CAMS Service Evolution

D6.4 Uncertainty in global CAMS products arising from meteorological forecast vs emissions/deposition processes

Due date of deliverable	30 June 2025
Submission date	30 June 2025
File Name	CAMEO-D6-4-V1.1
Work Package /Task	WP6 T6.4
Organisation Responsible of Deliverable	HYGEOS
Author name(s)	Samuel Remy and Thierry Elias
Revision number	V1.1
Status	Issued
Dissemination Level	Public

Funded by the European Union

The CAMEO project (grant agreement No 101082125) is funded by the European Union.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

The representation of atmospheric composition through numerical modelling is fraught with uncertainties. They arise from various causes: modelling errors originate from representing processes in a simplified manner, partially, or not at all, as not all chemical and physical processes occurring in the atmosphere are well known or described. Emissions from anthropogenic or biogenic sources are hard to estimate and can be another source of uncertainty. Finally, many atmospheric composition processes are largely influenced by meteorology: most reaction rates depend on temperature, gas-particle partitioning between gaseous species and secondary inorganic aerosols depend on temperature and relative humidity; emissions of sea-salt and desert dust aerosols depend primarily on wind speed, deposition depends on precipitation and wind speed. Uncertainties in atmospheric composition modelling thus also result from uncertainties of simulated meteorological fields.

Copernicus Atmosphere Monitoring Service (CAMS) products are subjected from all these sources of error and uncertainty. In this deliverable report, we apply an ensemble methodology, inspired by and built from the ECMWF meteorological ensemble, to estimate the uncertainties of CAMS products, using a configuration close to that of the currently operational cycle 49R1. Random perturbations have been applied to different components of IFS-COMPO, the atmospheric composition modelling system applied to produce global CAMS forecasts, and its inputs. The uncertainties introduced in this way are then propagated in the perturbed air quality forecasts, and the ensemblist approach allows to quantify the mean uncertainty of the simulated fields through the ensemble spread. The resulting uncertainties are presented in this report for a selection of products and time. They also provide a measure of how sensitive CAMS products are to errors from different causes (meteorology, emissions, modelling errors). These values should be used with care: the presented uncertainties depend on a lot of assumptions made about the original uncertainties. For the temporal uncertainties of emissions, input from work package 5 was used; for meteorology, we rely on the existing perturbation framework of the ECMWF meteorological ensemble, while for modelling uncertainties, a set of assumptions were made. Also, the uncertainties presented here are for 2021; however, it was shown that the uncertainties of CAMS products depend a lot on the meteorological and atmospheric composition of the day, which limits the validity and usefulness of monthly values.

For a realistic assessment of uncertainties, a well-balanced ensemble is needed. Well-balanced in this case means an ensemble that is able to capture the observed variability while not producing forecasts that are outside the observational space (over dispersion). A number of evaluation tools have been developed in order to produce metrics to show how well-balanced the ensemble is. Those metrics are commonly used and presented for meteorological ensembles; however, their use in atmospheric composition is quite new. We'll present in detail these tools and how they were used. Most of the ensemblist diagnostics points to an under-dispersion or too small spread of the ensemble simulations that have been performed. In some cases (PM2.5 and surface ozone), this appears to be largely a consequence of model biases, which makes it harder for the ensemble to capture the observational variability. This is also a sign that perhaps the perturbations applied are too small or don't explore enough degree of freedom, such as correlated perturbations, or perturbations varying with forecast time.

The propagation and relative importance of the different sources of uncertainties vary a lot depending on the species considered. In general, anthropogenic emissions was found to have a relatively smaller impact on the uncertainty of AOD and surface ozone than other factors, while for PM, the impact of emissions is quite significant. For all parameters, meteorological and model uncertainties are the highest source of uncertainty.

Finally, the work presented in this deliverable can be seen as an early prototype of a global atmospheric composition ensemble. While the computing costs of such an ensemble approach are very large, the expected benefits are also significant in terms of the forecast skills, as shown by the regional CAMS ensemble. Such an ensemble could also provide an online estimate of atmospheric composition forecast uncertainties, as well as extreme scenarios.

Table of Contents

2 Introduction

2.1 Background

Monitoring the composition of the atmosphere is a key objective of the European Union's flagship Space programme Copernicus, with the Copernicus Atmosphere Monitoring Service (CAMS) providing free and continuous data and information on atmospheric composition.

The CAMS Service Evolution (CAMEO) project will enhance the quality and efficiency of the CAMS service and help CAMS to better respond to policy needs such as air pollution and greenhouse gases monitoring, the fulfilment of sustainable development goals, and sustainable and clean energy.

CAMEO will help prepare CAMS for the uptake of forthcoming satellite data, including Sentinel-4, -5 and 3MI, and advance the aerosol and trace gas data assimilation methods and inversion capacity of the global and regional CAMS production systems.

CAMEO will develop methods to provide uncertainty information about CAMS products, in particular for emissions, policy, solar radiation and deposition products in response to prominent requests from current CAMS users.

CAMEO will contribute to the medium- to long-term evolution of the CAMS production systems and products.

The transfer of developments from CAMEO into subsequent improvements of CAMS operational service elements is a main driver for the project and is the main pathway to impact for CAMEO.

The CAMEO consortium, led by ECMWF, the entity entrusted to operate CAMS, includes several CAMS partners thus allowing CAMEO developments to be carried out directly within the CAMS production systems and facilitating the transition of CAMEO results to future upgrades of the CAMS service.

This will maximise the impact and outcomes of CAMEO as it can make full use of the existing CAMS infrastructure for data sharing, data delivery and communication, thus supporting policymakers, business and citizens with enhanced atmospheric environmental information.

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

In this deliverable an atmospheric composition ensemble has been developed, evaluated and used to provide quantitative estimates of the uncertainties of CAMS products.

2.2.2 Work performed in this deliverable

In this deliverable the work as planned in the Description of Action (DoA, WP6 T6.4.1, 6.4.2 and 6.4.3) was performed.

2.2.3 Deviations and counter measures

No deviations have been encountered.

2.2.4 CAMEO Project Partners:

ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
Met Norway	METEOROLOGISK INSTITUTT
BSC	BARCELONA SUPERCOMPUTING CENTER-CENTRO NACIONAL DE SUPERCOMPUTACION
KNMI	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMi
SMHI	SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT
BIRA-IASB	INSTITUT ROYAL D'AERONOMIE SPATIALEDE BELGIQUE
HYGEOS	HYGEOS SARL
FMI	ILMATIETEEN LAITOS
DLR	DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV
ARMINES	ASSOCIATION POUR LA RECHERCHE ET LE DEVELOPPEMENT DES METHODES ET PROCESSUS INDUSTRIELS
CNRS	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS
GRASP-SAS	GENERALIZED RETRIEVAL OF ATMOSPHERE AND SURFACE PROPERTIES EN ABREGE GRASP
CU	UNIVERZITA KARLOVA
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
MF	METEO-FRANCE
TNO	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO
INERIS	INSTITUT NATIONAL DE L ENVIRONNEMENT INDUSTRIEL ET DES RISQUES - INERIS
IOS-PIB	INSTYTUT OCHRONY SRODOWISKA - PANSTWOWY INSTYTUT BADAWCZY
FZJ	FORSCHUNGSZENTRUM JULICH GMBH
AU	AARHUS UNIVERSITET
ENEA	AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

3 Implementation of a global atmospheric composition ensemble forecasting system

ECMWF is producing ensemble weather forecasts since 1992. An ensemble weather forecast is a set of forecasts that present the range of future weather possibilities. Multiple simulations are run, each with a slight variation of its initial conditions and with slightly perturbed weather models (Figure 1). These variations represent the inevitable uncertainty in the initial conditions and approximations in the models. They produce a range of possible weather conditions. Here we would like to extend the existing weather ensemble to atmospheric composition, by using the IFS-COMPO (Integrated Forecasting System with atmospheric composition extensions) instead of IFS in the existing ensemble architecture, and by introducing optional perturbations of initial conditions, emissions and model processes that are specific to atmospheric composition, and aim to represent the impact of uncertainties not related to meteorology.

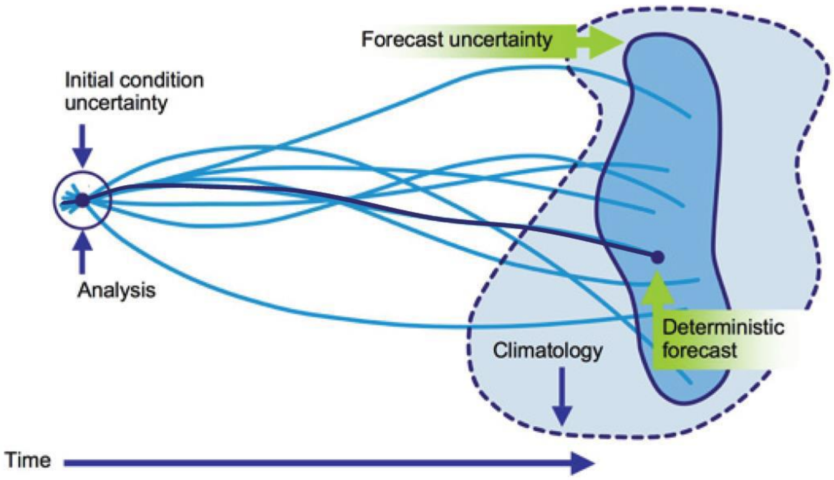

Figure 1: Schematic showing the principles of ensemble forecast.

Several atmospheric composition ensembles are currently operationally active; some of them multi model, and some of them with a single model, namely:

- The CAMS regional ensemble with 11 models over Europe,
- The MERRA2-AMIP dataset, which consists of a 10-member ensemble of free-running simulations with the GEOS atmospheric model. The model version, along with imposed boundary conditions, is configured identically to the GEOS model used in the MERRA-2 reanalysis dataset. The MERRA2-AMIP dataset supplements the MERRA-2 reanalysis, providing a suite of model simulations that are identical to the analyses, with the exception that they do not ingest the observations. The 10-member ensemble provides information on the natural variations inherent in any free-running model simulation.
- The ICAP (International Cooperative for Aerosol Prediction) Multi Model Ensemble (ICAP-MME, Sessions et al, 2015, Xian et al 2019), which includes 9 operational global atmospheric composition forecasting systems including CAMS and GEOS-5, NAAPS, MASINGAR, NGAC, MOCAGE, SILAM, and two dust-only global models: BSC MONARCH (former BSC-CTM) and UKMO Unified Model.
- The WMO SDS-WAS dust ensemble (https://dust.aemet.es/) which provides in near-real time dust related products from 13 models: CAMS, BSC-MONARCH, DREAM8,

NASA-GEOS, MetOffice-UM, NCEP-GEFS, EMA-RegCM4, SILAM, LOTOS-EUROS, ICON-ART, NOA-WRF-CHEM, ZAMG-WRF-CHEM and MOCAGE.

The evaluations carried out of the ensemble products from the CAMS regional ensemble, the SDS-WAS and the ICAP-MME ensembles all show that products from the multi model ensemble outperform all the deterministic forecasts that compose the ensemble. In the work shown here, multi model ensemble is impossible to implement as we use only the IFS-COMPO atmospheric composition forecasting system. However, we ambition to emulate a multi model approach by adding stochastics perturbations of relevant model parameterizations.

3.1 Description

The IFS-COMPO ensemble are built on the NWP ensemble in a cycle 48R1 configuration: they include a single control or unperturbed run, and 50 perturbed members. The ensemble uses an analysis field as initial conditions, so including aerosol and chemistry data assimilation. The perturbations applied to the meteorological fields are those of the IFS meteorological ensemble, described in the cycle 48R1 documentation (https://www.ecmwf.int/en/elibrary/81371-ifs-documentation-cy48r1-part-v-ensemble-prediction-system). They consist of perturbations of the meteorological initial conditions, which are provided from an ensemble of data assimilations (EDA), and perturbations constructed from the leading singular vectors. Model uncertainties are represented in the meteorological ensemble with the Stochastically Perturbed Parameterization Tendencies scheme (SPPT). It simulates the effect on forecast uncertainty of random model errors due to the parametrized physical processes. The scheme perturbs the meteorological tendencies by terms that are given by a random pattern, times the net parameterized physics tendencies, less the diagnosed clear-sky heating rate. The random pattern varies horizontally and with time; each ensemble member uses a different realisation of the random pattern. On top of the meteorological perturbations, the following set of perturbations have been implemented:

- Perturbations of the initial conditions of all aerosol and selected (carbon monoxide, ozone, sulphur dioxide and nitrogen dioxide) chemistry tracers in the troposphere
- Perturbation of the inputs of the dust emission scheme: dust source function (DSF), sand/silt/clay fraction of the soil used for dust emissions and assumed size distribution at emissions
- Perturbation of the anthropogenic emission inputs of IFS-COMPO
- Perturbation of atmospheric composition specific model parameterizations:
    - Computation of dry deposition velocity
    - Rates of chemical reactions
    - Photolysis rates
    - Wet deposition and re-evaporation rates
    - Production rate of sulphate and nitrate aerosols
    - Production rate of secondary organic aerosols

Ensemble simulations can be carried out using all or a selection of the perturbations presented above. Also, we developed the possibility to run ensemble simulations without perturbations of meteorological initial conditions and parameterizations, so as to allow to assess the impact of the uncertainties of atmospheric composition specific inputs and processes. It should be noted that **perturbations of different input/parameterizations are assumed to be uncorrelated, and perturbations are supposed to be constant for all forecasts time**, which is a strong (and probably wrong) assumption. The perturbations themselves consist of 2D fields that represent correlated gaussian noise, computed using a given correlation length and standard deviation. They are computed on-the fly using a python script that is called in the IFS-COMPO scripts, and are applied to each relevant model input (emissions, initial conditions, etc.). For the model parameterizations, two such stochastic perturbations are computed, loaded into IFS-COMPO, and used in the IFS-COMPO Fortran code to scale the output of the parameterizations listed above. Figure 2 shows an example of such a perturbation.

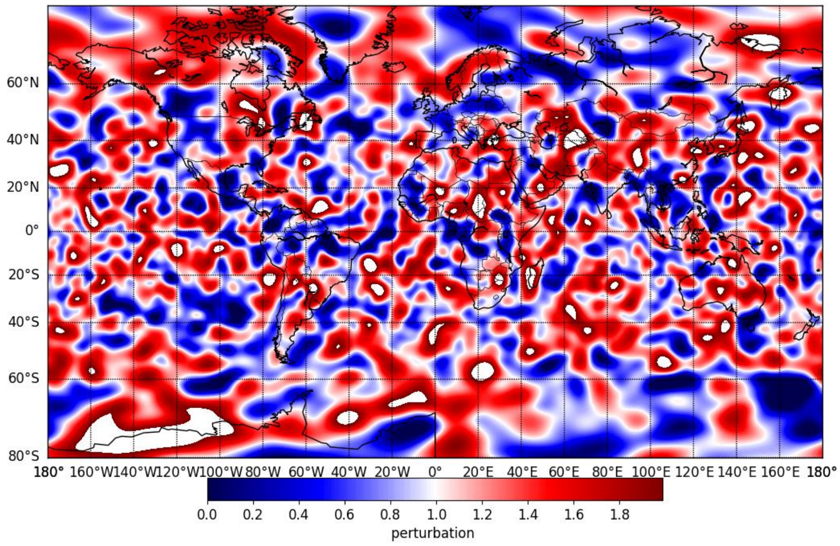

Figure 2: Example of a perturbation scaling factor generated with a correlation length of 500km and a standard deviation of 0.5.

Table 1 lists the specifics of the perturbations applied to the atmospheric composition initial conditions, the inputs of the dust emission scheme and the model parameterizations. These values have been obtained by evaluating how dispersive the ensemble is, depending on the specifics of the input perturbations. For the initial conditions, it was found that higher standard deviations could make ensemble simulations over dispersive for long lived species, such as carbon monoxide and ozone, and much lower values have been used for later simulations.

Table 1: Standard deviation and correlation length of the random perturbations applied to different inputs and parameterizations of the IFS-COMPO ensemble

Perturbed field	Standard deviation	Correlation length
Aerosol initial conditions	0.25	500
CO initial conditions	0.04	500
O_3 initial conditions	0.07	500
NO_2 initial conditions	0.125	500
SO_2 initial conditions	0.19	500
Model parameterizations	0.5	500
Inputs of the dust emission scheme	0.5	500

Simulations have been carried out with correlation lengths varying between 250 and 2000km. The impact on the ensemble spread, averaged monthly and over the globe, was found to be very small. Those runs led to the selection of correlation lengths presented in Table 1. A special treatment has been implemented for the perturbation of emission input, detailed in the next subsection.

