ECMWF COPERNICUS REPORT

Copernicus Atmosphere Monitoring Service

Documentation for the Evaluation and Quality Control (EQC) of the CAMS regional services

Status 2022

Issued by: Norwegian Meteorological Institute

Date: 05/08/2022

Ref: CAMS283_2021SC1_D83.4.1.1-2022_202206_v2

European Commission

Copernicus
Europe's eyes on Earth

IMPLEMENTED BY ECMWF

This document has been produced in the context of the Copernicus Atmosphere Monitoring Service (CAMS).
The activities leading to these results have been contracted by the European Centre for Medium-Range Weather Forecasts, operator of CAMS on behalf on the European Union (Contribution Agreement signed on 22/07/2021). All information in this document is provided "as is" and no guarantee of warranty is given that the information is fit for any particular purpose.
The users thereof use the information at their sole risk and liability. For the avoidance of all doubt, the European Commission and the European Centre for Medium-Range Weather Forecasts have no liability in respect of this document, which is merely representing the author's view.

Contributors

METEO-FRANCE
V. Petiot
M. Joly
E. Blot
A. Royer

MET NORWAY
M. Gauss
K. S. Karlsen
L. Blake

KNMI
J. Douros
A. Tsikerdekis
H. J. Eskes

CNRS
Y. Bennouna
V. Thouret

INERIS
F. Meleux
B. Raux

Table of Contents

Introduction

As part of its routine Evaluation and Quality Control (EQC), CAMS evaluates the CAMS regional models and the ENSEMBLE. A large number of statistical skill scores, graphs and tables are provided to evaluate the models on a daily, weekly and seasonal basis against different types of observations (surface measurements, remote sensing, and airborne measurements).

Currently the results (graphs and reports) are published on the following web pages:

- **Operational evaluation** at https://regional.atmosphere.copernicus.eu/ (see items in right menu under 'VERIFICATION RESULTS'). Evaluation is done and updated daily, with graphs illustrating model performance for the last day, the last week and the last (up to 12) quarters. The pages have been developed by Meteo France, using the software package *Evaltools*, and are maintained throughout the transition period of CAMS2 83 (i.e. until April 2023);
- **Prototype evaluation** at https://cams2-83.aeroval.met.no/. Evaluation is done and updated daily, with graphs illustrating model performance for the last day, the last week and the last quarters. These pages are under development at MET Norway and will be fully operational from 2023 (they will not be described further in this report);
- **Quarterly reports** on the evaluation of NRT products (forecast and analysis) are found at https://atmosphere.copernicus.eu/regional-services. They are created by MET Norway, Meteo France, KNMI, CNRS and FMI every three months for the previous meteorological season;
- **Annual reports** on the evaluation of the Interim and Validated Reanalysis are provided by INERIS (with contributions from KNMI).

This document provides information on how the various statistical skill scores are calculated and how the graphs are created. It will be updated regularly as the methodologies evolve, and based on user feedback.

Section 1 deals with the surface evaluation, while Section 2 describes the methodologies used for above-surface evaluation. Examples of plots are given in both sections to illustrate how the results are visualized on the web pages and in the *quarterly* reports.

1. Surface Evaluation

1.1 Observations used for statistics

The reference observations used for statistics are hourly in-situ surface observations acquired daily from the European Environment Agency (EEA). These observations can be acquired at https://discomap.eea.europa.eu/map/fme/AirQualityUTDExport.htm.

Only measurements that are considered representative of background air pollution, which is the scale that the models are able to simulate (i.e. not road-scale pollution), are kept. To operate such a filter, we select background stations that are classified from 1 to 7 according to the Joly and Peuch [2012] classification. The latest version of this selection is from March 2022 and can be downloaded at https://opensource.umr-cnrm.fr/attachments/4364/2022_update_listing.csv.

In addition, negative observations as well as observations above a certain threshold are considered aberrant and are removed. These thresholds differ according to the pollutant (O_3: 500, NO_2: 700, SO_2: 1200, CO: 15000, PM_{10}: 1000, $PM_{2.5}$: 700 µg.m^{-3}).

For the evaluation of analysis products, the split between the set of assimilation stations and the set of verification stations is performed by the CAMS regional service provider.

Special cases are CO and SO_2, for which *all* background stations that are classified from 1 to 7 according to the Joly and Peuch [2012] classification are used by models in the assimilation. Therefore, for these two species, verification is performed using also those observations that have been assimilated. For O_3, NO_2, PM_{10} and $PM_{2.5}$, the station list used for assimilation is based on the so-called *set14* built by INERIS for the CAMS interim reanalysis (last updated in March 2022).

1.2 Statistics used

Five statistical parameters are available:

- mean bias
- modified mean bias (MMB)
- root mean square error (RMSE)
- fractional gross error (FGE)
- correlation

The normalized scores (MMB, FGE and correlation) are independent of the mean concentrations of the pollutants. Therefore, these scores, as obtained for different pollutants and different seasons, are easier to compare.

In the following, let $M = (M_i)_{1 \leq i \leq n}$ be a vector of *n* modelled values and $O = (O_i)_{1 \leq i \leq n}$ the corresponding vector of *n* observed values (where the (O_j, M_j) pair is removed if M_j and/or O_j is missing).

1.2.1 Mean Bias

$$MeanBias = \bar{M} - \bar{O}$$

The mean bias is the average difference between the modelled and the observed values. A positive value, for example, means that the forecasts are on average higher than the observations. The aim is to be as close to 0 as possible.

1.2.2 Modified Mean Bias

$$MMB = \frac{2}{n} \sum_i \frac{M_i - O_i}{M_i + O_i}$$

The Modified (Normalized) Mean Bias (MMB, also called MNMB) is normalized by the mean of the observed and modelled values. This statistic ranges between -2 and 2. The aim is to be as close to 0 as possible.

1.2.3 Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_i (M_i - O_i)^2}$$

RMSE measures the standard deviation of the differences between the modelled and the observed values. The aim is to be as close to 0 as possible.

1.2.4 Fractional Gross Error (FGE)

$$FGE = \frac{2}{n} \sum_i \left| \frac{M_i - O_i}{M_i + O_i} \right|$$

This is a normalized version of the mean error, based on absolute values (instead of squared values, as in the normalized mean squared error). The FGE ranges between 0 and 2. The aim is to be as close to 0 as possible.

