# AIR QUALITY FORECASTS ON A KILOMETER-SCALE GRID OVER COMPLEX SPANISH TERRAINS

**M.T. Pay[1,*], F. Martínez[1], M. Guevara[1], J.M. Baldasano[1,2]**

[1]{Earth Sciences Department, Barcelona Supercomputing Center-Centro Nacional de Supercomputación, Barcelona, Spain}

[2]{Environmental Modeling Laboratory, Technical University of Catalonia, Barcelona, Spain}

[*]{now at: Laboratoire de Météorologie Dynamique, École Polytechnique, Palaiseau Cedex, France}

## Abstract

CALIOPE-AQFS represents the current state-of-the-art in air quality forecasting systems of high-resolution running on high-performance computing platforms. It provides a 48-h forecast of ~~the main pollutants (~~$NO_2$, $O_3$, $SO_2$, $PM_{10}$, $PM_{2.5}$, CO, and $C_6H_6$~~) found~~ at ~~a~~ 4-km horizontal resolution over all of Spain, and at 1 km over the most populated areas in Spain with complex terrains (Barcelona, Madrid and Andalucia domains). Increased horizontal resolution from 4 km to 1 km over the aforementioned domains leads to finer texture and more realistic concentration maps, which is justified by the increase in $NO_2$/$O_3$ spatial correlation coefficients from 0.79/0.69 (4 km) to 0.81/0.73 (1 km). High resolution emissions using the bottom-up HERMESv2.0 model are essential for improving model performance when increasing resolution on an urban scale, but it is still insufficient. Decreasing grid spacing does not reveal the expected improvement in hourly statistics, i.e., decreasing $NO_2$ bias only by ~2 μgm$^{-3}$ and increasing $O_3$ bias by ~1 μgm$^{-3}$. The grid effect is less pronounced for $PM_{10}$ because part of its mass consists of secondary aerosols, which are less affected than the locally emitted primary components by a decreasing grid size. The resolution increase has the highest impact over Barcelona, where air flow is controlled mainly by mesoscale phenomena and a lower PBL. Despite the merits and potential uses of the 1-km simulation, the limitations of current model formulations do not allow ~~confirming~~ their expected superiority close to highly urbanized areas and large sources. Future work should combine high grid resolution with techniques that decrease subgrid variability, and also include models that consider urban morphology and thermal parameters.

# 1 Introduction

The World Health Organization (WHO) has recently shown that there is sufficient evidence ~~supporting the belief~~ that particulate matter (PM), ozone ($O_3$) and nitrogen dioxide ($NO_2$) affect human health (WHO, 2013). Although $NO_2$ and PM concentrations improved from 2002 to 2011 in Europe, the situation is still far from matching the WHO air quality guidelines (AQG). The European annual limit values for $NO_2$ (annual) and $PM_{10}$ (daily) had been exceeded at 42-43% of the traffic stations in 2011. For the same year, about 33% of the European urban population was exposed to $PM_{10}$ concentration above the daily limit value, and nearly 88% ~~was~~ exposed to the respective WHO AQG (EEA, 2013). Air pollution legislation for the protection of the increasing city population has recently increased the demand for urban air pollution forecasting systems that can assess and understand ~~its~~ dynamics, alert the population when health-related issues occur, and develop emission abatement plans (EEA, 2011).

When applying an air quality modeling system, defining the grid resolution is an important consideration~~;~~ ~~the~~ potential benefits of higher-resolution modeling should be weighed against~~:~~ the increased complexity of the inputs, CPU time, and disk space requirements. In theory, higher resolution modeling is expected to yield more accurate forecasts because of better resolved model input fields (topography, land cover and emissions)~~,~~ and better mathematical characterization of physical and chemical processes. Furthermore, high resolutions (ranging from 1 to 5 km) are essential to reproduce mesoscale phenomena, e.g., those controlling $O_3$ transport along the mountainous northeastern Mediterranean coast (Fay and Neunhäuserer, 2006; Jiménez et al., 2006). Even at ~~the~~ fine~~st~~ scale, the modeled concentrations are not necessarily the best (Mass et al., 2002; Gego et al., 2005; Valari and Menut, 2008)~~,~~ because increasing emission and meteorology spatial resolution can also increase uncertainties, at the risk of reduced model performance. Nowadays, fine horizontal resolution is a persistent challenge when assessing health impact and population exposure studies (Thompson et al., 2013).

Several studies have evaluated the impact of increasing horizontal resolution on different scales over the eastern and southeastern USA using the Community Multiscale Air Quality (CMAQ) model and the Comprehensive Air Quality Model with Extensions (CAMx), which range from 32 km – 12 km – 4 km (Cohan et al., 2006; Tesche et al., 2006; Queen and Zhang, 2008). They found no significant changes for $O_3$ and PM (<5%

on average), and those changes were even ~~lower~~ at resolutions of between 12 km and 4 km (<3%). Concerning PM components, Fountoukis et al. (2013) found that increased resolution provides differences mostly for primary PM rather than secondary PM. Recently, a model intercomparison exercise, named ScaleDep, was performed to determine the effect of grid resolution on air quality modeling performance over Europe at a regional and urban scale (Cuvelier et al., 2013). The exercise, involving five ~~C~~hemical ~~T~~ransport ~~M~~odels (CTMs) (EMEP, CHIMERE, CMAQ, LOTOS-EUROS and RCGC) running under the same conditions over the full year 2009 and at four resolutions (56, 28, 14, and 7 km), showed that it is difficult to define a grid size that is adequate for resolving the urban signal under all conditions affecting Europe. Still, a 14-km resolution seems to be a good compromise between background applications and those reproducing most of the urban signals (7 km resolution). However, the ScaleDep exercise did not distinguish between the different topographies or complex meteorological patterns which are characteristic of the Iberian Peninsula.

Few studies have been performed over selected areas in Spain; and of those, the focus has been mainly on $O_3$ and $NO_2$. Vivanco et al. (2008) evaluated the annual impact by increasing ~~the~~ resolution (to 36, 19, and 7 km) over Madrid for $NO_2$ and $O_3$. They used the WRF-CHIMERE model, disaggregating the EMEP emission inventory based on land use information. Their evaluation showed that the model improve~~s~~ ~~greatly~~ for $NO_2$ than $O_3$, ~~and~~ the most significant improvement ~~is~~ achieved when resolution increases from 36 to 19 km rather than to 7 km, which ~~is~~ linked to increased uncertainty in the emission data introduced with the disaggregation techniques. Jiménez et al. (2006) used the MM5-CMAQ model along with a bottom-up emission model (EMICAT2000) to assess the influence of grid resolution on $O_3$ (at 8, 4, and 2 km) over the complex terrain of the northeastern Iberian Peninsula (Catalonia) during a summer pollution episode. They ~~indicate~~ that~~, due to~~ improved performance of the mesoscale phenomena and a better allocation of emissions~~, a~~ 2-km resolution improves the capability of the model to simulate exceedances of European limit values. An important issue in both studies is the emission modeling approach (top-down vs. bottom-up) when applying high resolution at the local scale (<10 km). As Fountoukis et al. (2013) and Timmermans et al. (2013) demonstrate, in the range of local scale (e.g., the greater Paris area), the grid resolution is not currently the major source of discrepancies in model performance~~;~~ but ~~in fact,~~ the predicted concentrations and corresponding gradients are more consistent with observed concentrations when provided by bottom-up emission inventories rather than down-scaled inventories. If local variation in input data (e.g., emission patterns or land use) cannot be properly characterized, modeling with a finer grid resolution may not provide any great advantages.

Increasing resolution is a technical challenge, since computational cost markedly increases in inverse proportion to grid spacing. The current progress in computation allows increased model resolution and for multiple spatial scales to be investigated with the aim of establishing adequate grid size for forecasting air quality at the local scale. Recently, Colette et al. (2014) evaluated the impact of increasing resolution up to 2 km over the European continent by using the CHIMERE model for an episode of air pollution in 2009. They ~~used~~ 2 million grid cells ~~at~~ over 2000 CPUs of a high performance computing system, which was hosted by the French Computing Centre for Research and Technology (CCRT/CEA).

In terms of computational resources, horizontal resolution is critical to an operational air quality forecast. In Europe, operational air quality systems use resolutions ~~of~~ between 12-25 km~~; mean~~while application to a single country can reach resolutions of between 4-10 km (Zhang et al., 2012). Over Spain, there are three systems providing air quality forecasts running at different horizontal resolutions. The lowest resolution system is the Technical University of Madrid's OPANA (OPerational Atmospheric Numerical model for urban and regional Areas), running at 27 km x 27 km and based on the MM5/CMAQ/EMIMO models (San José et al., 2009). It is followed by the Spanish meteorological office's system (AEMET, http://www.aemet.es/es/eltiempo/prediccion/calidad_del_aire), which forecasts at 10 km x 10 km using the HIRLAM-HRN/MOCAGE/GEMS-TNO models. The CALIOPE Air Quality Forecast System (CALIOPE-AQFS; Baldasano et al., 2011; Pay et al., 2012a; and references therein), of the Barcelona Supercomputing Center–Centro Nacional de Supercomputación (BSC-CNS), runs at the highest resolution, 4 km x 4 km, and it is based on the WRFv3.5/CMAQv5.0.1/HERMESv2.0/BSC-DREAM8bv2 models. Moreover, CALIOPE-AQFS provides 1-km x 1-km resolution forecasts for the Madrid and Barcelona metropolitan areas (since 2009)~~,~~ and the Andalusian region (since 2013). Such resolution has been possible thanks to both the high performance computing resources at the BSC-CNS and the availability of detailed emission data covering Spain.