3.1.1 Perturbation of emission input

For emissions, sectoral perturbations are used, i.e., the perturbations are the same for the emissions of all the species of a given emission sector.

Work package 5 delivered at the end of May 2024 estimates of the temporal uncertainties of anthropogenic emissions. For more details on how these estimates were computed, please refer to deliverable D5.1. These consist of global values representing the standard deviation for each hour for the diurnal cycle uncertainties, and gridded monthly standard deviation to represent the uncertainty of the seasonal cycle. These two sets of standard deviations are provided for each emission sector. An example of the monthly standard deviation is shown in Figure 3 for the energy sector.

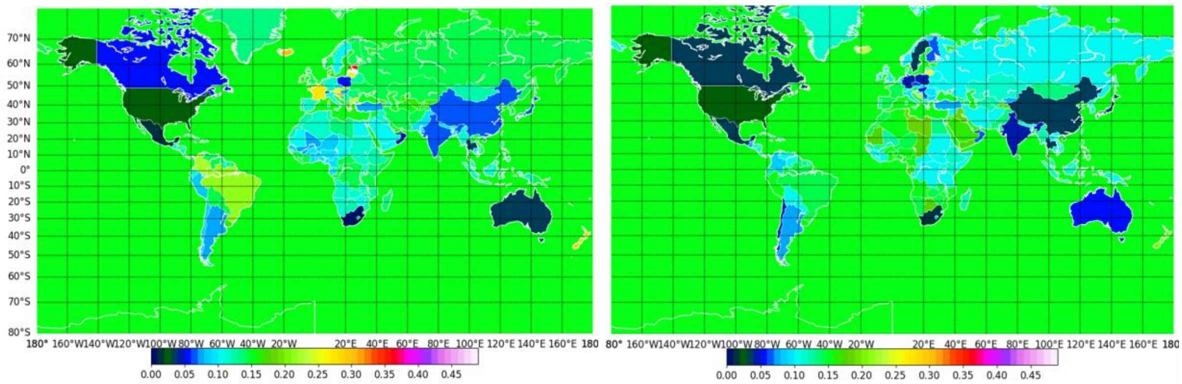

Figure 3: Standard deviation in emissions from the energy sector, January (left) and July (right).

Following the advice of the WP5 leader, the maximum value of the uncertainty standard deviation of the diurnal cycle was used. The values are summarized in Table 2.

Table 2: Standard deviation of the uncertainty of anthropogenic emissions, per sector.

Sector	Standard deviation
Ind	0.15
Res	0.487
Agl	0.451
Awb	1.233
Ags	0.451
Ene	0.063
Tro	0.397
Fef	0.25
Slv	0.375
Tnr	0.469
Swd	0.25

Shp	0.25

These values are used, with a correlation length of 500km, to generate perturbation files. On top of this, the gridded monthly values are used to modulate the perturbations, by scaling the perturbation over each grid cell by the gridded monthly value divided by the spatial average of the gridded monthly value. An example is shown in Figure 4, for the energy sector in January. The higher perturbations over Iceland and Estonia/Latvia correspond to areas where the gridded values that represent the seasonal uncertainty spread for this month are the highest, as shown in Figure 3.

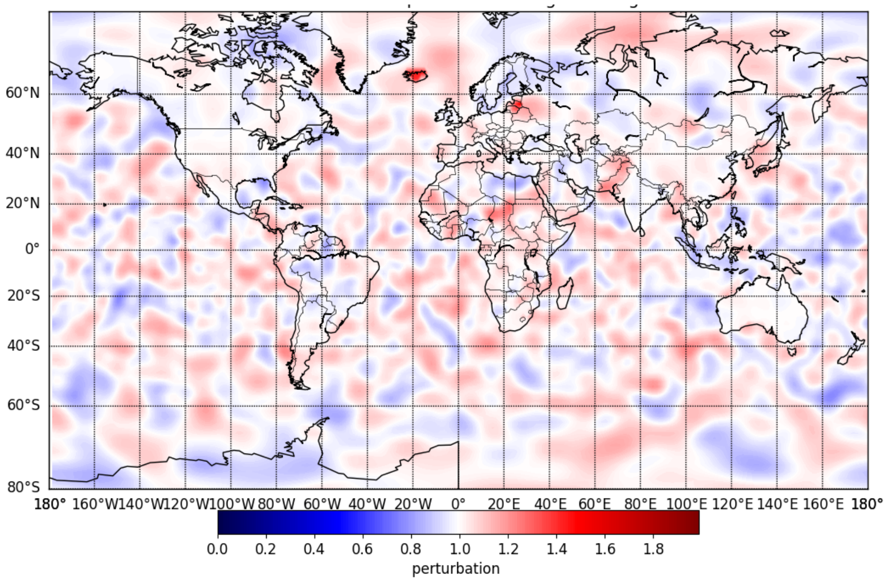

Figure 4: Example of a perturbation scaling factor for emissions from the energy sector in January.

The perturbations of emissions computed following this methodology only represent the temporal uncertainties as provided by WP5. They don't include other aspects of emissions uncertainties, relating to activity data used to compute the emissions, conversion factors etc. is not taken into account, which means that it is very possible that the emissions uncertainties, as represented by the estimated perturbations are underestimated.

3.2 Experiments

A number of experiments have been carried out in order to evaluate and adjust the perturbations applied to initial conditions, emissions and model parameterizations, which are not shown here. The ensemble simulations use initial conditions (before perturbation) from analysis simulations from cycle 48R1 for the year 2021. They use a cycle 48R1 branch, which includes modelling updates for cycle 49R1. As such, the uncertainties shown are representative of the uncertainties of the official cycle 49R1 products, but for the year 2021.

The experiment specifics are the following:

- T_L255L137 resolution (80km grid cell)
- 120h maximum forecast time, output every 12 hours
- Simulated year 2021
- Use of CAMS_GLOB_BIOv3.1 and CAMS_GLOB_ANT v5.3 emissions

Because of the high computing costs of the IFS-COMPO ensemble experiments, most of them didn't complete the year 2021. We chose to show results for experiments that perturb only meteorological fields, only anthropogenic emissions, only model parameterizations, and all uncertainties combined. The EMI_OLD and EMI_OLD2 are experiments that tested different perturbations of the anthropogenic emission inputs, before the uncertainty estimates of WP5 became available. The experiments are listed in Table 3.

Table 3: IFS-COMPO ensemble experiments shown in this report

Experiment	Characteristics
MET	Perturbations of meteorological initial conditions and processes
EMI	Perturbations of anthropogenic emissions input only, using WP5 input
EMI_OLD	Perturbations of anthropogenic emissions input only, using 0.25 standard deviation and 500km length scale.
EMI_OLD2	Perturbations of anthropogenic emissions input only, using 0.5 standard deviation and 500km length scale.
MODEL	Perturbation of atmospheric composition model parameterizations only
ALL	All perturbation applied: • Meteorological initial conditions and processes • Atmospheric composition initial conditions • Atmospheric composition model parameterizations • Inputs of the dust emission scheme • Anthropogenic emission input using WP5 input
ALL_NOMODEL	All perturbations applied except model parameterizations
INI	Perturbation of atmospheric composition initial conditions only

Experiments will be shown in the following sections based on their relevance, not all experiments will be included. In particular, the estimated uncertainty will be shown for MET, EMI, MODEL and ALL.

4 Evaluation of the ensemble forecasts

The raw ensemblist forecasts look like the stamp plot shown in Figure 5, which shows how a situation with high PM10 from a desert dust plume over South of France and Northern Italy is simulated by an IFS-COMPO ensemble simulation. These plots include a lot of information; however, it can be hard to compare simulations from one ensemble against another, or even two different simulations of the same ensemble. To do this, metrics that incorporate information from all ensemble members are needed, such as the ensemble standard deviation or spread and ensemble median. Examples of these two quantities are shown in Figure 6 for surface ozone from two IFS-COMPO ensembles: MET and ALL, for a single day that saw high ozone concentrations over most of Europe. The ensemble median is very similar for the two experiments, with an area of values above 100 µg/m^3 that covers most of Western Europe. The spread of MET is relatively small, between 2 and 5 µg/m^3 in general, which shows that the meteorological perturbations have relatively little impact on simulated ozone on that day. The spread is significantly higher for ALL, with values generally between 10 and 15 µg/m^3 over most of Europe.

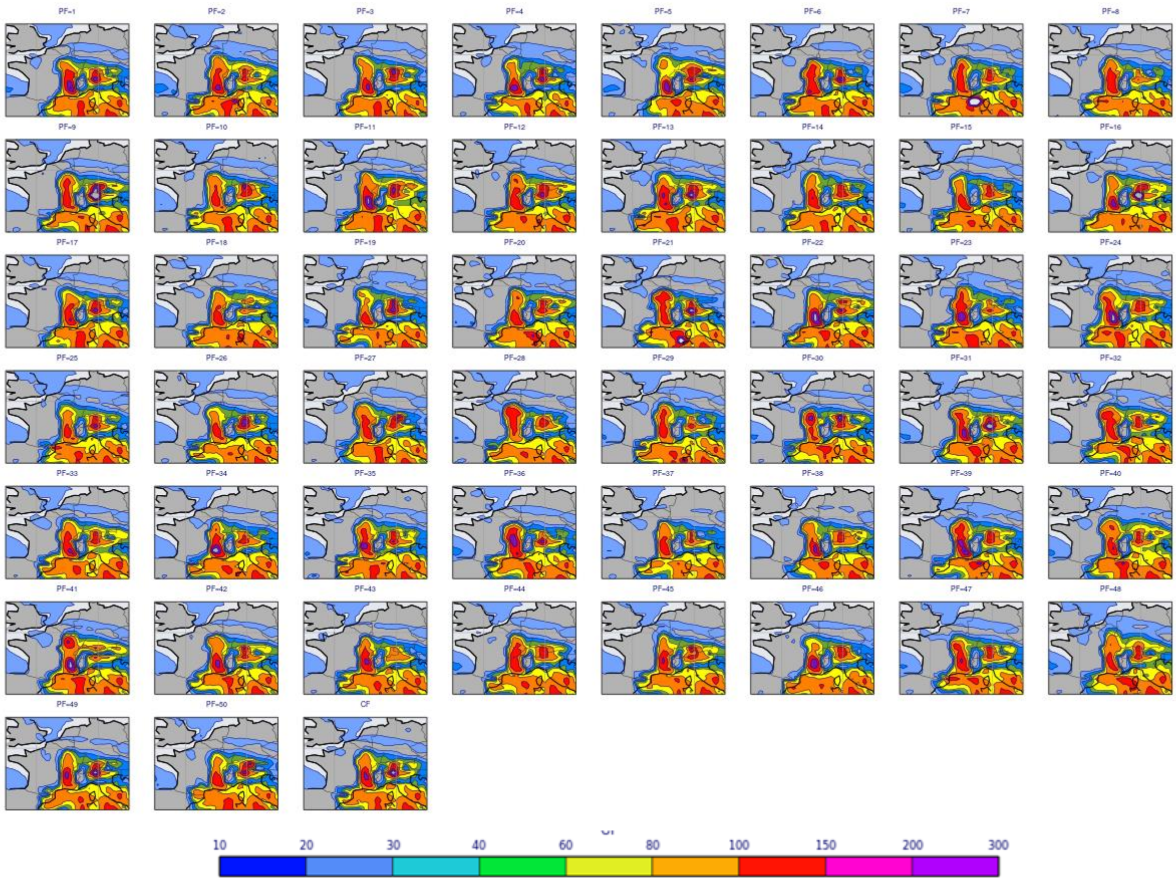

Figure 5: Simulated PM10 at 24h forecast time over Western Europe, simulation starting on 6/2/2021, IFS-COMPO ensemble perturbing meteorology only.

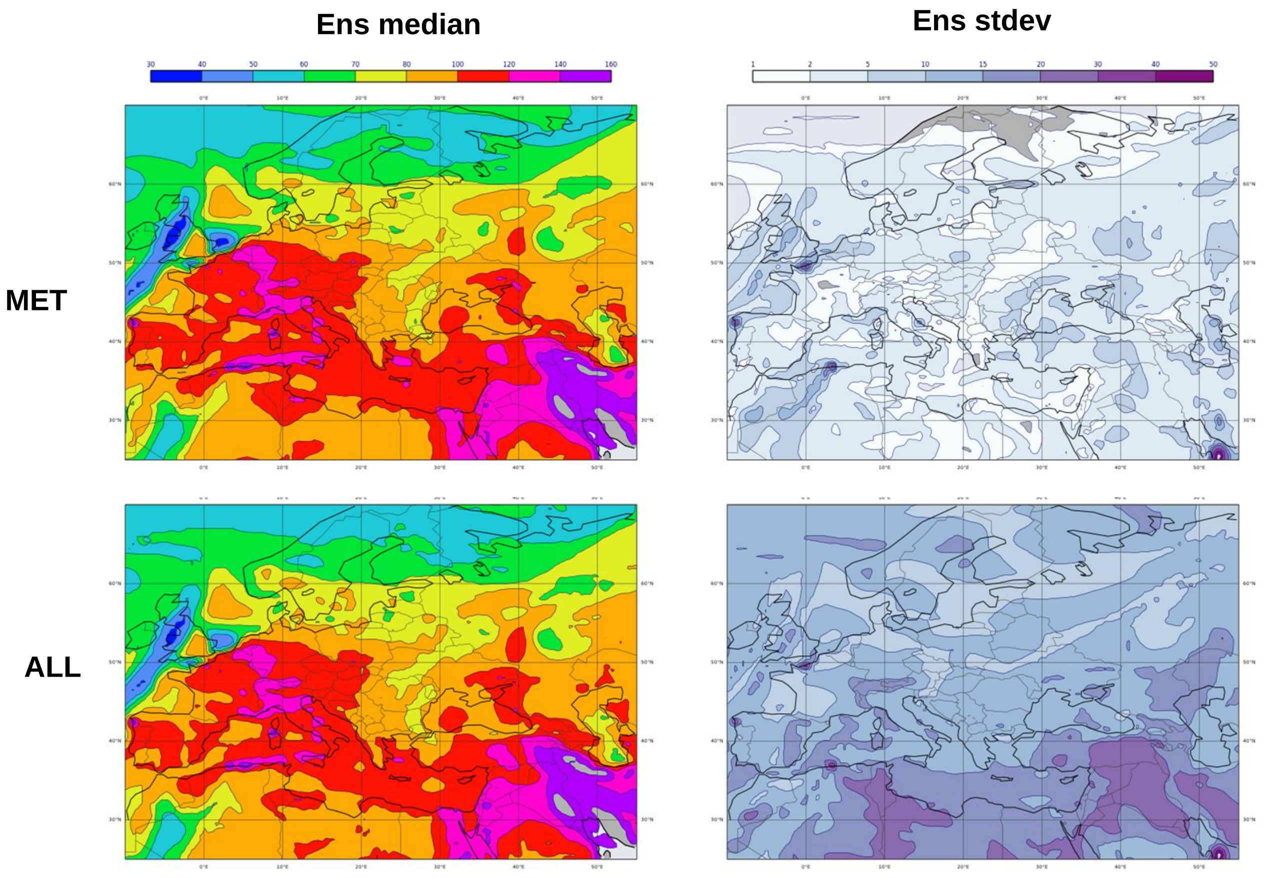

Figure 6: IFS-COMPO ensembles MET (top) and ALL (bottom), surface ozone at 36h forecast, run starting on 15/6/2021. Ensemble median (left) and standard deviation/spread (right).

Comparing the spread of two ensemble simulations is a good way to assess how sensitive the model is to perturbations of each kind of input. However, this doesn't give any information if the spread is "too low" or "too high". For this, skill scores that involve observational datasets are required. We implemented two different metrics

- A comparison of ensemble spread and the root mean square error (RMSE) of the ensemble mean,
- Rank histograms (Talagrand diagrams),
- Comparison of simulated and observed probability density functions.

It is common to assess the skill of an ensemble by comparing the ensemble mean RMSE and the ensemble standard deviation (ensemble spread). The former measures how accurate the ensemble mean is, i.e. how near the mean of the ensemble forecasts is to analysis fields or observations; the latter verifies whether the ensemble forecasts simulated a wide enough range of possible atmospheric states to reflect the error characteristics of the ensemble mean. Ideally, one would want the ensemble mean RMSE to be as small as possible and the spread to be equal to the ensemble mean RMSE on average over many cases.