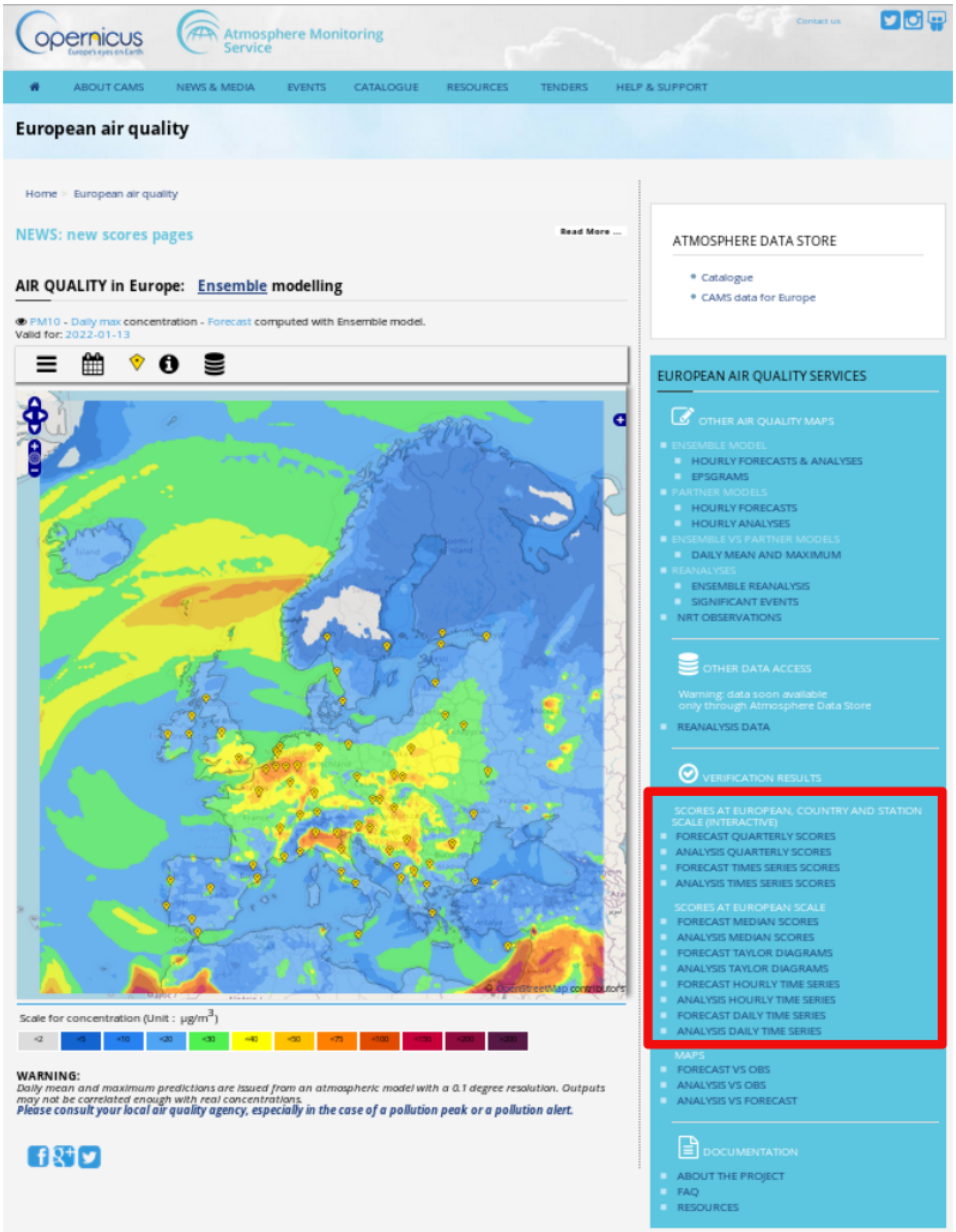
1.2.5 Correlation

$$Correlation = \frac{cov(O,M)}{\sigma_O \, \sigma_M} = \frac{\sum_i (O_i - \bar{O})\,(M_i - \bar{M})}{\sqrt{\sum_i (O_i - \bar{O})^2}\,\sqrt{\sum_i (M_i - \bar{M})^2}}$$

Correlation refers to the extent to which the modelled and the observed values have a linear relationship with each other. The correlation is between -1 and 1. The aim is to be as close to 1 as possible.

1.3 Graphs shown on the web

The *Verification Plots* provided on the CAMS regional web site – *Median Scores*, *Time Series*, and *Taylor Diagrams* – aim to help the user to assess at a glance the performance of the different models and the Ensemble (median of the models), using several statistical indicators. Verification Plots are calculated daily by comparing the forecasted values to in-situ observations at a selection of European monitoring sites (representative of the CAMS regional scale), using the various statistics described in Section 1.2.



European air quality

NEWS: new scores pages

AIR QUALITY in Europe: Ensemble modelling

PM10 - Daily max concentration - Forecast computed with Ensemble model.
Valid for: 2022-01-13

Scale for concentration (Unit : µg/m^3)

Warning: data soon available only through Atmosphere Data Store

WARNING:
Daily mean and maximum predictions are issued from an atmospheric model with a 0.1 degree resolution. Outputs may not be correlated enough with real concentrations.
Please consult your local air quality agency, especially in the case of a pollution peak or a pollution alert.

Figure 1 - Home page of the CAMS regional website with the verification section showed within the red box.

When looking at the graphs from *SCORES AT EUROPEAN SCALE*, the user can choose between the following parameters:

- Forecast Base Time: refers to the beginning of the last forecast taken into account to compute the statistics.
- Parameter: gives the choice between six pollutants (ozone, nitrogen dioxide, sulfur dioxide, carbon monoxide, particulate matter below 10 μm, particulate matter below 2.5 μm).
- Scores: refers to different statistical scores described in Section 1.2.
- Type (only for the *Time series*): to choose whether the statistics are computed from daily mean or daily maximum concentrations.

1.3.1 Median scores

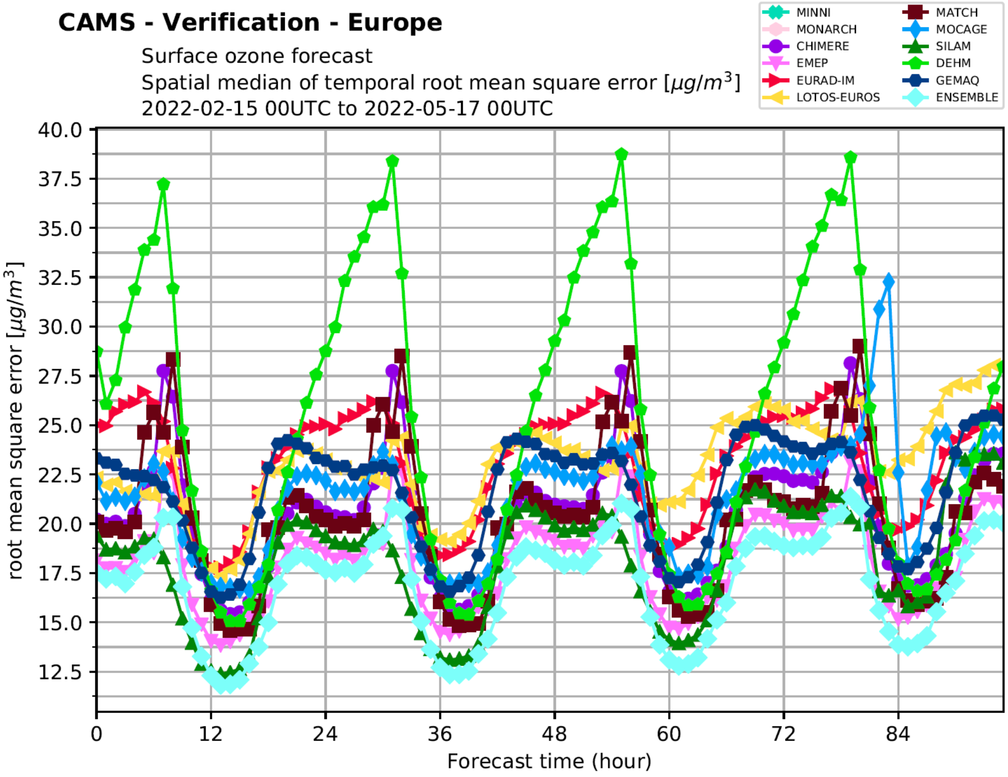


Figure 2 - Example of a median score plot. Temporal scores are shown for each model as function of forecast hour.

For each model, the different statistics used in the regional near real time evaluation are computed for each hourly time-step of the daily 4-day forecasts. *Median Scores* graphs thus display the hourly evolution of the chosen statistics as a function of the forecast hour from 0 to 95 on the graph. They are helpful to assess the daily cycle of models.

More precisely, scores are first computed for each station $s \in S$ and each forecast hour $h \in \{0, \dots, 95\}$

$$Score_{h,s} = Score\left(\left(M_{d,h,s}\right)_{d\in\{1,\dots,D\}}, \left(O_{d,h,s}\right)_{d\in\{1,\dots,D\}}\right)$$

where $\{1, \dots, D\}$ are all the days of the studied period[1].

Then, the median of the scores is computed for each forecast hour to display scores as a function of the forecast hour:

$$y(h) = median_s\left(Score_{h,s}\right)$$

In other words, *Median Score* graphs show the agreement between the observed and forecasted temporal patterns for each forecast hour.

1.3.2 Time series

Time Series (see example in Figure 3) shows – for each day of the considered period (last week or last 3 months) – the evolution of the chosen score between the observed and the forecasted fields of the first day of the model forecasts. In case of missing data (model or observations), there will be a gap in the time series. For this chart, the user can choose the type of values (daily mean or daily maximum concentrations) for which the statistics are computed. When choosing daily maximum, the *maximum* of the model is compared to the *maximum* of the observations even if they do not occur at the same time.

[1] The DJF season has 90 or 91 days, MAM has 92 days, JJA has 92 days, and SON has 91 days. But for the daily updated charts on the website, the last 91 days are used.

