

Applying the delta tool to support the Air Quality Directive: evaluation of the TCAM chemical transport model

C. Carnevale · G. Finzi · A. Pederzoli · E. Pisoni ·
P. Thunis · E. Turrini · M. Volta

Received: 28 October 2013 / Accepted: 22 January 2014 / Published online: 6 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract This paper presents an application of the DELTA evaluation tool V3.2 to support the EU Air Quality Directive (AQD 2008). This software, designed in the frame of the FAIRMODE project (Forum for Air Quality Modelling in Europe, <http://fairmode.ew.eea.europa.eu/>), is currently used as support to working groups of modelers across Europe in the diagnostics and assessment of air quality model performances under the AQD (2008). The skills of the DELTA tool V3.2 are tested by looking at the results of a 1-year (2005) simulation performed using the transport chemical aerosol model (Carnevale et al. 2008) at 6×6 -km² resolution over the Po Valley. The modeled daily PM₁₀ concentrations at surface level are compared to observations provided by approximately 50 stations distributed across the domain. The main statistical parameters (i.e., bias, root mean square error, correlation coefficient, standard deviation) as well as different types of diagrams (scatter plots, time series plots, Taylor and target plots) have been produced. A representation of the observation uncertainty in the target plot, used to derive model performance criteria for the main statistical indicators, is also presented and discussed.

Keywords PM₁₀ concentration · Model uncertainty · Performance criteria · Target diagram

Introduction

Present-day numerical air quality models are seen as important tools for the assessment and forecast of air pollutant concentrations and depositions, contributing to the development of effective strategies for the control and reduction of air pollutant emissions (Thunis et al. 2012a, b, c). These tools can simulate concentrations and deposition fluxes of pollutants on a wide range of spatial (global, regional, urban) and temporal (hourly, daily, monthly, yearly) scales. They are used for identifying critical polluted areas, integrating measurements and achieving a deeper scientific understanding of the physical and chemical processes involving air pollutants in the atmosphere (EPA 2005). The use of comprehensive air quality models started in the late 1970s (Daly and Zannetti 2007), and since then, their development has increased rapidly, together with the fast increase in computational resources. Today, more and more complex and computationally expensive numerical models are developed by the scientific community, and their results are made available to the environmental authorities dealing with the development of air quality plans and regulations. Actually, the AQD (2008) states that modeling techniques must integrate fixed measurements in order to provide suitable information about the spatial distribution of pollutant concentrations, and model results have to be taken into account for the assessment of air quality with respect to the limit values (AQD 2008). The role of air quality models in the frame of the AQD (2008) is multifold. First, they must be capable of assessing both the current air quality status (i.e., assessing the number of exceedances of the standard values, calculating population exposure to pollution and health impacts) as well as the future one (air quality forecasting) by providing next day and near-real-time information to national, regional, and local authorities warning the public about expected high pollution episodes. Furthermore, the use of air quality models is required for long-term air quality planning:

C. Carnevale · G. Finzi · A. Pederzoli (✉) · E. Turrini · M. Volta
Department of Mechanical and industrial Engineering,
University of Brescia, Via Branze 38, Brescia 25123, Italy
e-mail: Anna.Pederzoli@unibs.it

E. Pisoni · P. Thunis
Joint Research Centre, European Commission,
Via E. Fermi 2749, Ispra (VA) 21027, Italy

they must be designed in order to identify possible ways for reducing emissions by optimizing the costs of control strategy implementation. They must also be capable of predicting how the reduction or control of emissions may affect pollutant concentrations in the atmosphere (emission scenarios). Considering these motivating factors, air quality models may be considered as valid tools to support the AQD (2008), as long as their performances satisfy some minimum levels of efficiency. These performances are of course affected by model uncertainty. Uncertainties associated with air quality models can have a significant impact on their predictions. In addition to those related to the model input datasets (meteorology and emissions), they may also arise from the model parameterization schemes used to describe the physical and chemical processes involving air pollutants in the atmosphere. Thus, evaluation of model results plays a key role in air quality planning and assessment, and it needs to be addressed properly.

Air quality models are usually evaluated by comparing the modeled pollutant concentrations versus ground measurements, provided by monitoring networks, and by computing statistical indicators, function of both the modeled and measured values, in order to quantify the model's capability to reproduce observations both spatially and temporally. However, even if these statistics can give an estimate of the model's accuracy, they do not tell whether model results reach a sufficient level of quality for a certain application. This is the reason why model performance criteria (MPC) have been introduced as an indicator of the minimum levels of quality to be achieved by a model for policy use (Boylan and Russell 2006). A large number of studies suggested different sets of MPC usually differing from each other, as they depend on the type of application, the spatial scale, and/or the time period of interest. Boylan and Russell (2006) recommended specific values for the mean fractional bias and mean fractional error indicators for particulate matter (PM). Chemel et al. (2010) suggested using a similar approach for O₃ concentration, while Derwent et al. (2010) proposed a MPC based on the bias and factor of 2 of observations (FAC2) indicators. Jolliff et al. (2009) identified the root mean square error (RMSE) normalized by the standard deviation of the observations (σ_O) as the main statistical indicator and suggested a MPC equal to unity. A "less or equal to 1" value for this parameter means that the model is a better predictor of the monitoring data compared to the mean of the monitoring data (Stow et al. 2009). Denby (2010) proposed the relative directive error (RDE) as representative of a model's uncertainty. Up to now, this has also been the indicator officially recommended by the AQD (2008), stating that a MPC for RDE of 50 % needs to be fulfilled by models (AQD 2008) to be used. Some recent studies (Thunis et al. 2012a, b, c; Pederzoli et al. 2012) have pointed out some limitations of the RDE, strictly connected to its own definition. In particular, they highlight the fact that this

indicator is based on the difference between the modeled and observed values around the limit fixed by the AQD (2008) ($50 \mu\text{g m}^{-3}$ for daily average PM₁₀, $200 \mu\text{g m}^{-3}$ for NO₂ hourly maximum, and $120 \mu\text{g m}^{-3}$ for daily 8-h maximum mean O₃ concentration), without accounting for the timing of the events. This means that the temporal correlation between the modeled and observed data is not accounted for and a model fulfilling the 50 % MPC may actually perform in an uncorrelated way with respect to observations. As a consequence, the information provided by the RDE can be misleading and does not assure that models provide good results for the right reasons. Recently, Thunis et al. (2012b, 2013) suggested normalizing the RMSE by the uncertainty of observations (U). In this way, the resulting MPC on RMSE is independent of the pollutant, application, or time/spatial scale considered. One of the main advantages of this new formulation compared to the previous studies is that the MPC for the different statistics are derived by considering the same input (U). Furthermore, as RMSE can be expressed as a function of the main statistical indicators (bias, correlation, and standard deviation), this formulation provides an opportunity to identify which aspects of the model performance are the weakest and need to be improved. This methodology has been proposed in the frame of the FAIRMODE activities (Forum for Air Quality Modelling in Europe, <http://fairmode.ew.eea.europa.eu/>) and implemented inside the delta evaluation tool, V3.2 (Thunis et al. 2012a, b).

More specifically, in Thunis et al. (2012b), the measurement uncertainty U (i.e., the maximum uncertainty of instruments used for measuring pollutant concentration) is considered, as a first step, as constant regardless of the pollutant concentration level; it is set according to the data quality objective value of the AQD, i.e., 15, 15, and 25 % for O₃, NO₂, and PM₁₀, respectively. Later studies (Thunis et al. 2013; Pernigotti et al. 2013a, b) overcome the assumption about U 's independency from the concentration by identifying the main sources of uncertainty in O₃, NO₂, and PM₁₀ measurements. A new formulation of U as a function of pollutant concentration is presented and used to update the MPC proposed in Thunis et al. (2012b).