For the verification of ensemble forecasts the Rank histograms (also called Talagrand diagrams) are widely used. This type of diagram shows how often observations match different parts of an ensemble forecast distribution. To this end, the ensemble forecast distribution is divided into bins of equal size, matching the ensemble size (or number of members), for example going from low predicted to high predicted AOD at 550nm. The observations are then put in the appropriate bins forming a histogram. In a reliable ensemble forecast, the frequency of observations in each bin will be identical since each part of the ensemble forecast distribution is equally likely. Figure 7 show three examples of simulated and observed probability distributions along with rank histograms. When the ensemble is well balanced, the rank histogram is roughly flat. High values at the extremes diagnose an ensemble with too little spread, also called under dispersive. A biased ensemble will show a slope. Finally, an over-dispersive ensemble (not shown) will show high values in the middle ranks, and null or lower values at the extremes.

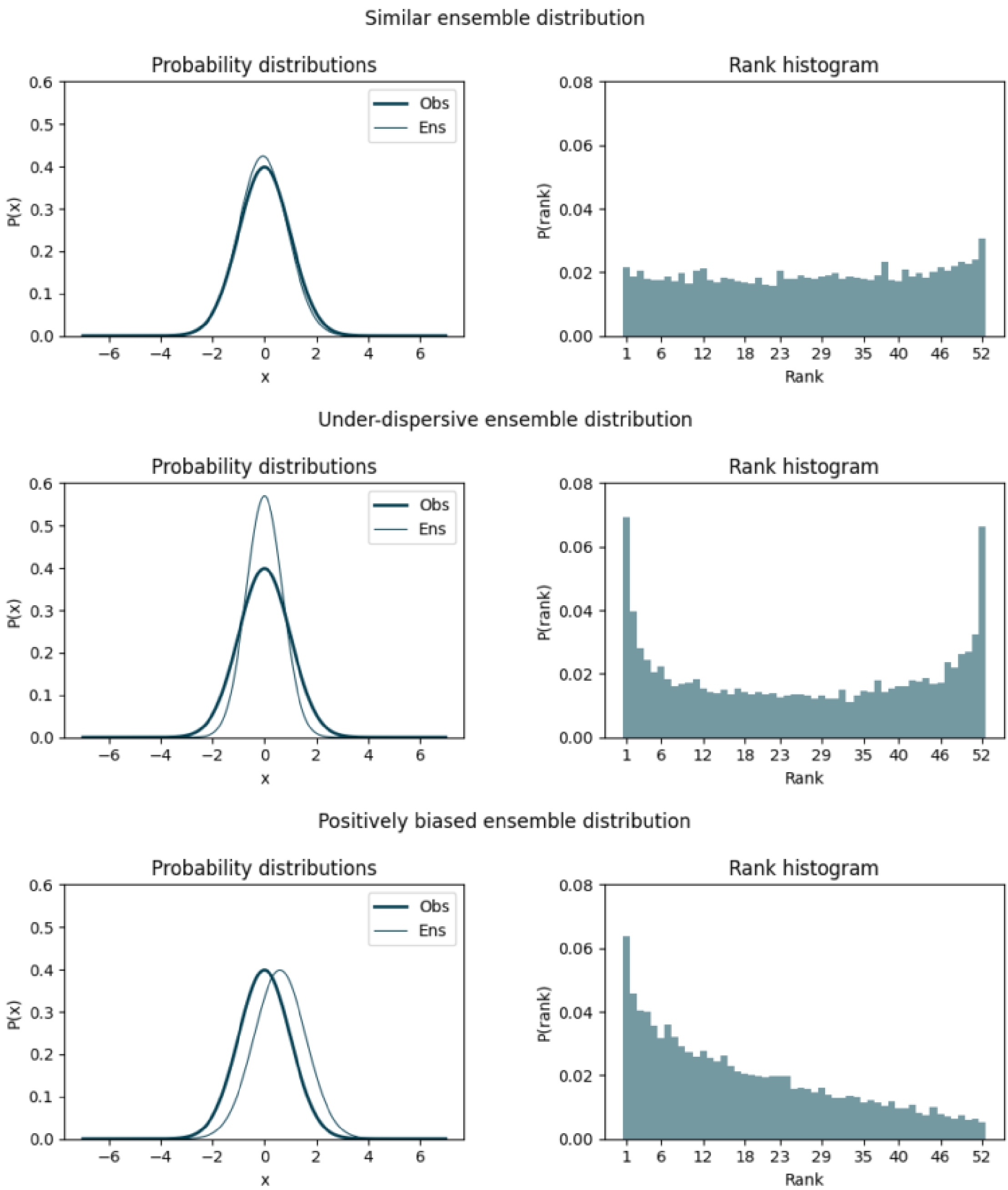

Figure 7: Three examples of probability distribution and rank histograms, showing well balanced (top row), under-dispersive (middle row) and biased situations (bottom row).

This ensembles scores can be computed using either an observational dataset, or against their own analysis. Each approach has its benefits and drawbacks. Using the ensemble analysis for verification is more adequate for assimilated quantities (such as AOD) than for non assimilated species, as the analysis is then deemed more reliable. Using this method for PM2.5 or surface ozone for example is not advised (at least as long as they are not assimilated), as the link between the analysis and observations is indirect in the global CAMS system. Using its own analysis for verification is also very convenient because the model space and the observation space are the same; there are no resolution or representativeness issues. Verification against observational datasets offer a better reference than against own analysis; however, the data can be sparse and there can be representativeness issues between observation and model. In this section, we present rank histograms computed against both observational datasets and own analysis.

4.1 Rank histograms (Talagrand diagrams)

4.1.1 AOD at 550nm

Figure 9 shows the observed and simulated frequency distribution, a rank histogram and a time series of observed versus simulated Aerosol Optical Depth (AOD) at 500nm for the ALL experiment in February 2021, for a 120h forecast time. The evaluation is done against all AERONET level 2 data available in February 2021. The AERONET network is shown in Figure 8.

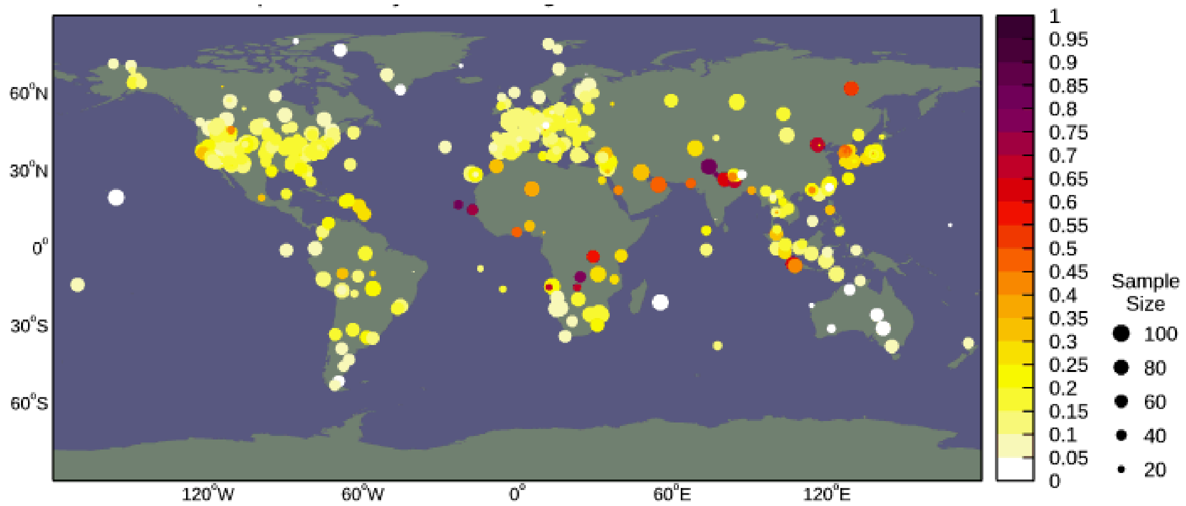

Figure 8: AERONET level 2 AOD at 500nm in the summer of 2020, values and sample size.

The rank histogram shows that for AOD and this experiment, the IFS-COMPO ensemble is clearly under dispersive: a significant amount of observations falls outside of the 50 members and are above the maximum values out of the 50 members (high value for entry 50 in the X axis). This is also shown by the lower ensemble standard deviation (0.15) as compared to the observation standard deviation (0.19). The ensemble suffers also from a negative bias, which is also apparent in the time series: both the control and ensemble mean are most of the time below the observational average because of a persistent low bias of the control run over the considered period. As such, the low bias shown here is rather a consequence of inherent model bias, also present in the control run. However, the ALL experiment lacks the ability to represent a significant fraction of the observed space for AOD at 550nm.

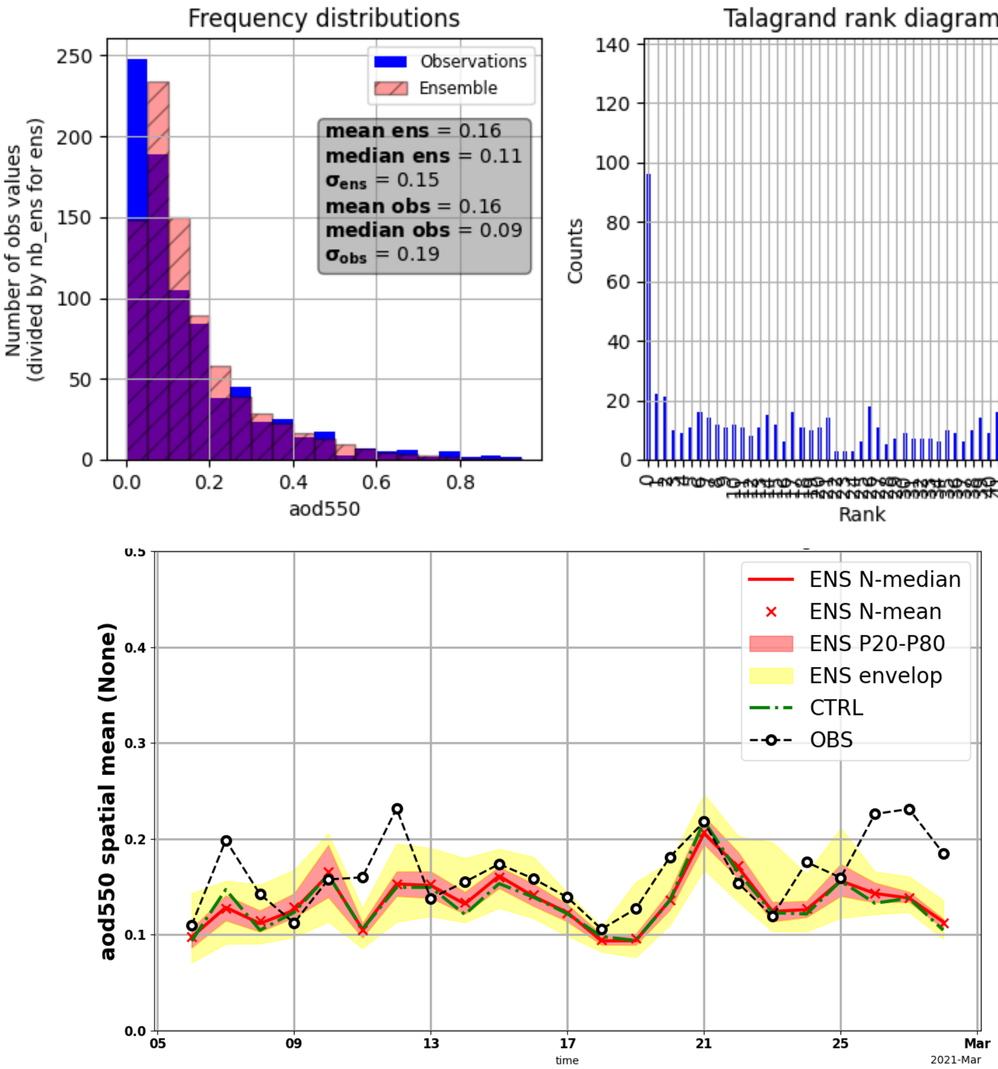

Figure 9: ALL experiment, February 2021, 120h forecast time, evaluation of AOD at 550nm versus all AERONET level 2 observational data. Top left, frequency distribution; top right, rank histogram. Bottom, time series of daily observed mean AOD at 550nm averaged over all AERONET stations together with the ensemble mean and median and values simulated by the control run. The ensemble 20% and 80% centiles as well as the envelope are shown.

Figure 10 shows a similar global evaluation of a series of IFS-COMPO ensemble simulations against their own analysis, for AOD at 550nm simulated at 120h forecast time. The diagnostic is similar to that reached against AERONET: all of the ensembles are more or less under dispersive for AOD. However, at the global scale, there is no sign of a positive or negative bias (this is not the case for the regional evaluation, not shown). The EMI and INI experiments are significantly more under dispersive than the MET one, showing that for AOD, the perturbations applied to the atmospheric composition initial conditions as well as anthropogenic emissions yield relatively little ensemble spread. The ALL_NOMODEL experiment is slightly less under dispersive than MET, but by a small margin: this shows that meteorological perturbations have the largest impact on ensemble spread for AOD (perturbations of the atmospheric composition model parameterizations were not evaluated).

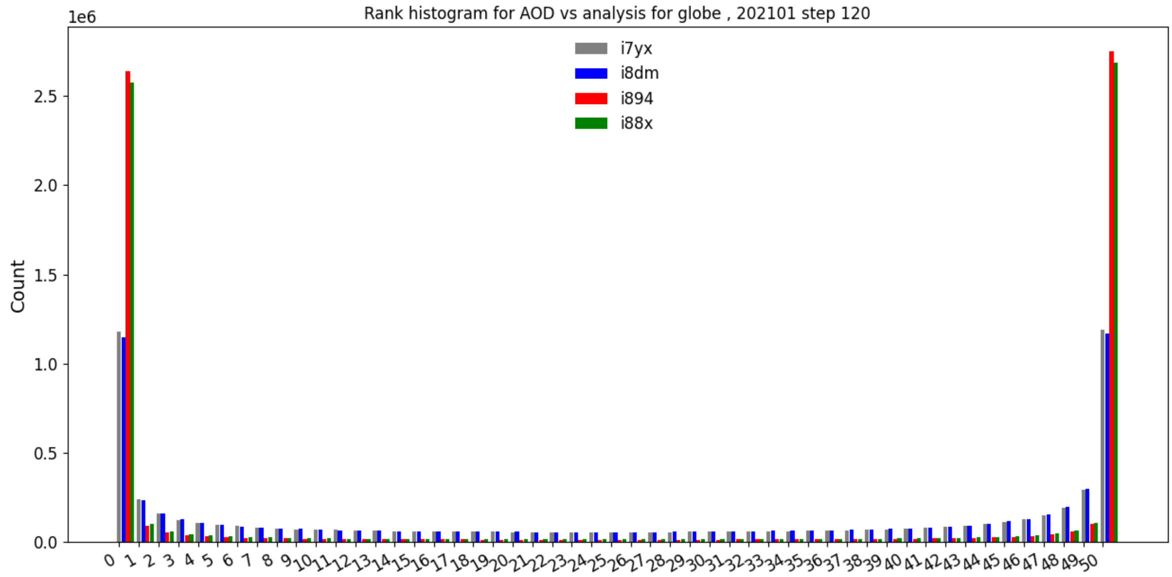

Figure 10: January 2021, 120h forecast time, evaluation of AOD at 550nm versus own analysis. Rank histogram of MET (grey), ALL_NOMODEL (blue), EMI (red), and INI (green).

4.1.2 PM2.5

Rank histograms have been built also for PM2.5 against European airbase/EEA stations, for simulated PM2.5 at 120h forecast time in February 2021 by the MODEL and ALL experiments (Figure 11). The two experiments are under-dispersive and negatively biased, although the under-dispersion is less pronounced for the ALL experiment. This negative bias is very apparent in the time series, which shows that up to 20/21 February, the control run and ensemble mean PM2.5 are significantly lower than the observational average. The MODEL experiment struggles to reach observed values even for the maximum ensemble values, while the ALL experiment sometimes manages to reach the observed value. The period 20-25 February 2021 was marked by a combined pollution and dust intrusion event over most of Western and Central Europe, which is reflected in the higher observed average PM2.5. The simulated values are less impacted by a low bias during this period.