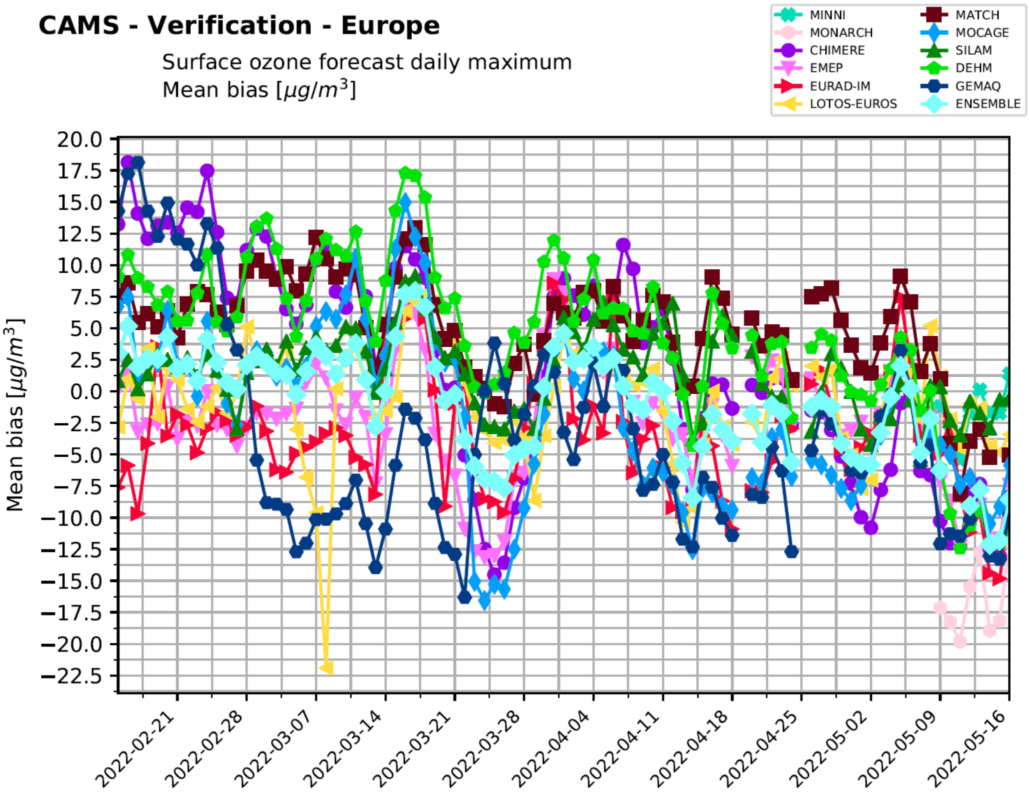


Figure 3 - Example of a time series for RMSE. Time series show spatial scores (for the 1st day of the forecast) as a function of calendar day.

1.3.3 Hourly time series

Hourly time series are useful to assess recent model forecasts hour by hour. The user can choose which forecast day to display (D0, D1, D2 or D4). In each of these charts, only a period of three days is available for more readability. In case of missing data (model or observations), there will be a gap in the time series.

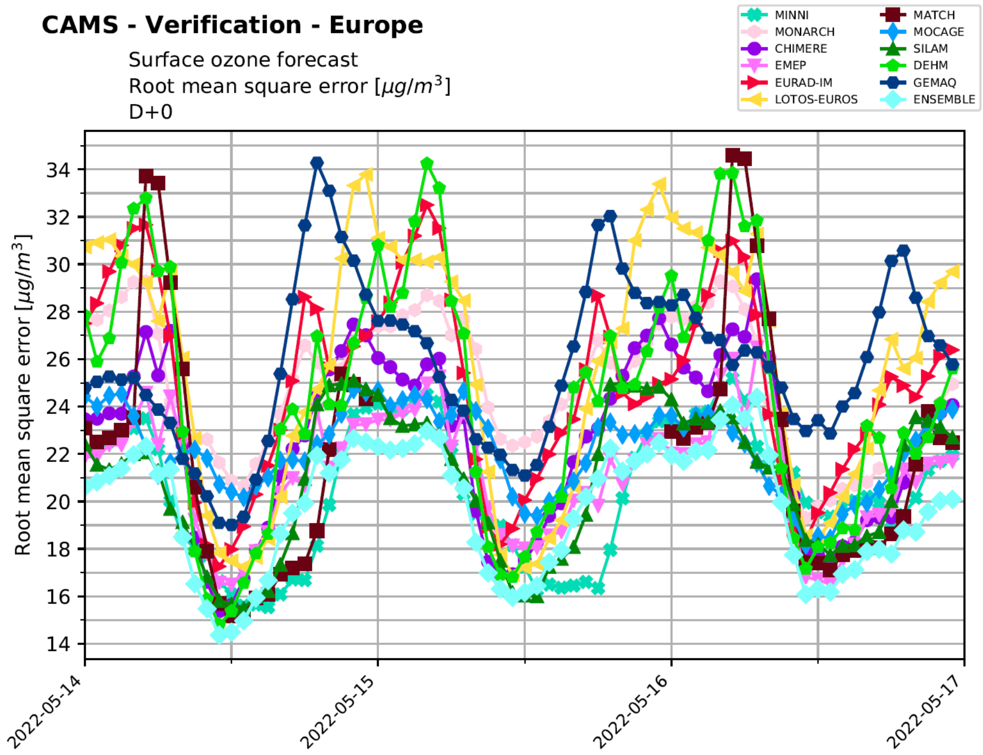


Figure 4 - Example of an hourly time series for RMSE. Time series show spatial scores as a function of calendar day are shown for the 1st, 2nd, 3rd or 4th day of the forecast (to be chosen by the user). The example shows scores for the 1st day (’D+0’).

1.3.4 Taylor diagrams

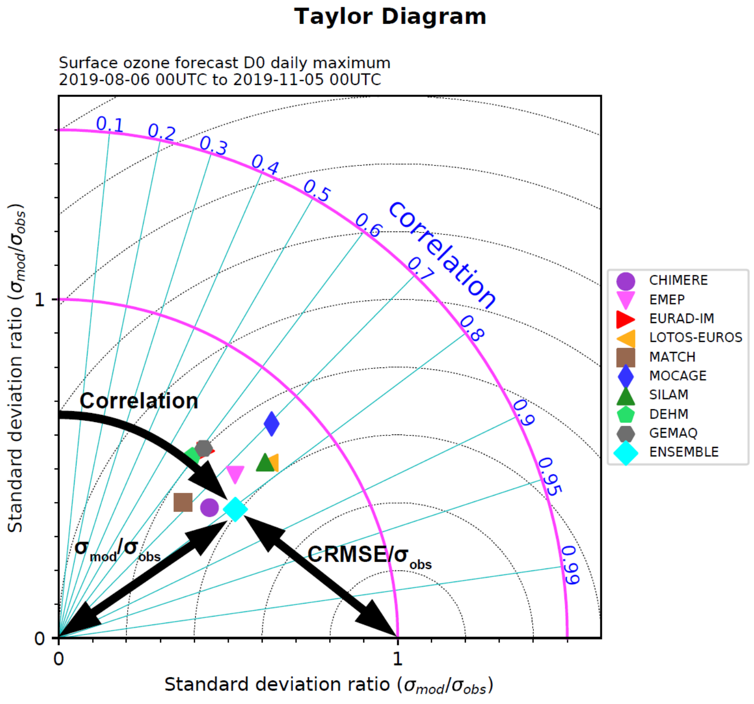


Figure 5 – Example of a Taylor diagram.

Taylor diagrams combine – for the chosen period (last week or last 3 months) – three statistics simultaneously (for the first day of forecast only): the CRMSE (centralized, or unbiased, RMSE), the correlation, and the ratio of the standard deviations of the observed and modelled values. These statistics are computed from daily maximum concentrations for ozone and nitrogen dioxide, and from daily mean concentrations for other species. Before computation, daily concentration time series obtained at each monitoring site over the considered period are concatenated.

1.3.5 Interactive scores

Quarterly scores

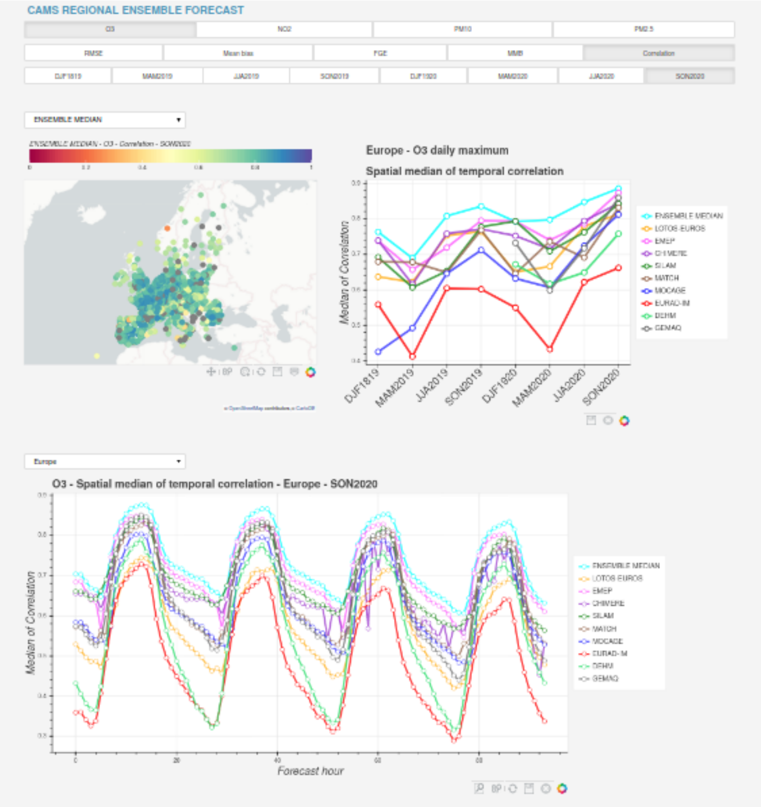


Figure 6 - CAMS regional interactive web page showing scores for the last eight quarters.

Since January 2021, a new interactive layout is available. Scores displayed here are recalculated with a consistent station listing on the last 8 quarters and updated each quarter.

