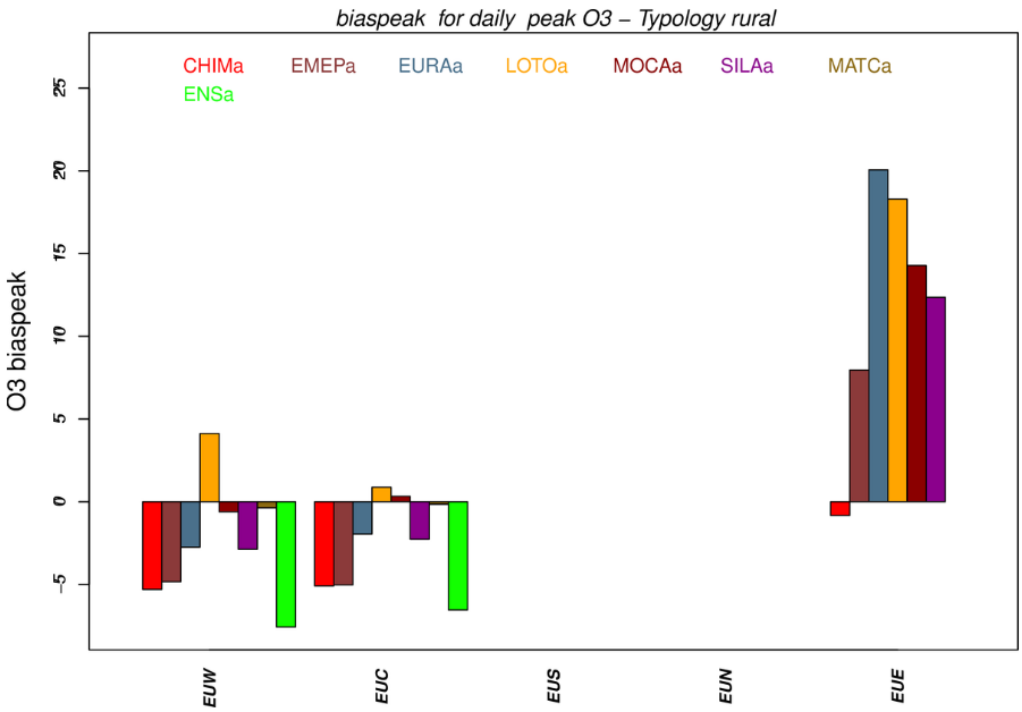


Figure 1 - Taylor diagram presenting performances of all CAMS regional models to simulate summer ozone daily maximum (hourly average) for various stations typologies (rural, urban, suburban).

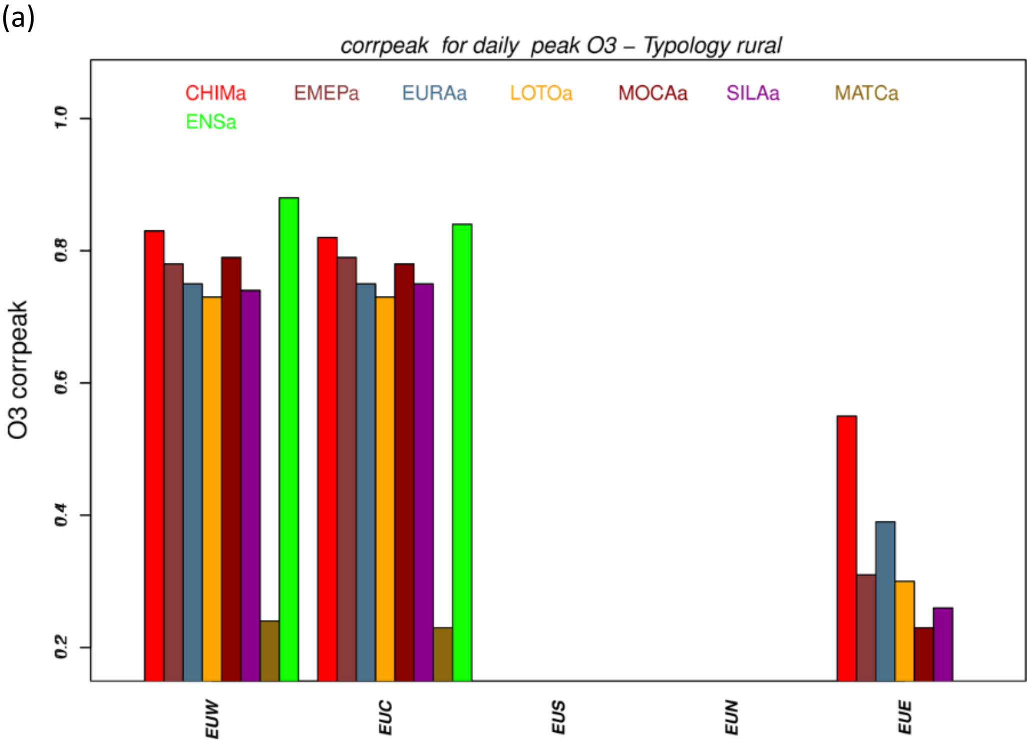


(a) ozone bias (daily maximum) (b) ozone RMSE (daily maximum) (c) ozone correlation coefficient (daily maximum)

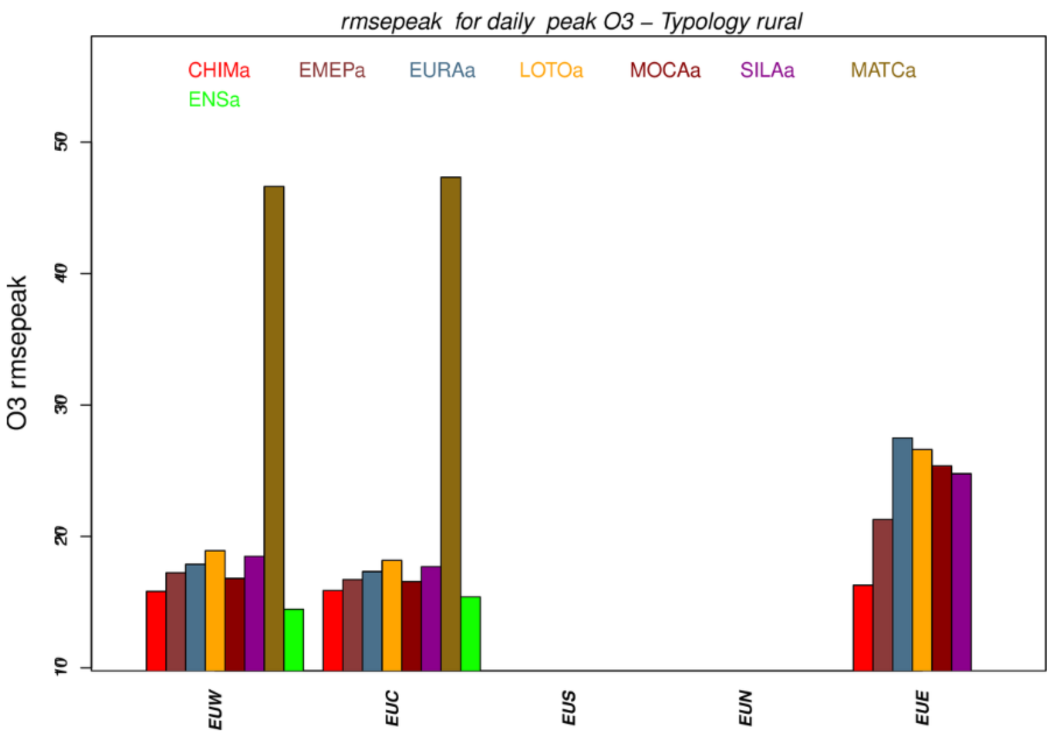
Figure 2 - Maps of Statistical scores of the ENSEMBLE interim re-analysis results against the observation validation dataset from the AQ e-reporting database, for the ozone daily maximum from 01/04/2017 to 30/09/2017 (a) Bias, (b) Root mean square error, (c) Correlation coefficient.


biaspeak for daily peak O3 – Typology rural
CHIMa
EMEPa
EURAa
LOTOa
MOCAa
SILAa
MATCa
ENSa
O3 biaspeak
25
20
15
10
5
0
−5
EUW
EUC
EUS
EUN
EUE

(a)

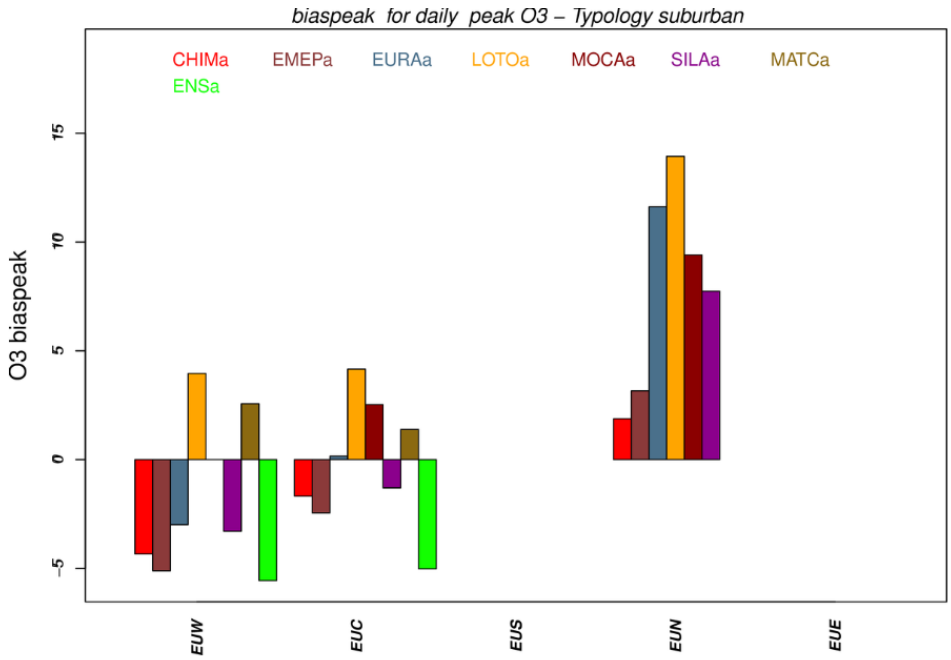

corrpeak for daily peak O3 – Typology rural
CHIMa
EMEPa
EURAa
LOTOa
MOCAa
SILAa
MATCa
ENSa
O3 corrpeak
1.0
0.8
0.6
0.4
0.2
EUW
EUC
EUS
EUN
EUE

(b)

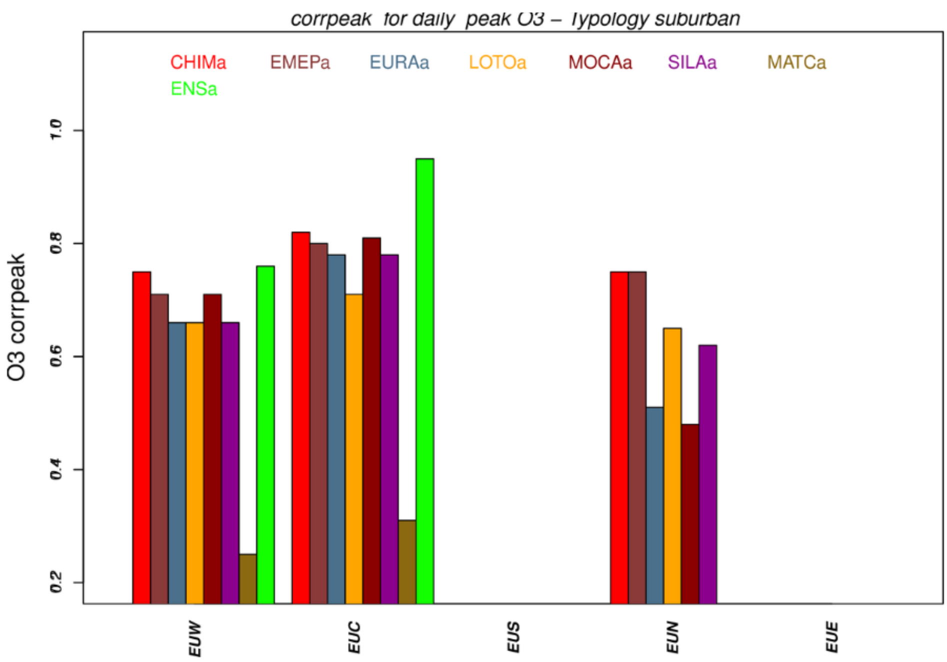


(c)