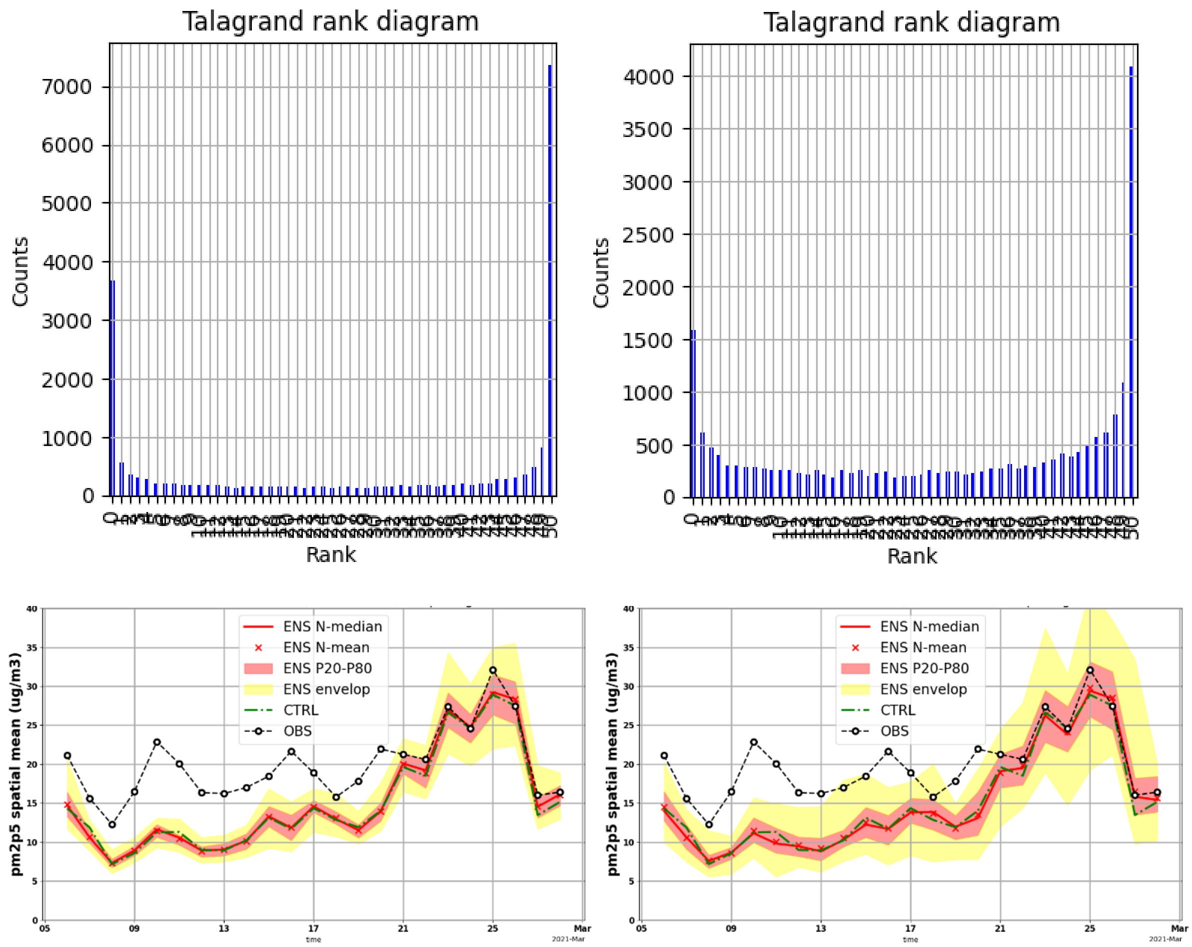

Figure 11: Top, rank histograms of MODEL experiment (left) and ALL experiment (right), February 2021, 120h forecast time, evaluation of PM2.5 versus all available airbase/EEA data over Europe. Bottom, time series of daily observed mean PM2.5 in February 2021 over European stations together with the ensemble mean and median and values simulated by the control run. The ensemble 20% and 80% centiles as well as the envelope are shown.

4.1.3 Surface ozone

For surface ozone, we focus on June 2021, a month that saw a short heat wave between 15-17 June and an associated surface ozone peak. In order to better assess the representation of the diurnal ozone peak, a forecast time of 108h instead of 120h is considered in this subsection. Figure 12 show frequency distributions and rank histograms for the MET, MODEL and ALL experiments, and Figure 13 show time series of simulated versus observed daily surface ozone averaged over all airbase/EEA stations. Similarly to AOD at 550nm and PM2.5, all experiments show under-dispersion and, to varying degrees, a low bias in surface ozone. This arises from a negative bias in the control run up to 20th of June around as shown in Figure 10. In particular, the ozone peaks of the first half of June are generally significantly underestimated. The meteorological perturbations bring relatively little spread in the time series; the perturbations of the model parameterizations have a slightly higher impact on ensemble spread, with a mean ensemble standard deviation of 17.3 versus 16.7 for meteorological perturbations (the observational standard deviation is 26.1). The ALL ensemble is still under-dispersive and low biased, but to a lesser extent, with a mean ensemble spread of 18.0.

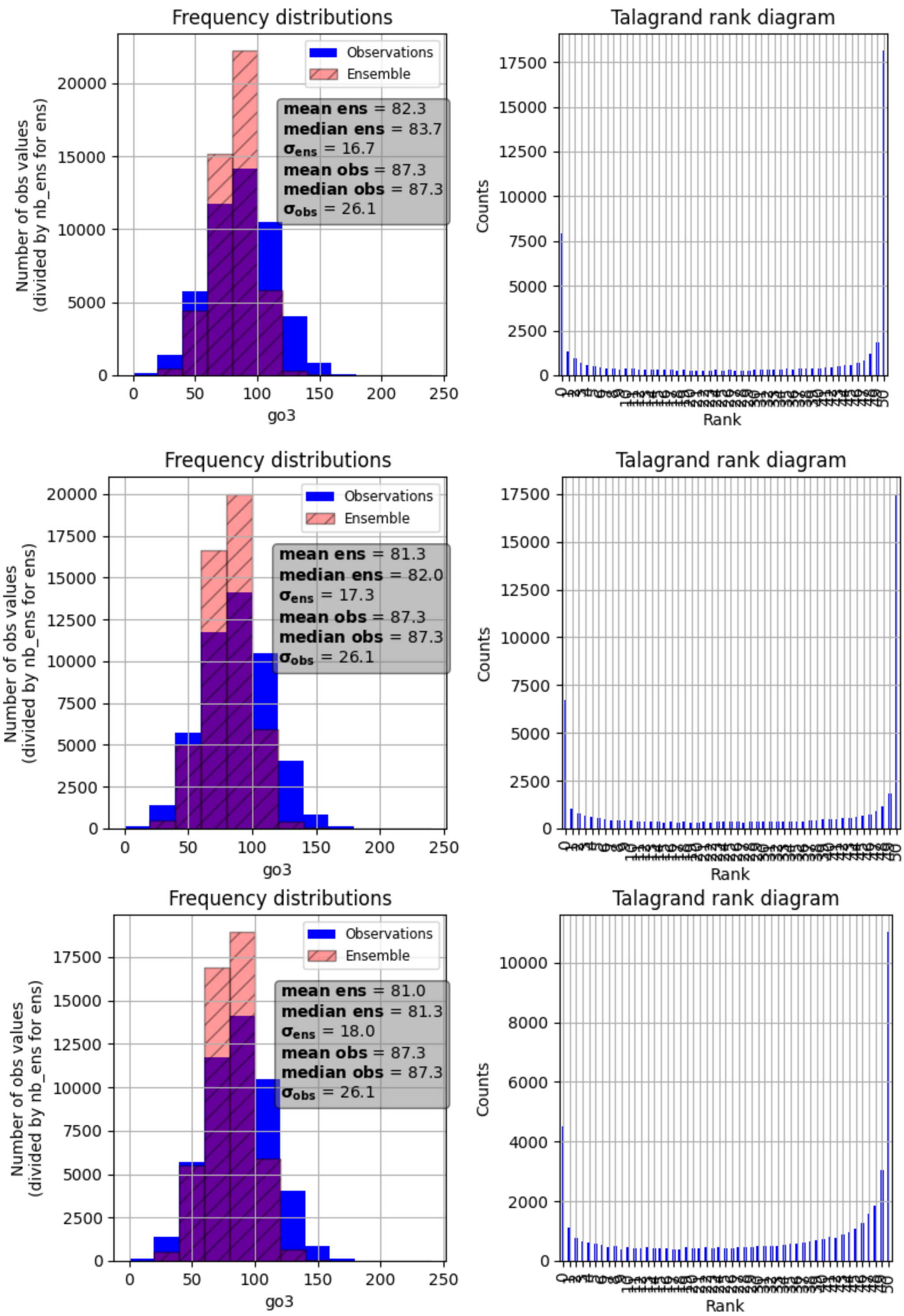

Figure 12: MET (top), MODEL (middle) and ALL (bottom) experiments, frequency distribution (left) and rank histograms of simulated surface ozone, June 2021, 108h forecast time, evaluation versus all available airbase/EEA data over Europe.

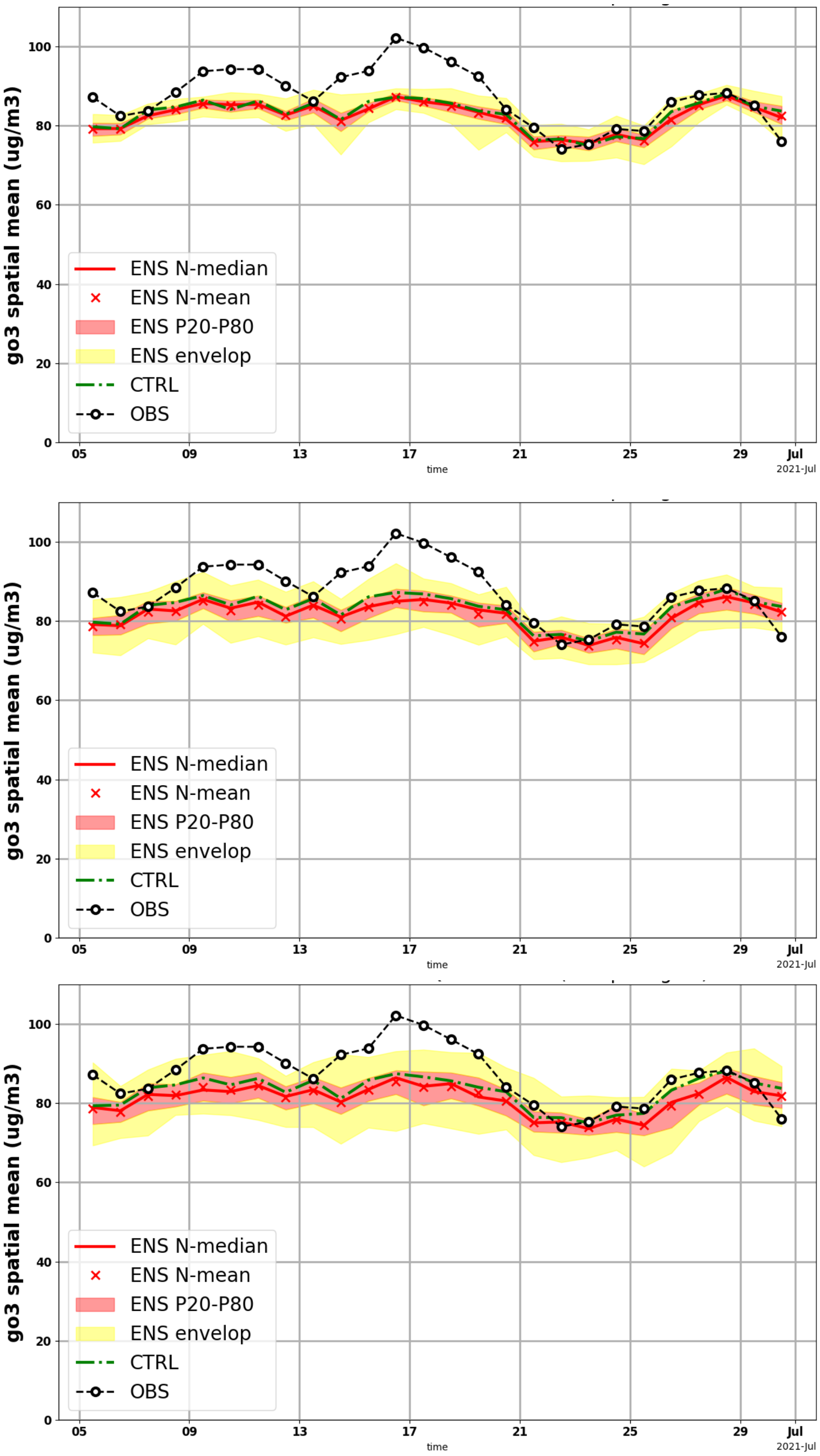

Figure 13: MET (top), MODEL (middle) and ALL (bottom) experiments, time series of daily observed mean surface ozone in June 2021 over European stations together with the ensemble mean and median and values simulated by the control run. The ensemble 20% and 80% centiles as well as the envelope are shown.

To summarize, it appears that most of the ensemblist diagnostics points to an under-dispersion or too small spread. In some cases (PM2.5 and surface ozone), this appears to be largely a consequence of model biases, which makes it harder for the ensemble to capture the observational variability. This is also a sign that perhaps the perturbations applied are too small or don't explore enough degree of freedom, such as correlated perturbations, or perturbations varying with forecast time.

5 Uncertainties of selected CAMS products

In this section, we present monthly average standard deviation and relative standard deviation (normalized by the ensemble mean value) of IFS-COMPO ensembles MET, EMI, MODEL and ALL. The standard deviation, or ensemble spread, is taken as a measure of the uncertainty of the simulated products. Other measures are possible, such as the difference between the 25 and 75% percentiles.

5.1 Monthly values

In this subsection we show the simulated uncertainty of AOD at 550nm, dust AOD at 550nm, PM2.5, PM10 and surface ozone averaged over January and June 2021. We chose to show monthly values, as the daily variability is very high. The values shown in these two months are different for absolute spread, but not to a very large extent for relative spread, and other months (not shown) show values in the same range. For relative spread Thus, the results shown here can qualitatively be extended to other months as far as relative spread is concerned.

5.1.1 AOD at 550nm

Figures 14, 15, 16 show the ensemble spread and relative spread of the MET, EMI, MODEL and ALL experiments, for January 2021 and for 24 and 120h forecast time, and for June 2021 only 120h forecast time except for EMI for which data is incomplete in June. For AOD at 550nm and 24h forecast, the uncertainty caused by meteorological factors and model perturbations is much higher than that from the emissions. The relative spread is generally comprised between 5 and 15% (higher over dust source regions for MODEL) for these two experiments, and is generally below 5% for EMI. The ALL experiment shows quite homogeneous spread of 15-25% almost everywhere. Over oceans this could be caused partly by the perturbations in initial conditions.

The uncertainty of simulated AOD at 550nm is unsurprisingly much higher for 120h forecasts, with values between 10 (over oceans) to more than 30% for MET and MODEL, while EMI perturbations are generally between 5 and 15% over continents (closer to emission sources) and less over most of the oceans. The very high relative spread over boreal regions correspond to areas with relatively small absolute spread, indicating that the small value of the ensemble mean contributes to the high value of the relative uncertainty. Also, meteorological perturbations could result in the transport of higher AOD values over areas with relatively low AOD, which mechanically provokes a very high relative spread. Also, it is clear that the propagation of uncertainties through the IFS-COMPO ensemble is a very non-linear phenomenon: the spread of ALL is smaller than the sum of the spread of MET, MODEL, EMI (and the other contributions such as the initial conditions perturbations which play a smaller role at 120h forecast time than at 24h).

In June 2021, the absolute uncertainty is much higher than in January, mostly because simulated AOD at 550nm is higher over many regions. The relative spread is quite similar to that of January for MODEL, but higher over many regions including Europe for MET. Interestingly, the spread of ALL is smaller than that of MET over several regions, such as over parts of China and Europe. Meteorological and model perturbations (particularly the perturbations impacting emissions of sea-salt aerosols and desert dust, and wet deposition) could partly cancel each other out.