The previous work~~s~~ demonstrate there is not a single answer which explains the merits of high-resolution modeling for all applications. The present work aims to assess the impact of increasing the horizontal resolution from 4 km to 1 km, specifically over areas affected by heterogeneous emission patterns and complex terrains such as the Barcelona and Madrid metropolitan areas (BCN and MAD) together with the Andalusian region (AND). For that purpose, CALIOPE-AQFS forecasts pollutant concentrations ($O_3$, $NO_2$, and $PM_{10}$) at two horizontal resolutions: first at a 4-km resolution covering Spain (IP4), and second at a 1-km resolution covering the AND, BCN, and MAD domains. ~~The~~ study is performed for the period April 2013, which ~~presented~~ seven day~~s of an~~ air pollution episode. We use observations from routine air quality monitoring networks to evaluate both resolutions.

Section 2 describes the configuration and computational setup of CALIOPE-AQFS~~; it analyses the domains and the period under study and it defines the methodology used to evaluate the resolution increase~~. Section 3 quantifies the impact of resolution increase on forecasting hourly concentrations (and exceedances) in terms of~~:~~ pollutant, domain, building density and major emission sources. Section 4 concludes with the main results and some recommendations.

## 2 Methodology

### 2.1 Domain and period under study

Figure 1 shows the ~~main~~ $NO_2$ emission patterns and topographic characteristics of the ~~domains: the Barcelona and Madrid metropolitan areas (BCN and MAD) and the Andalusian region (AND)~~. BCN is a coastal area characterized by several valleys perpendicular to the coastline and two main mountain ranges, one coastal (500 m height) and one pre-coastal (1000-1700 m height). These features induce mesoscale phenomena such as sea-breeze and mountain-valley winds. On the other hand, MAD is a continental region with a much simpler topography that includes the Tajo valley in the south~~ern of MAD~~ and the mountain range of the Central System ~~located~~ in the north~~western MAD~~, with summits reaching 2500 m~~ height. These features bring different locally-driven flows~~.

The urban contribution in BCN (3.1 million inhabitants) is accompanied by industrial and power generation emissions, the road network and the harbor~~; meanwhile~~ the

Spanish capital of MAD is mainly influenced by emission from the urban area (5.8 million inhabitants) and the road network that connects MAD with the surrounding commercial and industrial zones as well as the urban areas.

AND is the southern-most region in Spain, with complex topography characterized by the large depression of the Guadalquivir Basin (delimited by the Iberian Massif and the Betic Range), which crosses the region from NE to SW over a 60–km stretch. About three quarters of AND has a mountainous orography, including the Sierra Nevada (3481 m). AND includes one of the five biggest cities in Spain, Seville (~700 000 inhabitants), which hosts industrial and electric generation activities around the Algeciras bay, and it is affected by dense maritime traffic through the Strait of Gibraltar.

The study is performed ~~over~~ April 2013. At the beginning and end of the month, the synoptic circulation was controlled by a low pressure system ~~displaced~~ over the south of the British Isles and ~~which~~ affected Western Europe by leading to atmospheric instability over the IP. This pattern is typical of transitional months such as April and November (García-Valero et al., 2012; Valverde et al., 2014), which produce precipitation and ~~decreased~~ temperatures because of cold and humid winds ~~entering~~ from the Atlantic Ocean. In contrast, from 12-18 April there was a high pressure system crossing the Iberian Peninsula in a SW-NE direction, transporting dust from the Sahara Desert and increased temperatures ~~of~~ up to 25-28ºC. During the latter episode, available air quality stations ~~at~~ the study domains ~~displayed~~ several exceedances of the European limit values (8 exceedances of the $NO_2$ hourly limit value, 25 exceedances of the $O_3$ information threshold, and 31 exceedances of the $PM_{10}$ daily limit value).

## 2.2 CALIOPE-AQFS

CALIOPE-AQFS has provided 48-h air quality forecasts for Europe and Spain since October 2006 (www.bsc.es/caliope) and has been described and evaluated in detail elsewhere (Baldasano et al., 2008, 2011; Pay et al., 2011, 2012a). Briefly, it integrates a meteorological model (WRF-ARW v3.5; Skamarock and Klemp, 2008), an emission model (HERMESv2; Guevara et al., 2013), a chemical transport model (CMAQv5.0.1; Byun and Schere, 2006; Appel et al., 2013), and a mineral dust atmospheric model (BSC-DREAM8bv2; Pérez et al., 2006; Basart et al., 2012); together, all of these comprise an air quality forecast system.

Figure 1 shows the working domains of CALIOPE-AQFS. First, CALIOPE-AQFS was run over Europe ~~at a~~ 12-km x 12-km horizontal resolution using initial/boundary conditions from the Final Analyses of the National Centers of Environmental Prediction (FNL/NCEP). The analyses began at 12 ~~h~~ UTC, at intervals of 6 h (0.5ºx0.5º) for meteorology. The global model LMDz-INCA2 (3.75ºx2.5º, Szopa et al., 2009) was used for chemistry. Then, CALIOPE-AQFS was run at ~~a higher horizontal resolution (~~4 km x 4 km (IP4)~~)~~ over the ~~Iberian Peninsula~~ using one-way nesting. In the present work CALIOPE-AQFS runs at 1 km x 1 km over the ~~domains at hand (AND, BCN and MAD)~~, with nesting of over IP4 throughout. HERMESv2.0 forecasts anthropogenic emissions for the year 2009 by following a bottom-up methodology (point, linear and area), and biogenic emissions using the MEGANv2.0.4 model (Guenther et al., 2006). Emissions are aggregated into 1-km grids for AND, BCN and MAD 1-km simulations, and into 4 km for IP4.

Vertically, WRF-ARW is configured with 38 sigma layers up to 50 hPa, with 11 characterizing the planetary boundary layer (PBL)~~; meanwhile~~ CMAQ vertical levels are obtained by collapsing from the 38 WRF levels to a total of 15 layers that steadily increase from the surface up to 50 hPa. Six layers are within the PBL, and the first layer depth is 39 m.

The present WRF setup uses~~:~~ the Rapid Radiation Transfer Model (RRTM) and Dudhia for long- and short-wave radiation, respectively~~;~~ the Kain-Fritsch cumulus parameterization (Kain and Fritsch, 1990)~~;~~ the single-moment 3-class (WSM3) microphysics scheme~~;~~ and the Yonsei University PBL scheme (YSU). The Noah land-surface model (NoahLSM), based on the U.S. Geological Survey's (USGS) land-use data, is used by default in the present WRF configuration.

Currently, a new CMAQ version is being tested in the CMAS community, namely CMAQv5.0 (CMAQ, 2012). It includes substantial scientific improvements over Version 4.5 and is especially devoted to improving SOA formation as well as the dynamic interactions of fine and coarse aerosols. Based on the evaluation results from the previous CMAQ version within CALIOPE-AQFS (4.5 vs. 5.0) (Pay et al. 2012b), CMAQ has been updated to Version 5.0.1 using the CB05 chemical mechanism (Yarwood et al., 2005), the AERO5 for aerosol modeling, and the in-line photolysis calculation.

CALIOPE-AQFS considers desert dust contribution by means of the BSC-DREAM8bv2, which runs off-line at a 0.5º x 0.5º resolution covering Europe, North Africa and the Middle East. Its outputs are mass conservative interpolated to the CMAQ's Lambert conformal conic grids ~~and~~ at the required resolution and domain. After interpolating, the modeled $PM_{10}$ concentration is~~;~~ the sum of Aitken, accumulation and coarse-mode modes from CMAQ, and the corresponding BSC-DREAM8bv2 bins with a diameter of ≤ 10 µm (Pay et al., 2012a).

## 2.3 Computational strategy

Running CALIOPE-AQFS at 4 and 1 km is a technical challenge. The simulations are run on MareNostrum supercomputer (Intel Xeon E5-2670, 16 CPUs and 64 GB RAM memory per node) at BSC-CNS. Table 1 depicts the computational requirements for forecasting air quality at 48 h for each domain. The number of CPUs was chosen to maximize CPU efficiency. Thanks to the parallelization of meteorological and air quality models, MareNostrum uses up to 256 CPUs. Due to the variable nature and complex dependencies, the computational time for forecasting 48 h of air quality fields for the 4 domains is 8-9 hours. The most computationally demanding domain is the AND~~,~~ at 1-km resolution (366x358 cells, 256 CPU max., and 300 min). For the April 2013 simulation, times add up to 2880 CPU hours/day, or 86400 CPU hours in one CPU (9.86 years). The storage for the April 2013 output files was 6.13 TB (~200 GB/day).

## 2.4 Evaluating the increase in resolution

Comparing CALIOPE-AQFS ~~grid resolutions~~ and measurements was done ~~in terms of~~ gas-phase and aerosol concentrations ($O_3$, $NO_2$, and $PM_{10}$). Representativeness continue to be a challenge when comparing gridded simulations to observational data at a point in time and space, as modeled concentrations represent a volumetric average over an entire grid cell. Furthermore, the stochastic compound embedded in the observations is not accounted for. Concerning temporal representativeness in the present comparison, both modeled and measured concentrations are averaged hourly. CALIOPE-AQFS operationally receives air quality measurements from Spanish administrative networks in near real time (NRT) without any ~~quality data or~~ quality control. For the present study, NRT measurements are filtered by removing data before and after measurement interruptions or calibrations. Also, a minimum cut-off threshold of 1 µgm$^{-3}$ is applied to the observed concentrations in order to avoid unrealistic observations. After filtering, the number of stations is 48/30/36 for $O_3$, 51/42/42 for $NO_2$, and 52/15/33 for $PM_{10}$ ~~at AND/BCN/MAD~~.

The meteorological fields are evaluated for wind speed ~~at 10 m (U10), and wind direction (WD10)~~ and temperature at 2 m (T2M)~~, all of them at 10 METAR stations~~ located at airports (6/2/2 stations in AND/BCN/MAD)~~. They~~ are discussed in Sect. S1.

Figure 2 shows the location of the air quality and METAR (METeorological Aerodrome Report) stations over the respective domains. The spatial representativeness of the air quality network is highly variable. The influence of the station type is based on two classifications of air quality monitoring stations, the environment type (rural, R; suburban, S; and urban, U), and the dominant emission source (traffic, T; industrial, I; and background, B). These were derived from the Council decision 97/100/EC (Garber et al. 2002).