On the map, stations are coloured according to the current selected score. This score is computed temporally from hourly concentrations of the selected quarter.

The figure to the right of the map displays the spatial median of temporal scores of the first day of the forecast computed from daily maximum concentration values for ozone and nitrogen dioxide, and from daily mean concentration values for particulate matter.

If several stations are selected, scores displayed on the figure below the map correspond to median scores described in Section 1.3.1. If a single station is selected, they are simply scores computed for each hourly time-step of the daily 4-day forecasts.

Time series of concentration and scores

Since January 2022, time series of concentration and scores can be investigated at station and country level (see Figure 7).

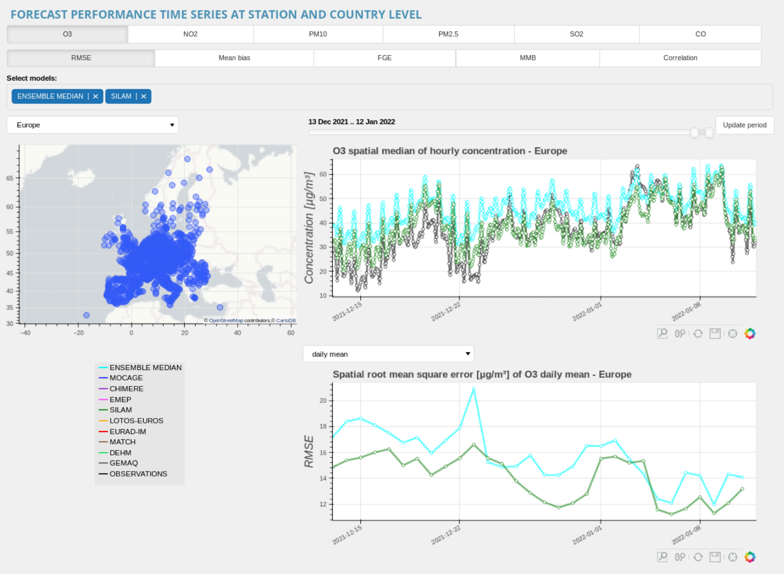


Figure 7 - CAMS regional interactive web page showing time series of scores.

Curves in the top figure show concentration values for observations and selected models. If a country is selected or all European station are selected, the curves correspond to the median of concentrations.

Curves in the bottom figure show scores computed from daily mean or daily maximum concentration values. These scores are computed with the same method as for the Time series plots (Section 1.3.2).

1.4 Graphs shown in the EQC reports

A large number of plots is created on a quarterly or annual basis for the Evaluation and Quality Control (EQC) reports. Some of them are of similar type as the ones shown on the web, while others are (as of 2022) shown only in the quarterly and annual reports. In the following sub-sections, we describe how they are produced.

1.4.1 Quarterly time series of RMSE

For each pollutant, the first chart presented in the EQC reports displays the root-mean square error of daily maximum (for ozone, NO_2, SO_2 and CO) or of daily mean (PM_{10} and $PM_{2.5}$) for the first-day forecasts with regard to surface observations, for every quarter during the last 3 years.

The RMSE is computed for each station along all days of the period:

$$\forall\, s \in S, err_s = RMSE\left(\left(M_{d,s}\right)_{d\in\{1,\dots,D\}}, \left(O_{d,s}\right)_{d\in\{1,\dots,D\}}\right)$$

and then the median is taken:

$$KPI = median((err)_{s\in S})$$

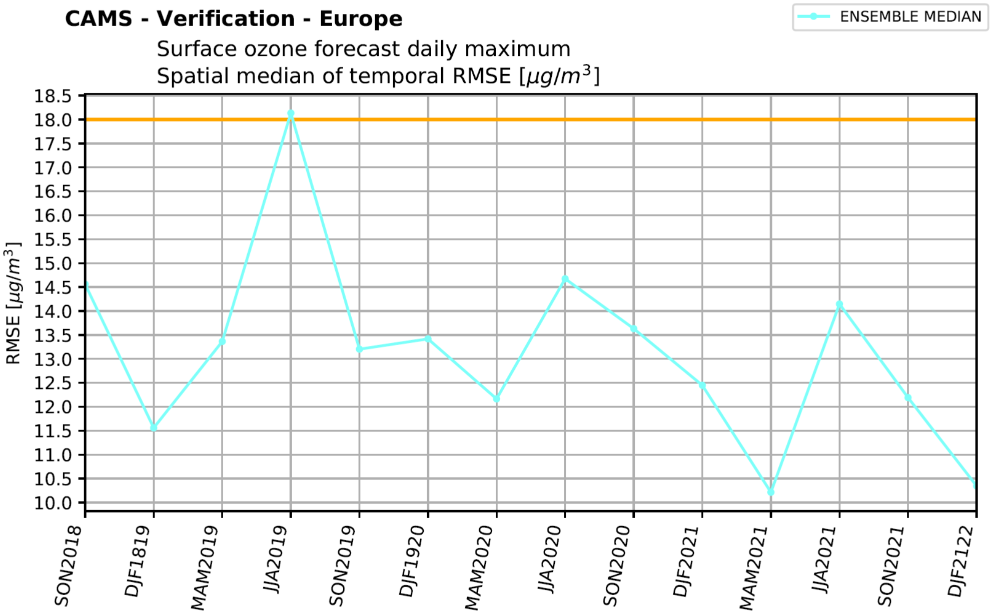


Figure 8 - Example of quarterly RMSE for the ENSEMBLE ozone forecasts as included in the EQC reports.

A target reference value is indicated as an orange line (which was defined for ozone, NO_2 and PM, but not yet for SO_2 and CO). These values are listed in Table 1.

Table 1: RMSE target reference values.

	Forecast	Analysis
O_3 daily max	18 µg/m³	16 µg/m³
NO_2 daily max	25 µg/m³	22 µg/m³
PM10 daily mean	18 µg/m³	16 µg/m³
PM2.5 daily mean	18 µg/m³	16 µg/m³

1.4.2 Median scores

The three next charts presented in the EQC reports correspond to RMSE, MMB and correlation as a function of forecast hour. The computation method used is the same as in Section 1.3.1.

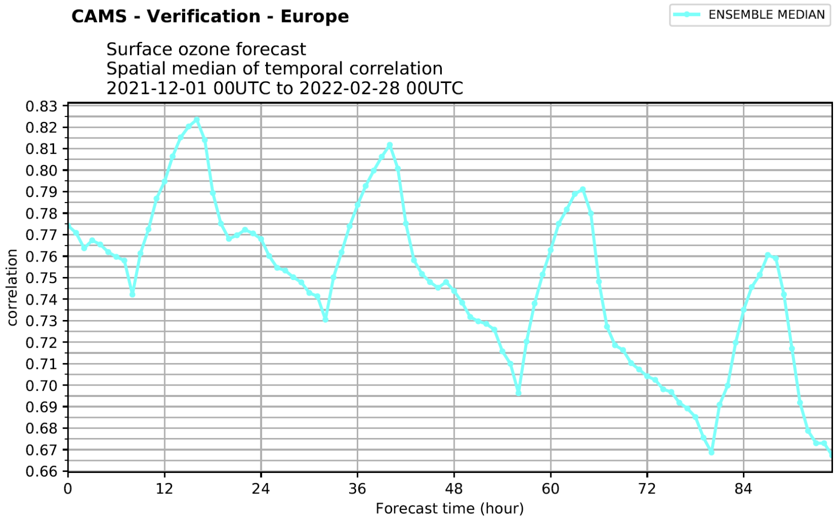


Figure 9 - Example of correlation as a function of the forecast hour for the ENSEMBLE ozone forecasts as included in the EQC reports.

1.4.3 Time series

In addition, EQC reports show the daily RMSE, computed over all stations of the domain from daily concentrations (daily max for ozone and NO_2, and daily mean for PM, CO and SO_2), for the ENSEMBLE forecasts, for all days of the period. This allows for an easier detection of potentially outlying days in the period. The computation method used is the same as used for the time series plots in Section 1.3.2.