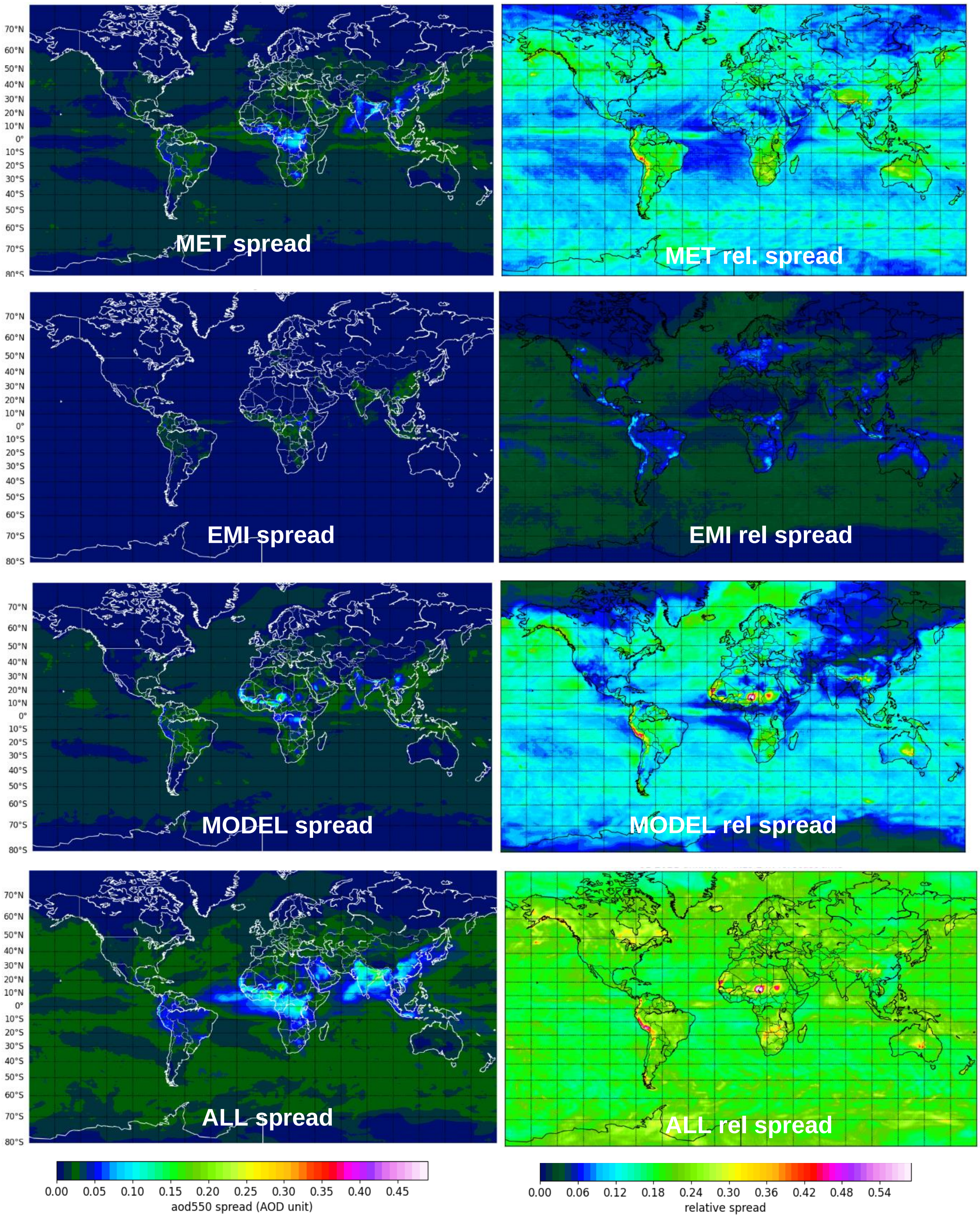

Figure 14: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated AOD at 550nm in January 2021, at 24h forecast time.

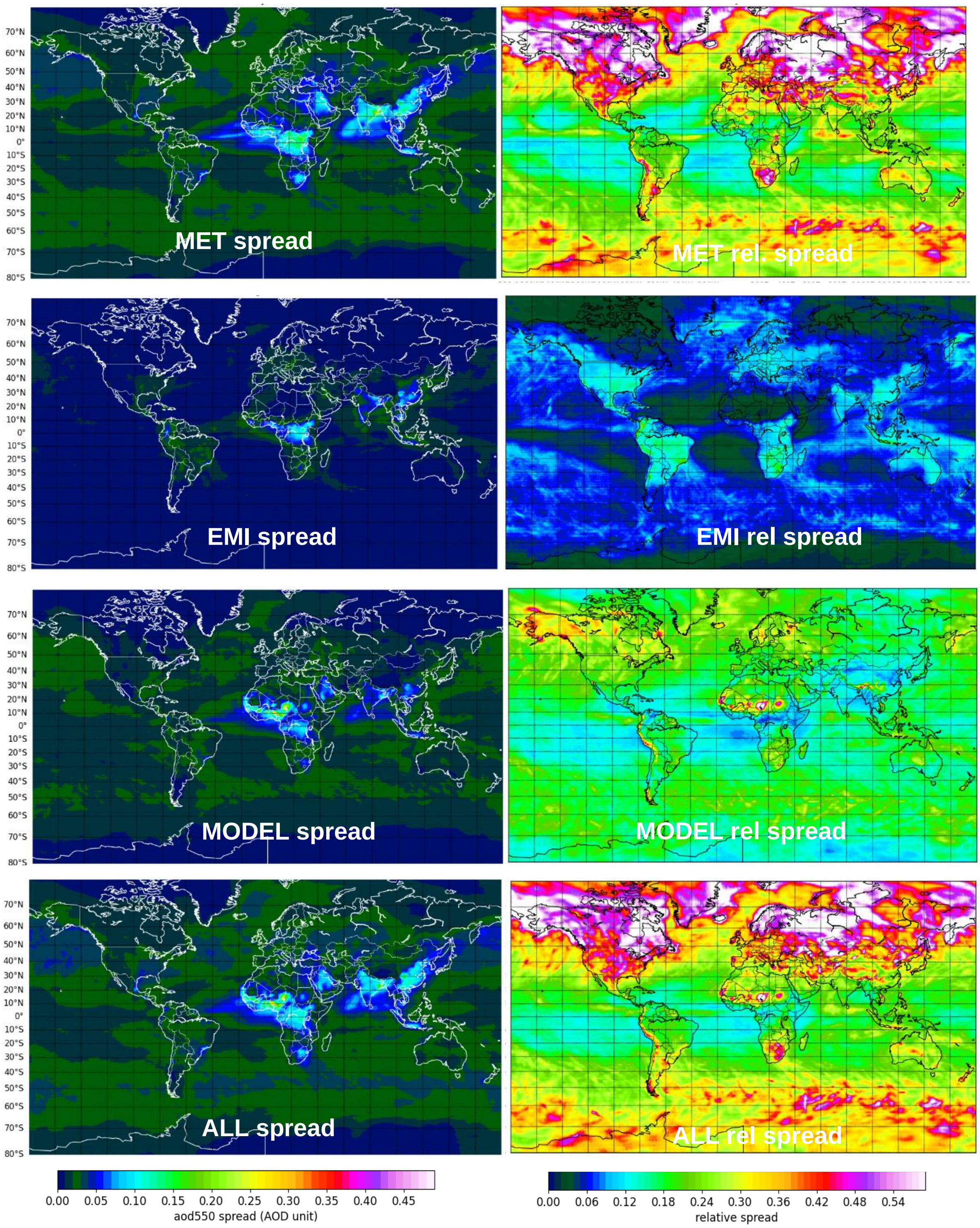

Figure 15: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated AOD at 550nm in January 2021, at 120h forecast time.

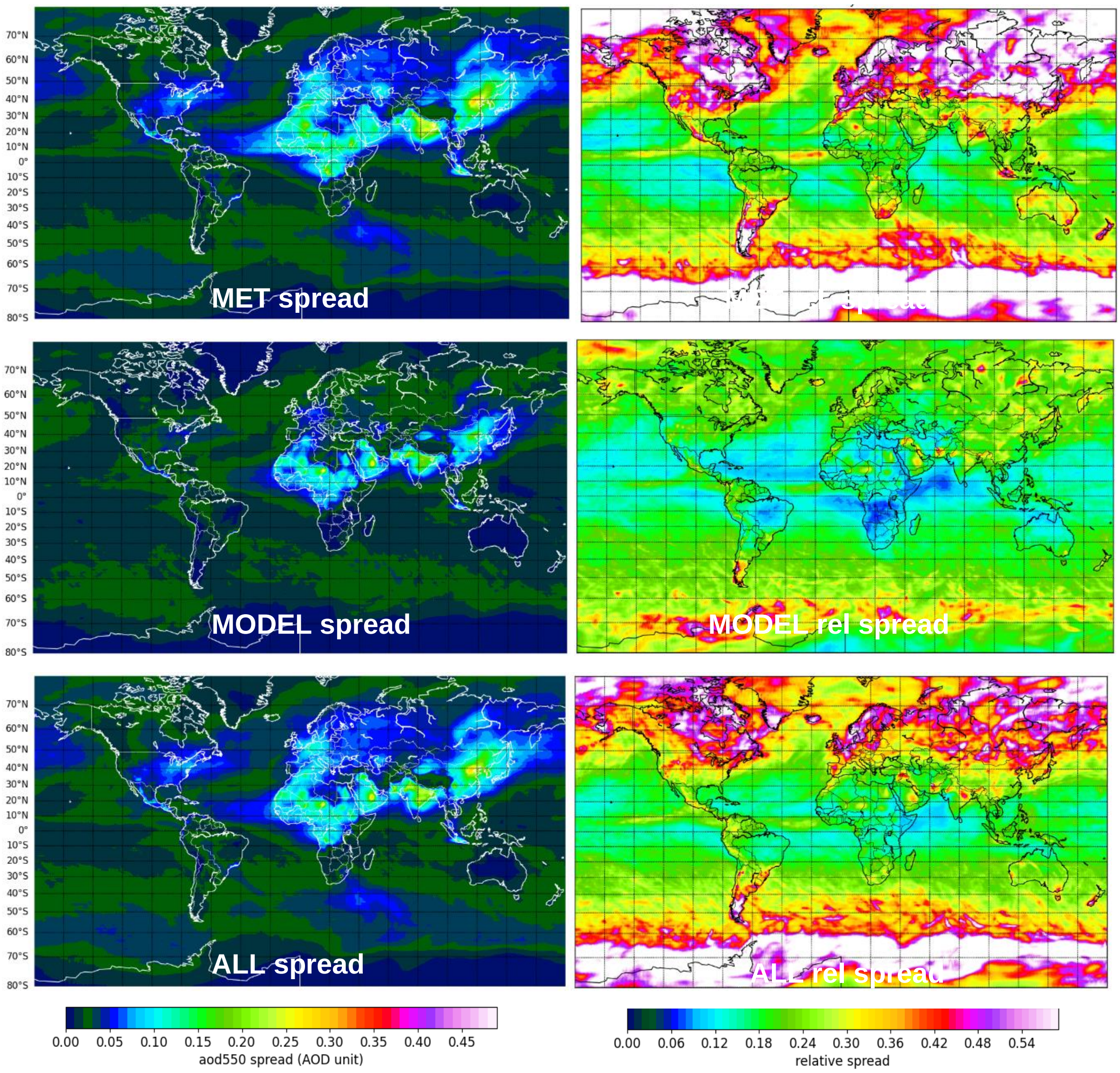

Figure 16: Spread (left) and relative spread (right) of the MET, MODEL and ALL experiments for simulated AOD at 550nm in June 2021, at 120h forecast time.

5.1.1 Dust AOD at 550nm

For dust AOD at 550nm (Figures 17 and 18) the EMI experiment is not shown, as its impact is quite small. Also, only June 2021 is shown, as desert dust sources are much more active than in January 2021. At 24h forecast time (Figure 17), the MET and MODEL absolute spread show different patterns: high values localized over dust source regions for MODEL arising from perturbations of the dust emissions, and high values over all Sahara/Middle East for MET. The relative spread also shows the impact of the wet deposition perturbation for MODEL (the higher values around the ITCZ. For MET, the relative spread is generally higher than for MODEL, which is also because MET perturbs the meteorological initial conditions, which perturbs in turn transport, emissions and deposition very quickly after the beginning of the simulation. At 120h forecast time, MET shows very high values over most of regions that are very remote from dust source regions, for the same reason as explained above: MET is the only experiment that perturbs transport processes, which means that relatively high values of dust AOD could be transported over areas where the ensemble mean is small, resulting in very high relative spread, while the absolute spread is not remarkable. The spread of MODEL is much higher everywhere at 120h than at 24h. The spread of ALL at 120h forecast time is again smaller than that of MET over several regions.

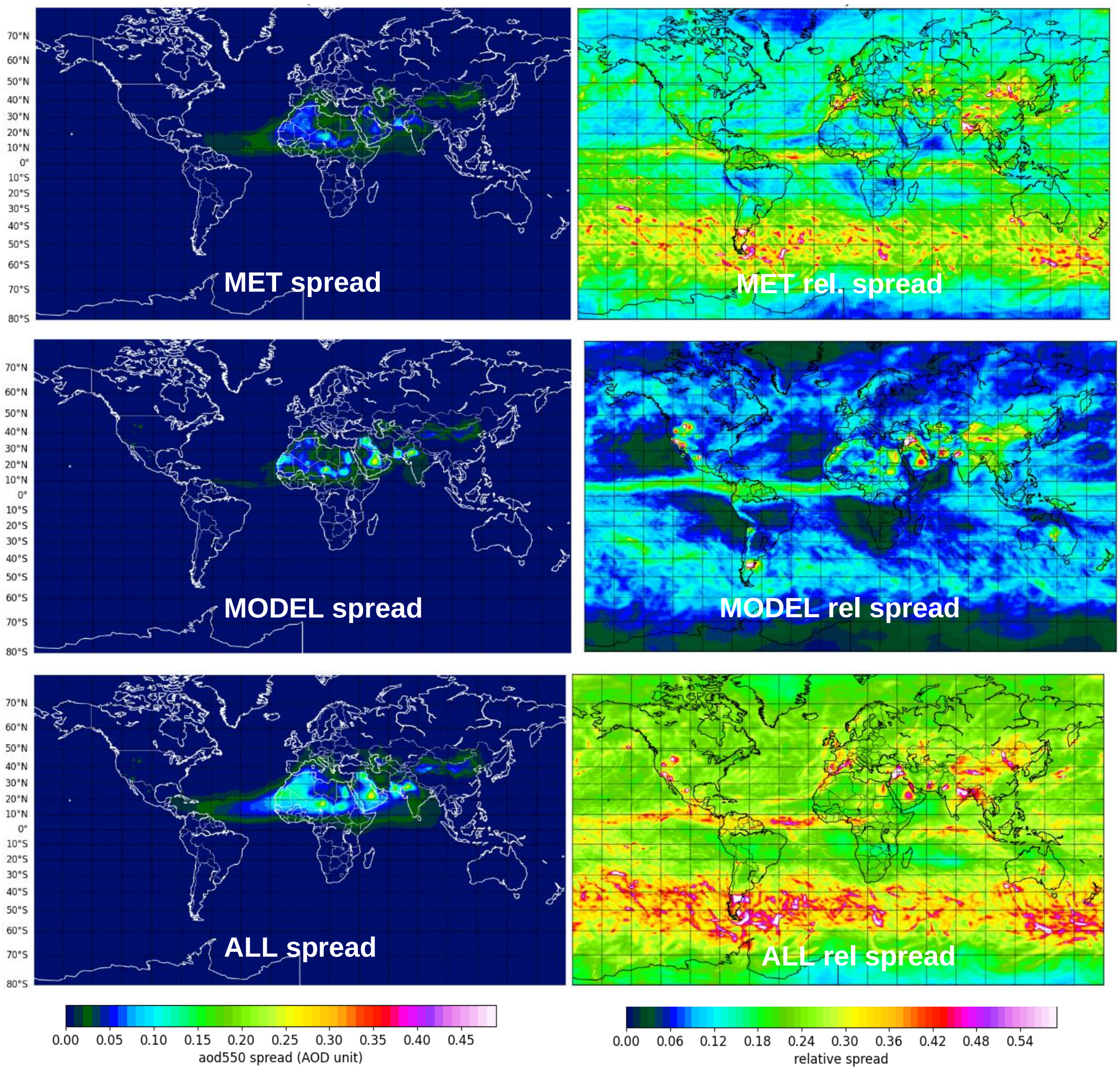

Figure 17: Spread (left) and relative spread (right) of the MET, MODEL and ALL experiments for simulated dust AOD at 550nm in June 2021, at 24h forecast time.