The evaluation is based on discrete statistics performed on an hourly basis. We consider the correlation coefficient (r), mean (absolute, relative, and fractional) biases (MB, MNBE, and MFB), and error (MAE, MNGE, and MFE). Root Mean Square Error (RMSE) is also calculated because it intensifies large differences between measured and observed concentrations (Table A1).

In order to evaluate the effect of ~~resolution increase~~ on forecast~~ed~~ exceedances and non-exceedances of limit values established by the European legislation, we calculate categorical statistics based on comparisons with fixed concentration thresholds (T). The calculated statistics are accuracy (A), bias (B), probability of detection (POD), critical success index (CSI), and false alarm ratio (FAR), whose formulas and descriptions are ~~explained~~ in Table A2 and ~~elsewhere~~ in Kang et al. (2005) and Eder et al. (2006). The 2008/50/EC directive sets an information threshold of 180 $\mu gm^{-3}$ for maximum daily $O_3$ concentrations (Max 1h $O_3$) and a target value of 120 $\mu gm^{-3}$ for the maximum daily 8-h running $O_3$ mean (Max 8h $O_3$), which should not be exceeded more than 25 days per year. It establishes a limit value of 200 $\mu gm^{-3}$ for maximum daily $NO_2$ concentrations (Max 1h $NO_2$), and 50 $\mu gm^{-3}$ for the daily $PM_{10}$ mean (Mean 24h $PM_{10}$), which should not be exceeded more than 35 times per year. Therefore, categorical evaluation will be performed for Max 1h $NO_2$, Max 1h and Max 8h $O_3$, and Mean 24h $PM_{10}$. Note that mean and maximum concentrations are calculated by considering at least 75% of the data in the corresponding time base, i.e., values of at least 18 hours per day for Mean 24 h, Max 1h, and Max 8h; and 6 hours for 8 h values, as established by 2008/50/EC.

## 3 Concentration maps and spatial representativeness

To analyze the spatial differences between resolutions, Figs. 3, 4, and 5 show the monthly mean concentration maps for April 2013 over MAD, BCN and AND domains at 4 km (left panels) and 1 km (right panels) for $NO_2$, $O_3$, and $PM_{10}$, respectively.

The maps of $NO_2$ and $PM_{10}$ at both resolutions display similar distribution along the MAD and BCN urban plumes. On-road traffic constitutes the main source of primary pollutants in MAD and BCN. HERMESv2.0 estimates that 75% and 59% of $NO_x$ emissions are produced by on-road traffic in both domains, respectively. Consequently, when the resolution increases, the monthly mean $O_3$ concentration maps are almost identical, although the $NO_x$ titration effect on $O_3$ is significant along highways and major point sources. In AND, $NO_2$ and $O_3$ concentrations are also conserved between resolutions along the shipping route crossing the Strait of Gibraltar towards the Mediterranean Sea.

However, the definition of $NO_2$ concentrations along highways connecting the biggest cities with the rest of the country and industrial sectors are more easily identified at 1-km simulations than at 4 km, especially along those roads from/to Barcelona (e.g., the AP7 Mediterranean highway and C32, which connects the harbor and the airport) and Madrid (the A-2 and A-6 in the north, and A-3, A-4 and A-5 in the south). In the same way, 1-km $O_3$ maps are more textured than those at 4 km along highways, because the titration effect is more significant at 1 km, due to less dilution within grid cells. The titration effect of $NO_x$ on $O_3$ over the main sources is more forceful in BCN than in MAD, given that BCN has a larger concentration gradient resulting from complex topography and recirculation flows that accumulate pollutants.

The improvement of the definition along roads in AND is lower than that observed in the MAD and BCN domains, due to the fact that the AND domain is bigger and displays lower traffic emission sources than the MAD or BCN domains. Regarding $PM_{10}$, the main component in AND is the desert dust (~40% in both resolutions) from North Africa. This is because there were two episodes on 14-19th and 25-26th April that affected the IP, as shown by the S-N $PM_{10}$ gradient (Fig.5e and f). The desert dust is transported from long-range simulation with BSC-DREAM8bv2.

Over complex terrains, the 1-km simulation ~~allows reproducing~~ more realistic $NO_2$ concentration maps because of its more detailed topographic information ~~input~~. For instance, the BCN 1-km simulation displays the lowest $NO_2$ concentrations (< 10 µgm$^{-3}$) along the coastal chain (500 m height) and pre-coastal chain (1000-1700 m height), except for the city's urban hill, where concentrations reach 20-40 µgm$^{-3}$. In contrast, the 4-km simulation provides smoother $NO_2$ concentrations without any concentration gradient. Thus, the 1-km simulation generates slightly higher $O_3$ background concentrations than ~~does~~ the 4-km simulation along the BCN pre-coastal chain (66-70 vs. 70-74 µgm$^{-3}$), as well as across the Iberian Massif (AND), where the $O_3$ ~~map~~ displays significant structure due to the higher resolution topography that ~~shapes~~ the basin and the ~~on-road traffic~~.

Figures 3, 4, and 5 include dots corresponding to mean concentrations at air quality stations that help to qualitatively evaluate the modeled spatial representativeness at both resolutions. Note the ~~high~~ ~~concordance~~ between $NO_2$ observations and the 1-km simulation near the primary suburban traffic roads in BCN (e.g., Vilafranca, Igualada, Manresa, and Mataró). Regarding $O_3$, although observed concentrations depict an overall tendency of the model to underestimate concentrations at both resolutions, the 1-km simulation ~~displays a higher accord~~ with measurements at rural background stations (e.g., El Atazar, San Martín, Villa de Prado, Villarejo and Orosco stations in MAD), and at suburban traffic stations (e.g., Manresa, Igualada and Vilafranca in BCN, with modeled $O_3$ concentrations of around 54-58 µgm$^{-3}$ at 1 km, and 60-66 µgm$^{-3}$ at 4 km). For $PM_{10}$, comparisons with measurements show that modeled concentrations are underestimated over background areas, mainly outside the urban/suburban area, as already discussed in Pay et al. (2012a). However, $PM_{10}$ measurements at the urban/suburban stations of Vilafranca, Sant Celoni, and Mataró in BCN (14-16 µgm$^{-3}$) show a higher ~~concordance~~ at 1 km than at 4 km (12-14 µgm$^{-3}$ vs. 8-10 µgm$^{-3}$).

The spatial variability of the increased resolution is quantitatively analyzed by means of concentration maps, shown in Fig. 6 for $NO_2$, $O_3$ and $PM_{10}$ over AND, BCN and MAD. Over all domains, the ~~explained spatial variability~~ improves as a function of the resolution increase for $NO_2$ and $O_3$, ~~which is sustained~~ by the increase in monthly r from 0.79 (4 km) to 0.81 (1 km) for $NO_2$, and from 0.69 to 0.73 for $O_3$. The slopes improve with the resolution increase, from 0.72 (4 km) to 0.77 (1 km) for $NO_2$, and from 0.50 (4 km) to 0.54 (1 km) for $O_3$. This results from the improved model performance at urban stations, indicating that CALIOPE-AQFS ~~explains better~~ the magnitude of the variability between urban regions at 1 km than at 4 km. In contrast, for $PM_{10}$, the monthly r decreases from 0.67 to 0.58 when the resolution increases. ~~Although~~ the PM10 spatial variability over BCN and MAD improves when the resolution increases (r increases by 0.01 and 0.04, respectively), the global correlation coefficient ~~deviates mainly~~ at the AND stations (52/100)~~, where r decreases~~ by 0.1 from 0.36 (4 km) to 0.26 (1 km). Despite the unfavorable effect of the resolution increase in $PM_{10}$ over AND, the $NO_2$ and $O_3$ concentrations show the highest absolute increase ~~in~~ the spatial r over this domain, from 0.62 (4 km) to 0.71 (1 km) for $NO_2$ and from 0.58 (4 km) to 0.64 (1 km) for $O_3$ (increasing r by 0.09 and 0.06, respectively).

## 4 Temporal evaluation

~~The present section discusses the temporal evaluation of the resolution increase by pollutant, environment, predominant emission sources, and study domain. Fig. 7 summarizes the statistical evaluation.~~

### 4.1 Pollutant

Table 2 ~~depicts~~ the statistical evaluation by pollutant, with a focus on the reproduction of high concentrations established by the European directive (2008/50/EC). Depending on the pollutant's lifetime and variability, as well as its dependency on precursors, increased resolution shows different impacts. The resolution increase has a positive effect on $NO_2$, decreasing its bias by 2.0 $\mu gm^{-3}$ (from -4.5 to -2.5 $\mu gm^{-3}$)~~;~~ but it also increases absolute (squared) error~~s~~ by 0.3 $\mu gm^{-3}$ ~~(~~0.9 $\mu gm^{-3}$~~)~~. This positive effect is sustained by the perceptual variability, where the MB (MFB) is reduced by 42% (19%)~~;~~ ~~whereas~~ MAE (MFE) only increases by 2% (1%). The ~~correlation coefficient~~ does not significantly change, which is obvious because emissions at both resolutions are modeled using the same approach. The bias improvement at 1-km resolution is justified, because the higher resolution leads to better emission allocation from point, linear or area sources~~; it~~ decreases the artificial dilution of emission compared to the larger grid area~~;~~ and, due to the decrease of artificial dilution, it treats chemistry more properly near large emission sources.

In contrast, the resolution increase has a negative effect on hourly and Max 8h $O_3$ concentrations~~ because it~~ increases~~,~~ bias~~es~~ and error~~s~~ by 0.1-0.8 $\mu gm^{-3}$. Relative (fractional) bias~~es~~ and error~~s~~ increase by 8% (15%) and 1% (1%), respectively, for hourly $O_3$; and 6% and 4% for Max 1h $O_3$. However, the statistical evaluation alone is not enough to explain the impact of the resolution increase on $O_3$ concentrations.