In this work, the methodology presented in Thunis et al. (2013) and Pernigotti et al. (2013a) has been applied for the first time to a specific case: a 1-year simulation (2005) over the Po Valley at $6 \times 6\text{-km}^2$ resolution performed using the transport chemical aerosol model (TCAM; Carnevale et al. 2008). The DELTA tool is used for analyzing the model results. Furthermore, the study presents for the first time a sensitivity analysis performed on the parameters used for the computation of U (based on the analytical PM measurement technique implemented by monitoring sites).

The main goal of the study was to show how the methodology proposed by Thunis et al. (2013) can be applied to a "real" modeling case and how the DELTA tool can provide

support to modelers for the evaluation of their models in the frame of the AQD (2008).

Methodology

TCAM configuration

The TCAM is a 3D multiphase air quality model developed at the University of Brescia (Carnevale et al. 2008). It contains specific parameterization schemes for simulating the main atmospheric processes that affect the concentration and deposition of atmospheric pollutants on a regional scale. It includes parameterizations for both convection (Pepper et al. 1979) and advection (Yamartino et al. 1992) of both gaseous and aerosol species, diffusion (Holstag and Nieuwstadt 1986; Scire et al. 1990), and deposition processes. Dry deposition is described by the full-resistance model (Yamartino et al. 1990), whereas for wet deposition, the scavenging coefficient is computed taking into account the rainfall rate, cloud water content, gas solubility, diffusivity, and particle size. Gas-to-particle conversion using a splitting operator technique (Marchuk 1975) is also included. Gas chemistry is simulated using a modified version of the SAPRC97 scheme (Carter 1990). The module ISORROPIA-II (Nenes et al. 1998; Fountoukis et al. 2009; <http://nenes.eas.gatech.edu/ISORROPIA>), for modeling the thermodynamic equilibrium of inorganic species, is integrated in the system. In ISORROPIA-II, the activity coefficients are computed by applying the Bromley method (Nenes et al. 1998), which takes into account the temperature effect. The aerosol module adopts a sectional approach and includes 21 chemical compounds: 12 inorganic species (H_2O , SO_4 , NH_4^+ , Cl^- , NO_3^- , Na^+ , H^+ , $\text{SO}_{2(\text{aq})}$, $\text{H}_2\text{O}_{2(\text{aq})}$, $\text{O}_{3(\text{aq})}$, elemental carbon, and others) and 9 organic species (a primary organic one and eight classes of secondary organic species). Each chemical species is speciated into ten size bins. The TCAM has been widely used in the frame of several European projects (Cuvelier et al. 2007; Thunis et al. 2009).

DELTA tool

The DELTA evaluation tool (Thunis et al. 2012a) is an IDL-based model evaluation software which has been designed in the frame of the FAIRMODE project and is currently used for supporting modeling groups across Europe in the diagnostics and evaluation of their air quality models according to the AQD (2008) guidelines. The tool makes use of paired data of modeled and observed surface pollutant concentrations and computes the main statistical indicators (i.e., bias, RMSE, correlation coefficient), which can be summarized in appropriate statistic tables. It can also visualize them on a wide range of plots (scatter, time series, Taylor and target diagrams), providing an overview of the quality of model results

with respect to observations. The tool, successfully tested on both Linux and Windows environments, applies a graphic user interface which guides the user through the main functionalities of the software. The package can be downloaded through the DELTA web site (<http://aqm.jrc.ec.europa.eu/DELTA>). All computational details, including data formats and installation procedure, are described in Thunis et al. (2012a, c).

The latest version of DELTA (3.2), which includes the normalization of all diagrams, MPC, and statistical indicators by the observation uncertainty U , is used in this work. A detailed description of this latest version of the software can be found in Thunis et al. (2012c).

Standard statistical indexes

Four standard statistical indicators (Eqs. 1–4) have been selected for this study: RMSE, correlation coefficient (R), normalized mean bias (NMB), and normalized mean standard deviation (NMSD). This set of parameters, similar to the one recommended by Borrego et al. (2008), is representative of the main aspects of model performances: amplitude (NMSD), phase (R), and bias (NMB). RMSE can also be expressed as a function of the random error (centered root mean square error, CRMSE) and the systematic one (bias; Eq. 5). The computation of these standard performance metrics is based on the comparison between the modeled and measured values. These standard indicators provide information about model responses, such as the tendency to under/overpredict concentrations (bias) and the intensity of the model deviation with respect to observations (e.g., standard deviation). However, the simple computation of these parameters presents some limitations. First of all, a model may perform well for some aspects, but not others. Pederzoli et al. (2012) showed how the analysis of a single indicator independently from others can be deceptive and lead to partial conclusions; in any case, it cannot be considered as exhaustive for evaluating the model's performance. Furthermore, it does not tell whether the overall model response is actually acceptable for regulatory purposes according to the AQD (2008). The approach suggested by Thunis et al. (2012b), based on the use of the target diagram, provides a more comprehensive indication of the model's response as all relevant information, including those coming from traditional indicators, is included. Furthermore, the introduction of MPC values in the diagrams permits establishing whether the model reaches the minimum level of efficiency in order to be used for regulatory decisions, according to the AQD (2008). The main aspects of this methodology are briefly recalled in the next section.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2} \quad (1)$$

$$R = \frac{\sum_{i=1}^N (M_i - \bar{M}) \cdot (O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (M_i - \bar{M})^2} \cdot \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}} \quad (2)$$

$$\text{NMSD} = \frac{\sigma_M - \sigma_O}{\sigma_O} \quad (3)$$

$$\text{NMB} = \frac{\text{Bias}}{\bar{O}} = \frac{\bar{M} - \bar{O}}{\bar{O}} \quad (4)$$

$$\text{RMSE}^2 = \text{CRMSE}^2 + \text{Bias}^2 = \sqrt{\frac{1}{N} \sum_{i=1}^N ((O_i - \bar{O}) - (M_i - \bar{M}))^2} + (\bar{M} - \bar{O})^2 \quad (5)$$

where O_i and M_i stand for observations and model results, respectively; the subscript i indicates the time interval, the overbars indicate the time average over N time intervals, and the symbol s indicates the standard deviation.

U formulation

As previously mentioned, recent studies (Thunis et al. 2012b, c, 2013; Pernigotti et al. 2013a, b) suggested the use of RMSE normalized by the measurement uncertainty U (also named “expanded uncertainty,” RMSE_U) as the main model quality objective (MQO) for air quality model evaluation.

$$\text{MQO} = \text{RMSE}_U = \frac{\text{RMSE}}{2\text{RMS}_U} = \frac{\sqrt{\frac{1}{N} \sum (O_i - M_i)^2}}{2\sqrt{\frac{1}{N} \sum U_i^2}} \leq 1 \quad (6)$$

where RMS_U is the root mean square error of the uncertainty.

$$\text{RMS}_U = 2\sqrt{\frac{1}{N} \sum U_i^2} \quad (7)$$

The “less than 1” condition in Eq. 1 is derived by making two main assumptions on the uncertainty of an individual observation O_i and its corresponding modeled value M_i (each one belonging to its confidence interval, $[O_i \pm U(O_i)]$ and $[M_i \pm U(M_i)]$, respectively): (1) an overlap between the two intervals exists and (2) the model uncertainty is less than or equal to the observation one: $U(M_i) < U(O_i)$. With these two assumptions, the following condition is achieved:

$$|O_i - M_i| \leq 2U(O_i) \quad (8)$$

For a time series of N measurements, Eq. 6 can be generalized to Eq. 8. $U(O_i)$ in Eq. 6 is the maximum measurement uncertainty, i.e., the one resulting from the maximum estimate

of all possible sources of errors arising from measurement techniques.