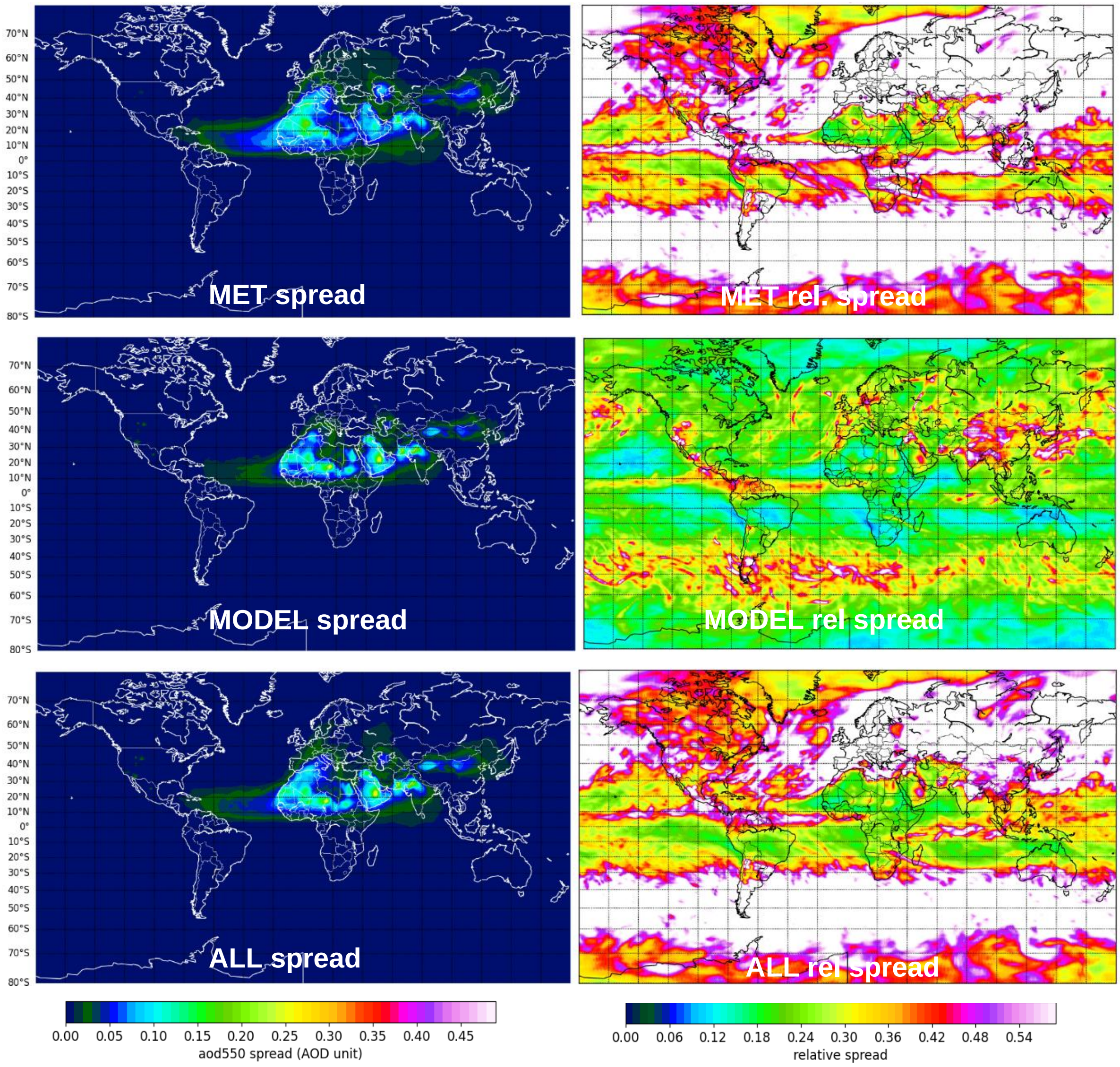

Figure 18: Spread (left) and relative spread (right) of the MET, MODEL and ALL experiments for simulated dust AOD at 550nm in June 2021, at 120h forecast time.

5.1.2 PM2.5

Figures 19, 20 and 21 show the ensemble spread and relative spread of PM2.5 simulated by the MET, EMI, MODEL and ALL experiments, for January 2021 and for 24 and 120h forecast time, and for June 2021 only at 120h forecast time except for EMI for which data is not fully available in June. The patterns are quite different from the ones shown for AOD. In January 2021 and at 24h forecast time, the uncertainty is much higher with MODEL over oceans as compared to MET and EMI, while over continents the uncertainties of MODEL and MET are close. For EMI, the values are much higher than those calculated for AOD, with uncertainties at 24h forecast time of around 10-15% over many regions with high anthropogenic emissions. The relative uncertainty of PM2.5 arising from model errors is generally between 20 and 25% already at 24h forecast time, higher than those calculated for AOD. For MET, the values are between 5% (over oceans) and 15-20% over most continent. As for AOD, the increase of the absolute and relative uncertainty is significant at 120h forecast time compared to 24h values. For EMI, most of the large anthropogenic sources and adjacent areas see uncertainties of around 20%. The uncertainty of MODEL is often higher than that of MET, particularly over oceans, with 20-30% values against 10-25% for MET. The uncertainty of ALL is significantly higher than that of MET or MODEL, with values between 30 and 60% over many continental areas except equatorial areas where the uncertainty is around 20%. The relative spread is quite similar for all experiments between January and June.

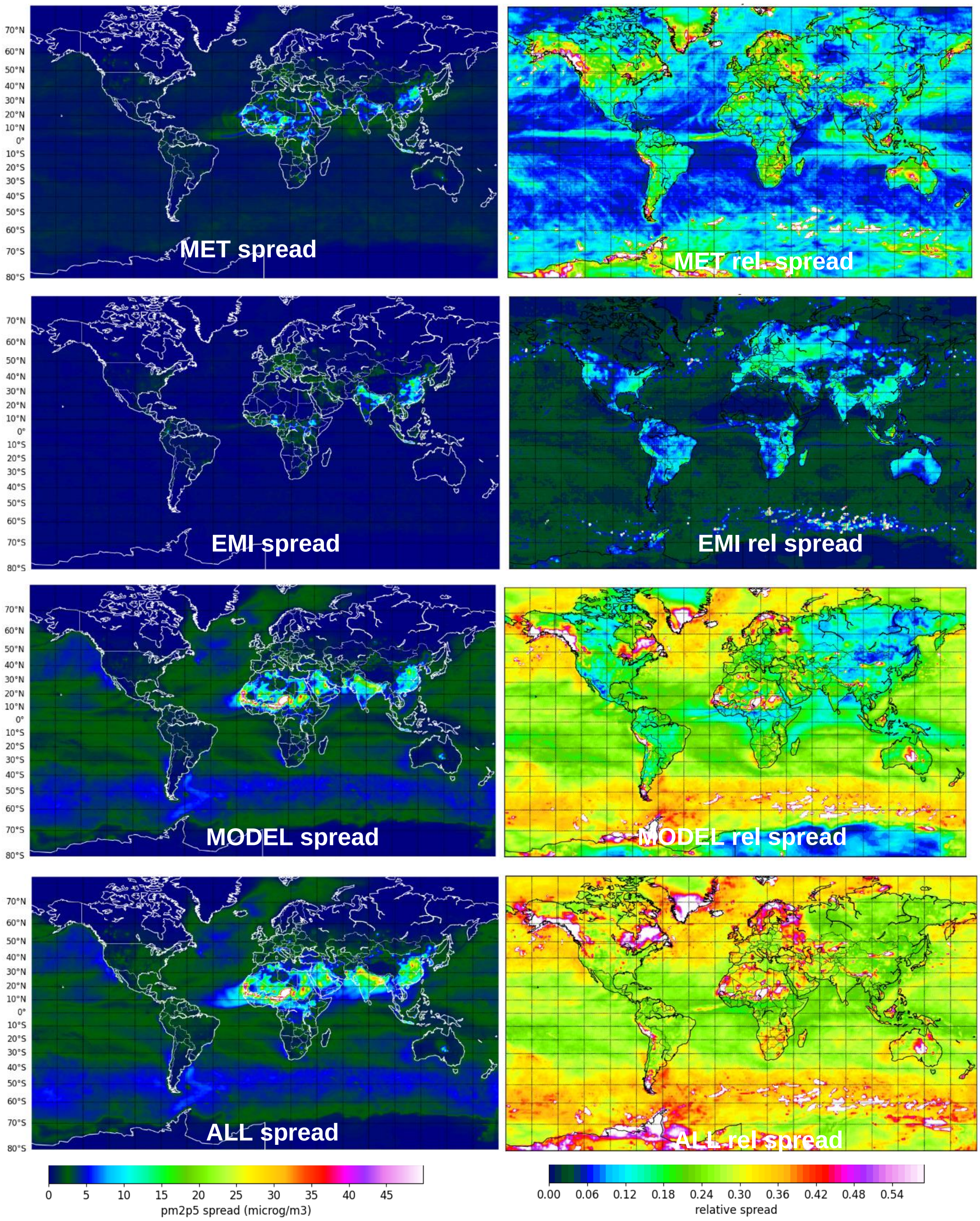

Figure 19: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated PM2.5 in January 2021, at 24h forecast time.

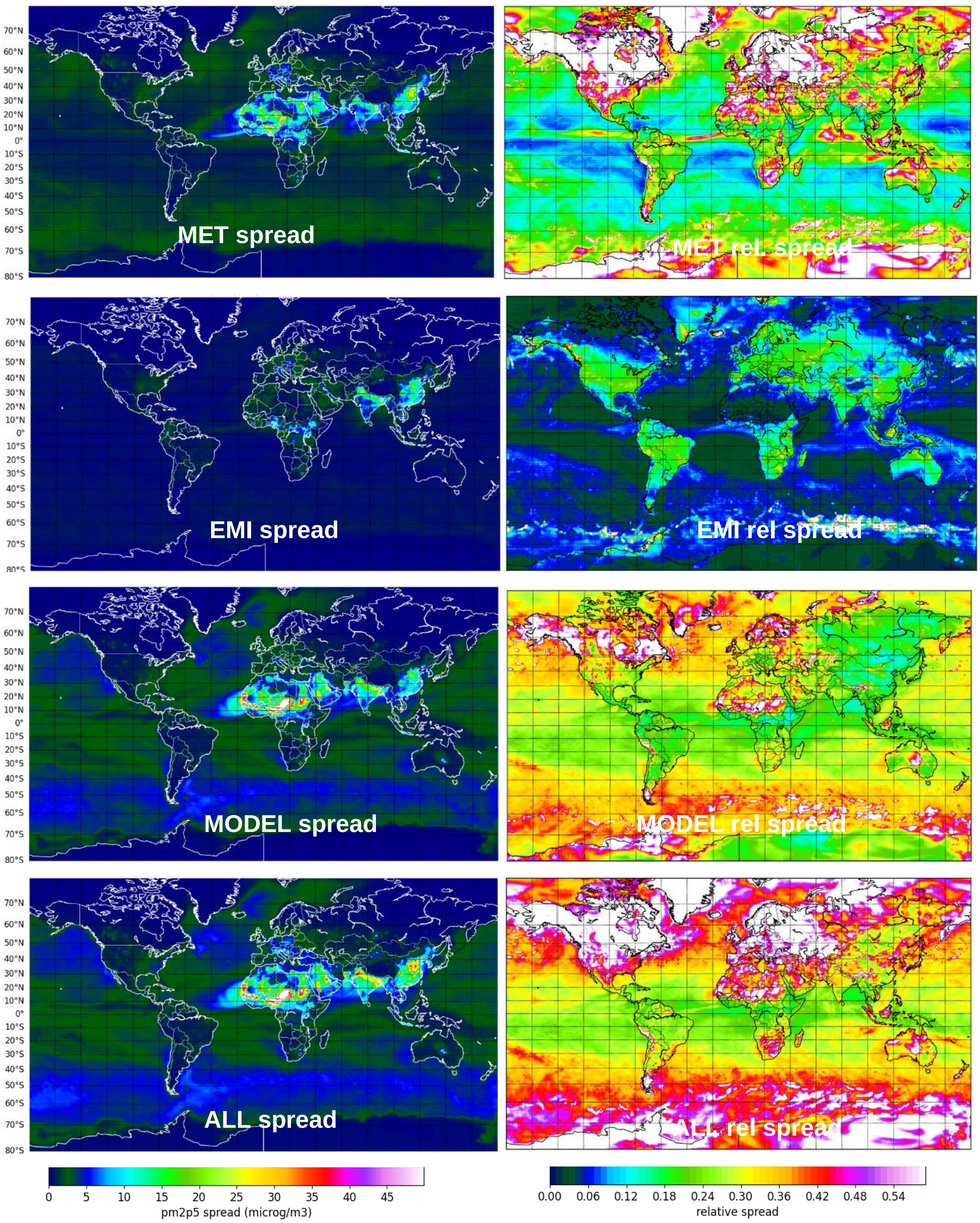

Figure 20: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated PM2.5 in January 2021, at 120h forecast time.

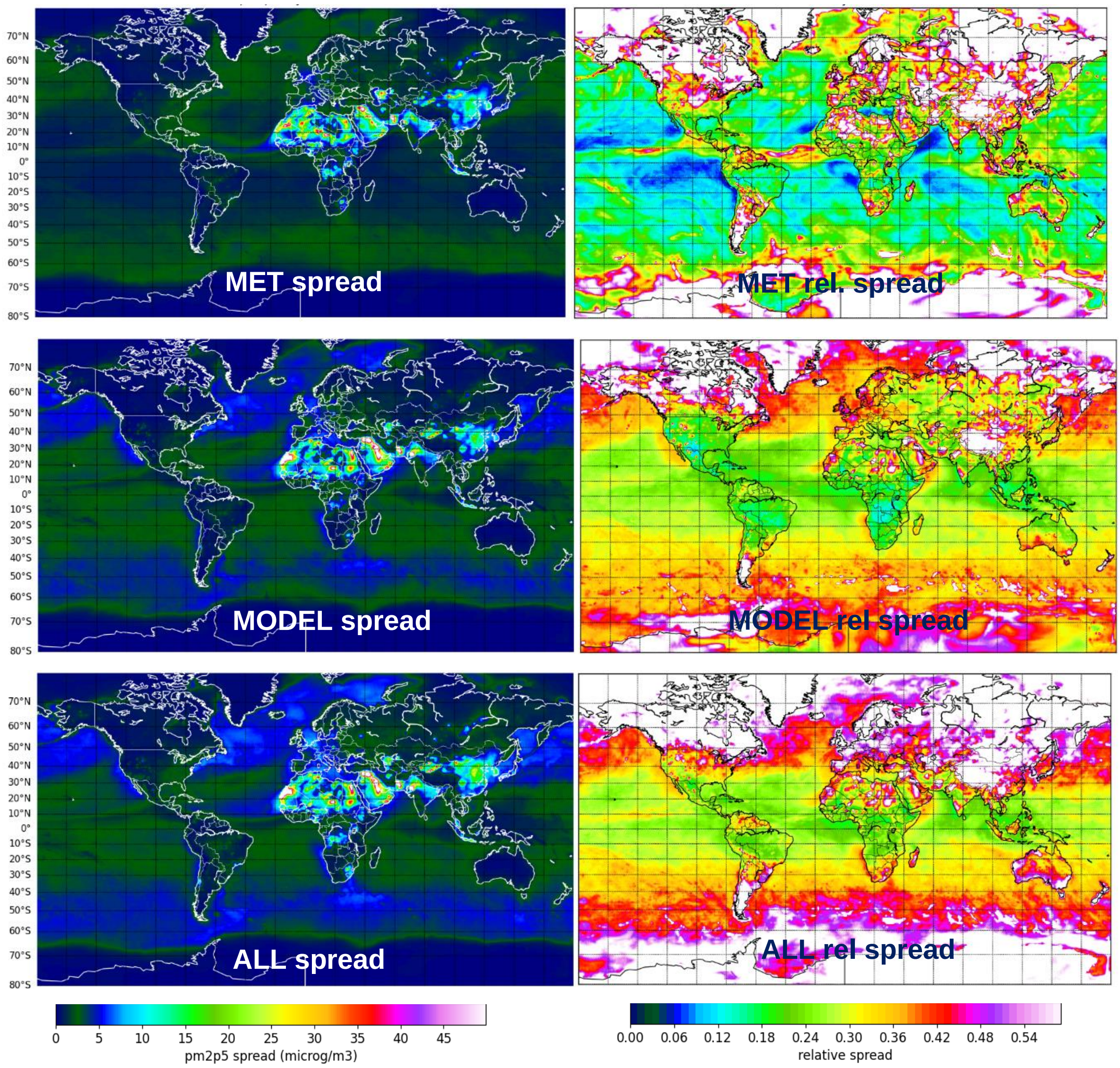

Figure 21: Spread (left) and relative spread (right) of the MET, MODEL and ALL experiments for simulated PM2.5 in June 2021, at 120h forecast time.