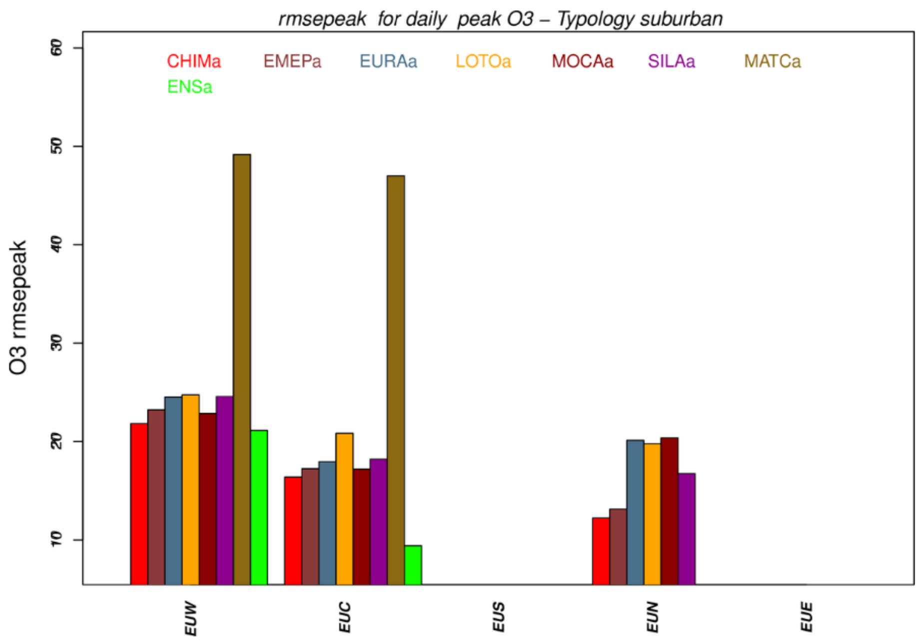
Figure 3 - CAMS Regional interim re-analyses for predicting daily ozone peak over the summer 2017 throughout European sub-regions (a) Bias, (b) Correlation coefficient, (c) Root mean square error at rural stations.


biaspeak for daily peak O3 – Typology suburban
CHIMa
EMEPa
EURAa
LOTOa
MOCAa
SILAa
MATCa
ENSa
O3 biaspeak
15
10
5
0
-5
EUW
EUC
EUS
EUN
EUE

(a)


corrpeak for daily peak O3 – Typology suburban
CHIMa
EMEPa
EURAa
LOTOa
MOCAa
SILAa
MATCa
ENSa
O3 corrpeak
1.0
0.8
0.6
0.4
0.2
EUW
EUC
EUS
EUN
EUE

(b)

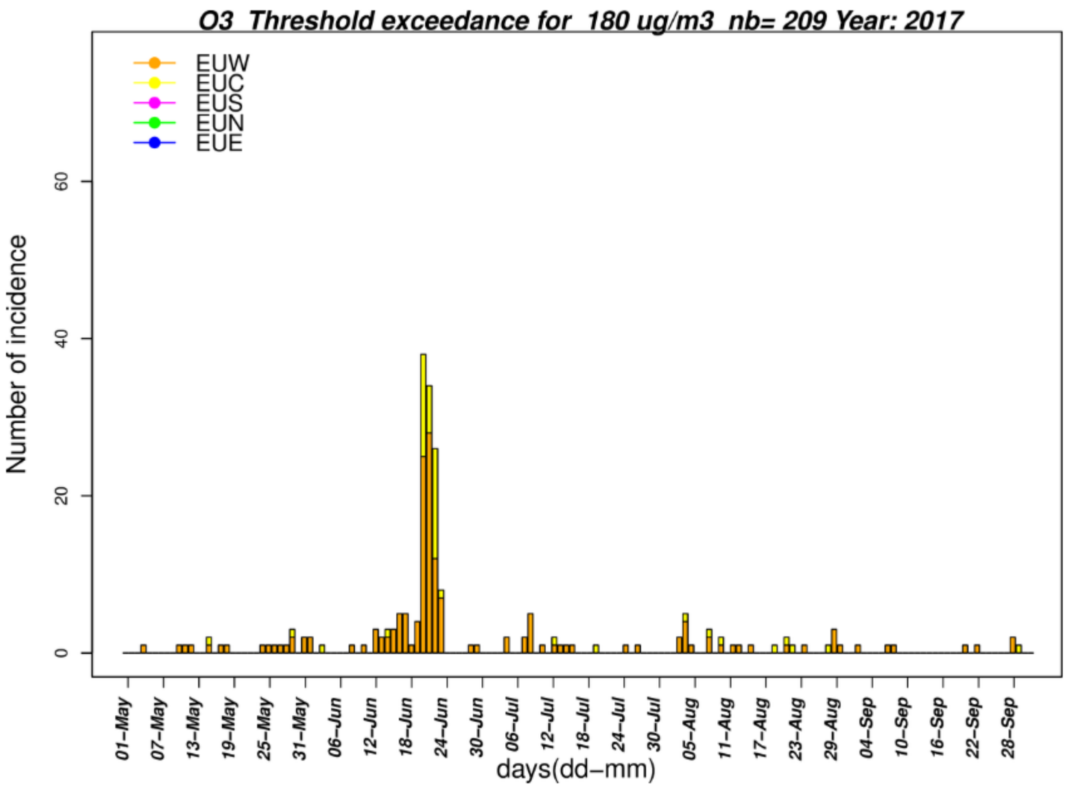


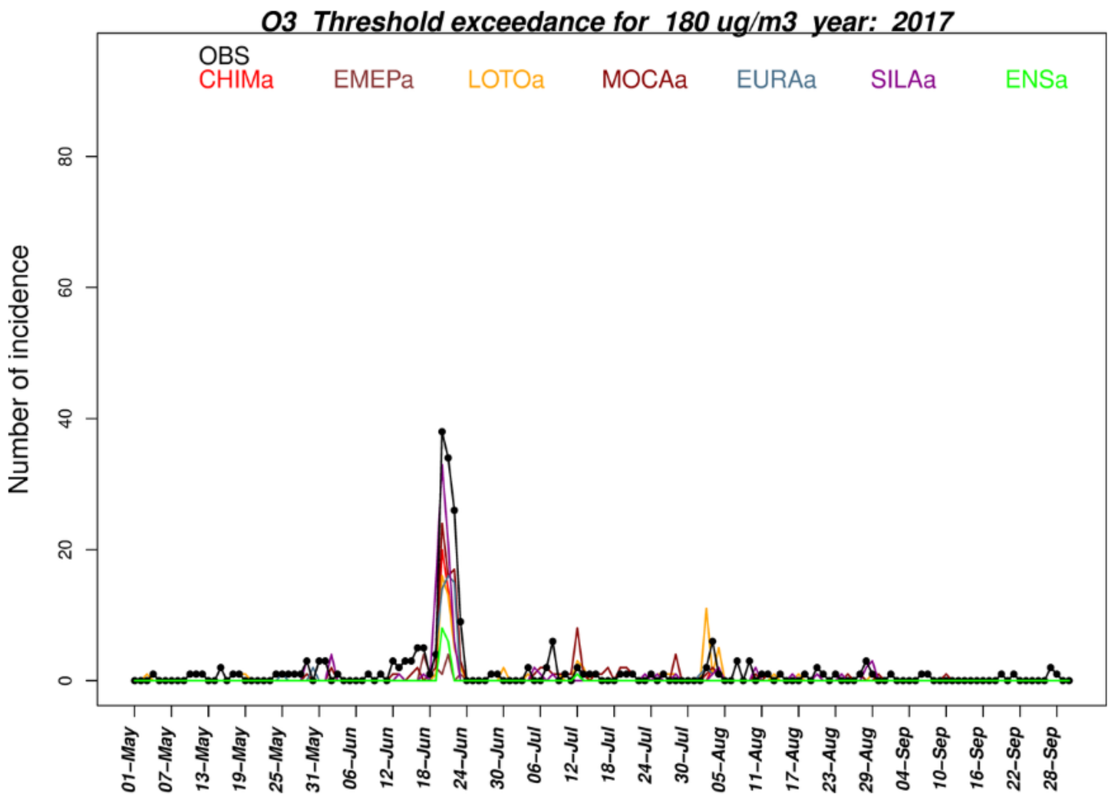
(c)

Figure 4 – CAMS Regional interim re-analyses for predicting daily ozone peak over summer 2016 throughout European sub-regions (a) Bias (b) Correlation coefficient (c) Root mean square error at suburban stations.

Finally, the models' ability to simulate the number of exceedances of a given threshold value has also been assessed. This is important for ozone, since the EU legislation (Directive 2008/50/EC) sets quality objectives with an information threshold (180 µg/m^3) and an alert threshold (240 µg/m^3), over which short-term action plans and communication towards the general public should be implemented by Member States. However, this kind of evaluation against threshold value is very stringent and not always representative of the model quality. Situations above and below the threshold value are counted, but to correctly take into account model uncertainty, it would be necessary to take a range of acceptable values around the threshold. This is not done in the present study. Therefore, the diagnosis can be seen as a pessimistic analysis of the models' performances.

Erreur ! Source du renvoi introuvable. below shows the number of situations when the hourly information threshold has been exceeded during the summer time in 2017 (time is presented on the x-axis), sorted per geographical region (various colors). The first set of histograms shows observed exceedances at ozone stations in Europe. Few exceedances of the information threshold were recorded: only 209 located in Central and Western Europe. Only one significant episode occurred at the end of June 2017. The figure in the middle panel displays exceedances modelled by all CAMS models and the ENSEMBLE. If the performances of the ENSEMBLE were good considering statistical indicators, they show very disappointing results for threshold indicators. The ENSEMBLE detected only 15 exceedances. This can be explained by the nature of the indicator (no range of uncertainty is taken into account), but also by the way the ENSEMBLE is built up. It is based on the median of individual model results with performance varying largely from a model to another (**Erreur ! Source du renvoi introuvable.**). The median smooths the indicator (evaluation against threshold values) and the obtained results cannot be considered as representative of the actual quality and accuracy of the models. The model which gave the best results was MOCAGE (107 exceedances simulated), as illustrated by the last graph on **Erreur ! Source du renvoi introuvable.**.


O3 Threshold exceedance for 180 ug/m3 nb= 209 Year: 2017
EUW
EUC
EUS
EUN
EUE
Number of incidence
days(dd–mm)


O3 Threshold exceedance for 180 ug/m3 year: 2017
OBS
CHIMa
EMEPa
LOTOa
MOCAa
EURAa
SILAa
ENSa
Number of incidence

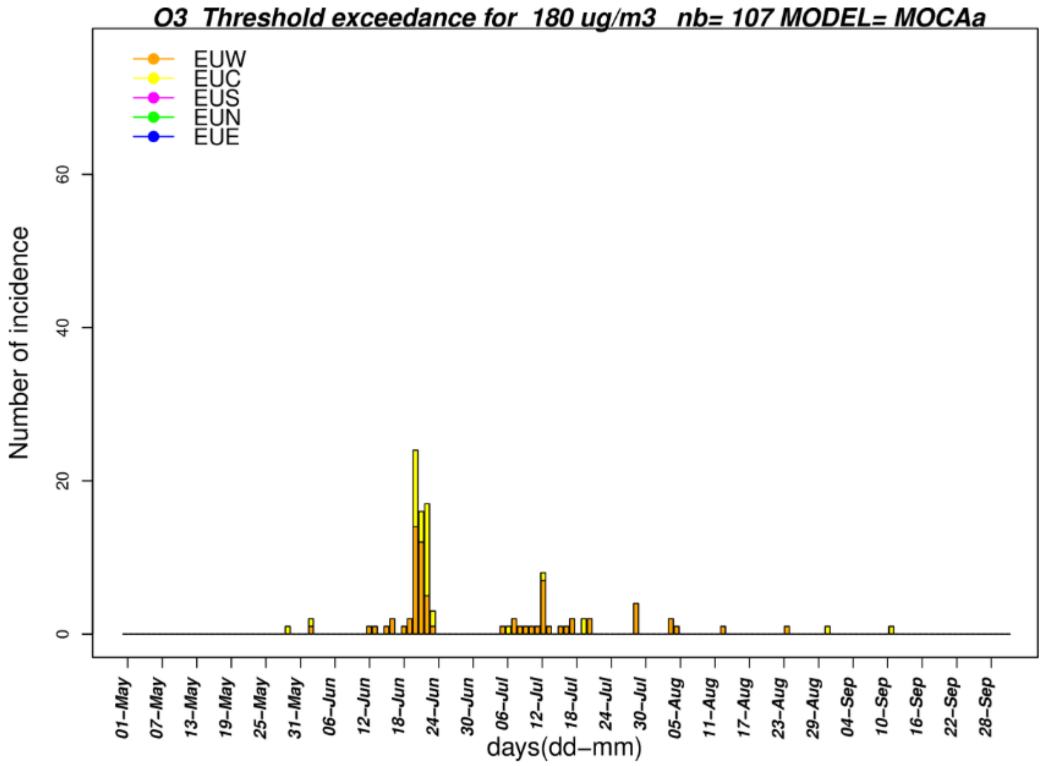


Figure 5 - Number of exceedances of the information threshold value for ozone in summer 2017 – observed (top), modelled by all the interim analyses (middle) and modelled by MOCAGE (bottom).