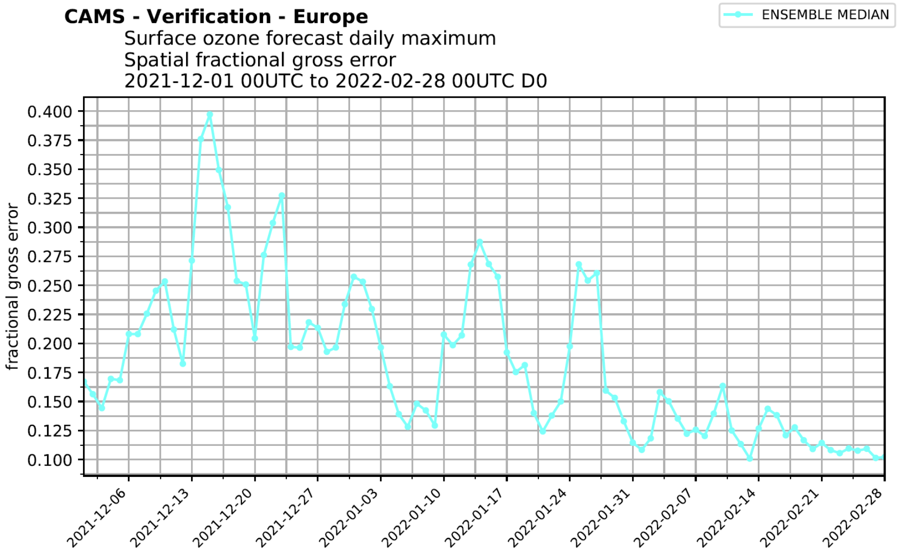


Figure 10 - Example of FGE time series for the ENSEMBLE ozone forecasts as included in the EQC reports.

1.4.4 FAIRMODE plots

In EQC reports, the chapter called *FAIRMODE diagrams* shows evaluation plots as suggested by the Forum for Air Quality Modeling (FAIRMODE). More specifically, skill scores of the regional forecasts and analyses of CAMS are visualized in so-called *Target diagrams* and *Summary reports*.

Details on how to calculate the metrics that form the basis of these plots (hereafter referred to as 'FAIRMODE plots') can be found in Janssen et al. [2022]. Here we only briefly summarize some basic information necessary to understand the FAIRMODE plots.

The *Model Quality Indicator* at a given station is defined as:

$$MQI = \frac{RMSE}{\beta RMS_u}$$

where RMS$_u$ is the measurement uncertainty, RMSE is the root mean square error of the model at the station, and β is a factor chosen (by FAIRMODE) to be equal to 2 in coordinance withThunis et al. [2013] and Pernigotti et al. [2013]. The *Model Quality Objective* (MQO) is considered fulfilled when MQI ≤1, i.e. when the error of the model is equal or lower than twice the uncertainty of the observations. The derivation of the measurement uncertainty RMS$_u$ is given by Janssen et al. [2022] and depends on several measurement parameters (their values are specified on each plot). Equation (36) of Janssen et al. [2022] provides the uncertainty of the observation O_i, however a more detailed explanation is provided below for clarity and completeness.

Computing observation uncertainties The RMS_u is defined as

$$RMS_u = \sqrt{\frac{1}{N}\sum_{i=1}^{N} U\,(O_i)^2}$$

where $U(O_i) = U_r(RV)\sqrt{(1-\alpha^2)O_i^2 + \alpha^2 RV^2}$ (Janssen et al., 2022, their Equation 37) is the uncertainty associated to observation O_i. This expression implies

$$\begin{aligned} U^2(O_i) &= U_r^2(RV)\alpha^2 RV^2 + U_r^2(RV)(1-\alpha^2)O_i^2 \\ &=: A^2 + B^2 O_i^2, \end{aligned}$$

where $A = U_r(RV)\alpha RV$, $B = U_r(RV)\sqrt{1-\alpha^2}$. The parameters have very clear physical meaning: they are absolute and relative errors assigned to the observations. Janssen et al. (2022, their Table 7) provide a list of parameters used to calculate the measurement uncertainty and is included below as Table 2.

Table 2: Parameters used to calculate the assigned measurement uncertainty.

	$U_r(RV)$	RV	α	N_p	N_{np}
NO_2	0.24	200 µg/m^3	0.20	5.2	5.5
O_3	0.18	120 µg/m^3	0.79	11	3
PM_{10}	0.28	50 µg/m^3	0.25	20	1.5
$PM_{2.5}$	0.36	25 µg/m^3	0.50	20	1.5

Target diagrams In the Target diagrams, each station is represented by a symbol (squares for *urban*, circles for *rural*, and triangles for *sub-urban* stations), with the bias and the centered (bias-free) RMSE (=CRMSE) shown on the vertical and horizontal axes, respectively. Both the bias and CRMSE are normalized by β times the measurement uncertainty.

As the centered RMSE is always positive, additional information can be conveyed by either placing the station to the left or the right-hand side of the origin. If the correlation error (phase error) dominates, the station is placed on the left, while if the standard deviation error (pattern error) dominates, the station is placed on the right.

It can be shown mathematically that the distance of the station from the origin, when plotted in this way, is equal to the RMSE normalized by β times the measurement uncertainty, i.e. the MQI as defined by FAIRMODE. The model quality objective (MQO) is thus fulfilled when the station is within the grey circle of the Target diagram, and the model is said to be fit for purpose (in the FAIRMODE context) if at least 90% of all stations have MQI ≤ 1, i.e. are placed within the grey circle. MQI_{90} (i.e. the 90th percentile of the MQI) is given to the right of the plot. When $MQI_{90} \leq 1$ the model is considered to be fit for purpose for the species in question (e.g. forecast of the ozone daily maximum).

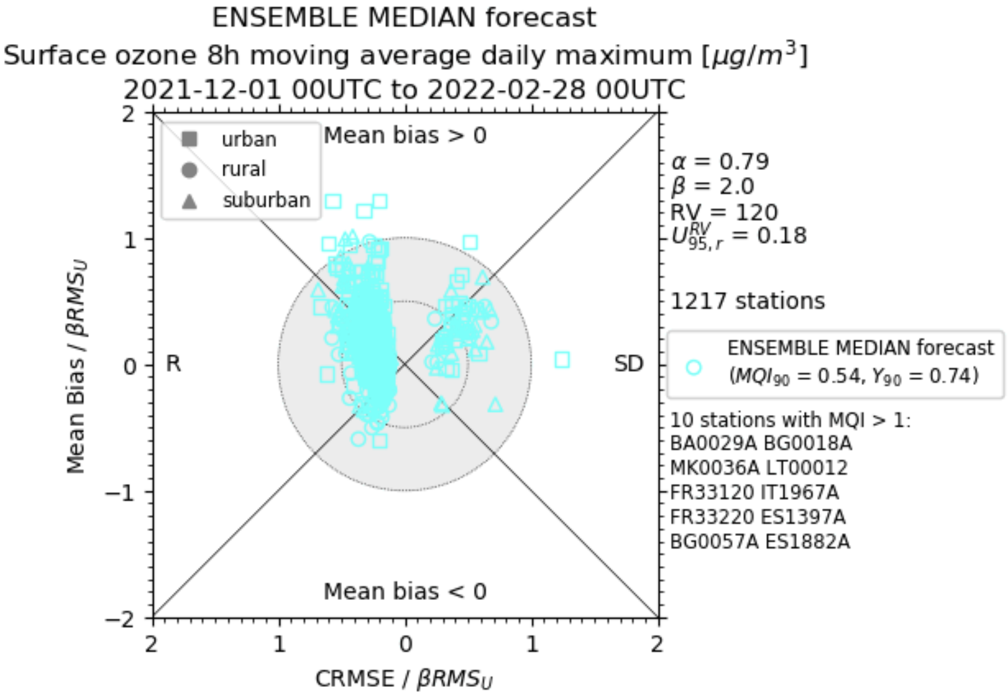


Figure 11 - Example of a FAIRMODE *Target Diagram* for the ENSEMBLE ozone forecasts.

Summary plots The summary plots provide complementary information:

An indicator computed on seasonally averaged model results is also shown on target plots, called Y_{90}. It is the 90th percentile of the MQI defined as the mean bias between modelled (M) and observed (O) seasonally averaged concentrations, normalized by β times the expanded measurement uncertainty, U_{95}, of the mean concentration:

$$MQI = \frac{|\bar{O} - \bar{M}|}{\beta U_{95}(\bar{O})}$$

As in the case of MQI_{90}, the value for Y_{90} should be lower or equal to 1.

A second view of results is provided by the summary report, describing several statistics to better determine the model performances. It is a complementary source of information to MQI.

The summary report is structured as follows (adapted from Janssen et al. [2022]):

- Rows 1 and 2 provide the measured observed seasonal means calculated from the hourly values, and the number of exceedances for the selected stations. The threshold values for calculating the exceedances are set to 50, 200 and and 120 g/m^3 for the daily PM_{10}, the hourly NO_2 and the 8h daily ozone maximum, respectively. For other variables ($PM_{2.5}$, etc.) no exceedances are shown.
- Rows 3 to 6 provide an overview of the temporal statistics for bias (row 3), correlation (row 4) and standard deviation (row 5) as well as information on the ability of the model to capture the highest range of concentration values (row 6). The fourth indicator represents the capability of the model to reproduce extreme events, i.e. concentrations above a certain percentile. It is calculated as

  $$H_{perc} = \frac{O_{perc} - M_{perc}}{\beta U(O_{perc})}$$

  with percentiles ('perc') chosen according to the legislation: 92.9% for ozone daily maximum 8-h mean (concentrations should remain below or equal the target value for at least 340 days) and 90.4% for PM_{10} daily mean (concentrations should remain below or equal the limit value for at least 330 days). Note that for the correlation a normalised indicator based on "1 – correlation" is plotted. A value close to zero is thus obtained at stations where the model is excellent. Each point represents a specific station. For each indicator, the model is considered to meet the performance criterion at stations that lie within the colored (green or orange) shaded areas. If a point falls within the orange shaded area the error associated with the particular statistical indicator is dominant. Note that fulfilment of the bias, correlation, standard deviation and high percentile related indicators does not guarantee that the overall MQO based on the MQI (or RMSE, visible in the Target diagram) is fulfilled.
- Rows 7 and 8 provide an overview of statistics for spatial correlation and standard deviation. Average concentrations over the selected time period are first calculated for each station and these values are then used to compute the averaged spatial correlation and standard

deviation. As a result, only one point representing the spatial correlation of all selected stations is plotted. Color shading follows the same format as for rows 3-5.