According to the categorical evaluation, only a few exceedances of the European target and limit values were detected for Max 1h $NO_2$ (9), Max 1h $O_3$ (25), and Mean 24 $PM_{10}$ (31) in April 2013. Thus, categorical evaluation is performed on the temporal basis established by the European legislation, but it uses a T based on the 75p of the observed concentrations in each case~~,~~ T ~~corresponds to~~ 71 $\mu gm^{-3}$ for Max 1h $NO_2$, 108 (101) $\mu gm^{-3}$ for Max 1h (Max 8h) $O_3$, and 27 $\mu gm^{-3}$ for Mean 24h $PM_{10}$.

Overall, CALIOPE-AQFS underestimates exceedances at both resolutions, indicating that errors of missing observed exceedances are not ~~totally~~ resolved by a ~~resolution increase~~ (a$<$d). The best performance is found for Max 1h $NO_2$, where bias (B) improves from 37% (4 km) to 40% (1 km).

For $NO_2$ Max 1h, there are 953 observed exceedances (b+d) of the threshold (T=47 $\mu gm^{-3}$). Increasing the resolution increases the POD from 49% (4 km) to 56% (1 km). As POD, CSI ~~studies~~ the exceedances, but in a more ~~coherent~~ way by considering both false alarms and missing events. ~~Both~~ POD and CSI increase by 14% and 20% when resolution is increased. ~~The opposite effect appears for $O_3$. Aside from the fact that the $O_3$ POD is relatively low,~~ the POD decreases as the resolution increases. Of the 1306 observed exceedances of the 108 $\mu gm^{-3}$ Max 1h, CALIOPE-AQFS detected 112 exceedances at 4 km and only 96 at 1 km. Increasing resolution decreases POD and CSI by 22% and 25% for $O_3$ Max 1h, whereas they do not significantly change for Max 8h $O_3$ and Mean 24h $PM_{10}$.

FAR increases for Max 1h $NO_2$ (from 40% to 42%) and decreases for Max 1h $O_3$ (from 27% to 17%) when the resolution increases. In relative terms, this variability is more significant for Max 1h $O_3$ (37 %) than for Max 1h $NO_2$ (5%), indicating that, in terms of failures, the resolution has a positive global effect by reducing false exceedances.

For various reasons, accuracy (A) remains almost constant when the resolution increases. ~~Regarding~~ $NO_2$ and $O_3$, ~~it is~~ due to a stable sum of b and c, increasing the b at the cost of c, and vice versa. For $NO_2$, the number of hits (b) to forecast Max 1h at 1 km is higher than 4 km (537 vs. 466), but the number of correct negatives at 1 km is lower than at 4 km (2439 vs. 2517). The resolution increase has the opposite effect on $O_3$ over b and c for both Max 1h and Max 8h.

## 4.2 $PM_{10}$ components

The resolution increase has the lowest effect on $PM_{10}$ hourly concentrations and its exceedances (<1%). $PM_{10}$ components are secondary inorganic aerosols (SIA), which include sulfate ($SO_4$), nitrate ($NO_3$) ammonium ($NH_4$), secondary organic aerosol (SOA), elemental carbon (EC), sea salt (SS), desert dust (DD), and primary PM (PPM).

Pay et al. (2012a) ~~already~~ evaluated the PM components at some ~~Spanish~~ urban and rural background stations using the CALIOPE-AQFS based on CMAQv4.5, ~~and~~ they showed that the model underestimated ~~the secondary inorganic aerosols~~ by a factor of 2-3. The highest underestimation was found for fine carbonaceous aerosols (a factor of 4), in part related to the state-of-the-science concerning ~~secondary organic aerosol~~ formation pathways. The updated version of CMAQ, v5.0.1 includes scientific improvements ~~concerning~~ SOA formation and aerosol dynamics, which could improve the modeled $PM_{10}$ performance ~~for~~ its components.

Figure 8a shows that the resolution increase does not significantly change the $PM_{10}$ composition. DD remains the main contributor (~40-41%), followed by PPM (22-24%), SIA (~21-22%), SS (9-11%), EC (~4%) and SOA (~0.6%). However, the effect of the increased resolution on $PM_{10}$ component concentrations is different (Fig. 8b), depending on their origin, atmospheric cycle and the way they are modeled. DD concentrations do not change between resolutions, because they are mass conservative when interpolated from 0.5ºx0.5º till 1 km x 1 km.

~~Regarding~~ SIA, increasing the resolution increases $NO_3$ and $NH_4$ concentrations by 4 and ~2%, respectively, and it decreases $SO_4$ by ~2%. The $NH_4$ increase means there are more primary precursors ($H_2SO_4$ or $HNO_3/NO_2$) available to neutralize $NH_3$ (gas) to $NH_4$ (aerosol). However, the variability between $SO_4$ and $NO_3$ is more difficult to explain~~,~~ due to the nonlinearity of photochemistry and aerosol formation, which is controlled to some extent by the ISORROPIA thermodynamic equilibrium. Furthermore, the absence of aerosol measurements for April 2013 does not allow us to ~~explain this situation~~.

~~The resolution increase displays the highest decrease for SS (~16%).~~ CMAQv5.0.1 simulates SS emission as a function of the wind speed and ~~the~~ relative humidity (Gong, 2003; Zhang et al., 2005). Although not shown here, when the resolution increases, the wind speed increases at the available $PM_{10}$ stations by ~1.4/0.4/0.2 $ms^{-1}$ over AND/BCN/ MAD, and also over the open ocean.

For primary PM components ~~(~~EC and PPM~~)~~, increasing resolution ~~presents the highest increase in~~ concentration ~~(~~by 10 and ~12%, respectively~~)~~. As for $NO_2$, the 1-km simulation leads to a reduced effect of artificial dilution of emissions in a grid cell, so concentration gradients are stronger than in the 4-km simulation.

### 4.3 Domain

~~Due~~ to differences in geographical location and emission patterns over the domains ~~under study, the resolution increase has different impacts~~ (Fig. 7). BCN shows the highest $NO_2$ bias decrease (73%) when the resolution increases, ~~yet with~~ no effect on the correlation (<7%). However, $O_3$ shows significant variability over BCN, increasing r ~~(~~by 4%~~)~~ and MB ~~(~~by 23%~~)~~. To a lesser extent, MB also increases over AND (by 8%). Meanwhile, the variability over MAD is reduced (<4%). MB decreases for $PM_{10}$ (< 1 $\mu gm^{-3}$) over the urban domains of MAD (3%) and BCN (16%), and increases over AND (7%).

Figure 9 analyzes the impact of the resolution increase on daily cycles. Although PBL measurements are not available, PBL daily cycles are displayed together in order to find some correlations with the daily pollutant variability. Due to the lamination of PBL growth by the Mediterranean sea breezes, the PBL reaches its maximum height at midday, being the highest in MAD (1600 m AGL) followed by AND (1000 m AGL) and BCN (900 m AGL).

As shown in Sect. S1, the pollutant transport at the BCN coastal domain is controlled by mesoscale phenomena such as sea-breezes (day) and land-breezes (night), which are a result of its complex topography and location (Baldasano et al., 1994; Millán et al., 1997; Gonçalves et al., 2009). The $NO_2$ daily cycle is highly influenced by traffic emissions (Fig. 9). Both resolutions show the highest underestimations ~~at~~ the morning peak (5-9 am) (~20 $\mu gm^{-3}$). Although the afternoon peak is well reproduced, there is excessive variability at both resolutions, ~~which results from~~ problems with wind direction. During the sea breeze period, the mean simulated wind was more easterly than westerly, as registered by measurements (Sect. S1). Several works indicate that WRF ~~does not faithfully reproduce~~ the morning and evening transition over the urban environment, possibly because it does not model the heat retention in cities (Makar et al., 2006; Appel et al., 2013). Increasing the resolution increases ~~the~~ $NO_2$ concentrations from 14 µgm$^{-3}$ (4 km) to 17 µgm$^{-3}$ (1 km) during the morning hours after sunrise (5-9 am) and in the evening hours after sunset (5-9 pm). This behavior could be explained by PBL variability when increasing the resolution, which decreases PBL height by ~33 m for these hours.

$NO_2$ performance impacts the $O_3$ daily cycles over BCN, showing that 4- and 1-km simulations underestimate maximum $O_3$ concentrations by ~20 µgm$^{-3}$ at midday (1-4 pm)~~,~~ and ~~it~~ overestimate~~s~~ minimum $O_3$ concentration by ~20 µgm$^{-3}$ in the morning hours after sunrise (5-9 am). The resolution increase ~~allows slightly decreasing~~ $O_3$ concentrations at night, which is perhaps ~~controlled by the PBL decreasing at 1 km~~ during the early morning and late afternoon, when PBL ~~reaches the minimum height~~. During these hours, the titration effect of $NO_2$ on $O_3$ is ~~more effective~~, improving the $O_3$ overestimation of the daily minimum, which allows a slightly ~~increasing~~ hourly r (2%). However, $O_3$ underestimation increases in the late afternoon, contributing to an increase in the hourly mean bias from ~9 µgm$^{-3}$ (4 km) to ~11 µgm$^{-3}$ (1 km).

In BCN, the $PM_{10}$ underestimation is not systematic throughout the daily cycle (Fig. 9), which shows a bias of ~20~~/10~~ µgm$^{-3}$ at day~~/night~~ time. The higher daytime underestimation as compared to the nighttime cannot be explained by the current results, but it could be a result of missing sources and problems with PBL overestimation and ~~emission dilutions~~. The resolution increase ~~allows reducing~~ the bias by ~1 µgm$^{-3}$ (16%), especially during early morning and late afternoon, when the highest PBL variability between resolutions is ~~detected~~. Although the evaluation of T2M, U10 and WD10 indicates that the resolution increase has a ~~low~~ effect over BCN (Sect. S1), the reduction of the artificial dilution of $NO_2$ emissions –together with a lower PBL height at 1 km than at 4 km during the night and early morning– allows improving $NO_2$, $O_3$ and $PM_{10}$ concentrations, which in turn decreases their biases.