3. Performance indicators for nitrogen dioxide

Warning note: *As already stated from the MACC projects, it should be reminded that the CAMS Regional mapping system is not fitted to deal with local hot spot situations, such as those that develop near busy roads or on industrial sites. Actually, the model resolution is about 20 to 10 km, and is not sufficient to catch actual NO_2 concentrations at traffic and industrial sites.*

Erreur ! Source du renvoi introuvable. presents the Taylor diagram for CAMS Regional ENSEMBLE interim re-analyses, for the daily maximum (hourly average) of NO_2 concentrations. It shows disperse model performances depending on station typology, and quite poor performances, which show significant discrepancies compared to the interim re-analyses simulated for the year 2016 (**Erreur ! Source du renvoi introuvable.**). For 2017 interim re-analyses, the correlation coefficient did not exceed 0.4 (while it ranged from 0.3 to 0.7 in 2016) and the root mean square error ranged between 20 and 30 µg/m^3. **Erreur ! Source du renvoi introuvable.** showing individual models' performances demonstrates that poor scores and lack of accuracy concern all the models, allowing us to suspect a problem with the number and the quality of interim observation data assimilated in the interim re-analyses chains.

Maps in 0 allow to better investigate these differences between performances for both years 2016 and 2017. Degradation of interim re-analyses is very clear for all stations, even in locations where models usually perform well (Western Europe). We also note that there are more locations where the performances (especially correlation coefficient and RMSE) are clearly insufficient. Further investigations are needed to understand such behavior. For nitrogen dioxide, the influence of local sources is very important, but they cannot be accurately taken into account by the adopted model resolution. However, this explanation is insufficient to interpret the degradation of the results compared to 2016. This point will be investigated in-depth while assessing the performances of the validated re-analyses for 2017 and the performances of the interim re-analyses for 2018.

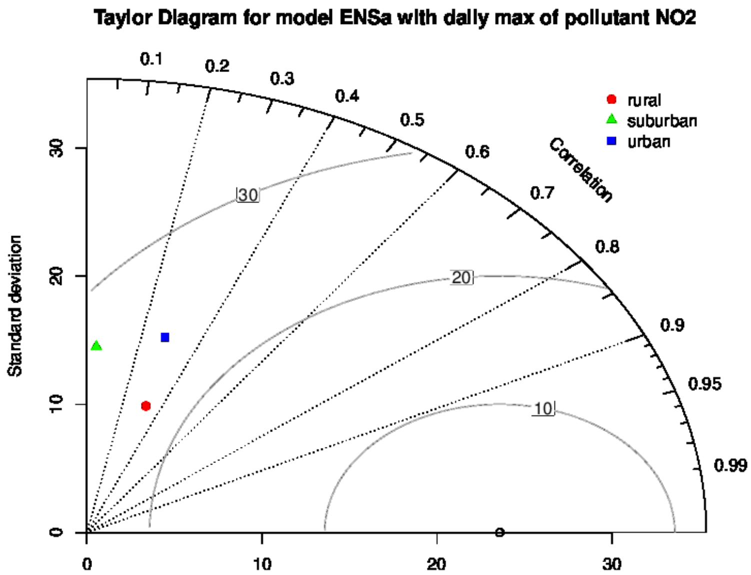


Figure 6 - Taylor diagram presenting the performances of the CAMS Regional ENSEMBLE interim re-analyses to predict NO_2 daily maxima in 2017.

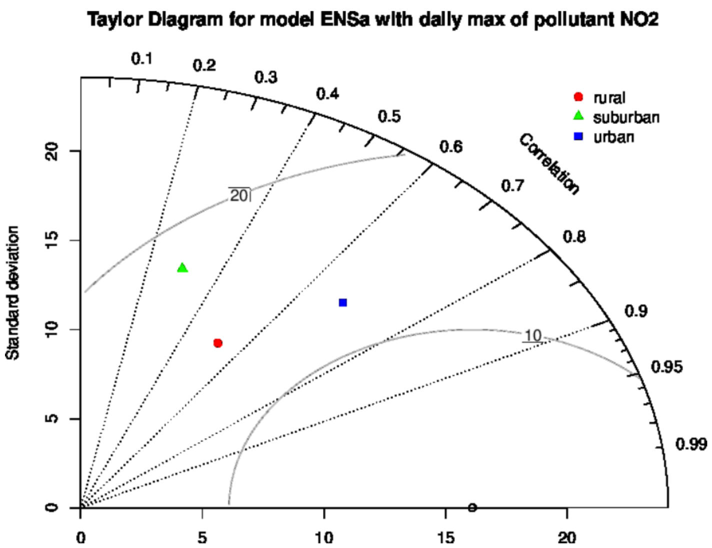


Figure 7 – Taylor diagram presenting the performances of the CAMS Regional ENSEMBLE interim re-analyses to predict NO_2 daily maxima in 2016.

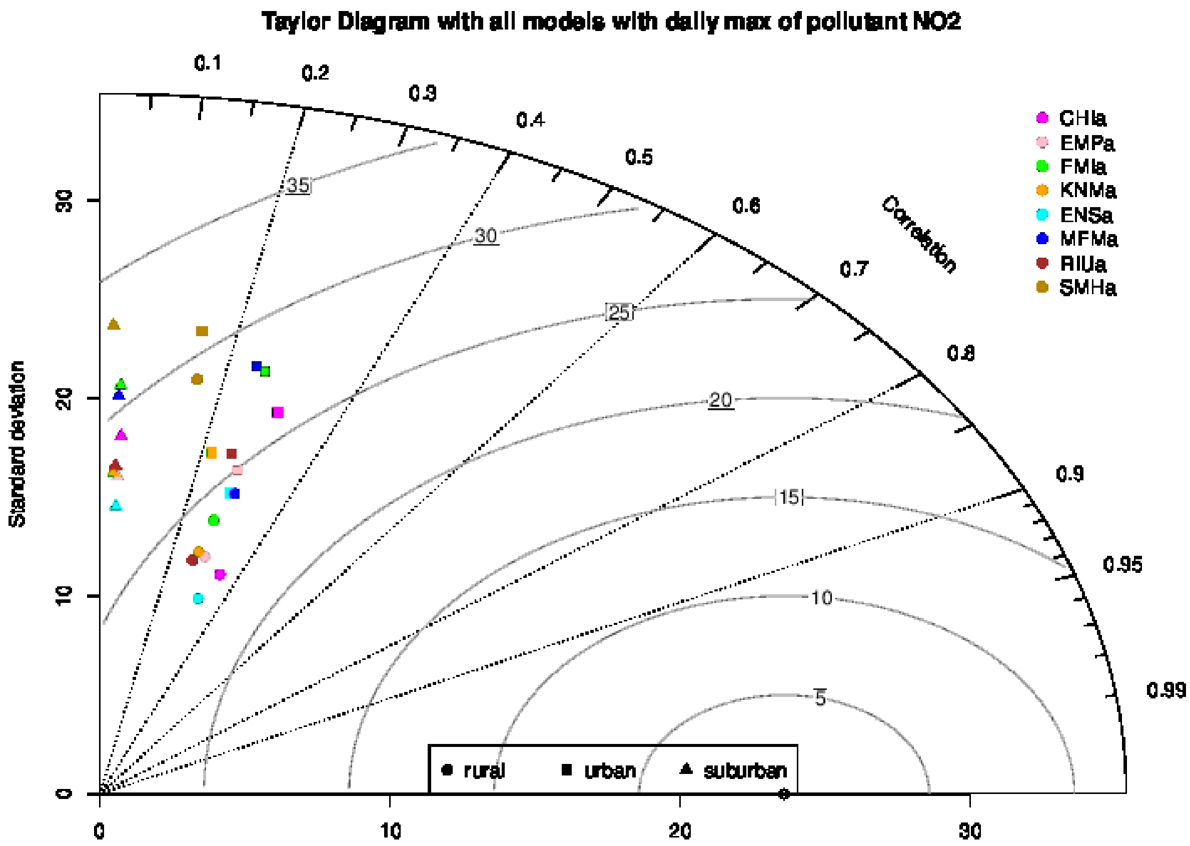


Figure 8 - Taylor diagram presenting the performances of all CAMS Regional re-analyses to predict NO_2 daily maxima in 2017.

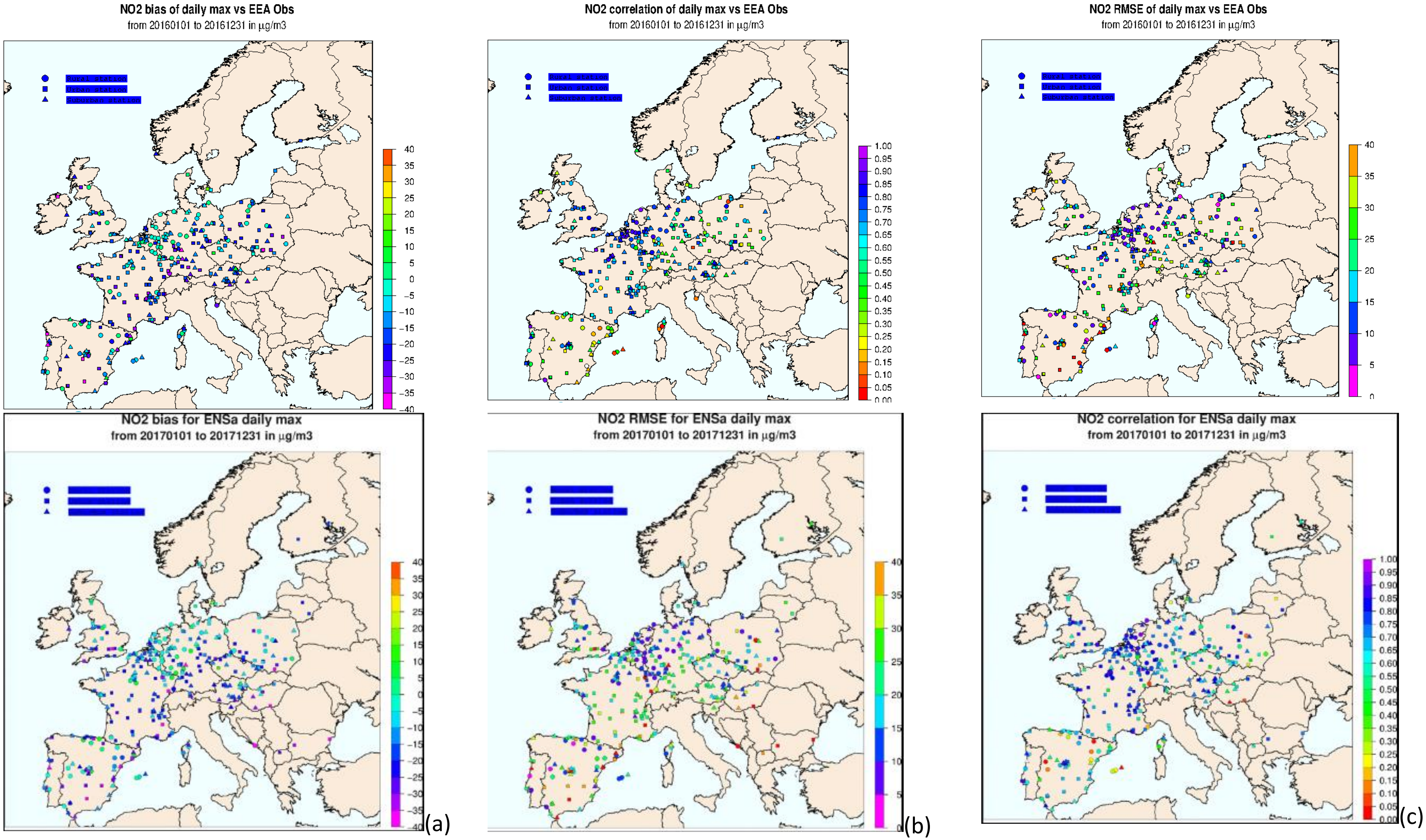


Figure 9 - Maps of Statistical scores of the ENSEMBLE interim re-analyses results against the observation validation dataset from the AQ e-reporting database for the NO_2 daily maximum, over 2016 (top) and 2017 (bottom): Bias (a), Correlation coefficient (b), Root mean square error (c).