In previous studies (Thunis et al. 2012b; Pederzoli et al. 2012), a simplified formulation for U was presented, considering U as pollutant-dependent only, i.e., independent of the observed concentration. This approach is not reliable at high and low concentrations where uncertainty is usually under- and overestimated. A new formulation for U , depending on the measured concentration, has been recently implemented in version 3.2 of the DELTA tool for NO_2 , O_3 , and PM_{10} and presented in Thunis et al. (2012c, 2013) and Pernigotti et al. (2013a, b). In this new approach, U is defined as

$$U = k u_r^{\text{RV}} \sqrt{(1-\alpha) \left(\bar{O}^2 + \sigma_O^2 \right) + \alpha \times \text{RV}^2} \quad (9)$$

Equation 9 is derived assuming U as a sum of two fractions, a proportional and a non-proportional one (i.e., independent of the measurement concentration). The non-proportional part is by definition independent of the concentration level; thus, it can be estimated around a reference value, RV. In the context of using the models to support air quality directive, Thunis et al. (2013) and Pernigotti et al. (2013a, b) set $\text{RV} = \text{LV}$, where LV is the daily/hourly limit values (LV) of the AQD for a reference time average (daily average PM_{10} concentration). α is the uncertainty fraction non-proportional to the concentration level around RV (between 0 and 1). u_r^{RV} represents the estimated relative measurement uncertainty around RV. Finally, U is multiplied by a coverage factor k , which is the level of confidence that the true value (i.e., the perfect measurement) can lie within the interval $[O_i \pm U]$. In other words, it provides an estimate (in percent) of the overlapping interval between O_i and M_i .

U-normalized statistical indexes

Three different possible conditions have been identified: if $\text{RMSE}_U \leq 0.5$, the RMSE between the observed and modeled values is less than U and the model results are, on average, within the range of the observation uncertainty for that station. This means that any attempt to improve the model performance further is unhelpful. If $0.5 < \text{RMSE}_U \leq 1$, the RMSE is, on average, greater than the range of U , but the model might still be closer to the “true value” (i.e., the perfect measurement) than observations. Finally, if $\text{RMSE}_U > 1$, observations are closer to the “true value” than the model results. Based on Eq. 9, Thunis et al. (2012b) derive a consistent set of MPC for three statistical indicators.

$$|\text{NMB}| < \frac{2U}{\bar{O}} \quad (\text{MPC}_{\text{NMB}}) \quad (10)$$

$$R > 1 - 2 \left(\frac{U}{\sigma_O} \right)^2 \quad (\text{MPC}_R) \quad (11)$$

$$|\text{NMSD}| < \frac{2U}{\sigma_O} \quad (\text{MPC}_{\text{NMSD}}) \quad (12)$$

Equations 10–12 represent necessary but not sufficient conditions to fulfill the MQO in Eq. 9. MPC_{NMSD} and MPC_R become dependent on σ_o (Eqs. 11 and 12), whereas MPC_{NMB} (Eq. 10) functions on the average of observations \bar{O} . The condition on RMSE_U in Eq. 9 is always 1 (MPC_{RMSE}) regardless of observations. This set of MPC has been introduced in the Delta tool: in this way, each diagram visualized by the software shows a colored green area where the values of RMSE, NMB, R , and NMSD fulfill the corresponding criteria.

Target diagram

The overall performance of a model can be represented in the DELTA tool through the target diagram (Jolliff et al. 2009). Figure 1 reports an example for this type of plot. The X -axis (CRMSE) and the Y -axis (bias) are normalized by U : for each dot reported on the plot, the radius between the point and the origin is the RMSE_U (Eq. 6) and the green area identifies the fulfillment of MPC_{RMSE} (Eq. 9). According to Thunis et al. (2012b), the CRMSE can be rewritten as

$$\text{CRMSE}^2 = \sigma_O^2 + \sigma_M^2 - 2\sigma_O\sigma_MR \quad (13)$$

which leads to

$$\frac{\text{CRMSE}(R = 1)}{\text{CRMSE}(\sigma_O = \sigma_M)} = \frac{\text{NMSD}}{\sqrt{2(1-R)}} \quad (14)$$

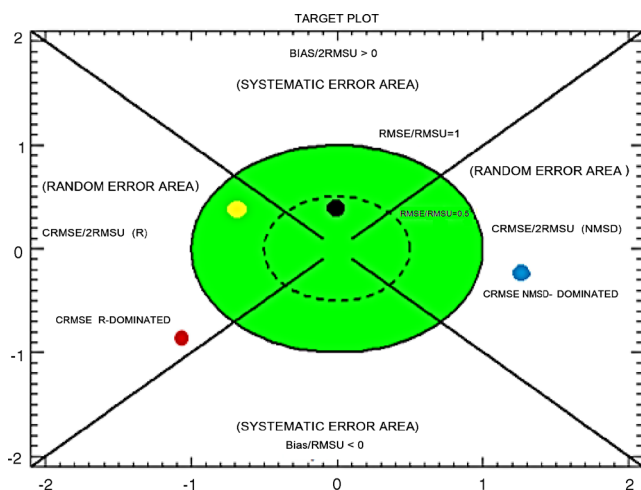


Fig. 1 Example of the target plot normalized by RMSU

In the target diagram, the ratio between the two terms is used to determine on which side of the X axis (negative or positive) a given points is located. If the correlation error dominates, the point is placed on the negative side of the diagram, whereas if NMSD is dominant, the point lies on the positive side.

In this way, four main zones in the diagram are identified: the lower and top zones identify systematic errors (positive and negative bias), whereas the left and right areas identify CRMSE (random) errors, with model performances dominated by R (red dot lying onto the left) or by NMSD (blue dot lying onto the right). If a point (yellow dot) lies inside the green area, it fulfills the MPC_{RMSE} (“less than 1” condition); if it lies in the green area but $\text{RMSE}_U \leq 0.5$ (black dot inside the dashed line circle), there is no chance for any model improvement as the RMSE between the observed and modeled values is less than U . This graphical representation has the advantage of providing, through RMSE, a general overview about the model’s performance; moreover, it helps in identifying the aspects (in terms of bias, correlation, and standard deviation) which can potentially be improved.

Results

Model domain

The model domain used in this study is a 95×65 -cell grid at 6×6 -km² resolution covering the Po Valley. This area, located in northern Italy, has already been identified by the scientific community as one of the most critical areas in Europe for air pollution issues: by 2020, ozone and particulate matter concentrations are expected to reach harmful levels despite the application, by national authorities, of the European legislation for emission control and reduction (Thunis et al. 2009).

The severe air pollution episodes affecting the area, especially during wintertime, are due to very unique conditions, a combination of several factors: high anthropogenic emissions, a complex topography, and frequent stagnant meteorological conditions.

First of all, the valley hosts one of the most important industrial and populated regions in Italy. High anthropogenic emissions, associated with traffic, industrial and residential heating, and combustion processes, lead to high concentrations of ozone and particulate matter in the whole area. Pollutant dispersion is strongly prevented by the complex terrain orography. Actually, the valley is located at the foot-steps of the Alps in the north and surrounded by the Apennines in the south; these mountain ranges protect the region from strong winds, which might favor air pollution dispersion. Pollutants remain trapped inside the valley and long-range transport outwards is prevented. Furthermore, during wintertime, meteorological conditions are typically characterized by low wind speeds and surface-based temperature

inversions, associated with a very stable planetary boundary layer (PBL). During nighttime, the PBL height is low, typically between 100 and 300 m (Rossa et al. 2012), favoring the formation of fog and low clouds covering the valley for several days. All these conditions facilitate the entrapment of polluted plumes and contribute to the formation of smog haze.