~~In~~ AND, the model ~~at both resolutions~~ underestimates observed $NO_2$ concentrations throughout the daily cycle ~~(~~~5 µgm$^{-3}$~~)~~, with the highest underestimation ~~at~~ the morning peak (~25 μgm$^{-3}$) and the lowest ~~at~~ the afternoon peak (~10 μgm$^{-3}$). The resolution increase reduces the bias from -3.5 to 2 μgm$^{-3}$ (by 43%) and increases r by 7% (from 0.39 to 0.41). As in BCN, the $NO_2$ underestimation directly impacts the $O_3$ daily cycle (which the resolution increase cannot resolve), increasing the bias by ~1 μgm$^{-3}$, a phenomenon that is predominant in the morning hours. In the case of $PM_{10}$, the daily cycle indicates that the biases are almost systematic throughout the ~~daily cycle~~ (~22 μgm$^{-3}$). Increasing the resolution increases the bias by less than 4% in the late afternoon, which is perhaps dominated by the ~~PBL~~ decrease. When the resolution is increased, $NO_2$ performs better because of the improved model performance for the temperature and wind speed (Sect. S1), as well as the lower ~~nocturnal~~ and higher ~~diurnal~~ PBL. Meanwhile the $O_3$ and $PM_{10}$ performance do not significantly change.

During April 2013, the main flow over MAD was controlled by S-SW synoptic winds channeled by orographic barriers in the NW domain and the Tajo valley (Valverde et al., 2014). The $NO_2$ daily cycle depicts a high influence of traffic emissions (Fig. 9), showing significant model underestimation at both resolutions for the morning/afternoon peaks (~15/10 μgm$^{-3}$). Note that, in terms of mean and variability resulting from southeastern winds, the model performs well at the afternoon peak. $NO_2$ performance leads to more accurate $O_3$ daily cycles than in AND and BCN, especially in the early morning, when the titration effect of $NO_2$ is ~~more efficient~~ because the $NO_2$ morning peak underestimation is ~~lower~~ when compared to the other domains. Meanwhile, the modeled $PM_{10}$ at both resolutions presents a profile controlled by traffic emissions. Observed concentrations display a flatter daily cycle, in which the model underestimation reaches 40 μgm$^{-3}$ in the morning. Increasing resolution shows a positive effect for $NO_2$, and $PM_{10}$ increases r by 0.01 and reduces MB and RMSE by 0.1-0.2 μgm$^{-3}$. However, it ~~depicts~~ the lowest variability when compared to the other domains (<5 % for bias, error and r), ~~which is~~ the result of ~~a~~ relatively simpler topography and meteorological patterns.

## 4.4 Environment and major sources

Figure 7 shows that the resolution impact also depends on the type of area and the dominant emission source. Theoretically, the meteorological fields of urban areas differ from those of surrounding rural areas because of their different morphology (radiation trapping and wind profiles), surface materials (heat storage) and variable energy consumption (heat release).

Increasing resolution reduces the $NO_2$ bias at suburban and urban stations by 1.8-2 μgm$^{-3}$, and, to a lesser extent, by 1.2 μgm$^{-3}$ at rural stations. The correlation coefficients also improve at suburban stations (from 0.48 to 0.52) and rural stations (from 0.34 to 0.35). ~~That is not surprising, because the 1-km grid allows~~ better allocation of land-use categories (urban vs. rural) and of their fraction in a grid cell ~~than does a 4-km grid~~. The $NO_2$ biases ~~exhibit a relative~~ 39% (65%) ~~decrease~~ at urban (background) stations, but $O_3$ biases increase by 9% (5%). For $PM_{10}$, the resolution increase does not significantly change as a function of area type, ~~and it depicts variation in biases and errors~~ of less than ±4% (<0.5 μgm$^{-3}$).

The low improvement at urban stations is obviously because the NoahLSM land-surface model does not consider the effect of urban morphology or thermal parameters in order to accurately model meteorological fields. Modeling air quality on an urban scale over cities requires a description of the heat/momentum exchange between buildings and the lower atmospheric layers. For instance, the impact of using an urban model on meteorological fields over the greater Paris area was studied by Kim et al. (2013) using WRF with the Urban Canopy Model, demonstrating that~~,~~ ~~below a 1000-m height,~~ overestimations of wind speed were significantly reduced.

The ~~r effect of~~ increasing resolution is positive for primary pollutants near important emission sources. For example, it reduces $NO_2$ biases at traffic (industrial) stations by ~3 μgm$^{-3}$ (2 μgm$^{-3}$), but it increases $O_3$ biases by ~2 μgm$^{-3}$ (1 μgm$^{-3}$). However, the resolution increase in the range of 4-1 km does not exhibit the expected improvement on the hourly statistics that are based on the constraints of the current model formulation. In other words, it cannot resolve the subgrid air quality variability merely by increasing resolution. For instance, although on-road traffic emissions are estimated by following a bottom-up approach along highways and routes, heterogeneity is lost in the CTM volume averaging process, which artificially dilutes emission rates over the grid cells. The resolution effect is ~~the~~ lowest at background stations, which are not influenced by any single source, but rather by the integrated contribution from all sources upwind of the stations where variations are less than 1% for $O_3$ and $PM_{10}$ (< 1

$\mu gm^{-3}$). However, background $NO_2$ levels increase by ~1 $\mu gm^{-3}$ (48%) from 4 km to 1 km.

Figure 10 shows the temporal series and daily cycles for $NO_2$ and $O_3$ at traffic and background stations throughout the episode of 12-18th April, 2013. At traffic stations, the temporal series show a remarkable $O_3$ daily cycle (observed 25p = 23.2 $\mu gm^{-3}$ and 75p = 77.5 $\mu gm^{-3}$), due to $O_3$ destruction caused by high $NO_x$ levels (observed 50p = 34.5 $\mu gm^{-3}$). In contrast, the $NO_2$–limited regime at background sites (observed 50p = 19 $\mu gm^{-3}$) allows higher $O_3$ concentrations (observed 25p = 38 $\mu gm^{-3}$ and 75p = 89 $\mu gm^{-3}$) than in high $NO_2$ environments.

During the episode mentioned above, the resolution increase at traffic stations had a positive effect by increasing the ~~correlation coefficient~~ for $NO_2$ (from 0.73 to 0.76) and $O_3$ (from 0.83 to 0.86), and also ~~by~~ decreasing the $NO_2$ mean bias by ~5 $\mu gm^{-3}$ (from 6 to 1 $\mu gm^{-3}$). The $NO_2$ daily cycle improves in the morning hours after sunrise, reducing bias by 5-10 $\mu gm^{-3}$ and contributing to a reduction in $O_3$ overestimations (~5 $\mu gm^{-3}$). In contrast, at background stations, where the $NO_x/O_3$ chemistry is less dominant, the resolution effect is not significant. Such behavior indicates that finer resolution improves the performance, because horizontal resolution affects the representation of chemical processes near large emission sources, such as the ~~efficient~~ formation of $O_3$ and nighttime $O_3$ titration (Mathur et al., 2005). However, the loss of subgrid variability and improved meteorological fields (transport and temperature) are required.

## 5 Conclusions

The present work shows the effects of increasing the horizontal resolution from 4 km to 1 km using the CALIOPE-AQFS on pollutant concentrations ($NO_2$, $O_3$, and $PM_{10}$) over three Spanish domains (AND, BCN and MAD) in April 2013.

The global features of concentration maps at both resolutions are quite similar, with zones of high/low concentration identically located, which is ~~obviously because~~ both simulations are based on the same emission dataset. Further comparisons demonstrate that increasing the resolution provides better-defined and more realistic concentration structures over large sources (roads and industries) and complex terrains (more sharply defined orographic hills). The titration effect on $O_3$ ~~concentrations~~ along highways and major point sources is more evident in 1-km simulations than at 4 km, ~~because the latter~~

~~is affected by higher dilution within the grid cells~~. This improvement is quantified by an increase in spatial correlation coefficients of 3% (6%) for $NO_2$ ($O_3$).

However, the resolution increase in the range of 4-1 km does not exhibit the expected improvement in hourly statistics for any pollutant. Hourly correlation coefficients do not significantly change, and absolute (relative) errors and biases vary < 2 µgm$^{-3}$ (9%). The merit of the resolution increase may be underrated when classical statistics are applied at measurement stations (Mass et al., 2002; Gego et al., 2005). For instance, although the structure of important $NO_2$ urban plume features (> 40 µgm$^{-3}$) often become more realistic (stronger and more defined plumes) as resolution increases, statistics are deeply degraded by even small timing and spatial errors.

The resolution increase has a significant impact on reducing $NO_2$ hourly bias (by 42%, 2 µgm$^{-3}$), without any significant change in the error and the r (<2%)~~, but it increases~~ $O_3$ hourly biases (<1 µgm$^{-3}$). The main differences between resolutions appear at daytime and nighttime traffic peaks, when the mixing height experiences rapid changes, allowing the 1-km simulation to slightly reduce $NO_2$ underestimation in the morning by ~5-10 µgm$^{-3}$. The $O_3$ daily cycles at large sources depict a high influence of hourly $NO_2$ concentrations, increasing the hourly $O_3$ bias by ~3 µgm$^{-3}$. ~~That~~ behavior is controlled by the daytime $O_3$ underestimation and, to a lesser extent, by the nighttime overestimation. The resolution increase ~~allows reducing~~ the $O_3$ overestimations at night (by ~5 µgm$^{-3}$), partly because of higher nocturnal $NO_2$ concentrations.

~~Concerning the capability of~~ forecast~~ing~~ 75p exceedances in the observed maximum 1h concentrations~~, the increased resolution has a positive effect: it increases the~~ number of hits that forecast 75p exceedances in the observed Max 1h $NO_2$ ~~(537 vs. 466 over 953 exceedances), and it reduces the false alarms for Max 1h $O_3$ exceedances (FAR improves by 37%)~~.