4. Performance indicators for PM_{10}

Erreur ! Source du renvoi introuvable. shows the Taylor diagram obtained for PM_{10} daily averages over the year 2017, for CAMS Regional individual and ENSEMBLE re-analyses. The results are very encouraging with the ENSEMBLE correlation coefficient ranging from 0.6 to 0.8 (a bit lower than what is achieved with the MOCAGE and CHIMERE models). Results are poorer in rural areas than at other locations. The ENSEMBLE RMSE ranges from 14 to 11 µg/m³, which is better than what was obtained for the 2016 re-analysis (16 to 14 µg/m³), but less good than the validated re-analyses usual performances (RMSE lower than 10 µg/m³). It shows the sensitivity of the maps and the associated performance indicators to the quality of the observation datasets.

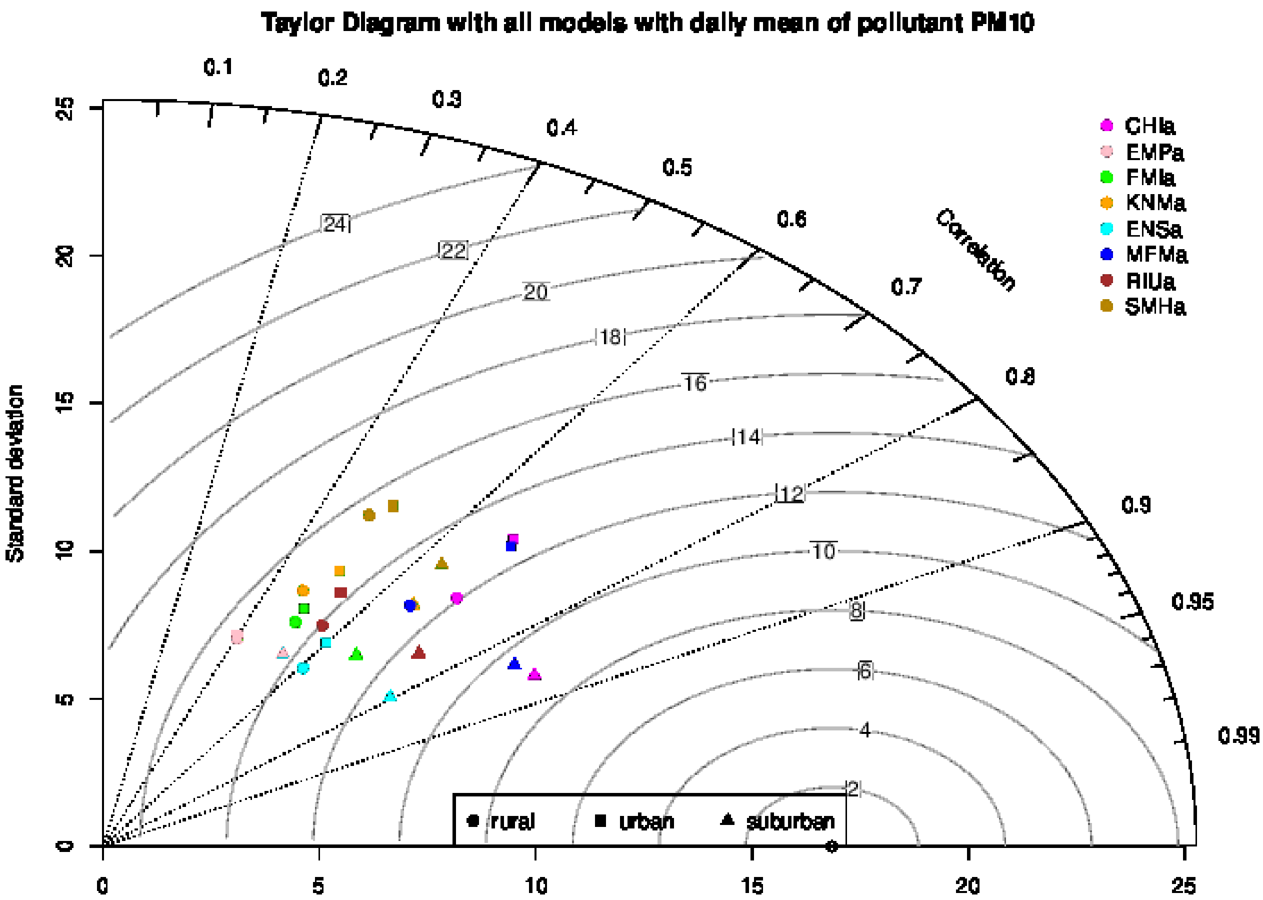


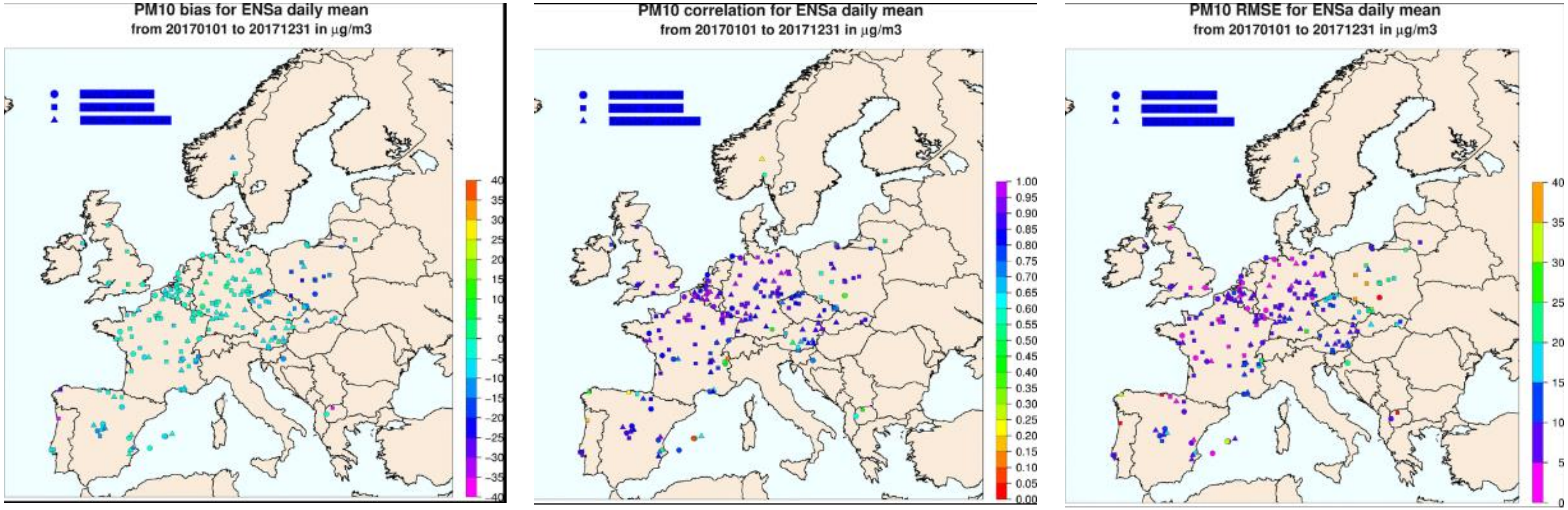
Figure 10 - Taylor diagram presenting the performances of the CAMS Regional ENSEMBLE interim re-analyses to predict PM_{10} daily average in 2017.

0 details the geographical distribution of statistical scores (bias, correlation coefficient and RMSE), for the ENSEMBLE interim re-analyses for the year 2017. Lowest correlation scores are obtained for stations located in Portugal, Spain and in the Alps (so in mountainous areas) and in the Central-Eastern parts of Europe. In several countries (France, Germany, Benelux and the UK), the RMSE ranges

between 1 and 5 µg/m^3, which is very encouraging. As observed in 2016, PM_{10} concentrations are always underestimated (see bias), but within an acceptable range (between 5 and 10 µg/m^3).

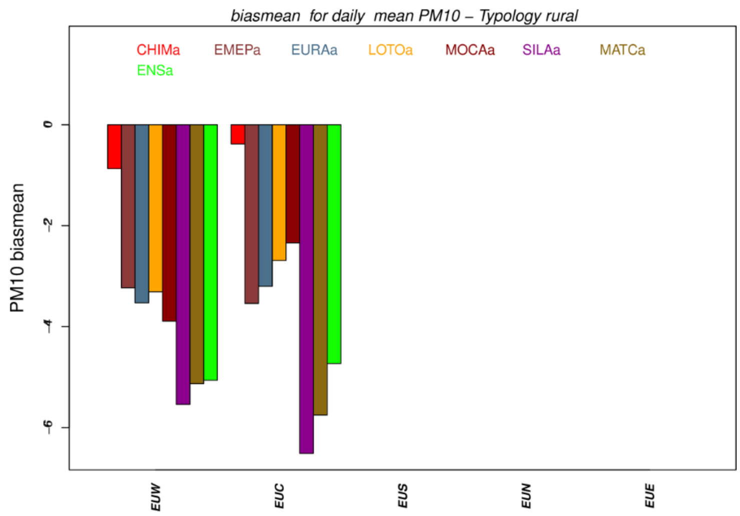
Local discrepancies can be explained by the complexity (in meteorological terms) of certain areas (mountainous regions), but also by uncertainties in the emission inventories (especially in Eastern Europe), and by the lack of available observation data for data assimilation and evaluation.

Differences between model results can be further investigated considering histograms of scores per region and for each model. 0, 0 and 0 show these results for rural, suburban and urban stations respectively. They confirm the low number of stations available for the verification of interim PM_{10} re-analyses, with huge gaps in some areas (Southern, Northern and Eastern Europe for rural and suburban stations, and Southern Europe for urban stations). The model performances vary largely from a region to another (for example for rural stations with the RMSE lower than 10 µg/m^3 in Western Europe and larger than 25 µg/m^3 in Central Europe). Because of the quite large variability of model responses, the ENSEMBLE is not always the best model, but behaves fairly well. In general, the CHIMERE model gives the best results followed by EURAD-IM and MOCAGE. Surprisingly, the models' performances are not so good at urban stations in Western Europe (correlation coefficient lower than 0.4 and RMSE higher than 40 µg/m^3) whatever the model. This pattern was already noted for the 2016 interim re-analyses and should be further investigated in the future and when validated re-analyses are available.