Several intercomparison studies (Cuvelier et al. 2007; Thunis et al. 2010) tried to describe the spatial and temporal distribution of PM and O₃ concentrations over Po Valley by applying different models and comparing the results; however, these attempts highlight how state-of-the-art air quality models are not capable of simulating pollutant concentrations (PM₁₀ in particular) properly, due to the complexity of the area and the peculiar meteorological conditions. Thus, estimating the model performances over this region area is important, as well as identifying which aspects of the model's performance need to be improved for achieving better results.

Setup

Observations from approximately 50 monitoring sites located in the Po Valley have been used in this study (Fig. 2). For the purposes of this work, the sites have been classified in terms of station type (suburban, urban, and rural). The orography (hilly, plane, valley) is also specified. Monitoring data are the same as those used in the model intercomparison exercise (POMI) performed for year 2005. Details about the POMI exercise can be found

at the POMI web site <http://aqm.jrc.it/POMI/index.html> as well as in Thunis et al. (2009) and Pernigotti et al. (2013b).

Daily PM₁₀ surface concentrations simulated by TCAM over this domain for the meteorological year 2005 are extracted at single-grid cells and compared to the measured data provided by the monitoring sites distributed across the area.

PM₁₀ uncertainty sources

Pernigotti et al. (2013a) provide estimates of α , k , and u_r^{RV} for PM₁₀ measurement uncertainties. The method suggested in the Guide for the Demonstration of Equivalence (ECWG 2010) is used for computing uncertainty budgets. This approach takes into account the fact that monitoring stations usually collect PM data by using different onboard instruments. For field gravimetric and TEOM techniques, data for computing the budgets come from specific monitoring campaigns carried out in the European PM QA/QC program. For beta-ray instruments, data from 30 airbase stations for year 2009 in Austria and Germany are used. The ECVG method for U budgets is based on the computation of the residual sum of square between a measurement value and a reference value, RV. $U(O_i)$ values are derived for each daily measured concentration. These values are then linearly fitted and u_r^{RV} and α are derived from the intercept and slope of the regression. Values of k , α , and u_r^{RV} for maximum uncertainties (Pernigotti et al. 2013a) are reported in Table 1.

A value of $k=2$ for PM₁₀ is set, which assures an interval confidence (overlapping interval between O_i and M_i) of 95 %.

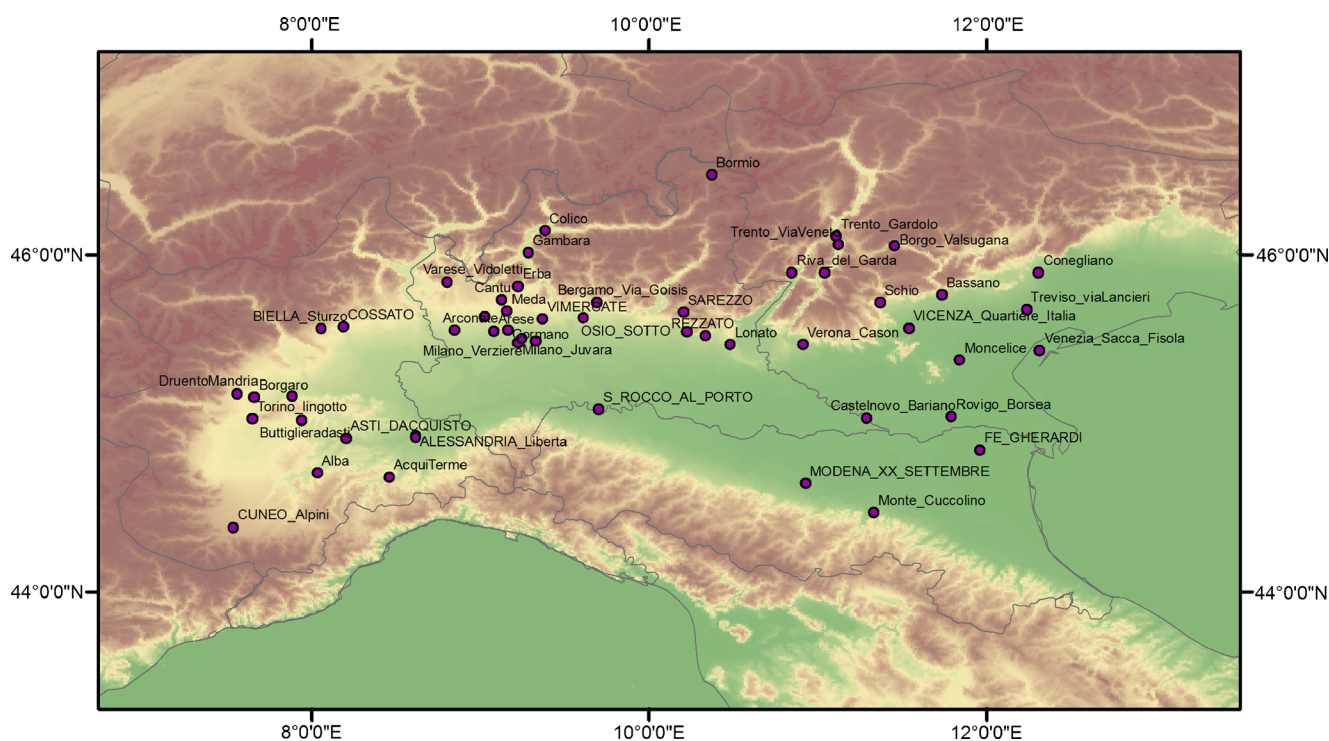


Fig. 2 Monitoring sites measuring PM₁₀ concentrations for year 2005 across the Po Valley

Table 1 Parameters for the computation of maximum source contribution uncertainties

	K	$RV=LV$	α	u_r^{RV}
PM_{10}	2.00	50	0.027	0.138

Source: Adapted from Pernigotti et al. (2013a, b)

As already mentioned, the reference value RV is set equal to the AQD (2008) Limit value is $LV=50 \mu g m^{-3}$. α and u_r^{RV} are derived from the linear fittings. Equation 9 is then applied inside the DELTA tool with these parameters, providing U estimates as a function of the measured concentration.

Diagrams

The DELTA tool V3.2 is used to visualize TCAM results onto different types of diagrams.

Figure 3 reports the target plot for PM_{10} TCAM concentrations, which provides an immediate overview about the overall model performance. Of the points, 43 % are outside the green circle, so MPC_{RMSE} is not fulfilled by the model at these sites. Furthermore, dots lie both in the CRMSE zone and in the bias zone. Thus, the model’s uncertainty is related to both the random component (CRMSE) and the systematic one (bias). Inside CRMSE, the correlation between the modeled and observed data is the main source of error (almost all points lie in the R zone). Only seven sites lie in the NMSD area. The bias is always negative, indicating a general underestimation of the PM_{10} concentration by the model. All dots are also

outside the dashed circle, which represents the observation uncertainty U : in this case, $0.5 < RMSE_U < 1$.