The grid effect is less pronounced for $PM_{10}$ than for $NO_2$ and $O_3$. When the resolution increases, the low increment of $PM_{10}$ mean concentrations (<0.1 µgm$^{-3}$) is the result of compensating biases of $PM_{10}$ components, which is controlled mainly by the PPM and EC increase as well as the SS decrease.

BCN is the domain where the resolution increase has the ~~highest~~ effect, with changes in bias (error) of 16-73% (< 5%), followed by AND with 4-43% (< 5%) and MAD < 3-5%

(< 1%). In BCN, as in the western Mediterranean Basin, the transport of $O_3$ and its precursors is governed by mesoscale circulation. In that sense, the resolution increase has a ~~great~~ impact over BCN, where induced mesoscale phenomena control the air flow~~;~~ ~~meanwhile~~ synoptic transport is more prominent in MAD and AND. The benefits of increasing the resolution to 1 km over rural areas (Mass et al., 2002) are that it increases the accurate representation of mesoscale meteorological structures such as orographic wind and circulation. Over urban areas along the western Mediterranean coast (Toll and Baldasano, 2002; Jiménez et al., 2006; Fay and Neunhäuserer, 2006), further improvements and urbanization steps are required before seeing any benefits in increasing the resolution to 1 km.

In urban areas ~~or~~ near large emission sources (industrial and traffic stations), $NO_2$ and $O_3$ concentrations are more sensitive to changes in the grid resolution. The concentration increase in primary anthropogenic pollutants ($NO_2$, PPM and EC) is obvious because the high resolution allows better allocation of emissions at point, linear and area sources. ~~What is more, it decreases~~ the artificial dilution of emissions when compared to the larger grid area. However, the 1-km simulation attempt to more accurately describe the chemical formation of $O_3$ and dilution of $NO_2$ concentrations over those areas was not generally successful.

This analysis demonstrates weaknesses in the current model formulations that cannot be resolved with only high-resolution modeling. The subgrid air quality variability at 1-km resolution is not reproduced over large emission sources or urban areas, because a finer spatial structure is expected but unresolved. ~~There are some underlying problems.~~ First, there is a loss of subgrid emission heterogeneity. Emission inputs to CTM are an average rate, which accounts for the volume averaged quantity of mass released per unit of time. No other information regarding emission allocation (e.g., point, linear or per area) is considered; for instance a large amount of mass can be emitted by a small portion of the grid surface or by several sources scattered around it (Galmarini et al., 2008; Cassiani et al., 2010; Ching and Majeed, 2012). Despite the fact that emissions are estimated by following a bottom-up approach emission model, emission heterogeneity is lost in the volume averaging process performed within CTM. The loss is even higher when resolution decreases (from 1 km to 4 km). Second, there is a low degree of complexity in flow and dispersion details at urban scales, where most of the pollutants come from street canyons and/or tree canopies, where they are transported until mixing conditions allow the pollutants to disperse above these urban canopy levels (Kim et al., 2013; Ching, 2013). Third, the USGS land-use data used in the WRF model is based on 1993 data, and urban changes in MAD and BCN over the last 20 years are significant.

Since temperature and wind speed are very sensitive to the ratio of building width to road width, ~~the next improvement~~ should focus on using an urban canopy model that considers effects on the transfer of energy and momentum between urban structures and the lower atmosphere. This is crucial for modeling meteorology and air quality. However, it requires an urban canopy scheme and a canopy parameter database (urban fraction, building height and area). Furthermore, in order to gain any benefits from increasing resolution, the meteorological modeling ~~should include an improved description of the land~~ instead of relying on USGS data from the year 1993. To this end, the Coordination of Information on the Environment (CORINE) provides a high resolution (100 m) land use database, which was developed by the European Environmental Agency and updated to the year 2006 (CLC2006) (EEA, 2007). This could be implemented in the WRF model following the methodology described in Pineda et al. (2004).

## Appendix A

Table A1. Definition of the discrete statistics used in the evaluation. Where $C_m(x, t)$ and $C_o(x,t)$ are the modeled and observed concentrations at a location (x) and time (t); N is the number of pairs of data. $\overline{C_m}$ and $\overline{C_o}$ are the modeled and observed mean concentrations over the whole period, respectively.

| Statistic | Formula | |
|---|---|---|
| Mean bias | $MB = \frac{1}{N}\sum_{i=1}^{N}(C_m(x,t) - C_0(x,t))$ | A1 |
| Mean normalized bias error | $MNBE = \frac{1}{N}\sum_{i=1}^{N}\frac{(C_m(x,t) - C_0(x,t))}{C_o(x,t)}\cdot 100$ | A2 |
| Mean fractional bias | $MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{(C_m(x,t) - C_0(x,t))}{(C_o(x,t) + C_m(x,t))/2}\cdot 100$ | A3 |
| Mean 3rror | $ME = \frac{1}{N}\sum_{i=1}^{N}\lvert C_m(x,t) - C_0(x,t)\rvert$ | A4 |
| Mean normalized gross 3rror | $MNGE = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert C_m(x,t) - C_o(x,t)\rvert}{C_o(x,t)}\cdot 100$ | A5 |
| Mean fractional error | $MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{\lvert C_m(x,t) - C_o(x,t)\rvert}{(C_m(x,t) + C_o(x,t))/2}\cdot 100$ | A6 |
| Root mean squared error | $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_m(x,t) - C_o(x,t))^2}$ | A7 |
| Correlation coefficient | $r = \frac{\sum_{i=1}^{N}(C_m(x,t) - \bar{C}_m)(C_o(x,t) - \bar{C}_o)}{\sqrt{\sum_{i=1}^{N}(C_m(x,t) - \bar{C}_m)^2}\sqrt{\sum_{i=1}^{N}(C_o(x,t) - \bar{C}_o)^2}}$ | A8 |

Table A2. Definition of the categorical statistics used in the evaluation. Exceedance analysis is based on a comparison with a fixed threshold concentration (T), where a is the number of false alarms, b is the number of hits, c is the number of correct negatives, and d is the number of misses.

| Statistic | Formula | |
|---|---|---|
| Accuracy | $A = \left(\frac{b + c}{a + b + c + d}\right)\cdot 100$ | A9 |
| Bias | $B = \left(\frac{a + b}{b + d}\right)\cdot 100$ | A10 |

| | | |
|---|---|---|
| Critical success index | $CSI = \left( \frac{b}{a+b+d} \right) \cdot 100$ | A11 |
| Probability of detection | $POD = \left( \frac{b}{b+d} \right) \cdot 100$ | A12 |
| False alarm ratio | $FAR = \left( \frac{a}{a+b} \right) \cdot 100$ | A13 |


## Acknowledgements

The Spanish administrations “Generalitat de Catalunya”, “Junta de Andalucia”, and “Comunidad de Madrid” are acknowledged for providing air quality measurements. The CALIOPE-AQFS team (G.Arévalo, K. Serradell, D. Carrió, M. Castrillo, A. Soret, S. Basart and S. Gassó) and F. Benincasa are also thanked for their technical support. This work is funded by the post-doctoral grant held by M.T. Pay in the Beatriu de Pinós programme (2011 BP-A 00427), Andalusian contract (NET838690), and the Severo Ochoa Program awarded by the Spanish Government (SEV-2011-00067).

## References

Appel, K.W., Pouliot, G.A., Simon, H., Sarwar, G., Pye, H.O.T., Napelenok, S.L., Akhtar, F., and Roselle, S.J.: Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0, Geosci. Model Dev., 6, 883-899, doi:10.5194/gmd-6-883-2013, 2013.

Baldasano, J. M., Cremades, L., and Soriano, C.: Circulation of Air Pollutants over the Barcelona Geographical Area in Summer, in: Proceedings of Sixth European Symposium Physico-Chemical Behaviour of Atmospheric Pollutants, Varese (Italy), 18–22 October, 1993, Report EUR 15609/1 EN, 474–479, 1994.

Baldasano J.M, Jiménez-Guerrero, P., Jorba, O., Pérez, C., López, E., Güereca, P., Martin, F., García-Vivanco, M., Palomino, I., Querol, X., Pandolfi, M., Sanz, M.J., and Diéguez, J.J.: CALIOPE: An operational air quality forecasting system for the Iberian Peninsula, Balearic Islands and Canary Islands- First annual evaluation and ongoing developments, Adv. Sci. Res. 2, 89-98, 2008.

Baldasano, J.M., Pay, M.T., Jorba, O., Gassó, S., and Jiménez-Guerrero, P.: An annual assessment of air quality with the CALIOPE modeling system over Spain, Sci. Total Environ., 409, 2163-2178, 2011.

Basart, S., Pérez, C., Nickovic, S., Cuevas, E., and Baldasano, J.M.: Development and evaluation of the BSC-DREAM8b dust regional modelo ver Northern Africa, the Mediterranean and the Middle East, Tellus Series B – Chemical and Physical Meteorology, 64, 1-12, 2012.

Byun, D. W. and Schere, K. L.: Review of the governing equations, computational algorithms and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, App. Mech. Rev. 59, 51-77, 2006.

Cassiani, M., Vinuesa, J.F., Galmarini, S., and Denby, B.: Stochastic fields method for sub-grid scale emission heterogeneity in mesoscale atmospheric dispersion models, Atmos. Chem. Phys., 10, 267-277, 2010.

Ching, J.K.S., and Majeed, M.A.: An approach to characterized within-grid concentration variability in air quality models, Atmos. Environ., 49, 348-360, 2012.

Ching, J.K.S.: A perspective on urban canopy layer modeling for weather, climate and air quality applications, Urban Climate, 3, 13-39, 2013.

CMAQ: Technical Documentation, Available at : http://www.airqualitymodeling.org/cmaqwiki/index.php?title=CMAQ_version_5.0_%28February_2012_release%29_Technical_Documentation, last access 8 January 2014.

Colette, A., Bessagnet, B., Meleux, F., Terrenoire, E., and Rouïl, L.: Frontiers in air quality modelling, Geosci. Model Dev., 7, 203-210, 2014.