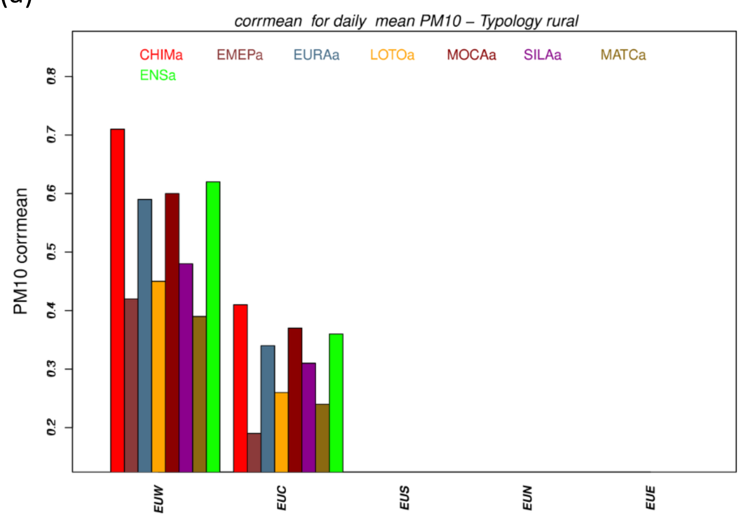


(a) (b) (c)

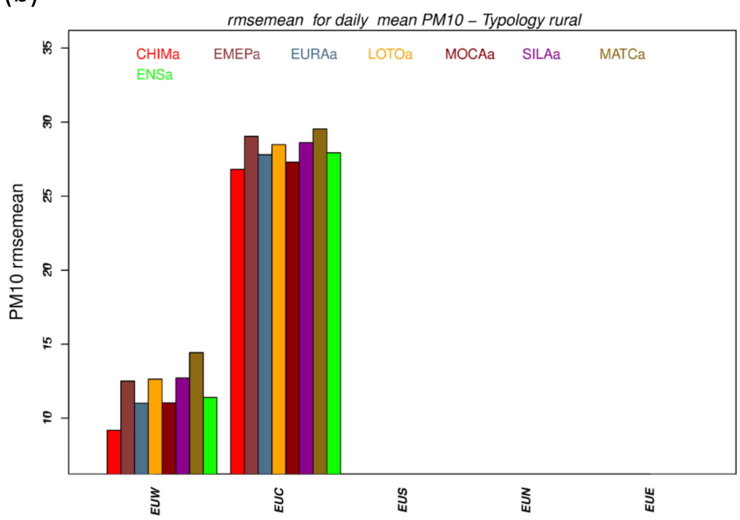
Figure 11 - Maps of Statistical scores of the ENSEMBLE interim re-analyses results against the observation validation dataset from the AQ e-reporting database for the PM_{10} daily average over the year 2017 Bias (a), Correlation coefficient (b), Root mean square error (c).



(a)

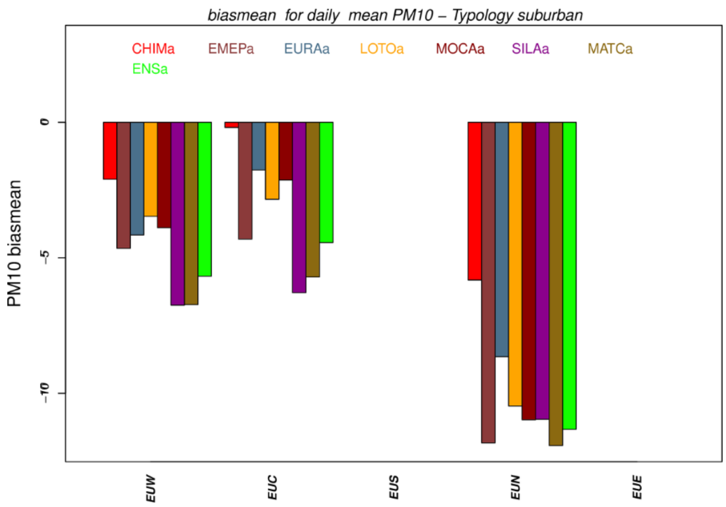


(b)

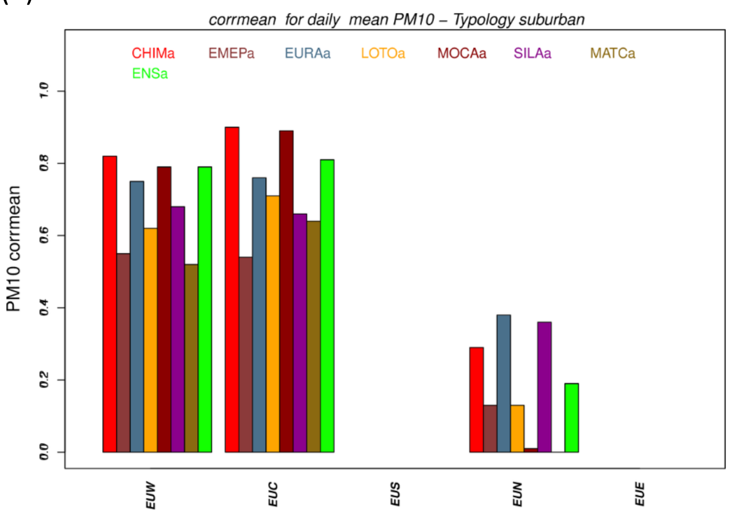


(c)

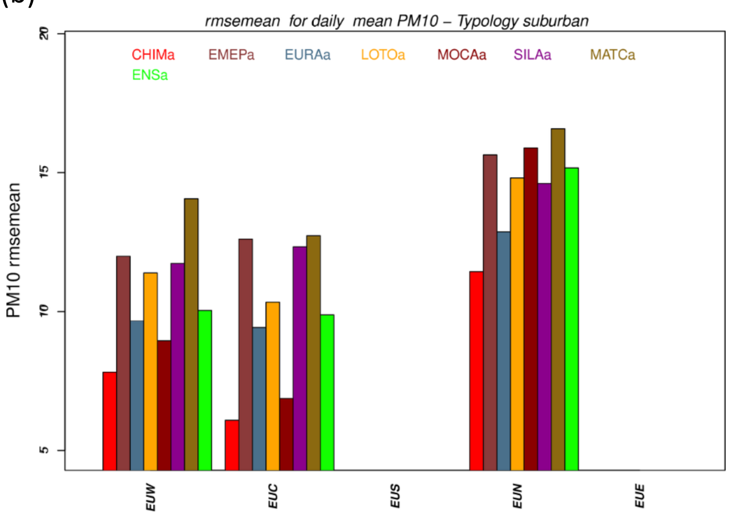
Figure 12 - CAMS Regional interim re-analyses for predicting PM_{10} daily average over the year 2017 throughout European sub-regions Bias (a), Correlation coefficient (b), Root mean square error at rural stations (c).



(a)

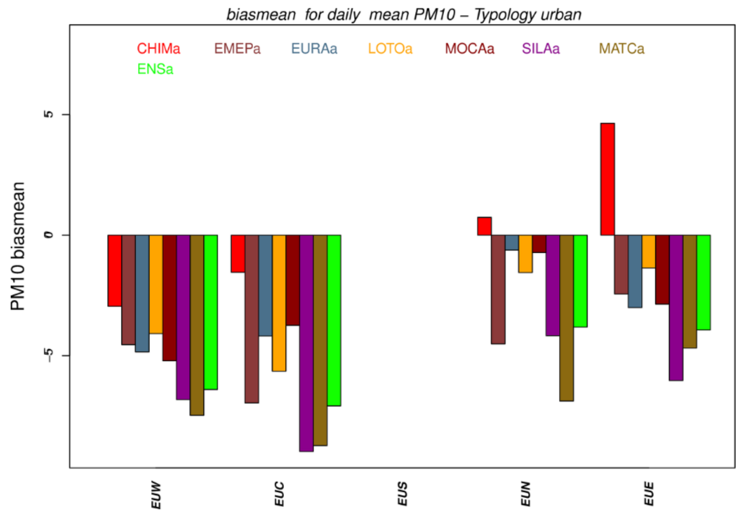


(b)

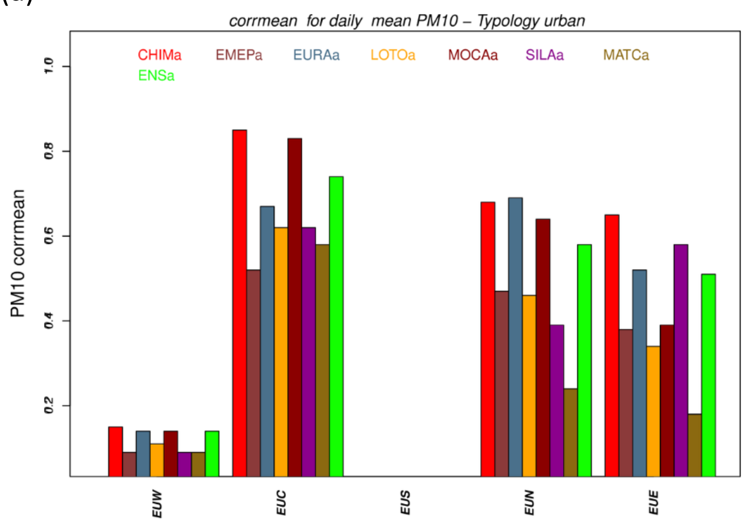


(c)

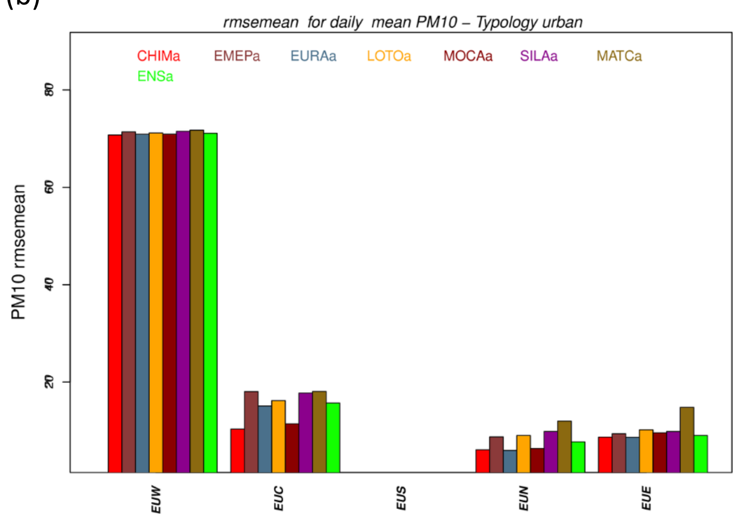
Figure 13 - CAMS Regional interim re-analyses for predicting PM_{10} daily average over the year 2017 throughout European sub-regions Bias (a), Correlation coefficient (b), Root mean square error (c) at suburban stations.



(a)

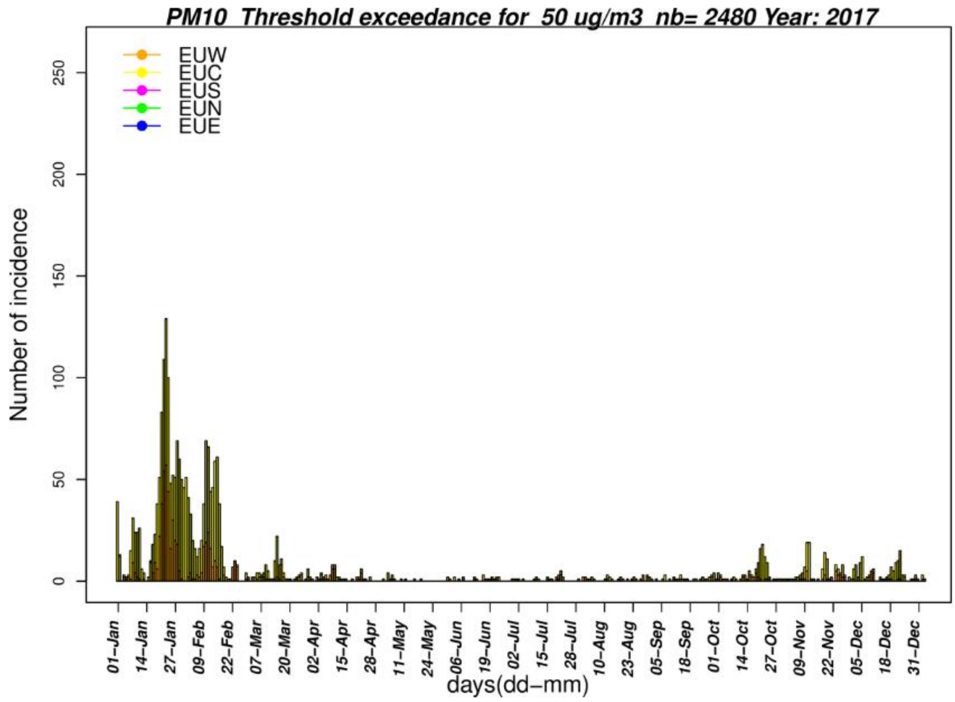


(b)



(c)

Figure 14 - CAMS Regional interim re-analyses for predicting PM_{10} daily average over the year 2017 throughout European sub-regions Bias (a), Correlation coefficient (b), Root mean square error (c) at urban stations.