The MPC diagrams (Fig. 4a, b) help in understanding the model behavior by plotting R and NMSD as functions of U/σ_o . The fulfillment of MPC_R and MPC_{NMSD} is identified by the green area. The correlation MPC plot (Fig. 4a) confirms the weakness of the model in terms of the correlation coefficient for PM_{10} concentration. R is between 0.2 and 0.7; MPC_R is respected by 70 % of sites (six dots are outside the green area, corresponding to 14 % of the total). For NMSD, all stations respect the criterion; NMSD is always negative (in the range from -10 to -60 %), apart from one site (Monte Cuccolino), with a NMSD of $+25$ %. For a significant number of stations in Fig. 4b, NMSD is, on average, within the range of the observation uncertainty; thus, there is no margin for any model improvement at these sites. Concerning NMB, Fig. 5 reports the scatter plot of the modeled versus observed PM_{10} averaged concentrations at each site. Dashed and solid lines indicate $RMSE/2RMS_U$ ratios of 0.5 and 1, respectively. The NMB is always negative (up to -30 %), indicating a general underestimation of PM_{10} concentration by TCAM. The MPC_{NMB} (function of \bar{O}) is respected by 75 % of sites.

Comparison to RDE

The target analysis helps identify potential model weaknesses which may not be revealed by the computation of the RDE only. For example, Fig. 4a shows that a certain number of sites (14 %) do not reach the criterion set for PM_{10} correlation coefficient: R looks “responsible” for the points outside the green circle. Whereas the identification of this weakness is straightforward by looking at the target and MPC plots, the computation of RDE alone is not helpful and, in some cases, misleading. Actually, for each of these outlying stations (six in total), the RDE is computed and reported in a bar plot (Fig. 6). According to this, the 50 % criterion for RDE set by the AQD (2008) (green area) is actually fulfilled at three sites (Alessandria Nuova Orti, Rovigo Borsea, Alba). This means that at these locations (already identified as weak spots, with a R approximately 0.4, as shown in Fig. 4a), the RDE wrongly suggests an acceptable performance of the model.

The reason for this deceptive behavior is attributable to the RDE definition itself Denby (2010).

$$RDE = \frac{|O_{LV} - M_{LV}|}{LV} \tag{14}$$

where O_{LV} is the closest observed concentration to the limit value concentration (LV) and M_{LV} is the correspondingly ranked modeled concentration. This definition is mainly related to the exceedances of the limit values and does not consider the timing of events (Pederzoli et al. 2012). In general, the use of the target methodology is a step forward compared to the

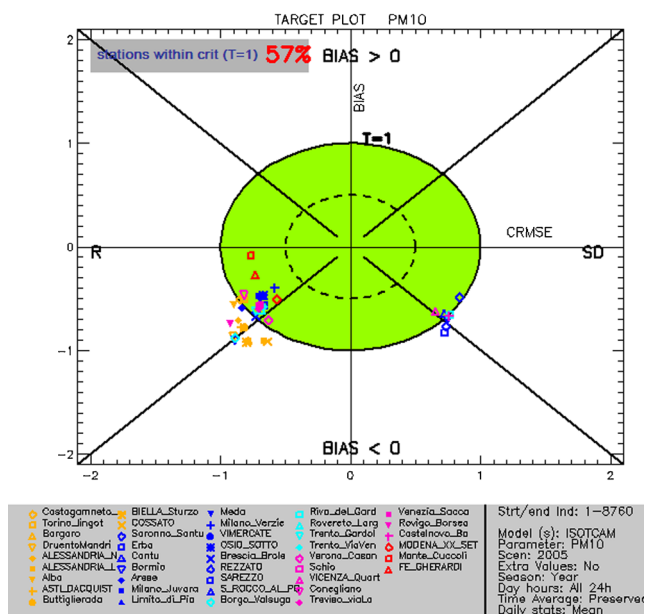


Fig. 3 Target diagrams for daily average PM_{10} concentrations. Model: TCAM. Po Valley dataset, year 2005

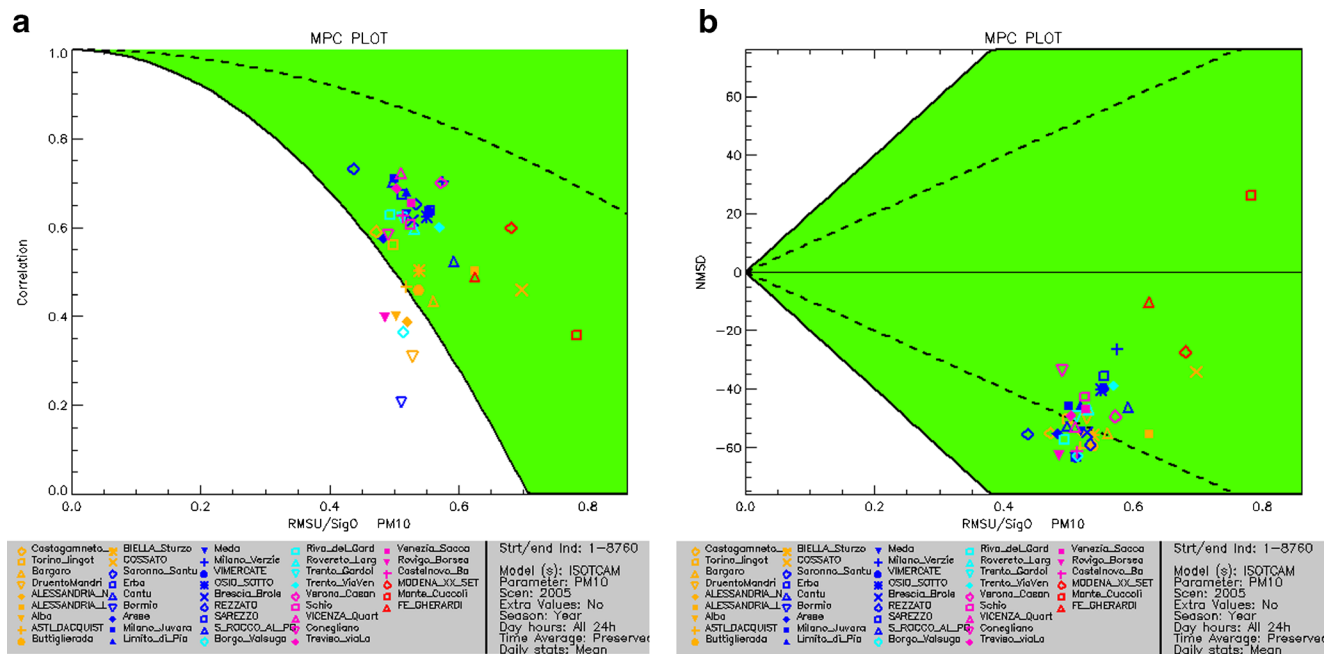


Fig. 4 MPC plots of R (a) and NMSD (b) for PM_{10} concentration. TCAM, year 2005

RDE approach, and it provides a more comprehensive overview about a model’s strengths and weaknesses.

Seasonal analysis

In order to identify the reasons for the weaknesses shown in Fig. 3 in terms of bias and correlation, a seasonal analysis has been performed; two time periods representing winter

(December–January–February) and summer (June–July–August), respectively, have been considered. The target plots (Fig. 7a, b) highlight the good model response during summer as well as a significant underestimation in wintertime for PM_{10} . Whereas in summer 93 % of sites fulfill the MPC_{RMSE} , about 50 % dots are outside the green area in the winter season. This is a well-known behavior: the POMI