Cuvelier, C., Thunis, P., Karam, D., Schaap, M., Hendriks, C., Kranenburg, R., Fagerli, H., Nyíri, Á., Simpson, D., Wind, P., Schulz, M., Bessagnet, B., Colette, A., Terrenoire, E., Rouïl, L., Stern, R., Graff, A., Baldasano, J. M., and Pay, M. T.: ScaleDep: performance of European chemistry-transport models as function of horizontal spatial resolution, EMEP Report 1/2013, 63 pp., available at: http://emep.int/publ/reports/2013/MSCW_technical_1_2013.pdf, last access 4 April 2013.Cohan, D.S., Hu, Y., and Russel, A.G.: Dependence of ozone sensitivity analysis on grid resolution, Atmos. Environ., 40, 126-135, 2006.

Eder, B., Kang, D., Mathur, R., Yu, S., and Schere, K.: An operational evaluation of the Eta-CMAQ air quality forecast model, Atmos. Environ., 40, 4894-4905, 2006.

EEA: CLC2006 technical guidelines. EEA Technical Report 17/2007. ISBN 978-90-9167-968-3. doi 10.2800/12134, 70 pp., 2007.

EEA: The application of models under the European Union's Air Quality Directive: A technical reference guide, EEA Technical report 10/2011, Publication Office of the European Union, Luxembourg, ISSN Technical report series 1725-2237, ISBN 978-92-9213-223-1, doi: 10.2800/80600, 76 pp., 2011.

EEA: Air quality in Europe- 2013 report, EEA Report 9/2013, ISSN 1725-9177, 112 pp., 2013.

Fay, B. and Neunhäuserer, L.: Evaluation of high-resolution forecast with the non-hydrostatic numerical weather prediction model Lokal modell for urban air pollution episodes is Helsinki, Oslo and Valencia, Atmos. Chem. Phys., 6, 2107-2128, 2006.

Fountoukis, C., Koraj, Dh., Denier van der Gon, H.A.C., Charalampidis, P.E., Pilinis, C., and Pandis, S.N.: Impact of grid resolution on the predicted fine PM by a regional 3-D chemical transport model. Atmos. Environ., 68, 24-32, 2013.

Galmarini, S., Vinuesa, J.F., and Martilli, A.: Modeling the impact of sub-grid scale emission variability on upper-air concentration, Atmos. Chem. Phys. 8, 141-158, 2008.

Garber, W., Colosio, J., Grittner, S., Larssen, S., Rasse, D., Schneider, J. and Houssiau, M.: Guidance on the Annexes to Decision 97/101/EC on Exchange of Information as revised by Decision 2001/752/EC, Technical Report, European Commission, DG Environment, 2002.

García-Valero, J.A., Montávez, J.P., Jérez, S., Gómez-Navarro, J.J., Lorente-Plazas, R., and Jiménez-Guerrero, P.: A seasonal study of the atmospheric dynamics over the Iberian Peninsula based on circulation types, Theor. Appl. Climatol., 110, 219-310, 2012.

Gego, E., Hogrefe, C., Kallos, G., Voudouri, A., Irwin, J., and Rao, S.T.: Examination of model predictions at different horizontal grid resolutions, Environ. Fluid Mech., 5, 63-85, 2005.

Gong, S.L.: A parameterization of sea-salt aerosol source function for sub- and super-micron particles, J. Geophys. Res., 17, 197, 2003.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P.I., and Geron, C.: 2006. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, 2006.

Gonçalves, M., Jiménez-Guerrero, P., and Baldasano, J.M.: Contribution of atmospheric processes affecting the dynamics of air pollution in South-Western Europe during a typical summertime photochemical episode, Atmos. Chem. Phys., 9, 849-864, 2009.

Guevara, M., Martínez, F., Arévalo, G., Gassó, S. and Baldasano, J.M.: An improved system for modelling Spanish emissions: HERMESv2.0, Atmos. Environ., 81, 209-221, 2013.

Jiménez, P., Jorba, O., Parra, R., and Baldasano, J.M.: Evaluation of MM5-EMICAT2002-CMAQ performance and sensitivity in complex terrains: High-resolution application to the northeastern Iberian Peninsula, Atmos. Environ., 40, 5056-5072, 2006.

Kain, J.S. and Fritsch, J.M.: A one-dimensional entraining/detraining plume model and its application in convective parameterization, J. Atmos. Sci., 47 (23), 2784-2802, 1990.

Kang, D., Eder, B.K., Stein, A.F., Grell, G.A., Peckham, S.E., and McHenry, J.: The new England air quality forecasting pilot program: development of an evaluation protocol and performance benchmark, J. Air Waste Mange. Assoc., 55, 1782-1796, 2005.

Kim, Y., Sartelet, K., Raut, Jean-Christophe, and Chazette, P.: Evaluation of the weather research and forecast/urban model over Greater Paris, Boundary-Layer Meteorol., 149(1), 105-132, 2013.

Makar, P. A., Gravel, S., Chirkov, V., Strawbridge, K. B., Froude, F., Arnold, J., and Brook, J.: Heat flux, urban properties, and regional weather, Atmos. Environ., 40, 2750–2766, 2006.

Mathur, R., Shankar, U., Hanna, A.F., Odman, M.T., McHenry, J.N., Coats, C. J., Alapaty, K., Xiu, A., Arunachalam, S., Olerud Jr., D. T., Byun, D. W., Schere, K. L., Binkowski, F. S., Ching, J. K. S., Dennis, R. L., Pierce, T. E., Pleim, J. E., Roselle, S. J., and Young, J. O.: Multiscale Air Quality Simulation Platform (MAQSIP): Initial applications and performance for tropospheric ozone and particulate matter, J. Geophys. Res., 110 (D13308), doi:10.1029/2004JD004918, 2005.

Mass, C., Ovens, D., Albright, M., and Westrick, K.: Does increasing Horizontal resolution Produce better Forecasts? The results of two years of Real-time Numerical Weather prediction in the Pacific Northwest, B. Am. Meteorol. Soc., 83, 407-430, 2002.

Millán, M., Salvador, R., Mantilla, E., and Kallos, G.: Photooxidant dynamics in the Mediterranean basin in summer: results from European research projects, J. Geophys. Res., 102, 8811– 8823, 1997.

Pay, M.T., Jiménez-Guerrero, P., and Baldasano, J.M.: Implementation of resuspension from paved roads for the improvement of CALIOPE air quality system in Spain, Atmos. Environ., 45, 802-807, 2011.

Pay, M.T., Jiménez-Guerrero, P., Jorba, O., Basart, S., Pandolfi, M., Querol, X., and Baldasano, J.M.: Spatio-temporal variability of levels and speciation of particulate matter across Spain in the CALIOPE modeling system, Atmos. Environ., 46, 376-396, 2012a.

Pay, M. T., Gassó, S., and Baldasano, J.M.: Evaluation of the CMAQ5.0 in the framework of the CALIOPE air quality forecasting system over Europe, in: 11$^{th}$ Annual CMAS Conference, Chapel Hill, North Carolina, USA, 15-17 October 2012, 2012b.

Pérez, C., Nickovic, S., Baldasano, J.M., Sicard, M., Rocadenbosch, F., and Cachorro, V.E.: A long Saharan dust event over the western mediterranean: Lidar, sun photometer observations, and regional dust modeling, J. Geophys. Res. 111 (D15214), 1-16, 2006.

Pineda, N., Jorba, O., Jorge, J., and Baldasano, J.M.: Using NOAA AVHRR and SPOT VGT data to estimate surface parameters: application to a mesoscale meteorological model, Int. J. Remote Sens. 25, 129–143, 2004.

Queen, A. and Zhang, Y.: Examining the sensitivity of MM5-CMAQ predictions to explicit microphysics schemes and horizontal grid resolutions, Part III - the impact of horizontal grid resolution, Atmos. Environ., 42, 3869-3881, 2008.

San José, R., Pérez, J.L., Morant, J.L., and González Barras, R.M.: The use of Modern third-generation air quality models (MM5-EMIMO-CMAQ) for real-time operational air quality impact assessment of industrial plants, Water Air Soil Pollu., 9, 27-37, 2009.

Skamarock, W.C., and Klemp, J.B.: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications, J. Comput. Phys. 227 (7), 3465-3485, doi:10.1016/j.jcp.2007.01.037, 2008.

Szopa, S., Foret, G., Menut, L., and Cozic, A.: Impact of large scale circulation on European summer surface ozone and consequences for modelling forecast, Atmos. Environ., 43, 1189-1195, 2009.

Thompson, T.M., Saari, R.K., and Selin, N.E.: Air quality resolution for health impacts assessment: influence of regional characteristics, Atmos. Chem. Phys. Discuss., 13, 14141-14161, doi:10.5194/acpd-13-14141-2013, 2013.

Tesche, T.W., Morris, R., Tonnesen, G., McNally, D., Boylan, J., and Brewer, P.: CMAQ/CAMx annual 2002 performance evaluation over the eastern US, Atmos. Environ. 40, 4906-4919, 2006.

Timmermans, R.M.A., Denier van der Gon, H.A.C., Kuenen, J.J.P., Segers, A.J., Honoré, C., Perrussel, O., Builtjes, P.J.H., and Schaap, M.: Quantification of the urban air pollution increment and its dependency on the use of down-scaled and bottom-up city emission inventories, Urban Climate, 6, 44-62, 2013.

Toll, I. and Baldasano, J. M.: Modeling of photochemical air pollution in the Barcelona area with highly disaggregated anthropogenic and biogenic emissions, Atmos. Environ., 34, 3060– 3084, 2000.

Valari, M. and Menut, L.: Does an increase in air quality models' resolution bring surface ozone concentrations closer to reality?, J. Atmos. Oceanic Technol., 25, 1955-1968, 2008.

Valverde, V. V., Pay, M. T., and Baldasano, J. M.: Climatic synoptic classification over the Iberian Peninsula oriented to air quality dynamic characterization, Int. J. Climatol., submitted, 2014.