For the indicators in rows 3 to 8, values beyond the proposed scale will be represented by the station symbol being plotted in the middle of the dashed zone on the right/left side of the proposed scale.

For all indicators, the colored circle on the right-hand side provides information on the number of stations fulfilling the performance criteria: the circle is colored green if more than 90% of the stations fulfil the criterion (model fit for purpose), and red if the number of stations is lower than 90%.

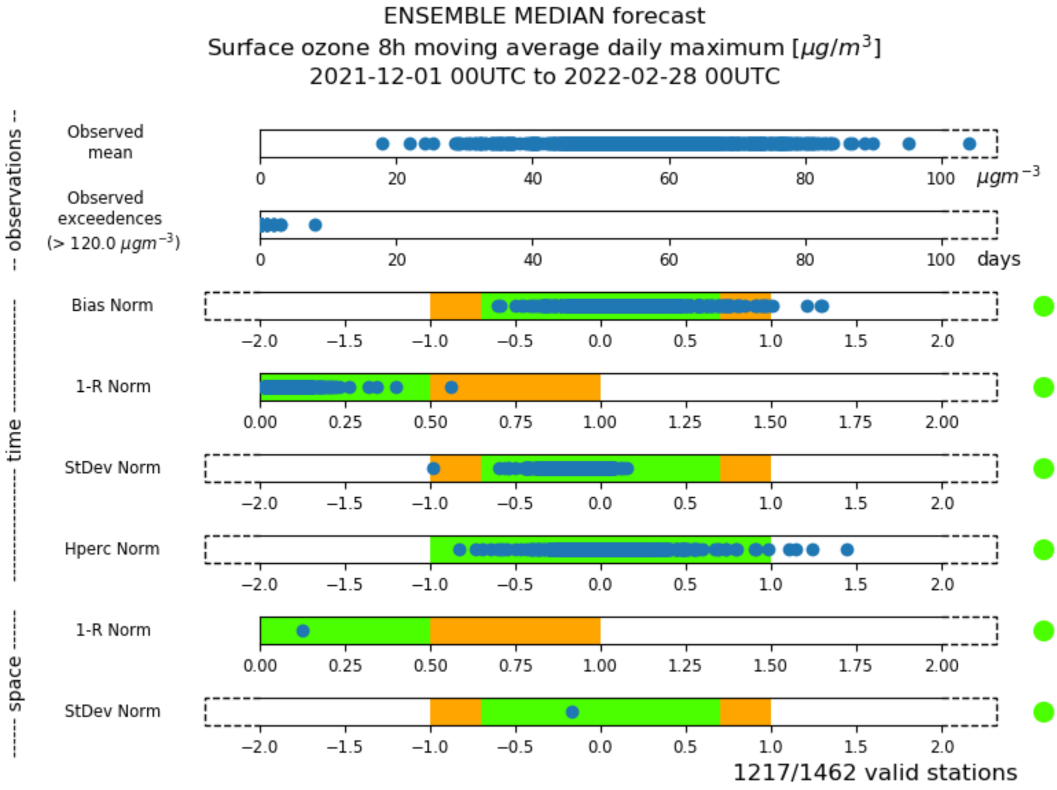


Figure 12 - Example of a FAIRMODE *Summary Report* for the ENSEMBLE ozone forecasts.

Station screening For the FAIRMODE diagrams produced for the quarterly evaluation reports, a minimum of data availability is required for statistics to be produced at a given station. Presently the requested percentage of available data over the selected period is 75%. Statistics for a single station are only produced when data availability of paired modelled and observed data is for at least 75% of the time period considered. When time averaging operations are performed the same availability criteria of 75% applies. For example, daily averages will be performed only if data for at least 18 hours are available. Similarly, an 8-hour average value for calculating the O_3 daily maximum 8-hour means is only calculated for the 8-hour periods in which at least 6 hourly values are available.

1.4.5 Performance diagrams

In the annual evaluation reports (for the interim and the validated reanalyses), the ability of the models to detect threshold exceedances is assessed via performance diagrams. Those plots are a convenient way to show the contents of the contingency table, which compares observations and simulation forecasts regarding a threshold value.

	obs>thr	obs<thr	Total
sim>thr	a	b	a+b
sim<thr	c	d	c+d
Total	a+c	b+d	a+b+c+d

The “a” cell is called “Good Detections”, and represents the number of times where observations and simulations agreed on a threshold exceedance.

The “b” cell is called “False Alarms”, and represents the number of times where simulations wrongly detected a threshold exceedance.

The “c” cell is called “Missed Detection”, and represents the number of times where simulations wrongly detected a threshold non-exceedance

The “d” cell is sometimes called “Good Detections (-)”, and represents the number of times where observations and simulations agreed on a threshold non-exceedance.

Considering these values, several indicators may be computed:

- POD, or Probability of Detection: It is the ratio of good detections above the total number of observed exceedances.

$$POD = \frac{a}{a+c}$$

- SR, or Success Ratio: It is the ratio of good detections above the total number of exceedances detected by the simulations.

$$SR = \frac{a}{a+b}$$

- FB, or Frequency Bias: It is the forecast bias of threshold exceedances. When inferior to 1, the simulation tends to produce more missed detections than false alarms.

$$FB = \frac{a+b}{a+c}$$

- CSI, or Critical Success Index: It is the ratio of good detections above the total number of predicted and missed events.

$$SR = \frac{a}{a + b + c}$$

The performance diagram combines these indicators on a single scatter plot, with Success Ratio as abscissa, Probability of Detection as ordinate, Critical Success Index as colored background, and Frequency Bias as dashed lines originating from zero.

The closer a model or simulation is to the upper right corner, the better skills it shows for threshold exceedances.

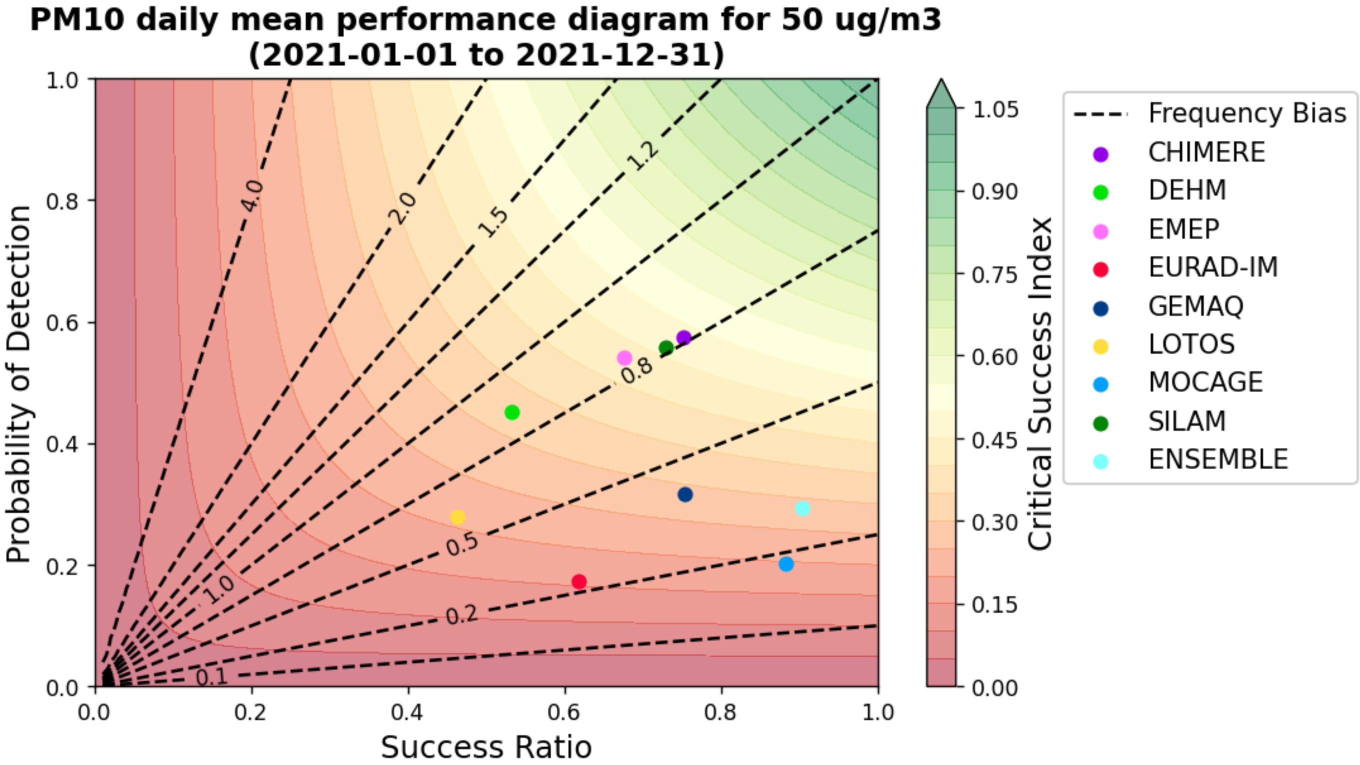


Figure 13 - Example of a performance diagram.