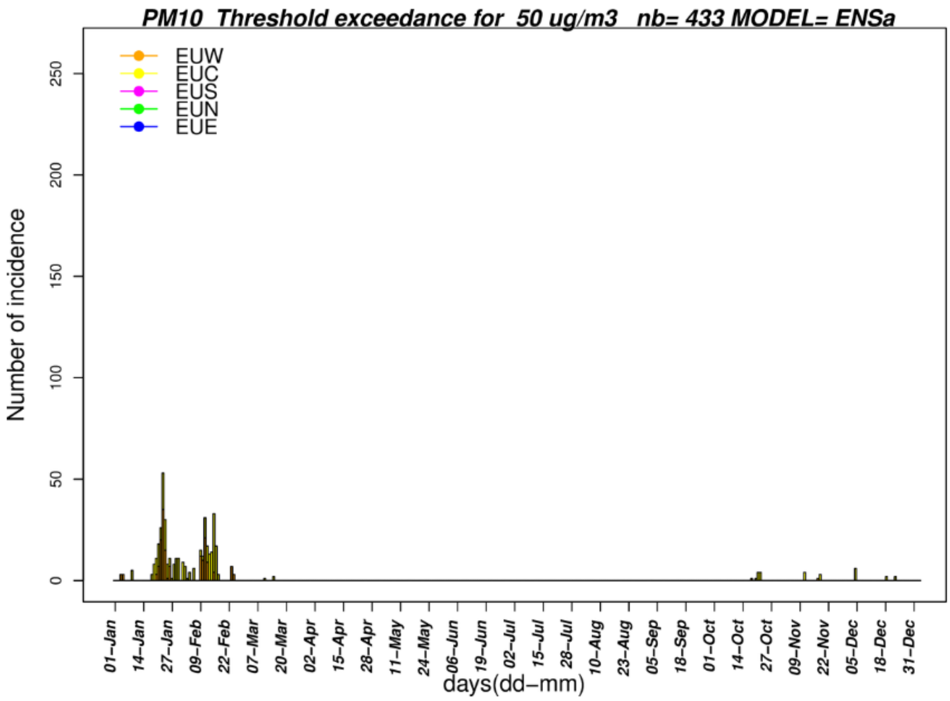


Figure 15 - Number of exceedances of daily limit value for PM_{10} in 2017 – observed (top) and modelled by the ENSEMBLE interim re-analyses.

Erreur ! Source du renvoi introuvable. shows the number of exceedances of the PM_{10} daily limit value (50 µg/m^3) sorted per region. Both observed and re-analysed data (from the ENSEMBLE) are presented and compared. Logically, a large number of exceedances are missed by the model, only

17% (19% in 2016) of the exceedance situations are correctly captured by the ENSEMBLE. As mentioned for ozone, the indicator does not allow accounting for intrinsic uncertainty around the threshold value and is very severe for model evaluation. It can be noted that all early winter exceedances were generally missed while models performed better for situations that occurred in February. The CHIMERE model managed to catch 80% of the exceedances (0), demonstrating, as for ozone, that the adopted ensemble approach in CAMS services is not really appropriate to monitor episodes of high concentration values.

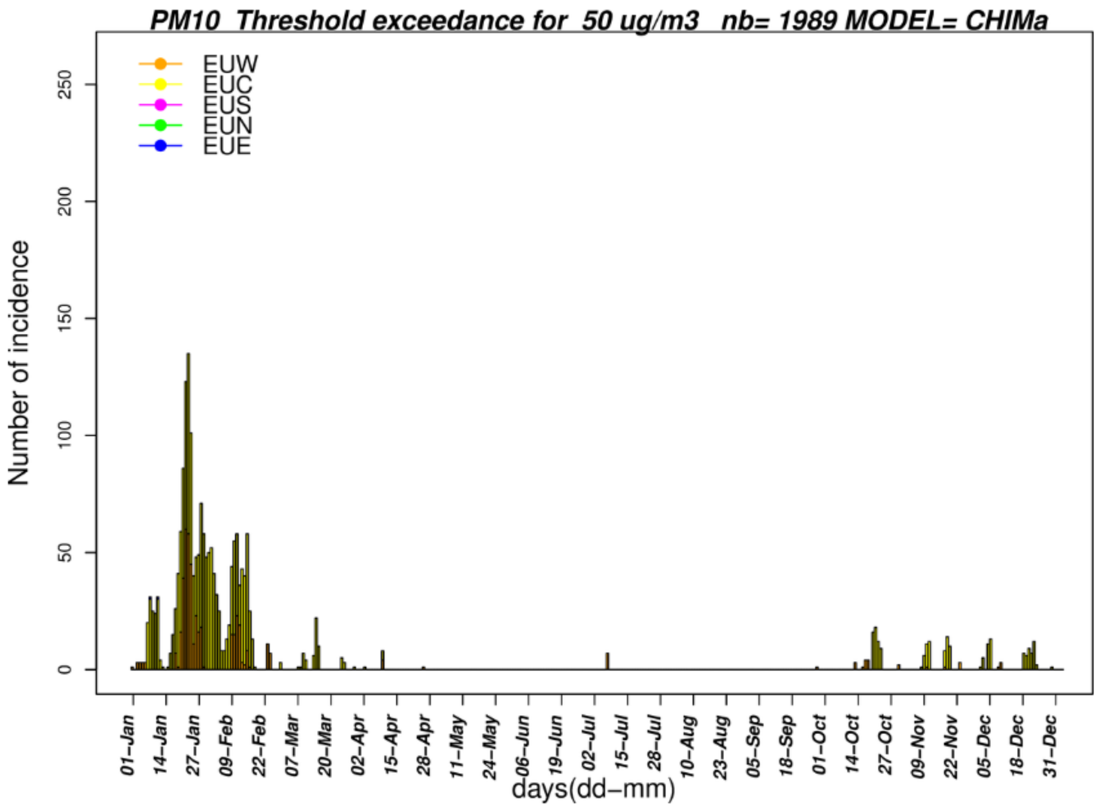
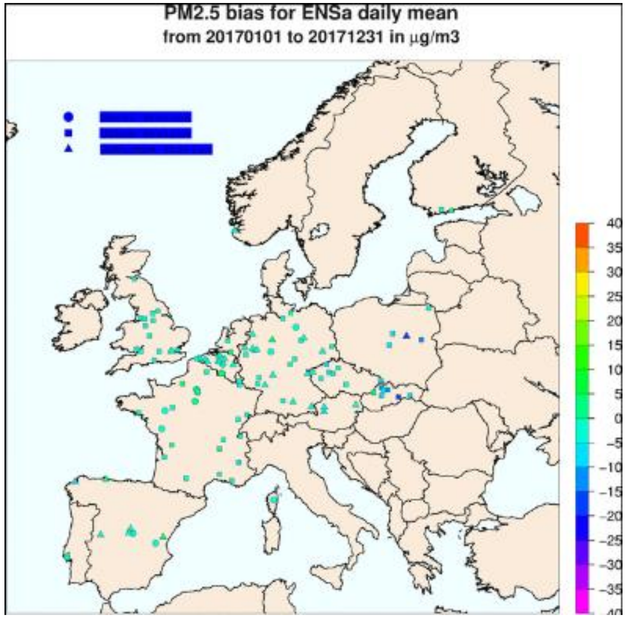


Figure 16 - Number of exceedances of daily limit value for PM_{10} in 2017 modelled by the CHIMERE interim re-analyses.

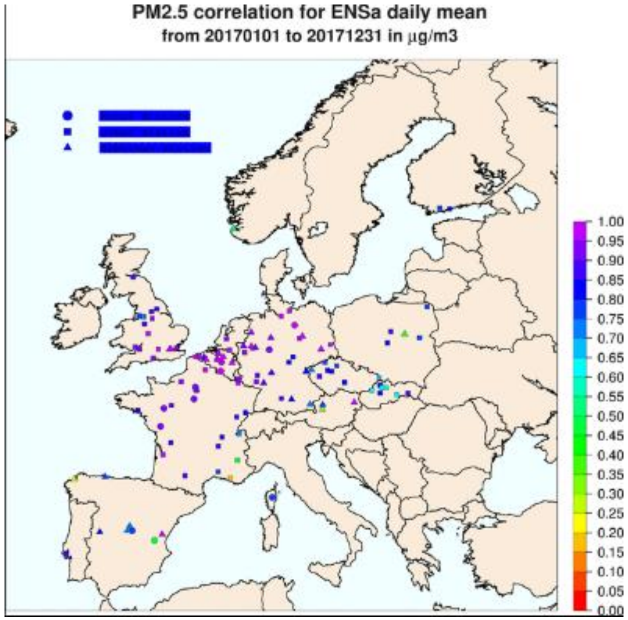
5. Performance Indicators for $PM_{2.5}$

The evaluation of models' performances for $PM_{2.5}$ was constrained by the low number of stations available, which is even lower than the number we had for the 2015 interim re-analyses verification. This limit is clearly highlighted considering the maps on 0. However, where there are some measurements, the results are rather good: the bias ranges from -5 to 10 $\mu g/m^3$, the correlation coefficient can exceed 0.8 except in some specific locations, and the RMSE stays generally below 10 $\mu g/m^3$ except in very few locations in Southern and Eastern Europe. Even if some concerns about the representativeness of these scores can be raised considering the low number of stations, we can consider those figures as encouraging. The values are remarkably homogeneous regarding the geographical location of the stations.

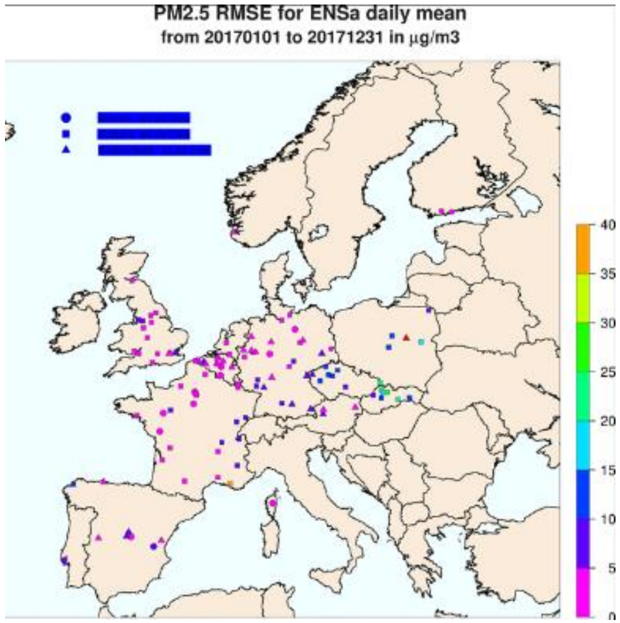
Those conclusions are confirmed by the analyses of the histograms by sub regions showing the correlation coefficient and the RMSE estimated for each model and for the various station typologies (rural, suburban and urban), respectively on 0, 0 and 0. Model responses are generally more homogeneous than for PM_{10} and the ENSEMBLE model is amongst the best models. The statistical scores are quite satisfactory with the correlation coefficient higher than 0.5 generally and the RMSE generally lower than 10 $\mu g/m^3$, except in Central Europe. However, those results are very difficult to interpret considering the low number of stations available for the evaluation.