5.1.3 PM10

Figures 22, 23 and 24 show the ensemble spread and relative spread of PM10 simulated by the MET, EMI, MODEL and ALL experiments, for January 2021 and for 24 and 120h forecast time, and for June 2021 only 120h forecast time except for EMI where data is not fully available in June.. The relative spread is quite similar for all experiments between January and June. While the absolute uncertainties are higher than for PM2.5, the patterns for relative uncertainty are quite similar to that of PM2.5 except for the MODEL experiment over dust emitting regions, where the relative spread/uncertainty is higher for PM10.

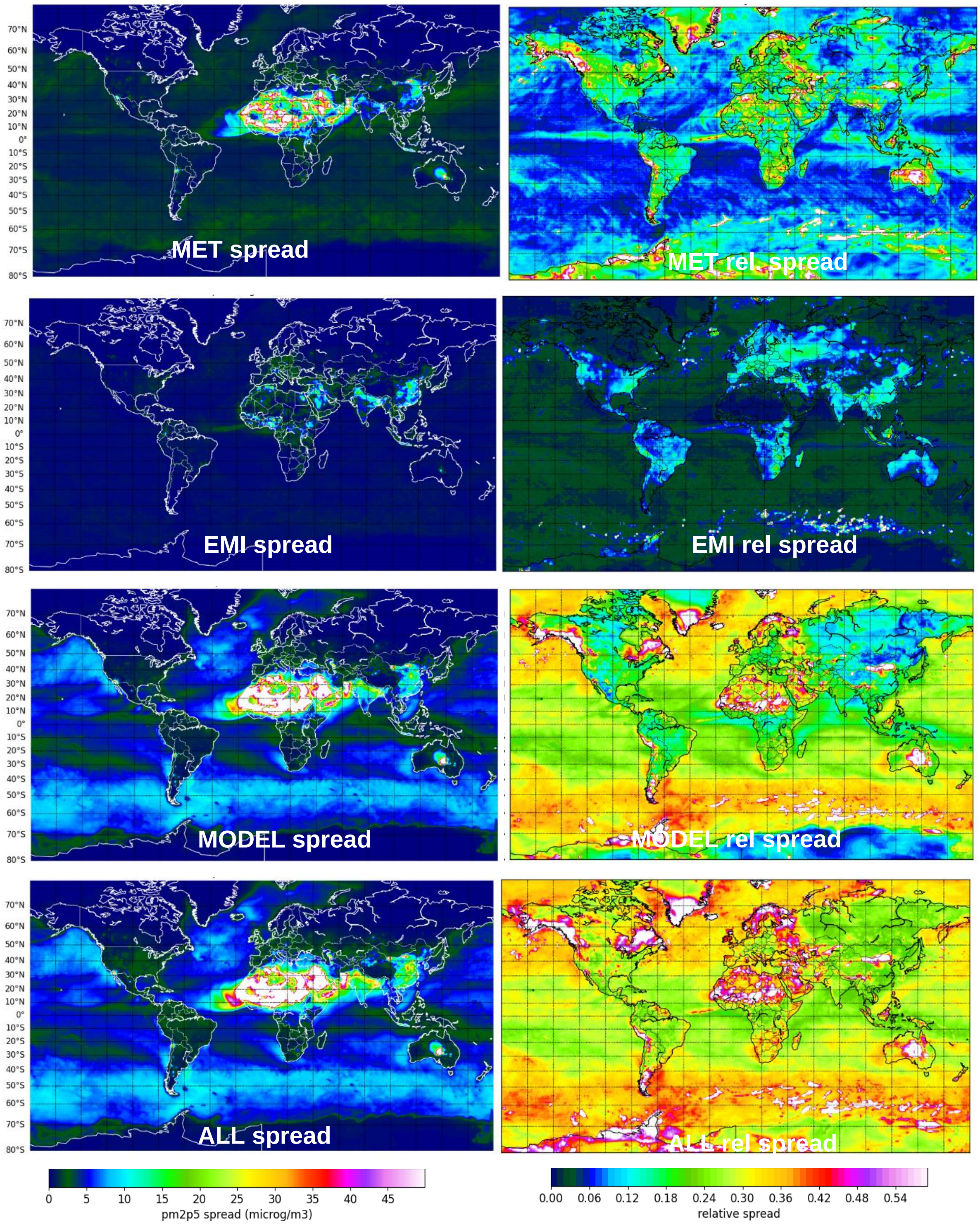

Figure 22: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated PM10 in January 2021, at 24h forecast time.

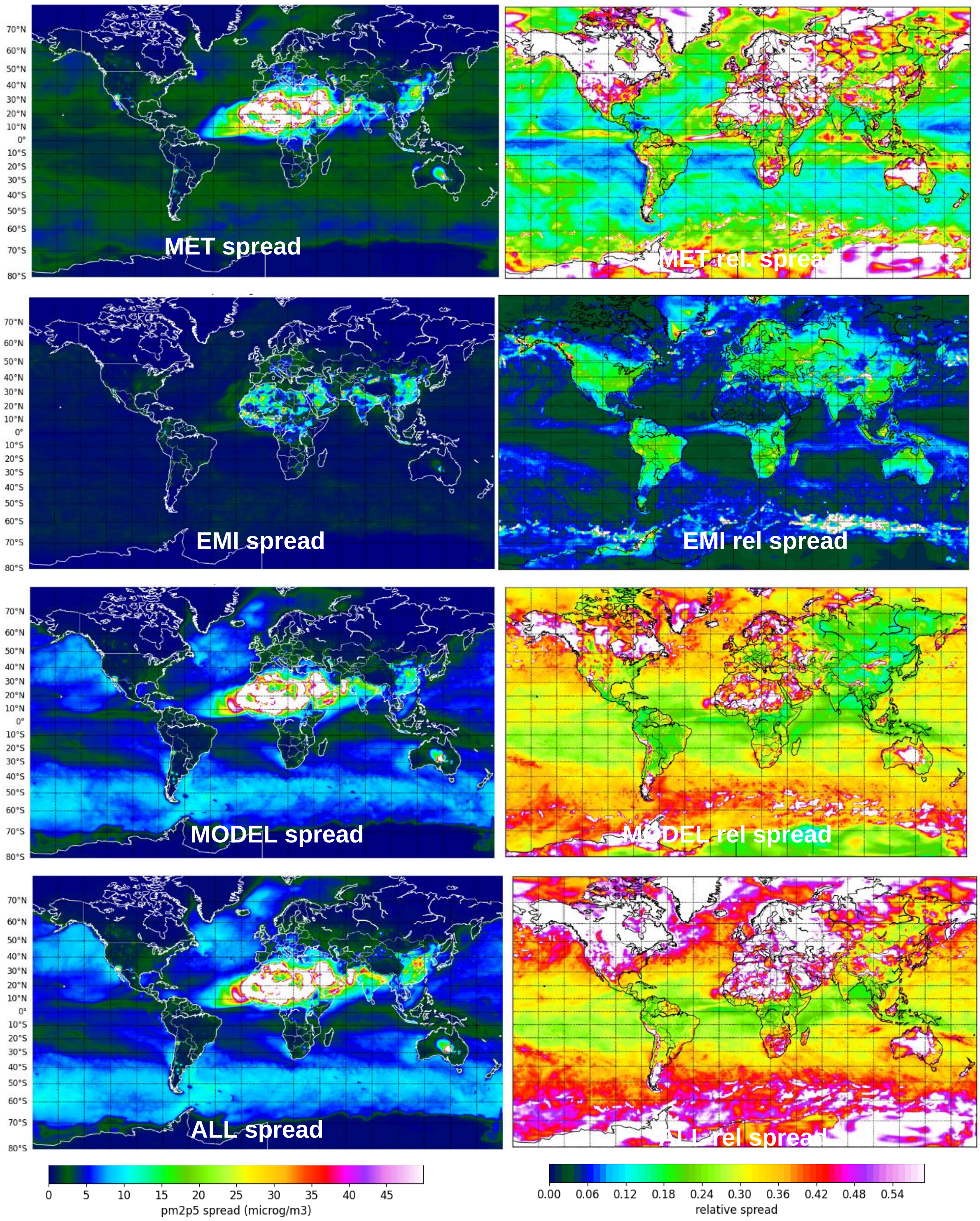

Figure 23: Spread (left) and relative spread (right) of the MET, EMI, MODEL and ALL experiments for simulated PM10 in January 2021, at 120h forecast time.

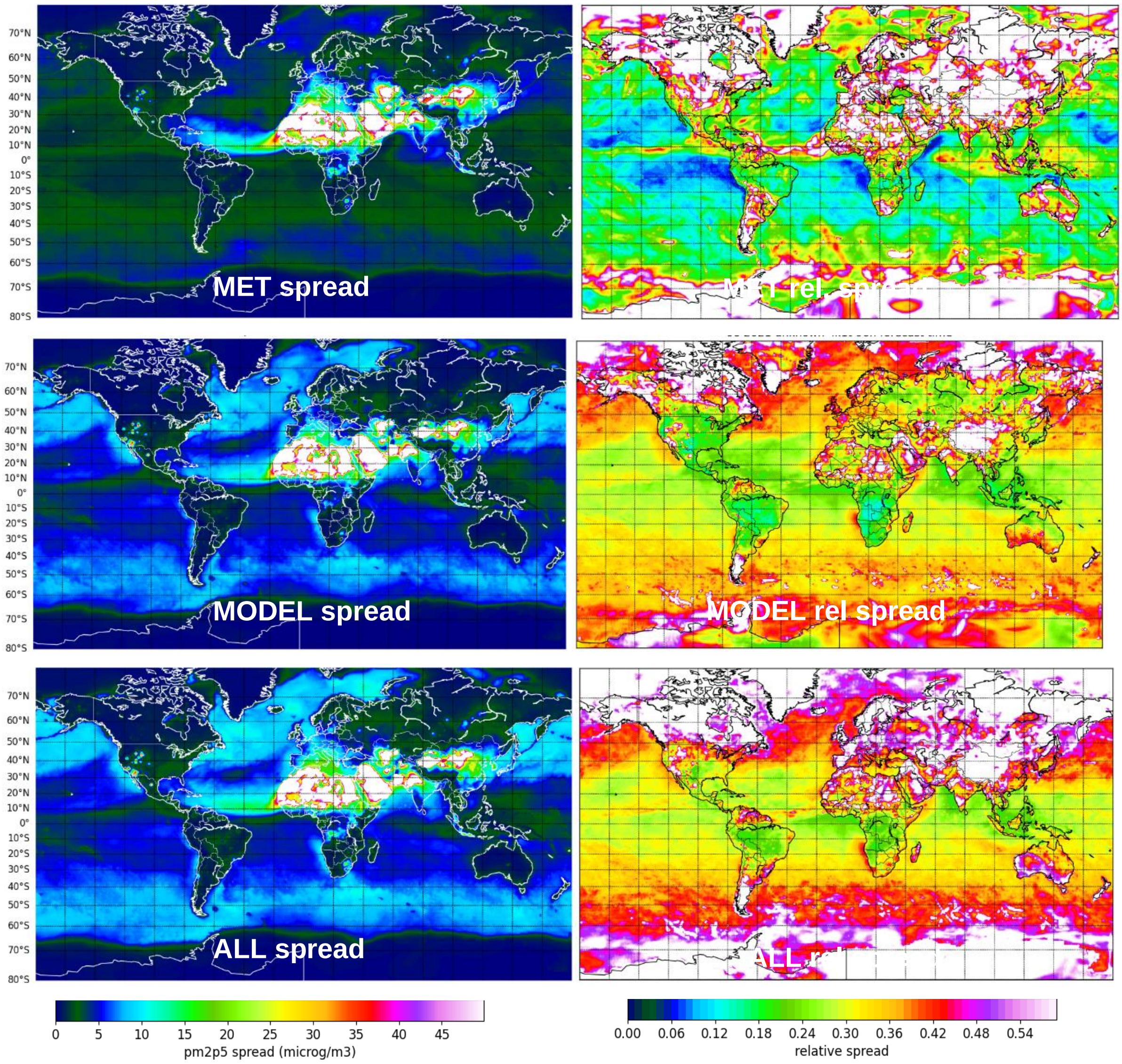

Figure 24: Spread (left) and relative spread (right) of the MET, MODEL and ALL experiments for simulated PM10 in June 2021, at 120h forecast time.

5.1.4 Surface ozone

Figures 25 and 26 show the ensemble spread and relative spread of surface ozone simulated by the MET, EMI_OLD, MODEL and ALL experiments, for June 2021 and for 24 and 120h forecast time. EMI was not available for June 2021.

At 24h forecast time, the absolute and relative spread are quite similar for MET and EMI_OLD, with values between 5 and 20% over ozone impacted regions. This could be because perturbing the anthropogenic emissions (of NOx in particular) could have similar impact on surface ozone to perturbing the meteorological parameters that drive the production of ozone, temperature in particular. The MODEL experiment shows a much higher uncertainty than the others, with values between 15 and 35% over many continental areas. ALL shows values quite similar to that of MODEL over continents, but much higher over oceans (5-10% versus less than 5%), most likely because of the perturbation of ozone initial conditions done in ALL and not in MODEL.

At 120h forecast time, the growth in uncertainty is again significant, and impacting downstream regions in addition to source of ozone precursors regions. “Plumes” of higher uncertainty occur over the Atlantic and Pacific, arising from transport out of source of ozone precursors regions in North America and East Asia. It is interesting also to note that the uncertainty decreases within the ALL experiment over most of Southern oceans, as the impact of the perturbation of the initial conditions is much smaller at 120h forecast time than at 24h. This indicates that unlike for some meteorological variables, there is no error growth of surface ozone arising from perturbed initial conditions, at least over Southern oceans, which is far from ozone producing regions.

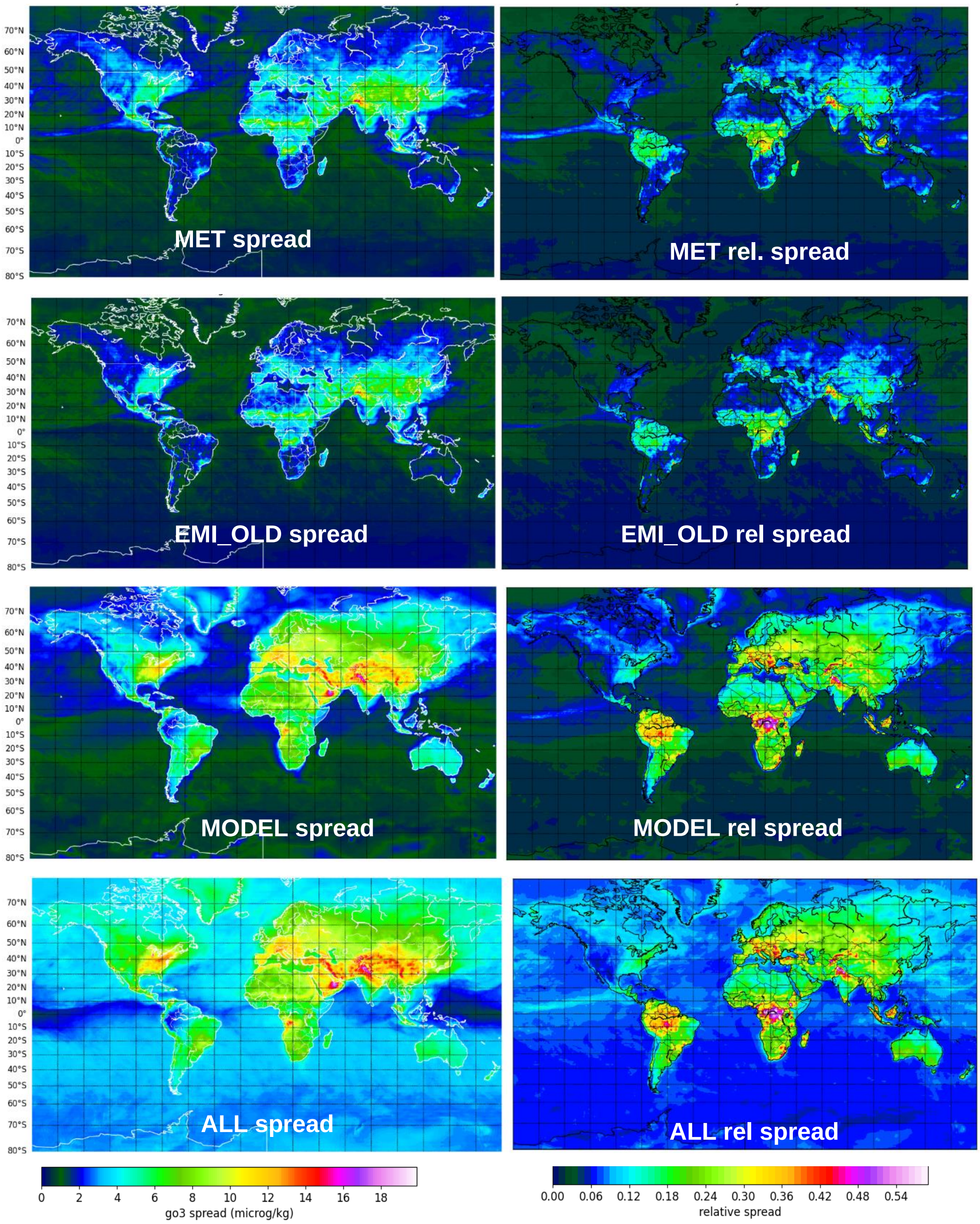

Figure 25: Spread (left) and relative spread (right) of the MET, EMI_OLD, MODEL and ALL experiments for simulated surface ozone in June 2021, at 24h forecast time.