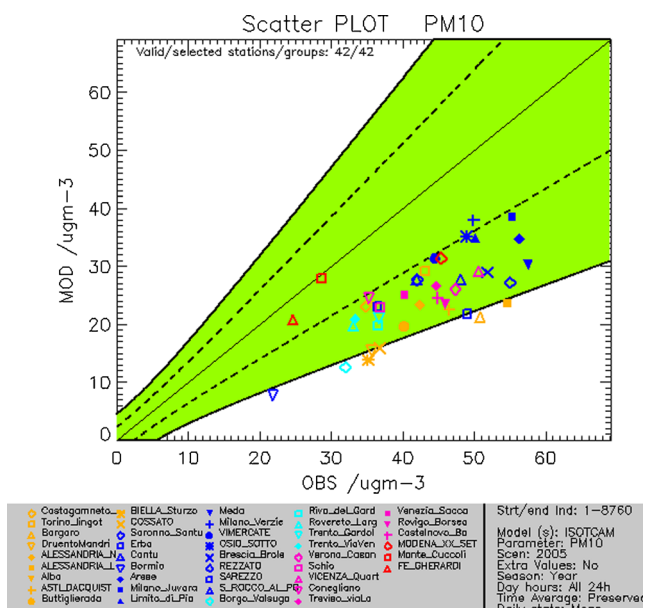


Fig. 5 Scatter plot of the modeled versus observed PM_{10} concentrations. TCAM, year 2005

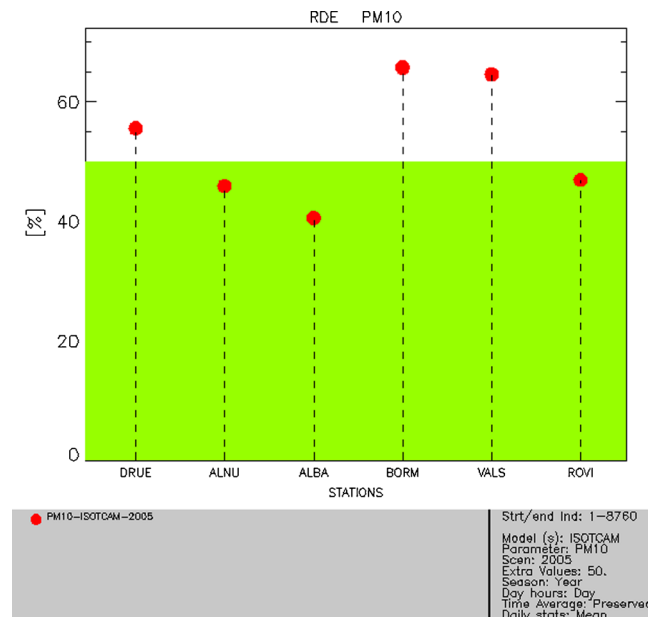


Fig. 6 Bar plot of the RDE for the modeled PM_{10} concentration at eight sites: Druento (*DRUE*), Alesandria Nuova Orti (*ALNU*), Alba (*ALBA*), Bormio (*BORM*), Borgo Val Sugana (*VALS*), and Rovigo Borsea (*ROVI*). TCAM, year 2005

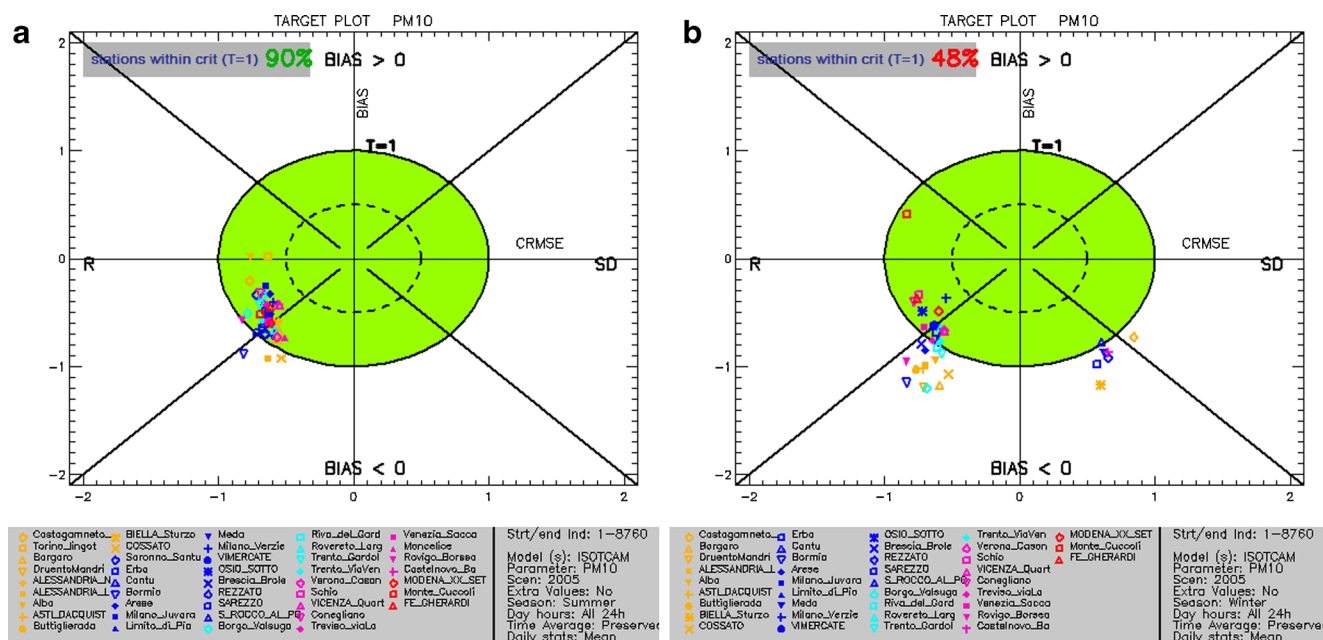


Fig. 7 Target plots of PM₁₀ concentrations for winter (a) and summer (b) seasons. TCAM, year 2005

intercomparison over the Po Valley (Thunis et al. 2013) showed that the tendency to underestimate PM concentration during wintertime is common to most air quality models. According to Thunis et al. (2009), this underestimation typically ranges between 25 and 30 μg m⁻³. This is due to several factors, including the under-prediction of primary organic compound emissions, the difficulties in modeling the secondary organic aerosol (SOA) formation (Camevale et al. 2012), and the sensitivity of air quality models to the meteorological input data. In particular, concerning organic compounds, more efforts are required in the treatment of traffic and residential heating emissions inside models for achieving a more detailed speciation between organic and elemental carbon (Thunis et al. 2009). A recent study (Pernigotti et al. 2013b) shows the analysis of speciated PM₂₅ for year 2005. Aerosol mass and organic matter (OM) concentrations have been collected at the Ispra background monitoring station (45.82° N, 8.62° E) and compared to results from six air quality models including TCAM. The comparison reveals that all models underestimate OM by 76 % up to 95 %. During the cold season, the primary OM fraction related to biomass burning is 63 % (Gilardoni et al. 2011). A poor characterization, mainly due to lack of knowledge, of wood burning emissions from domestic heating is identified by Pernigotti et al. (2013b) as one of the possible reasons for the OM underestimation. In summer, 54 % of the OM measured at the Ispra site has a secondary origin. In this case, the same study suggests that the OM under-prediction may be related to the lack of information on SOA sources from biogenic emissions.

Another study (Pernigotti et al. 2012) highlights how meteorological models tend to overestimate the wind intensity in

areas of complex orography such the Po Valley (bias between 0.6 and 1.4 ms⁻¹): as they are not able to simulate severe wind stagnation events, this significantly contributes to the PM₁₀ underestimation in air quality models; the use of observation’s nudging techniques can help to better simulate modeled surface wind fields, improving PM₁₀ model predictions.