(a)

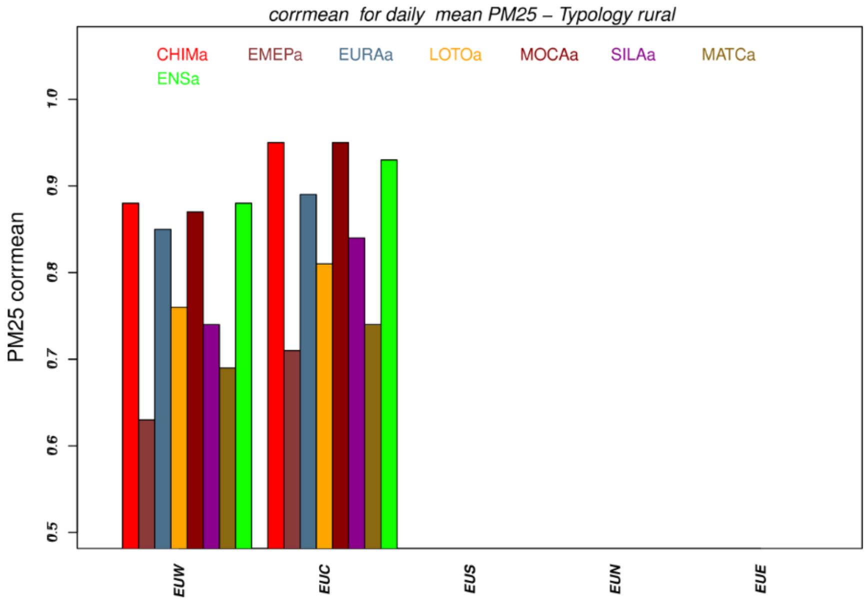


(b)

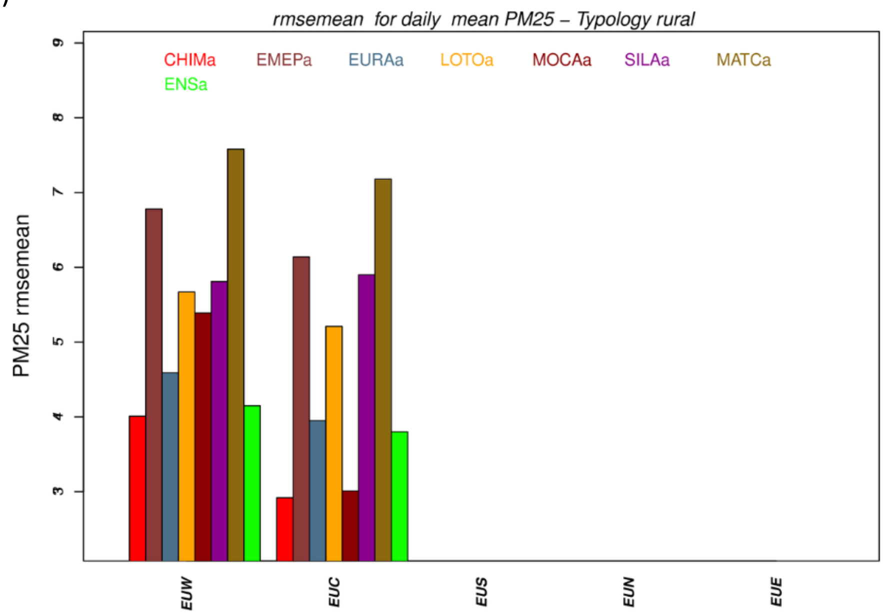


(c)

Figure 17 - Maps of Statistical scores of the ENSEMBLE interim re-analyses results against the observation validation dataset from the AQ e-reporting database for the $PM_{2.5}$ daily average over the year 2017 Bias (a), Correlation coefficient (b), Root mean square error (c).

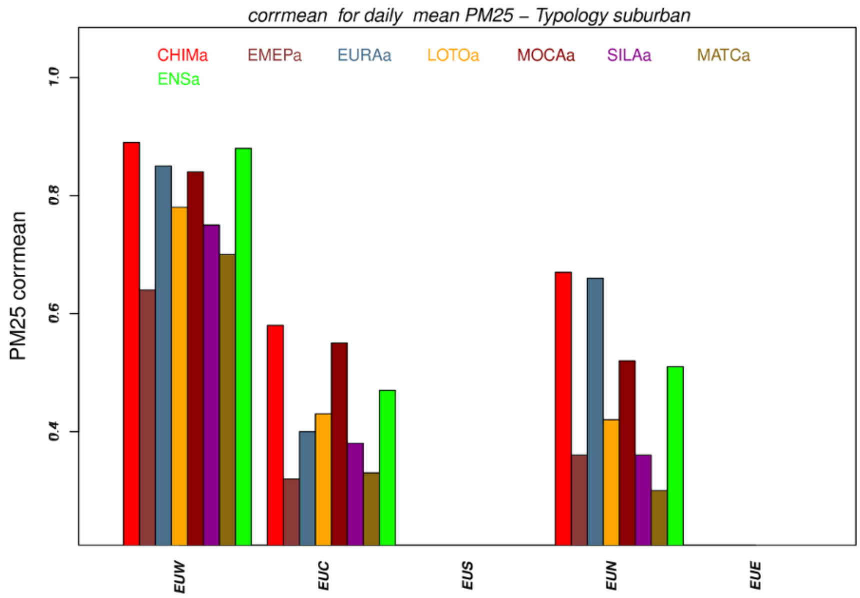


(a)

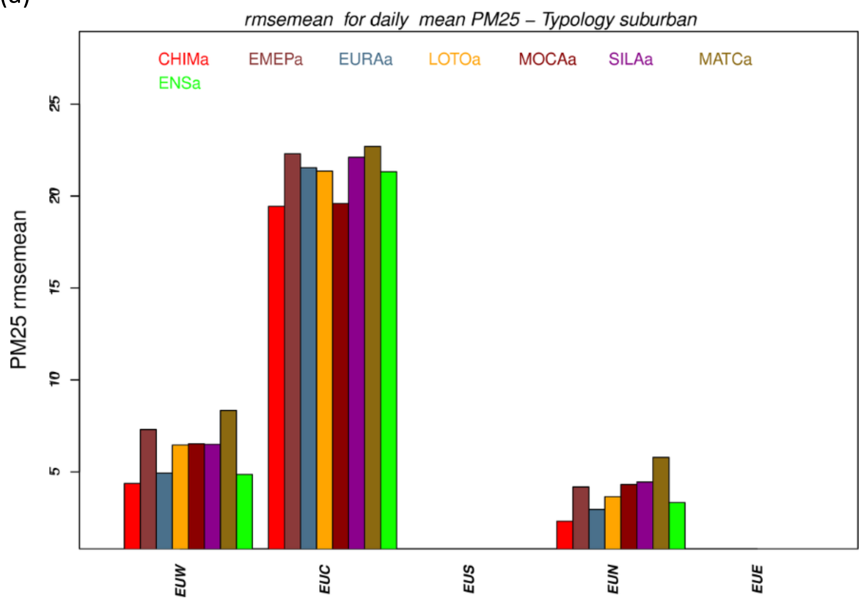


(b)

Figure 18 - CAMS Regional interim re-analyses for predicting $PM_{2.5}$ daily average over the year 2017 throughout European sub-regions Correlation coefficient (a), Root mean square error (b) at rural stations.

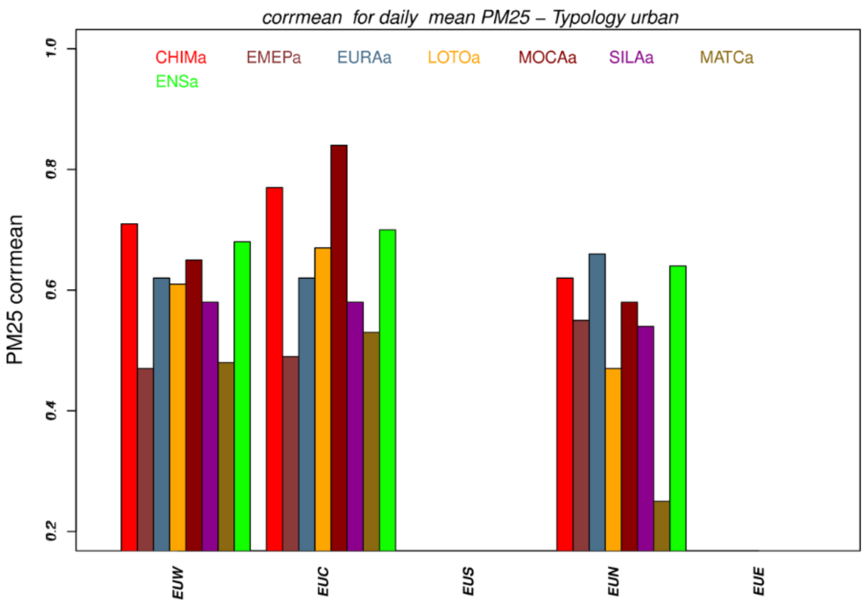


(a)

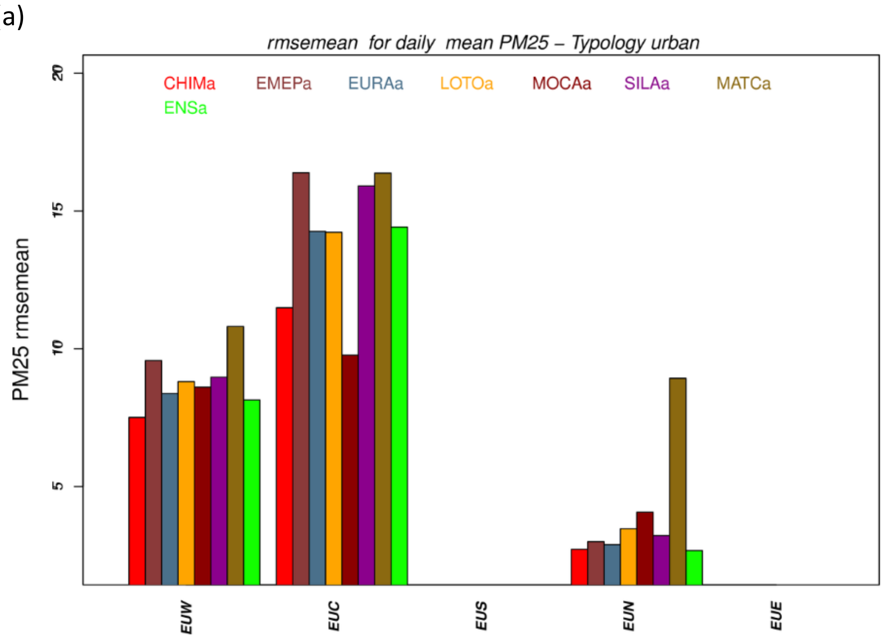


(b)

Figure 19 - CAMS Regional interim re-analyses for predicting $PM_{2.5}$ daily average over the year 2017 throughout European sub-regions Correlation coefficient (a), Root mean square error (b) at suburban stations.



(a)



(b)

Figure 20 - CAMS Regional interim re-analyses for predicting $PM_{2.5}$ daily average over the year 2017 throughout European sub-regions Correlation coefficient (a), Root mean square error (b) at urban stations.

Conclusion

The present report provides an analysis of the performances of the interim air quality re-analyses throughout Europe, produced by the CAMS Regional service for the year 2017. It focuses on ENSEMBLE air quality re-analyses resulting from the combination of seven well-validated and documented chemistry-transport models results. We call here "interim" re-analyses data assimilated fields of air pollutant concentrations based on up-to-date observation data. Because such data is quickly available after their production, the validation process it undergoes is not necessarily achieved and the data should be considered as "interim" data. Nevertheless, we found it interesting to elaborate interim re-analyses as first guess of air pollution patterns and levels that developed in Europe in 2016. Such information can be used to support Member States for the regulatory reporting duty on air quality (according to Directive 2008/50/EC). This is the reason why it is important to carefully evaluate the simulations against observations that are not used for the re-analyses production.

INERIS run this process and computed a number of performance indicators and scores for ozone, nitrogen dioxide, PM_{10} and $PM_{2.5}$ concentrations. They are presented in this report using maps, Taylor diagrams and histograms. The main conclusions arising from this analysis are the following:

- Too few up-to-date observation data was available to perform an extensive evaluation of interim re-analyses over the whole of Europe. Eastern and Southern European regions are not correctly covered, which is a pity since they correspond to areas where there are more uncertainties (especially because of emissions).
- In Western and Central Europe, where there are more stations for the evaluation of the models' performances, results are generally more representative and correct despite an overall degradation compared to the previous years. The reasons of this discrepancy have not been demonstrated, but lack of robustness of the process consisting in gathering interim observations can partially explain this situation. The European Environment Agency should strengthen quality assurance procedures in the coming years and more countries, especially in Southern Europe, are supposed to deliver up-to-date data, which will impact positively the interim re-analyses production process.
- Whatever the pollutant, the performances are always of lower quality than what can be achieved with the validated re-analyses process for which more stations are available and observation datasets are validated.
- For ozone, the ENSEMBLE model gives the best results. Ozone daily maxima are generally underestimated. The correlation coefficient ranges between 0.8 and 0.9 and the RMSE between 15 and 17 $\mu g/m^3$ at rural and suburban locations. At urban locations, these scores decrease significantly, the correlation coefficient ranging between 0.5 and 0.6 and the RMSE between 20 and 25 $\mu g/m^3$. This difference with respect to station typologies is more pronounced than in the previous assessments.
- Good model scores for simulating ozone in Western Europe are hampered by inferior performances at few stations in Eastern and Southern Europe.