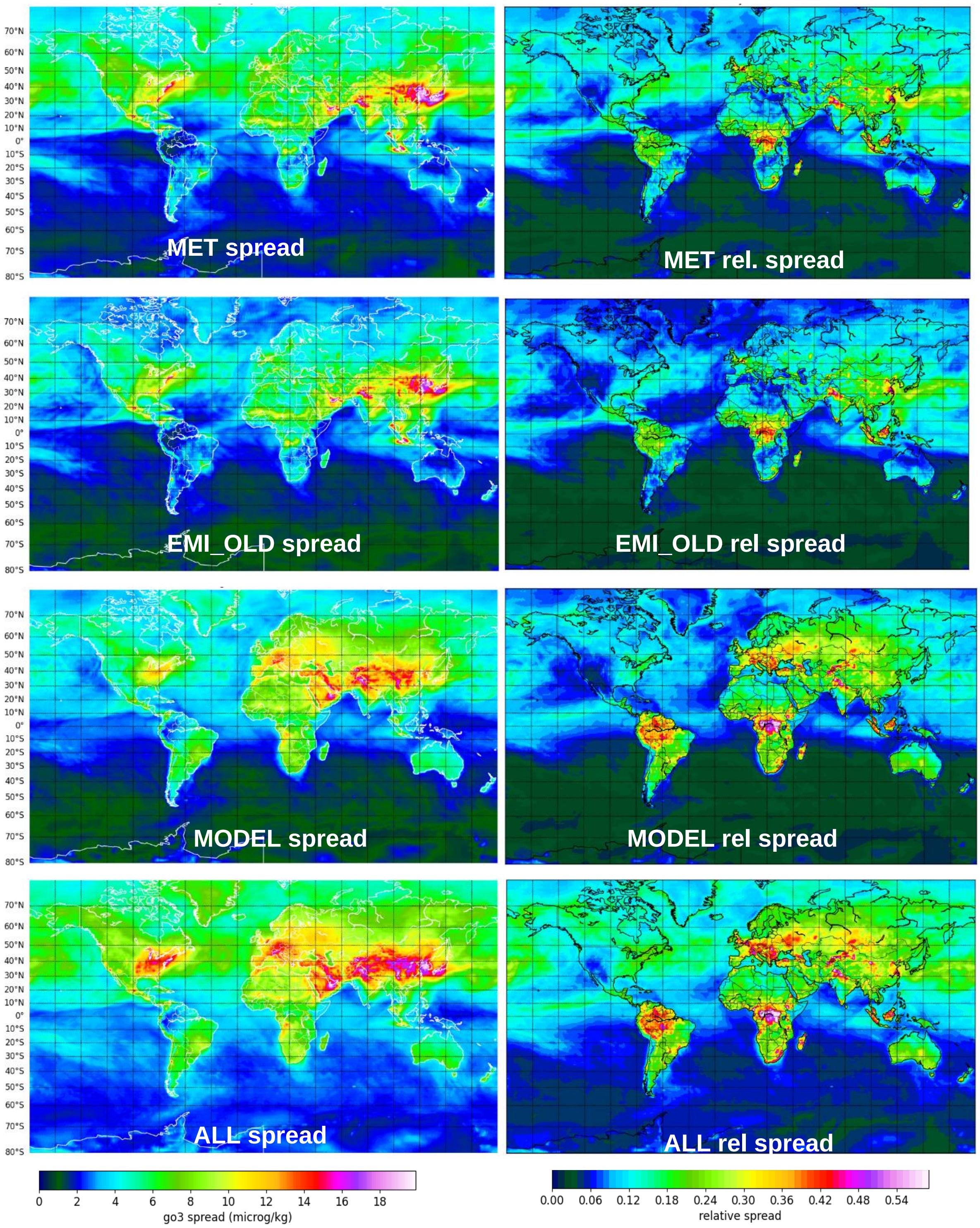

Figure 26: Spread (left) and relative spread (right) of the MET, EMI_OLD, MODEL and ALL experiments for simulated surface ozone in June 2021, at 120h forecast time.

5.2 Focus on the impact of emission perturbations

Figure 27 shows the relative spread of PM10 from three experiments perturbing only the anthropogenic emissions input:

- EMI, using the WP5 input to estimate emissions uncertainty,
- EMI_OLD, which uses a 0.25 standard deviation and 500km correlation length for perturbations,
- EMIS_OLD2, which uses a 0.5 standard deviation and 500km correlation length for perturbations.

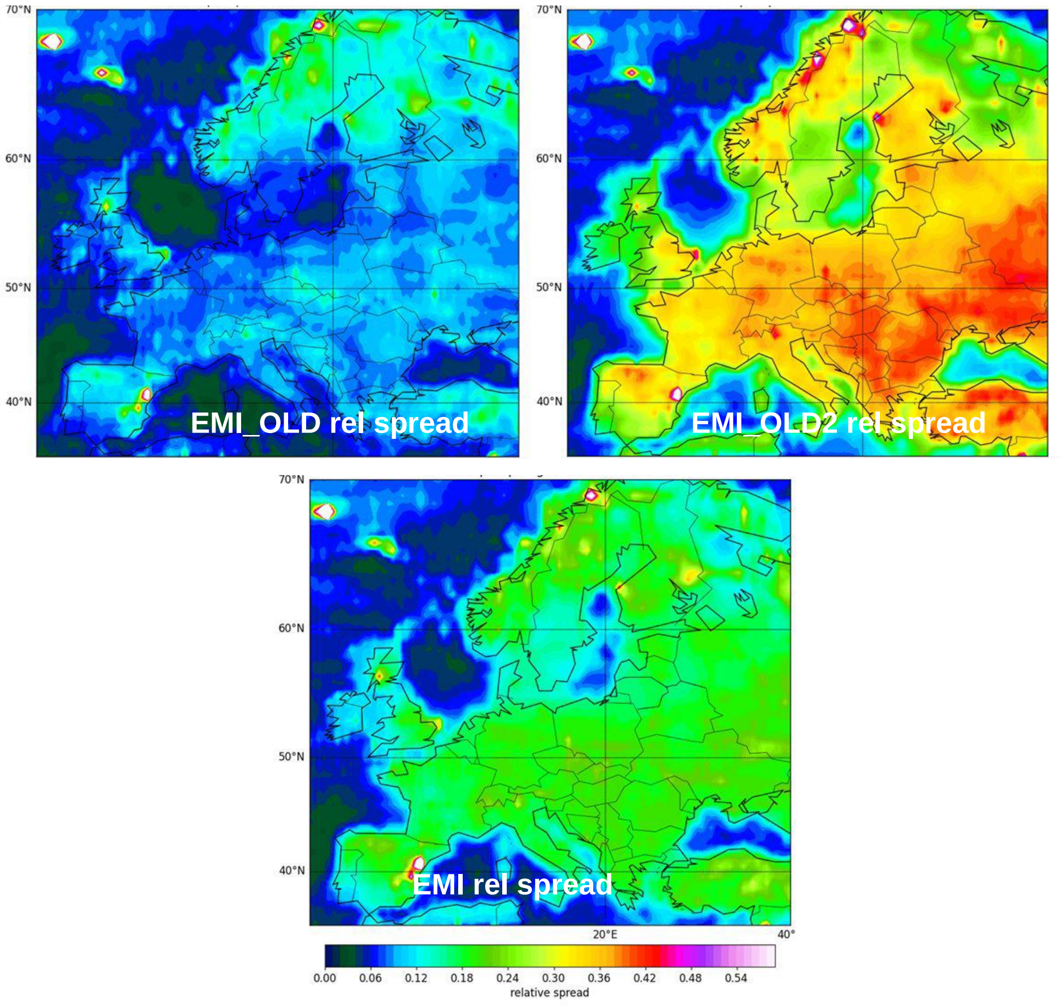

Figure 27: Relative spread of PM10 at 120h forecast time in February 2021 simulated by the EMI_OLD, EMI_OLD2 and EMI experiments.

The resulting spread of the three experiments shows the strong dependence between the characteristics of the perturbations applied to anthropogenic emissions and the resulting estimated uncertainty of simulated PM10. It is interesting to note that the standard deviation of PM10 after 120h forecast is lower than the standard deviation of the perturbations applied for all three experiments: for EMIS_OLD, values are between 10 and 20% over most of Europe, for EMIS_OLD2, between 30 and 40%, and for EMI as well, for which the standard deviation used to compute the perturbations was above 20% for the most emitting emission sectors, where 20% is the mean standard deviation of the ensemble over most of Europe from

Figure 27. Also, the simulated uncertainty appears much more homogeneous, at least for this particular month, when using the input of WP5. Finally, while the magnitude of the perturbations of the anthropogenic emissions is doubled between EMIS_OLD and EMIS_OLD2, the resulting uncertainty in PM10 is on average more than doubled.

5.3 Focus on the impact on selected meteorological variables

Finally, we decided to document the impact of some perturbations affecting atmospheric composition on the uncertainty/spread of two meteorological variables: the downward flux at surface of solar (visible) radiation (SSRD) and 2m temperature. This can help in quantifying the “mean” uncertainty in simulated 2m temperature arising from atmospheric composition uncertainties. Figure 28 shows the simulated uncertainty of SSRD and 2m temperature for MODEL, EMI and ALL. ALL includes meteorological perturbations which directly impact SSRD and 2m temperature, and is shown to provide a reference for the spread of 2m temperature and SSRD.

One one had, the impact of uncertainties of anthropogenic emissions is rather small in general, less than 2 W/m^2 for SSRD in general and less than 0.3-0.5K for 2m temperature. On the other hand, the impact of MODEL is quite significant, with large areas showing SSRD uncertainties of 5-10 W/m^2, and a 2m temperature uncertainty of 0.4-0.8K over most continents.

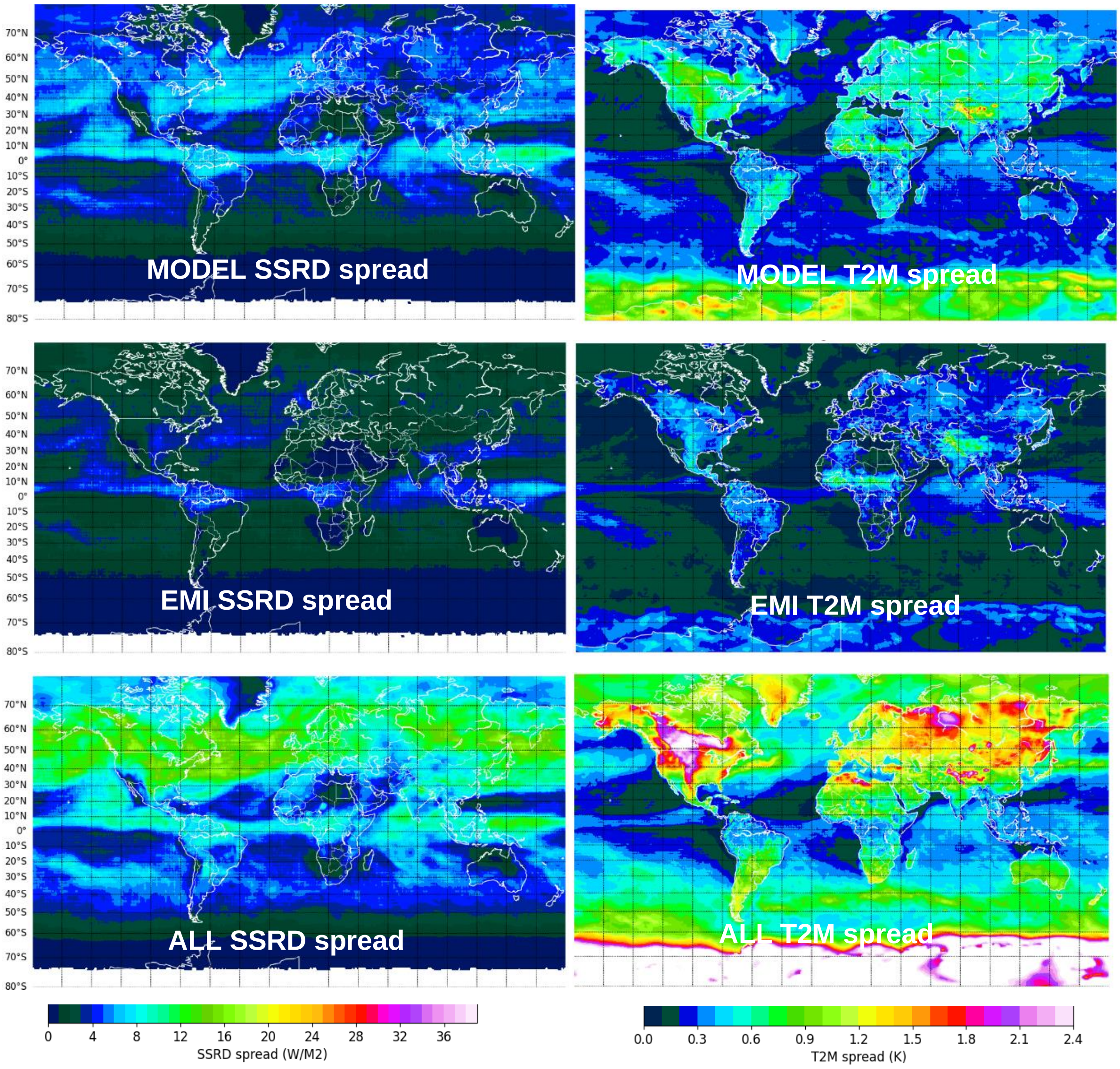

Figure 28: May 2021 average of relative spread of surface solar radiation downward flux (SSRD, left) and 2m temperature (T2M, right), of the MODEL, EMI and ALL experiments.

6 Conclusion

In this report we showed quantitative estimates of the uncertainties of a selection of CAMS global products, valid for 2021, which can give a qualitative assessment of uncertainties outside of 2021. The uncertainties vary a lot between species, forecast time, and the type of perturbations applied.

All sources of error combined, the uncertainty of most of products after 5 days forecasts is often in the range of 20-40%. Except for PM2.5, PM10 and surface NO2 (not shown), emissions in general bring relatively less uncertainty than model errors. This could also arise from the choice of perturbations applied.

This study also helps in evaluating the sensitivity of the IFS-COMPO system to uncertainties of different inputs and components of the model. The uncertainty of CAMS products was shown to be in general non-linear, i.e., the uncertainty from a combination of factors cannot be derived from the uncertainties of each factor taken separately.

These results were obtained with a first prototype of an IFS-COMPO ensemble. Diagnostics were developed to assess the skill of the ensemble as compared to observations and showed that all our ensemble simulations are under-dispersive: they fail to capture the whole variability of the observations, partly because of systematic model biases. This means that improving the skill of the model (debiasing, but also adding more degree of freedoms so as to better represent extremes) would allow for a lower uncertainty in the sense that fewer and or smaller perturbations would be needed to capture the observational variability More work is needed in order to derive perturbations that have more impact; inspiration could be drawn from the meteorological ensemble and the singular vector approach could be used for atmospheric composition perturbations.

Document History

Version	Author(s)	Date	Changes
1.0	Samuel Remy and Thierry Elias	12/06/2025	Initial version
1.1	Samuel Remy and Thierry Elias	27/06/2025	Corrected version after internal reviews

Internal Review History

Internal Reviewers	**Date**	**Comments**
Ingrid Super and Renske Timmermans (TNO)	19/6/2025	Mostly some textual remarks and clarifying questions; no major changes required.
Ana Carvalho (SMHI)	18/06/2025	Comments and suggestions.