Sensitivity analysis

A sensitivity analysis has been carried out by considering different values of u_r^{RV} and α in the computation of U . TCAM performances for PM₁₀ daily mean concentrations are discussed in this section. The sites have been classified into three subgroups based on the analytical measurement technique they implement for measuring PM: B (for beta-ray instruments), T (TEOM instruments), and G (gravimetric instruments). The values for u_r^{RV} and α currently used in the DELTA tool are those derived from the gravimetric data of the European PM QA/QC program (Table 1), and they are applied to all stations independently of their type of instrument. Different values of u_r^{RV} and α specific for TEOM and beta-ray systems, as provided by Pernigotti et al. (2013a), have

Table 2 u_r^{RV} and α values for U computation, specific for monitoring sites implementing TEOM (T) and beta-ray (B) instruments

	T	B
α	0.027	0.009
u_r^{RV}	0.192	0.192

Adapted from Pernigotti et al. (2013a)

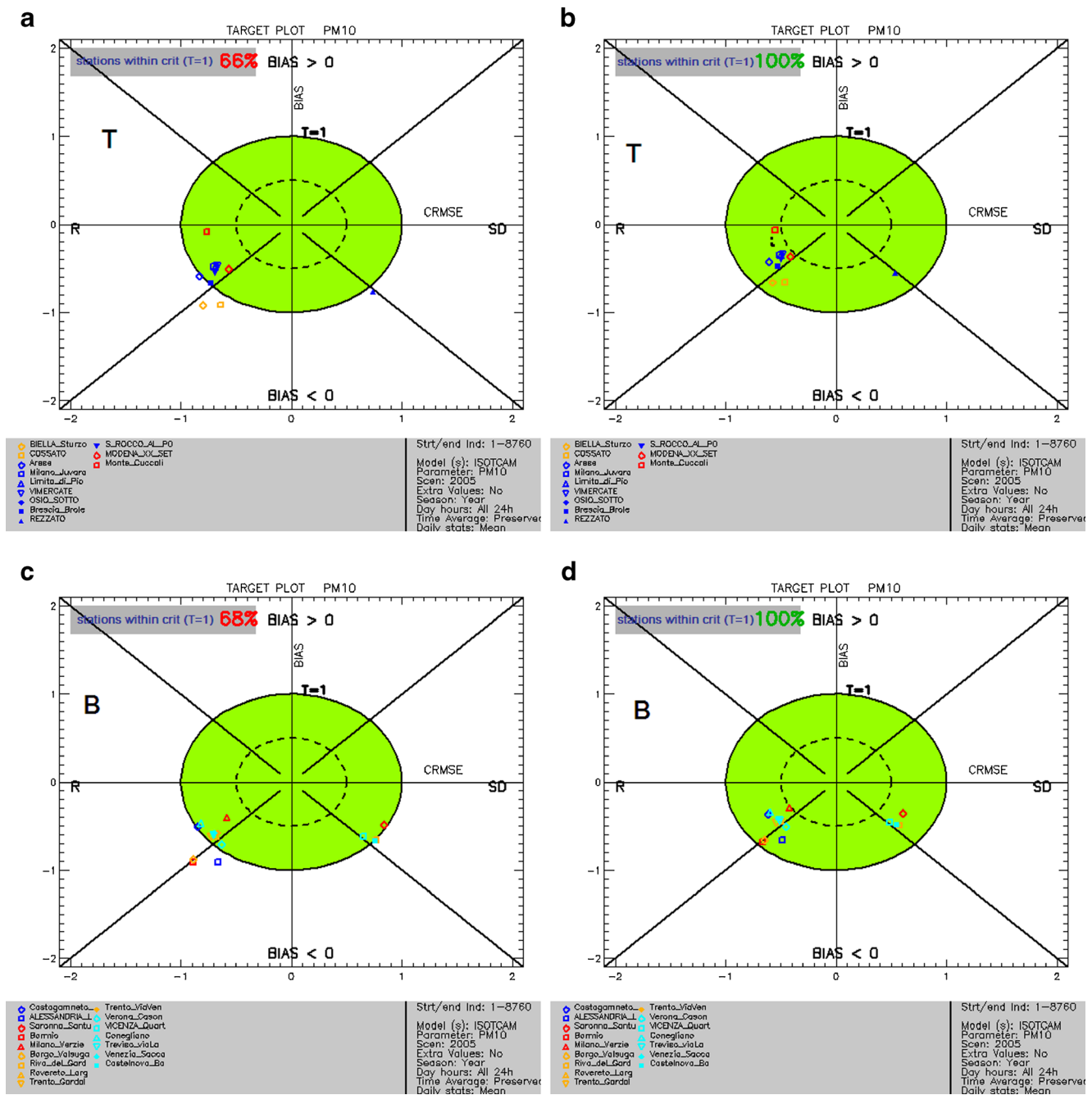


Fig. 8 Target plots of PM₁₀ daily mean concentrations for station groups T (top) and B (bottom). Targets on the left (a, c) are computed using default (gravimetric) values for u_r^{RV} and α , whereas targets on the right (b, d) make use of beta-ray and TEOM u_r^{RV} and α parameters. TCAM, year 2005

been used for groups T and B (Table 2). Figure 8 shows the improvement for both groups by using the new values: 100 % of sites are inside the green area for T and B compared to respectively 66 and 68 % of the “default” case. U values have also been computed (by applying Eq. 9) around LV for two sites of group T (Biella Sturzo and Cossato) and two sites of group B (Bormio, Alessandria Libertà). Performances at these monitoring stations are outside the green area for the default case (Fig. 8a, c). U values are reported in Table 3.

Table 3 U values around LV for Biella Sturzo, Cossato (T), Bormio, and Alessandria Libertà (B) computed using default parameters (Table 2) and new parameters (Table 2)

Group	D	Default	New
T	Biella Sturzo	11.40	15.89
T	Cossato	10.92	15.51
B	Bormio	7.53	10.20
B	Aless. Libertà	16.76	23.38

Conclusions

This study presents an application of the method of Thunis et al. (2012b) which makes use of the observation uncertainty (U) for normalizing the main statistical indicators (RMSE, NMB, NMSD, and R) used in model evaluation and computing the correspondent model performance criteria (MPC). In this way, MPC for the different statistics are derived by considering the same input (U).

Furthermore, the same margin of uncertainty (“less than 1” condition) is used for both model results and measurements.

Model results provided by the TCAM (Carnevale et al. 2008) for PM_{10} concentrations have been compared to measured data provided by 50 monitoring sites distributed across the Po Valley for year 2005. The target plot analysis reveals that TCAM generally performs better during the summer than winter; during the summer season, the model fulfills all criteria for correlation, systematic error, and standard deviation. In winter, the model uncertainty is associated with both the random (CRMSE) and the systematic components (NMB) of the error. The model shows a poor correlation with observations (between 0.2 and 0.7). The criterion is not fulfilled by 14 % of stations. Furthermore, the model tends to underestimate PM_{10} concentration with a bias (up to –30 %); NMSD is always negative (in the range from –10 to –60 %). In particular, the under-prediction of PM_{10} concentration in winter is attributable to the well-known difficulties of air quality models performing over the Po Valley, such as the sensitivity to the meteorological conditions and complex topography, the underestimation of primary organic compound emissions, and the modeling of SOA formation and growth.

A sensitivity analysis has also been performed by considering different values of u_r^{RV} and α in U computation. Parameters specific to the measurement instrument implemented at each station (gravimetric, TEOM, and beta-ray systems) have been introduced. TCAM performances for PM_{10} daily mean concentrations improve significantly (100 % of sites respect the MQO) with respect to the default case (gravimetric parameters for all sites). The use of more specific U parameters, based on the measurement instrument implemented at each monitoring station, is therefore recommended.

The study shows for the first time a real application of the methodology by Thunis et al. (2012a, b, c) as implemented inside the DELTA tool; it highlights how this approach helps identify potential model weaknesses which may not be revealed by the computation of the RDE only, providing a useful way for classifying air quality model performances with respect to the AQD (2008).