1.4.6 Summary plots

When handling similar simulations, especially for the runs of a model over several years, evaluation reports use summary plots to compare results in a global view. They consist of a bar plot of RMSE, completed by lollipop plots of correlation and mean bias. RMSE and bias share a common y-axis, while correlation has its own y-axis on the right side of the plot.

In the annual evaluation reports (for the interim and the validated reanalyses), all indicators are computed for each model – first for each station over the whole period, then averaged over all stations.

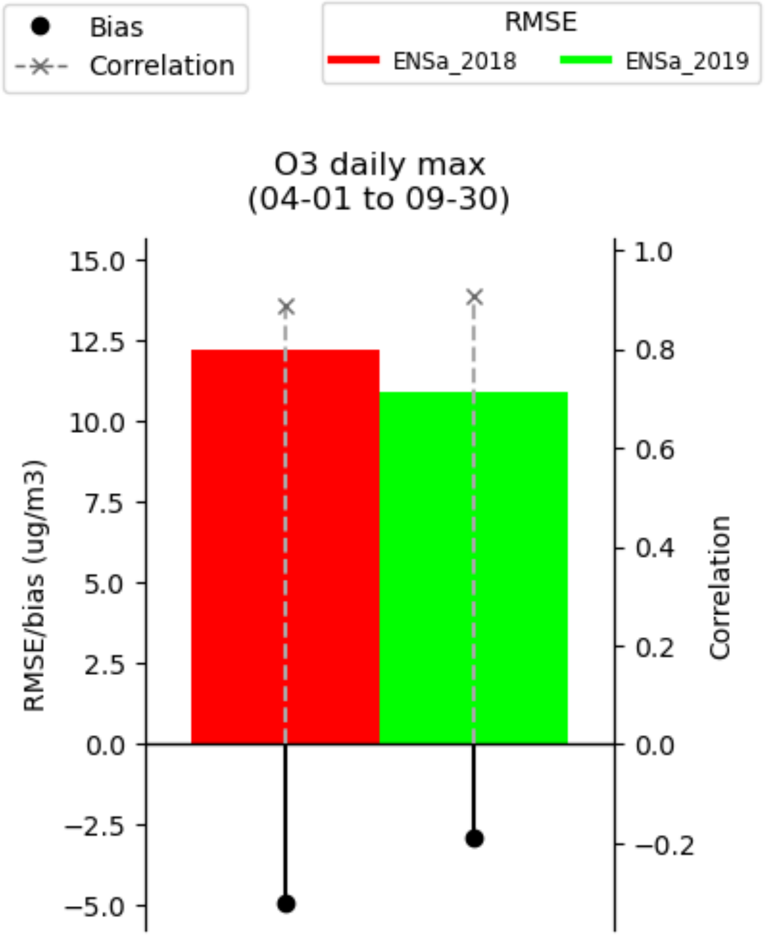


Figure 14 - Example of a summary plot.

2. Above-surface Evaluation

2.1 Ozone sondes

Modelled ozone profiles from the regional forecast models and the ENSEMBLE are routinely compared with ozone sonde profiles at eleven (maximum, depending on availability) European locations that are launched by several different institutes and national meteorological agencies (Table 3). The ozone sondes are downloaded regularly and checked, following the objectives of the GAW quality assurance system [Smit, 2013], from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC), the Network for the Detection of Atmospheric Composition Change (NDACC) and the Southern Hemisphere ADditional OZonesondes (SHADOZ). The vast majority of sondes use an electrochemical concentration cell, except the Hohenpeissenberg that uses Brewer-Mast. More information can be found in Eskes et al. [2021a].

Table 3: Ozone sondes metadata.

Station/Location	Lon (°)	Lat (°)	Altitude (m)	Launched by
Barajas	3.8	40.5	625	AEMET
Payerne	6.95	46.82	490	MeteoSwiss
Hohenpeissenberg	11.2	47.8	976	DWD
Prague	14.45	50	302	CHMI
Uccle	4.36	50.8	100	RMI
De Bilt	5.18	52.1	2	KNMI
Lindenberg	4.12	52.21	98	DWD
Legionow	20.97	52.4	94	IMGW
Lerwick	-1.19	60.14	76	Met Office
Jokioinen	23.5	60.8	104	FMI
Sodankyla	26.62	67.37	179	FMI

Modelled data are selected in space, using the grid cell the launch location is found in, and in time, at the hour closest to the launch time. The horizontal drift during the ascend of the sonde is considered negligible. The most typical launch time is approximately 11 UTC for most of the stations (Lerwick, Uccle, De Bilt, Barajas, Legionow, Lindebeberg and Prague), while in some sites, like Hohenpeissenberg, sondes are launched approximately at 5 UTC.

Observed concentrations corresponding to the modelling data at the different height levels (0, 50, 250, 500, 1000, 2000, and 5000m above the ground) are calculated by averaging sonde observations in the following intervals: (0, 20), (20, 80), (200, 300), (400, 600), (900, 1100), (1800, 2200), (2700, 3300), (4500, 5500) and by converting from mass mixing ratios using pressure and temperature values taken from the sonde observations.

2.2 IAGOS aircraft measurements

The regional EQC reports present comparisons of modelled profiles with aircraft in-situ profiles from the European Research Infrastructure IAGOS (https://www.iagos.org). The IAGOS program [Petzold et al., 2015, Thouret et al., 2020] uses sensors mounted on commercial aircraft to obtain in situ measurements of various chemical species in the atmosphere. All IAGOS-CORE aircraft are equipped with a package which provides volume mixing ratios of the trace gases O_3, CO, and water vapour, cloud particle number concentration, and meteorological measurements including temperature, pressure and winds. Further details of the O_3 and CO instruments and their operation can be found in Nédélec et al. [2015] and Blot et al. [2021]. The representativeness of airborne measurements over international airports have been documented by Petetin et al. [2018]. The comparisons of the CAMS regional data with IAGOS ozone and CO observations are possible at the European airports visited by the IAGOS fleet. The two European-based carriers Lufthansa and Air France provide regular profiles at their home airports, at Frankfurt and Paris (CDG) respectively. Occasionally other airports have been visited by Lufthansa over Europe and the Middle East: Berlin, Leipzig, Rome, Lisbon, Vienna and Beirut. Moreover, the Asian-based carrier China Airlines fly regularly from Taipei to Amsterdam, or to Vienna or Rome.

The aircraft takes about ten minutes to climb or descend the 5000m vertical extent covered by the regional models. During this time, travelling at up to 166 ms^{-1}, it covers about 120 km and therefore traverses many grid-boxes of a regional model. A spatial interpolation from the grid of the regional models to the aircraft's trajectory is performed using a radius distance of 10 km and the closest time step from the model. Volume mixing ratio data from IAGOS measurements (in ppbv) are then converted to mass concentration using pressure and temperature measured by the aircraft. The data used in these comparisons are validated by the Principal Investigator but are not yet calibrated (i.e. Level 1 data, in this context equivalent with near-real-time data), as calibration takes place only after an operational period of about six months (corresponding to Level 2 data).

2.3 TROPOMI satellite measurements

Satellite total column retrievals are implicitly dependent on an a priori tracer profile. The retrieval algorithm accounts for the fact that the sensitivity of the instrument is different at different altitudes (higher sensitivity in the free troposphere and lower in the boundary layer). This information is encoded in the averaging kernel which is proportional to the measurement sensitivity and depends on the viewing geometry, cloud properties, aerosols and surface albedo.

A direct comparison of the NO_2 tropospheric column as provided in the TROPOMI product and a model-generated column would thus introduce an unwanted bias, as the TROPOMI vertical column densities depend on the retrieval a priori profile, which in the case of the standard TROPOMI product originates from the TM5-MP CTM (Williams et al., 2017).