- The situation of nitrogen dioxide re-analyses is much more worrying with performances significantly degraded compared to the previous years. Since such poor results are observed for all individual models, the limited number of available stations and perhaps the lack of quality of observations can explain this degradation of the performances for NO_2.
- For PM_{10}, even if the results are quite satisfactory considering the state of the art, the statistical scores remain lower than what is usually achieved with validated re-analyses, and what was observed in the previous years. Up-to-date PM_{10} observation datasets need to be more consolidated in the future.
- PM_{10} is the pollutant for which model responses range in the largest interval: the correlation coefficient from 0.4 to 0.9 and the RMSE from 16 to 9 $\mu g/m^3$ depending on the model and the station typology. The results are the best for suburban stations in Western and central Europe.
- Moreover, the evaluation demonstrates how the ENSEMBLE approach, based on a median average of involved models is not appropriate to simulate exceedances of threshold values. Only 17% (2 point less than for 2016) of the exceedances of the PM_{10} daily limit values were correctly caught by the ENSEMBLE (against about 80% for some individual models).
- Finally, despite the fact that only little $PM_{2.5}$ measurement data was available for the evaluation, the results obtained for this pollutant are promising. The individual models' responses are quite consistent and the ENSEMBLE median gives generally the best results. The correlation coefficient ranges from 0.4 to 0.9 according to the location and the station typology and the RMSE from 3 to 15 $\mu g/m^3$, which is very reasonable. Once again, the conclusions are limited by the low number of stations available in some geographical areas and should be consolidated and improved in future interim assessments when the up-to-date data gathering process at the EEA is strengthened.

Appendix: modelling set-up for the 2017 interim reanalyses

Modelling system: CHIMERE	
Horizontal resolution	0.15°x0.1°
Vertical resolution	Variable, 8 levels from the surface up to 500 hPa
Gas phase chemistry	MELCHIOR2, comprising 44 species and 120 reactions (Derognat, 2003)
Heterogeneous chemistry	NO_2, HNO_3, N_2O_5
Aerosol size distribution	9 bins from 10 nm to 40 µm
Inorganic aerosols	Primary particle material, nitrate, sulphate, ammonium
Secondary organic aerosols	Biogenic, anthropogenic
Aqueous phase chemistry	Sulphate
Dry deposition/sedimentation	Classical resistance approach
Mineral dust	Dusts are considered
Sea Salt	Inert sea salt
Boundary values	Values provided by CAMS global
Initial values	24h forecast from the day before
Anthropogenic emissions	MACC-TNO inventory 2011
Biogenic emissions	MEGAN
Assimilation system	
Assimilation method	Kriging-based analysis
Observations	Surface ozone and PM_{10}
Frequency of assimilation	Every hour over the day before
Meteorological driver	00:00 UTC operational IFS forecast for the day before

Modelling system: EMEP	
Horizontal resolution	0.25° x 0.125° lon-lat (native model grid; downscaling to 0.1° x 0.1° is done in post-processing)
Vertical resolution	20 layers (sigma) up to 100 hPa, with approximately 10 in the Planetary Boundary layer
Gas phase chemistry	Evolution of the 'EMEP scheme', comprising 70 species and 140 reactions (Andersson-Sköld and Simpson, 1999; Simpson et al. 2012)
Heterogeneous chemistry	MARS (Binkowski and Shankar, 1995), oxidation of NO_2 by ozone on aerosols (night and winter)
Aerosol size distribution	2 size fractions $PM_{2.5}$ and $PM_{10-2.5}$
Inorganic aerosols	Thermodynamic equilibrium for the H^+-NH_4^+-SO_4^{2-}-NO_3^--H_2O system
Secondary organic aerosols	EmChem09soa (Simpson et al., 2012, Bergström et al, 2012)
Aqueous phase chemistry	SO_2 oxidation by ozone and N_2O_2
Dry deposition/sedimentation	Resistance approach for gases and for aerosol, including non-stomatal deposition of NH_3
Mineral dust	Boundary conditions from global C-IFS are used
Sea Salt	Boundary conditions from global C-IFS are used
Boundary values	Boundary conditions from global C-IFS are used
Initial values	24h forecast from the day before
Anthropogenic emissions	TNO-MACC emission data for 2011
Biogenic emissions	Included
Assimilation system	
Assimilation method	Intermittent 3d-var
Observations	NO_2 columns from OMI, NO_2 and O_3 surface concentrations from in situ data distributed by Meteo-France (with option to assimilate SO_2 surface concentrations)
Frequency of assimilation	6-hourly
Meteorological driver	12:00 UTC operational IFS forecast (yesterday's)

Modelling system: EURAD-IM	
Horizontal resolution	15 km on a Lambert conformal projection
Vertical resolution	23 layers up to 100 hPa Lowest layer thickness about 35 m About 15 layers below 2 km
Gas phase chemistry	RACM-MIM
Heterogeneous chemistry	N_2O_5 hydrolysis: RH dependent parameterization
Aerosol size distribution	Three log-normal modes: two fine + one coarse, fixed standard deviation
Inorganic aerosols	Thermodynamic equilibrium for the H^+-NH_4^+-SO_4^{2-}-NO_3^--H_2O system
Secondary organic aerosols	Updated SORGAM module
Aqueous phase chemistry	10 gas/aqueous phase equilibria 5 irreversible S(IV) -> S(VI) transformations
Dry deposition/sedimentation	Resistance approach/size dependent sedimentation velocity
Mineral dust	DREAM model
Sea Salt	Included
Boundary values	C-IFS forecast
Initial values	3d-var analysis for the previous day
Anthropogenic emissions	TNO MACC-III (2011) inventory with 0.125° x 0.0625° resolution
Biogenic emissions	MEGAN V2.10 (Guenther et. al, 2012) GFAS wild fire emission data
Assimilation system	
Assimilation method	Intermittent 3d-var
Observations	NRT surface in situ data distributed by Meteo-France, NO_2 column retrievals from AURA/OMI and METOP/GOME-2, MOPITT CO profiles, IASI CO data
Frequency of assimilation	Hourly
Meteorological driver	WRF forced by the operational IFS analysis for the previous day

Modelling system: LOTOS-EUROS	
Horizontal resolution	0.25° (longitude) x 0.125° (latitude)
Vertical resolution	4 layers, top at 3.5 km above sea level
Gas phase chemistry	Modified version of the original CBM-IV
Heterogeneous chemistry	N_2O_5 hydrolysis
Aerosol size distribution	Bulk approach: $PM_{2.5}$ and $PM_{2.5-10}$
Inorganic aerosols	ISORROPIA-2
Secondary organic aerosols	Not included in this version
Aqueous phase chemistry	Linearized
Dry deposition/sedimentation	Resistance approach, following Erisman et al. (1994). Zhang (2001) deposition scheme is used for particles, explicitly including particle size and sedimentation
Mineral dust	Emissions after Marticorena &Bergametti (1995) with soil moisture inhibition as described by Fécan et al (1999).
Sea Salt	Parameterized emissions based on wind speed at 10m following (Monahan et al., 1986) and sea-surface temperature (Martensson et al., 2003).
Boundary values	C-IFS forecast (lateral and top).
Initial values	24h forecast from the day before
Anthropogenic emissions	TNO-MACC-III (2011) inventory
Biogenic emissions	Following Guenther et al. (1993) using 115 tree types over Europe
Assimilation system	
Assimilation method	Ensemble Kalman filter
Observations	Ensemble Kalman filter
Frequency of assimilation	In-situ surface observations (O_3, NO_2, PM_{10}, $PM_{2.5}$) distributed by Meteo-France and OMI NO_2 columns
Meteorological driver	Hourly, performed once a day for the previous day

Modelling system: MATCH	
Horizontal resolution	0.2° (since 10 January 2011)
Vertical resolution	26 levels (using reduction of IFS levels)
Gas phase chemistry	Based on EMEP (Simpson et al., 2012), with modified isoprene chemistry (Carter, 1996; Langner et al., 1998)
Heterogeneous chemistry	HNO_3-formation from N_2O_5; equilibrium reactions for NH_3-HNO_3
Aerosol size distribution	2bins: 0.01–2.5, 2.5–10 µm
Inorganic aerosols	Sulphate, Nitrate, Ammonium
Secondary organic aerosols	Not in operational version yet (VBS-scheme based on Bergström et al., 2012, is implemented)
Aqueous phase chemistry	SO_2 oxidation by H_2O_2 and O_3
Dry deposition/sedimentation	Resistance approach/size dependent sedimentation velocity
Mineral dust	Yes
Sea Salt	Yes
Boundary values	C-IFS forecast for the day before (zero boundaries for sea-salt)
Initial values	MATCH 24h forecasts from the day before
Anthropogenic emissions	TNO MACC-III emission inventory for the year 2011 (Kuenen et al., 2014; Denier van der Gon et al., 2015a), with 1/8° × 1/16° resolution
Biogenic emissions	Isoprene (Simpson, 1995; updated biogenic emissions of isoprene and monoterpenes, based on Simpson et al., 2012, implemented but not yet in operational version)
Assimilation system	
Assimilation method	3Dvar
Observations	NRT in-situ observations (O_3, NO_2, CO, SO_2, PM_{10}, $PM_{2.5}$) distributed by Meteo-France
Frequency of assimilation	Hourly, performed once a day for the previous day
Meteorological driver	IFS forecast and analyses 00Z for the same day (0.2°, 78 levels)

Modelling system: MOCAGE	
Horizontal resolution	0.2° regular lat-lon grid
Vertical resolution	47 layers up to 5 hPa Lowest layer thickness about 40 m About 8 layers below 2 km
Gas phase chemistry	RACM (tropospheric) and REPROBUS (stratospheric)
Heterogeneous chemistry	Only reactions on Polar Stratospheric Clouds (stratosphere) yet
Aerosol size distribution	Bins
Inorganic aerosols	Included: ISORROPIA module (Guth et al, 2016)
Secondary organic aerosols	Not implemented in current CAMS version
Aqueous phase chemistry	Aqueous reactions for sulphate production
Dry deposition/sedimentation	Resistance approach (Michou et al., 2004) for gases, (Nho-kim et al., 2005) for aerosol
Mineral dust	Included: see evaluation by Sic et al. (2014)
Sea Salt	Included: see evaluation by Sic et al. (2014)
Boundary values	Global CAMS chemical and aerosol fields
Initial values	24h forecast from the day before
Anthropogenic emissions	TNO (2011) inventory binned at 0.2° resolution
Biogenic emissions	Fixed monthly biogenic emission, based upon Simpson approach.
Assimilation system	
Assimilation method	3d-var
Observations	O_3, NO_2 and PM_{10} in situ data provided by INERIS are assimilated
Frequency of assimilation	Hourly
Meteorological driver	00:00 UTC operational IFS forecast

Modelling system: SILAM	
Horizontal resolution	0.1° regular lat-lon grid
Vertical resolution	69 layers for meteorological pre-processor (IFS hybrid levels 69 to 137, covering the troposphere), 9 layers for chemistry and vertical sub-grid-scale mixing calculations
Gas phase chemistry	CBM-4 gas-phase transformation, inorganic chemistry scheme with input to heterogeneous transformations (Sofiev, 2000)
Heterogeneous chemistry	Sofiev (2000)
Aerosol size distribution	Bins. Varies: for anthropogenic source, follows the emission of $PM_{2.5-10}$ split, for sea salt uses 5 bins from 10nm up to 30µm, dust is split into 4 bins from 10nm up to 30 µm
Inorganic aerosols	SO_4, NO_3, NH_4, Primary BC, OC, sea salt, desert dust
Secondary organic aerosols	Volatility Basis-Set. Being validated, not used in operational runs
Aqueous phase chemistry	SO_2 oxidation, nitrate formation (Sofiev, 2000)
Dry deposition/sedimentation	Resistance approach (Wesely et al., 1989) for gases, (Kouznetsov & Sofiev, 2012) for aerosol
Mineral dust	Taken from the C-IFS boundary conditions
Sea Salt	Updated source term Sofiev et al (2011)
Boundary values	C-IFS values for all available species, except for sea salt, which is taken from SILAM global forecasts
Initial values	24h forecast from the day before
Anthropogenic emissions	MACC-2011 inventory binned at 0.1° resolution
Biogenic emissions	Dynamic biogenic emissions, based upon Poupkou et al. (2010)
Assimilation system	
Assimilation method	Operational intermittent 3d-var for analysis; 4dvar for pollen reanalysis
Observations	In-situ surface data O_3, NO_2, $PM_{2.5}$ operational; and vertically integrated columns in research mode (NO_2, AOD)
Frequency of assimilation	Hourly
Meteorological driver	00:00 UTC operational IFS forecasts up to +24h

Copernicus Atmosphere Monitoring Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-atmosphere.eu