References

- AQD (2008) Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe (no. 152). Official Journal
- Borrego C, Monteiro A, Ferreira J, Miranda AI, Costa AM, Carvalho AC, Lopes M (2008) Procedures for estimation of modelling uncertainty in air quality assessment. *Environ Int* 34:613e–620e
- Boylan JW, Russell AG (2006) PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. *Atmos Environ* 40:4946e–4959e
- Carnevale C, Decanini E, Volta M (2008) Design and validation of a multiphase 3D model to simulate tropospheric pollution. *Sci Total Environ* 390:166–176
- Carnevale C, Finzi G, Pisoni E, Thunis P, Volta M (2012) The impact of thermodynamic module in the CTM performances. *Atmos Environ* 61:652–660
- Carter WPL (1990) A detailed mechanism for the gas-phase atmospheric reactions of organic compounds. *Atmos Environ* 24A:481–518
- Chemel C, Sokhi RS, Yu Y, Hayman GD, Vincent KJ, Dore AJ, Tang YS, Prain HD, Fisher BEA (2010) Evaluation of a CMAQ simulation at high resolution over the UK for the calendar year 2003. *Atmos Environ* 44:2927e–2939e
- Cuvelier C, Thunis P, Vautard R, Amann M, Bessagnet B, Bedogni M, Berkowicz R, Brandt J, Brocheton F, Builtjes P, Carnevale C, Denby B, Douros J, Graf A, Hellmuth O, Hodzic A, Honore C, Jonson J, Kerschbaumer A, de Leeuw F, Minguzzi E, Moussiopoulos N, Pertot C, Peuch VH, Pirovano G, Rouil L, Sauter F, Schaap M, Stern R, Tarrason L, Vignati E, Volta M, White L, Wind P, Zuber A (2007) CityDelta: a model intercomparison study to explore the impact of emission reductions in European cities in 2010. *Atmos Environ* 41(1):189–207
- Daly A, Zannetti P (2007) Air pollution modeling—an overview. In: Zannetti P, Al-Ajmi D, Al-Rashied S (eds) *Ambient air pollution*, chapter 2.
- Denby B (2010) Guidance on the use of models for the European Air Quality Directive. A working document of the Forum for Air Quality Modelling in Europe FAIRMODE (ETC./ACC No. version 6.2)
- Derwent D, Fraser A, Abbott J, Willis P, Murrells T (2010) Evaluating the performance of air quality models (No. Issue 3). Department for Environment and Rural Affairs
- ECWG (2010) Guide to the demonstrations of equivalence of ambient air monitoring methods
- EPA (2005) EPA's guideline on air. Quality models (“guideline”). http://www.epa.gov/scram001/guidance/guide/appw_05.pdf
- Fountoukis C, Nenes A, Sullivan A, Weber R, VanReken T, Fischer M, Matias E, Moya M, Farmer D, Cohen R (2009) Thermodynamic characterization of Mexico City aerosol during MILAGRO 2006. *Atmos Chem Phys* 9:2141–2156
- Gilardoni S, Vignati E, Cavalli F, Putaud JP, Larsen BR, Karl M, Stenström K, Genberg J, Henne S, Dentener F (2011) Better constraints on sources of carbonaceous aerosols using a combined ^{14}C -macro tracer analysis in a European rural background site. *Atmos Chem Phys* 11(1):2503–2547. doi:10.5194/acpd-11-2503-2011
- Holstag AAM, Nieuwstadt FTM (1986) Scaling the atmospheric boundary layer. *Boundary-Layer Meteorol* 36:201–209
- Jolliff JK, Kindle JC, Shulman I, Penta B, Friedrichs MAM, Helber R, Arnone RA (2009) Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J Mar Syst* 76:64e–82e
- Marchuk G (1975) *Methods of numerical mathematics*. Editor: Springer-Verlag, New York, NY
- Nenes A, Pandis S, Pilinis C (1998) ISORROPIA: a new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols. *Aquat Geochem* 4(416):123–152

- Pederzoli A, Thunis P, Georgieva E, Carruthers D, Borge R, Pernigotti D (2012) Performance criteria for the benchmarking of air quality model regulatory applications: the “target” approach. *International Journal of Environment and Pollution (IJEP)*, vol. 50, Nos. 1/2/3/4, 2012
- Pepper DW, Kern CD, Long PE (1979) Modelling the dispersion of atmospheric pollution using cubic splines and Chapeau functions. *Atmos Environ* 13:223–237
- Pernigotti D, Georgieva E, Thunis P, Cuvelier C, De Meij A (2012) The impact of meteorology on air quality simulations over the Po Valley in northern Italy. *Atmos Environ* 51:303–310
- Pernigotti D, Thunis P, Gerboles M, Belis CA (2013a) Model quality objectives based on measurement uncertainty. Part II: NO₂ and PM₁₀. *Atmos Environ* 79:869–878
- Pernigotti D, Thunis P, Cuvelier C, Georgieva E, Gsella A, De Meij A, Pirovano G, Balzarini A, Riva GM, Carnevale C, Pisoni E, Volta M, Bessagnet B, Kerschbaumer A, Viaene P, De Ridder K, Nyiri A, Wind P (2013b) POMI: a model inter-comparison exercise over the Po Valley. *Air Qual Atmos Health* 6(4):701–715
- Rossa AM, Ferrario ME, Sansone M, Monai M (2012) Climatology of the static stability of the night-time Po Valley pbl from radio sondes and passive microwave radiometers. *Proceedings of the 9th International Symposium on Tropospheric Profiling, L'Aquila, Italy, September 2012*. ISBN: 978-90-815839-4-7
- Scire JS, Insley EM, Yamartino RJ (1990) Model formulation and user's guide for the CALMET meteorological model. Technical report, California Air Resources Board, Sacramento
- Stow CA, Jolliff J, McGillicuddy DJ, Doney SC, Allen, JI, Friedrichs MAM, Rose KA, Wallhead P (2009) Skill assessment for coupled biological/physical models of marine systems. *J Mar Syst* 76, 4e15
- Thunis P, Triacchini G, White L, Maffei G, Volta M (2009) Air pollution and emission reductions over the Po-Valley: air quality modelling and integrated assessment. *Proceedings of the 18th World IMACS/ MODSIM Congress, Cairns, Australia, 13–17 July 2009*
- Thunis P, Cuvelier C, Roberts P, White L, Nyiri A, Stern R, Kerschbaumer A, Bessagnet B, Bergström R, Schaap M (2010) EURODELTA: evaluation of a sectoral approach to integrated assessment. *Modelling—second report*. http://www.eurosfair.prd.fr/7pc/doc/1301907090_lbna24474enc_002.pdf?PHPSESSID=75bb8ddcaca78c97a1777f06549d5065
- Thunis P, Georgieva E, Pederzoli A (2012a) A FAIRMODE benchmarking tool to evaluate air quality model performances in regulatory applications. *Environ Model Softw* 38:220–230 (December 2012)
- Thunis P, Pederzoli A, Pernigotti D (2012b) Performance criteria for evaluating air quality model applications for policy use. *Atmos Environ* 59:476–482 (November 2012)
- Thunis P, Pederzoli A, Georgieva E, Cuvelier C, Pernigotti D (2012c) DELTA V3.2 user's guide. <http://aqm.jrc.ec.europa.eu/DELTA/Publications.htm>
- Thunis P, Pernigotti D, Gerboles M, Belis C (2013) Model quality objectives based on measurement uncertainty. Part I: O₃. *Atmos Environ* 79:861–868
- Yamartino RJ, Scire JS, Carmichael GR, Chang YS (1992) The CALGRID mesoscale photochemical grid model—I. Model formulation. *Atmos Environ* 26A:1493–1512