As explained in the TROPOMI Product User Manual [Eskes et al., 2021b], the a priori profiles in the retrieval may be replaced by any other model NO2 profile information, resulting in a new retrieved tropospheric NO2 column which is better comparable with the TROPOMI columns. The recipe makes use of the tropospheric averaging kernel and the air-mass factors provided by the TROPOMI L2 datafiles as also explained in Eskes et al. [2021b].

For the comparisons presented in the regional EQC reports NO2 profiles below 3 km altitude come from the regional model in question, while above 3 km and up to the tropopause, the profile from the CAMS-global model is used, with the assumption that the global model gives a more realistic description of NOx abundance in the free troposphere and thus the modelled tropospheric columns can be compared more accurately with the TROPOMI retrieved columns.

In order to minimise representativeness errors during the comparison, certain considerations are taken into account so that the fields can be correctly sampled in space and time. Horizontally, all available gridded data are interpolated to the CAMS regional model grid (0.1×0.1 degrees) and subsequently averaged in time for the period in question. Source grids in this process are either the TROPOMI native grid which is different for each orbit, the CAMS global grid or the TM5-MP grid. Horizontal interpolation of retrieval columns is realised by means of a weighted average of all individual columns within a target grid cell. Variables such as temperature, pressure, averaging kernel, and the tropopause layer index are interpolated horizontally using bilinear regridding. Modelled fields are also interpolated in time, based on the satellite overpass time over Central Europe. All vertical levels of source data are linearly interpolated to the TM5-MP vertical levels and all subsequent integrations to columns are performed based on those levels. Pressures at each of those levels are calculated based on the surface pressure and the hybrid coefficients included in the TROPOMI product which is based on TM5-MP. For the column integrations, all concentrations are converted to densities based on temperature and pressure profiles provided by TM5-MP [Ialongo et al., 2020].

3. References

Blot, R., P. Nédélec, D. Boulanger, P. Wolff, B. Sauvage, J.-M. Cousin, G. Athier, A. Zahn, F. Obersteiner, D. Scharffe, H. Petetin, Y. Bennouna, H. Clark, and V. Thouret. Internal consistency of the iagos ozone and carbon monoxide measurements for the last 25 years. Atmospheric Measurement Techniques, 14(5):3935–3951, 2021. 10.5194/amt-14-3935-2021. URL: https://amt.copernicus.org/articles/14/3935/2021/

Eskes, H., S. Basart, A. Benedictow, Y. Bennouna, A. Blechschmidt, S. Chabrillat, E. Cuevas, Q. Errera, H. Flentje, K. Hansen, J. Kapsomenakis, B. Langerock, M. Ramonet, A. Richter, M. Schulz, N. Sudarchikova, A. Wagner, T. Warneke, and C. Zerefos. Observation characterisation and validation methods document. Technical report, 2021a. URL https://atmosphere.copernicus.eu/sites/default/files/publications/CAMS84_2018SC1_D6.1.1-2021_observations_v6_0.pdf.

Eskes, H., J. van Geffen, F. Boersma, K. Eichmann, A. Apituley, M. Pedergnana, M. Sneep, P. Veefkind, and D. Loyola. Sentinel-5 precursor/tropomi level 2 product user manual nitrogen dioxide. 2021b. URL: https://sentinel.esa.int/documents/247904/2474726/Sentinel-5P-Level-2-Product-User-Manual-Nitrogen-Dioxide.

Huijnen, V., H. J. Eskes, A. Poupkou, H. Elbern, K. F. Boersma, G. Foret, M. Sofiev, A. Valdebenito, J. Flemming, O. Stein, A. Gross, L. Robertson, M. D'Isidoro, I. Kioutsioukis, E. Friese, B. Amstrup, R. Bergstrom, A. Strunk, J. Vira, D. Zyryanov, A. Maurizi, D. Melas, V.-H. Peuch, and C. Zerefos. Comparison of omi no2 tropospheric columns with an ensemble of global and european regional air quality models. Atmospheric Chemistry and Physics, 10(7):3273–3296, 2010. 10.5194/acp-10-3273-2010. URL: https://acp.copernicus.org/articles/10/3273/2010/.

Ialongo, I., H. Virta, H. Eskes, J. Hovila, and J. Douros. Comparison of TROPOMI/Sentinel- 5 Precursor NO2 observations with ground-based measurements in Helsinki. Atmospheric Measurement Techniques, 13(1):205–218, 2020, doi:10.5194/amt-13-205-2020. URL: https://amt.copernicus.org/articles/13/205/2020/.

Janssen, S., P. Thunis, M. Adani, C. Carnevale, C. Cuvelier, P. Durka, E. Georgieva, C. Guer- reiro, L. Malherbe, B. Maiheu, F. Meleux, A. Monteiro, A.I Miranda, H. Olesen, F. Pfafflin, J. Stocker, G. Sousa-Santos, A. Stidworthy, M. Stortini, E. Trimpeneers, P. Viaene, L. Vi- tali, K. Vincent, and J. Wesseling. FAIRMODE guidance document on modelling quality objectives and benchmarking. (KJ-NA-31068-EN-N (online)), 2022. ISSN 1831-9424 (on- line). 10.2760/41988(online). URL https://publications.jrc.ec.europa.eu/repository/bitstream/JRC129254/guidance_mqo_bench_vs3.3_20220519.pdf.

Joly, M., and V.-H. Peuch. Objective classification of air quality monitoring sites over Europe. Atmospheric Environment, 47:111 – 123, 2012. ISSN 1352-2310. doi:10.1016/j.atmosenv.2011.11.025. URL: https://opensource.umr-cnrm.fr/projects/air_quality-classification/wiki/-

Nédélec, Ph., R. Blot, D. Boulanger, G. Athier, J.-M. Cousin, B. Gautron, A. Petzold, A. Volz- Thomas, and V. Thouret. Instrumentation on commercial aircraft for monitoring the atmospheric composition on a global scale: the iagos system, technical overview of ozone and carbon monoxide measurements. Tellus B: Chemical and Physical Meteorology, 67(1):27791, 2015. doi:10.3402/tellusb.v67.27791. URL https://doi.org/10.3402/tellusb.v67.27791.

Pernigotti, D., M. Gerboles, C. Belis, and P. Thunis. Model quality objectives based on measurement uncertainty. Part II: NO_2 and PM_{10}. ATMOSPHERIC ENVIRONMENT, 79:869–878, 2013. ISSN 1352-2310. doi:10.1016/j.atmosenv.2013.07.045. URL: http://www.sciencedirect.com/science/article/pii/S1352231013005761.

Petetin, H., M. Jeoffrion, B. Sauvage, G. Athier, R. Blot, D. Boulanger, H. Clark, J.-M. Cousin, F. Gheusi, P. Nédélec, M. Steinbacher, and V. Thouret. Representativeness of the IAGOS airborne measurements in the lower troposphere. Elementa: Science of the Anthropocene, 6, 03 2018. ISSN 2325-1026. doi:10.1525/elementa.280. URL: https://doi.org/10.1525/elementa.280.23.

Petzold, A., V. Thouret, Ch. Gerbig, A. Zahn, Carl A. M. Brenninkmeijer, M. Gallagher, M. Hermann, M. Pontaud, H. Ziereis, D. Boulanger, J. Marshall, Ph. Nédélec, H. G. J. Smit, U. Friess, J.-M. Flaud, A. Wahner, J.-P. Cammas, A. Volz-Thomas, and IAGOS TEAM. Global-scale atmosphere monitoring by in-service aircraft – current achievements and future prospects of the european research infrastructure iagos. Tellus B: Chemical and Physical Meteorology, 67(1):28452, 2015. doi:10.3402/tellusb.v67.28452. URL: https://doi.org/10.3402/tellusb.v67.28452.

Smit, H. G. J. Quality assurance and quality control for ozonesonde measurements in GAW. WMO, 2013. URL: https://library.wmo.int/doc_num.php?explnum_id=7167.

Thouret, V., H. Clark, A. Petzold, Ph. Nédélec, and A. Zahn. IAGOS: Monitoring Atmospheric Composition for Air Quality and Climate by Passenger Aircraft, pages 1–14. Springer Nature Singapore, Singapore, 2020. ISBN 978-981-15-2527-8. doi:10.1007/978-981-15-2527-8 57-1. URL: https://doi.org/10.1007/978-981-15-2527-8_57-1.

Thunis, P., D. Pernigotti, and M. Gerboles. Model quality objectives based on measurement uncertainty. Part I: Ozone. ATMOSPHERIC ENVIRONMENT, 79:861–868, 2013. ISSN1352-2310. doi:10.1016/j.atmosenv.2013.05.018. URL: http://www.sciencedirect.com/science/article/pii/S1352231013003610 .

Copernicus Atmosphere Monitoring Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: https://support.ecmwf.int/

atmosphere.copernicus.eu copernicus.eu ecmwf.int