Vivanco, M., Correa, M., Azula, O., Palomino, I., and Martín, F.: Influence of model resolution on ozone predictions over Madrid area (Spain), in: Computational Science and Its Applications–ICCSA 2008 (pp. 165-178), Springer Berlin Heidelberg, 2008.

WHO: Review of evidence on health aspects of air pollution – REVIHAAP Project Technical report, World Health Organization, Regional Office for Europe, Copenhagen, Denmark, 2013.

Yarwood, G., Roa, S., Yocke, M., and Whitten, G.: Updates to the carbon bond chemical mechanism: CB05, Final report to the US EPA, RT-0400675, 2005.

Zhang, K., Knipping, E., Wexler, A., Bhave, P., and Tonnensen, G.: Size distribution of sea-salt emissions as a function of relative humidity, Atmos. Environ., 39, 3373-3379, 2005.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, part I: History, techniques, and current status, Atmos. Environ., 60, 632-655, 2012.

Table 1. CALIOPE-AQFS computational requirements, in terms of Central Processor Units (CPU) and computational time (in min), for simulating 48h air quality forecasts as a function of the domain: IP-4km (D2), AND-1km (D3), BCN-1km (D5) and MAD-1km (D4), all of which are described in Figure 1. D-domains are described in Figure 1.

| | IP-4km (399x399 cells) | AND-1km (366x358 cells) | BCN-1km (146x146 cells) | MAD-1km (146x158 cells) |
|---|---|---|---|---|
| Meteorological Modeling | 128 CPU/15 min | 256 CPU/80 min | 128 CPU/20 min | 128 CPU/20 min |
| Emission Modeling | 1 CPU/ 4 min | 1 CPU/ 4 min | 1 CPU/1 min | 1 CPU/1 min |
| Air Quality modeling | 256 CPU/210 min | 256 CPU/220 min | 128 CPU/150 min | 128 CPU/110 min |

Table 2. Discrete and categorical statistics for $NO_2$, $O_3$, $O_3$-8h, and $PM_{10}$ for April 2013 as a function of horizontal resolution (4 km and 1 km). n indicates the number of pairs of data used in the discrete evaluation on an hourly basis. OM and MM depict the measured and modeled mean concentrations, respectively. T is the threshold applied in the categorical evaluation. Max 1h and mean 24h concentrations are calculated by considering ≥75% of the hours in a day, as established by Directive 2008/50/EC.

| | $NO_2$ | $O_3$ | $O_3$-8h | $PM_{10}$ |
|---|---|---|---|---|
| n (stations) | 90761 (135) | 76471 (114) | 3248 (114) | 66642 (100) |
| OM ($\mu gm^{-3}$) | 22.0 | 68.4 | 88.6 | 20.6 |
| EU LV/TV ($\mu gm^{-3}$) (temp basis) | 200 (Max 1h ) | 180 (Max 1h) | 120 (Max 8h) | 50 (Mean 24h) |
| EU LV/TV exceedances | 0 | 25 | 0 | 31 |
| T* ($\mu gm^{-3}$) (temp. basis) | 71 (Max 1h) | 108 (Max 1h) | 101 (Max 8h) | 27 (Mean 24h) |

Discrete evaluation

| | 4 km | 1 km | 4 km | 1 km | 4 km | 1 km | 4 km | 1 km |
|---|---|---|---|---|---|---|---|---|
| MM ($\mu gm^{-3}$) | 17.4 | 19.3 | 58.0 | 57.3 | 72.4 | 71.5 | 13.9 | 14.0 |
| r | 0.54 | 0.54 | 0.61 | 0.61 | 0.54 | 0.51 | 0.45 | 0.44 |
| MB ($\mu gm^{-3}$) | -4.5 | -2.6 | -10.5 | -11.3 | -16.3 | -17.2 | -6.7 | -6.6 |
| MAE ($\mu gm^{-3}$) | 12.9 | 13.2 | 19.7 | 19.8 | 18.4 | 19.2 | 12.6 | 12.7 |
| RMSE ($\mu gm^{-3}$) | 19.8 | 20.4 | 24.6 | 24.7 | 21.8 | 22.8 | 17.2 | 17.4 |
| MNBE (%) | -20.4 | -11.8 | -15.3 | -16.5 | -18.4 | -19.4 | -32.5 | -32.0 |
| MNGE (%) | 58.5 | 59.9 | 28.8 | 28.9 | 20.8 | 21.7 | 61.1 | 61.6 |
| MFB (%) | -28.5 | -23 | -13.1 | -15.1 | -19.3 | -20.5 | -63.3 | -64.1 |
| MFE (%) | 69.2 | 68.7 | 37.1 | 37.4 | 22.5 | 23.5 | 85.7 | 87.1 |

Categorical evaluation (Threshold = T*)

| | 4 km | 1 km | 4 km | 1 km | 4 km | 1 km | 4 km | 1 km |
|---|---|---|---|---|---|---|---|---|
| a (false alarm) | 306 | 384 | 41 | 19 | 17 | 6 | 131 | 133 |
| b (hits) | 466 | 537 | 112 | 96 | 6 | 4 | 331 | 334 |
| c (correct negative) | 2517 | 2439 | 1846 | 1868 | 2826 | 2837 | 1978 | 1976 |
| d (misses) | 487 | 416 | 1194 | 1210 | 399 | 401 | 334 | 331 |
| B (%, 100) | 37 | 40 | 8 | 7 | 1 | 1 | 42 | 42 |
| POD (%, 100) | 49 | 56 | 9 | 7 | 1 | 1 | 50 | 50 |
| CSI (%, 100) | 81 | 97 | 12 | 9 | 6 | 2 | 69 | 70 |
| FAR (%, 0) | 40 | 42 | 27 | 17 | 74 | 60 | 28 | 28 |
| A (%, 100) | 79 | 79 | 61 | 62 | 87 | 87 | 83 | 83 |

T* is defined as 75p of the observed concentrations estimated temporally, as established by EU Directive 2008/50/EC

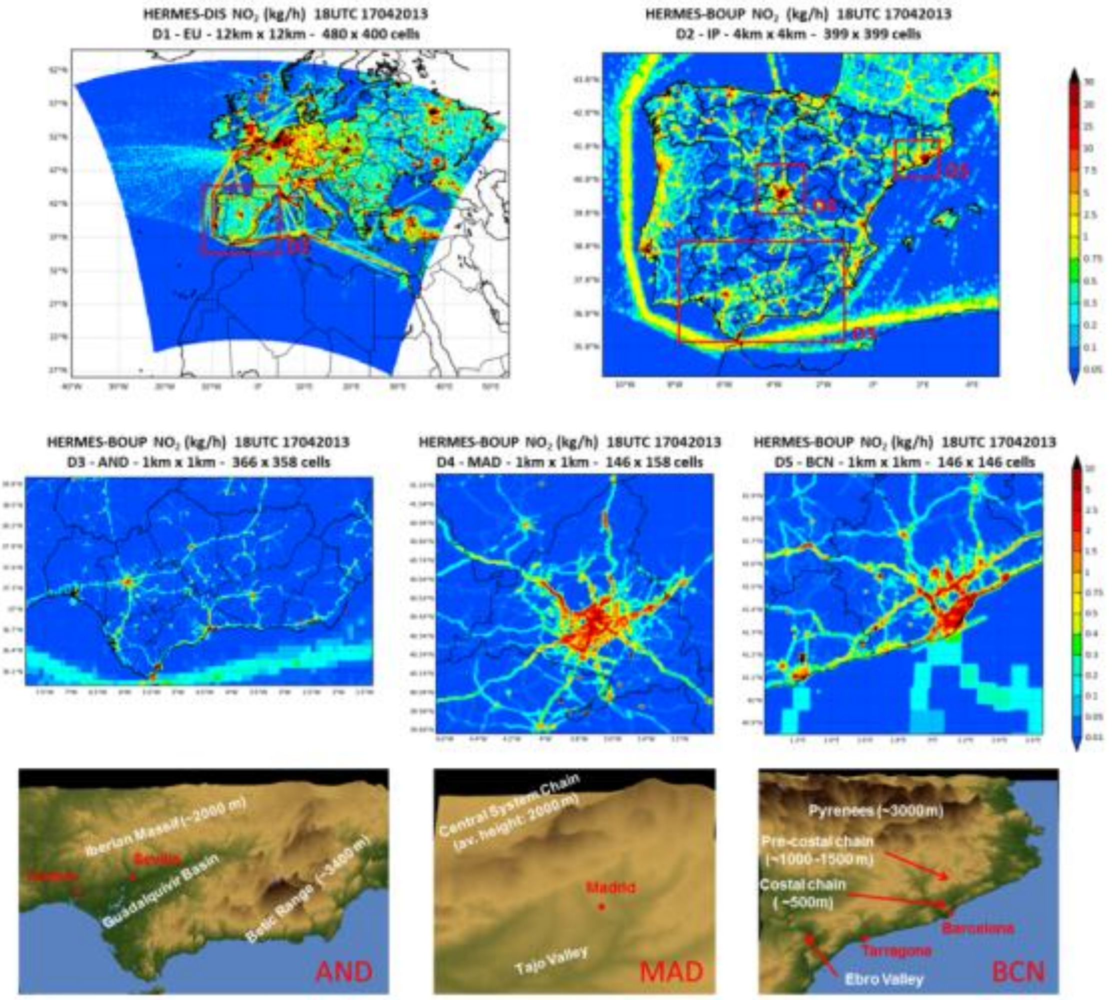



Fig. 1. CALIOPE-AQFS nesting strategy (D-domains) and study domains (Andalucia, AND; Madrid, MAD; and Barcelona, BCN). Colour chart at D-domains shows $NO_2$ emission rate ($kgh^{-1}$) for 17th April, 2013 at 18UTC. HERMES-DIS model generates emissions at 12 km x 12 km over Europe (the mother domain, D1) by performing disaggregation from the EMEP database. HERMES-BOUP model estimates emissions at 1 km x 1 km, following a bottom-up approach.

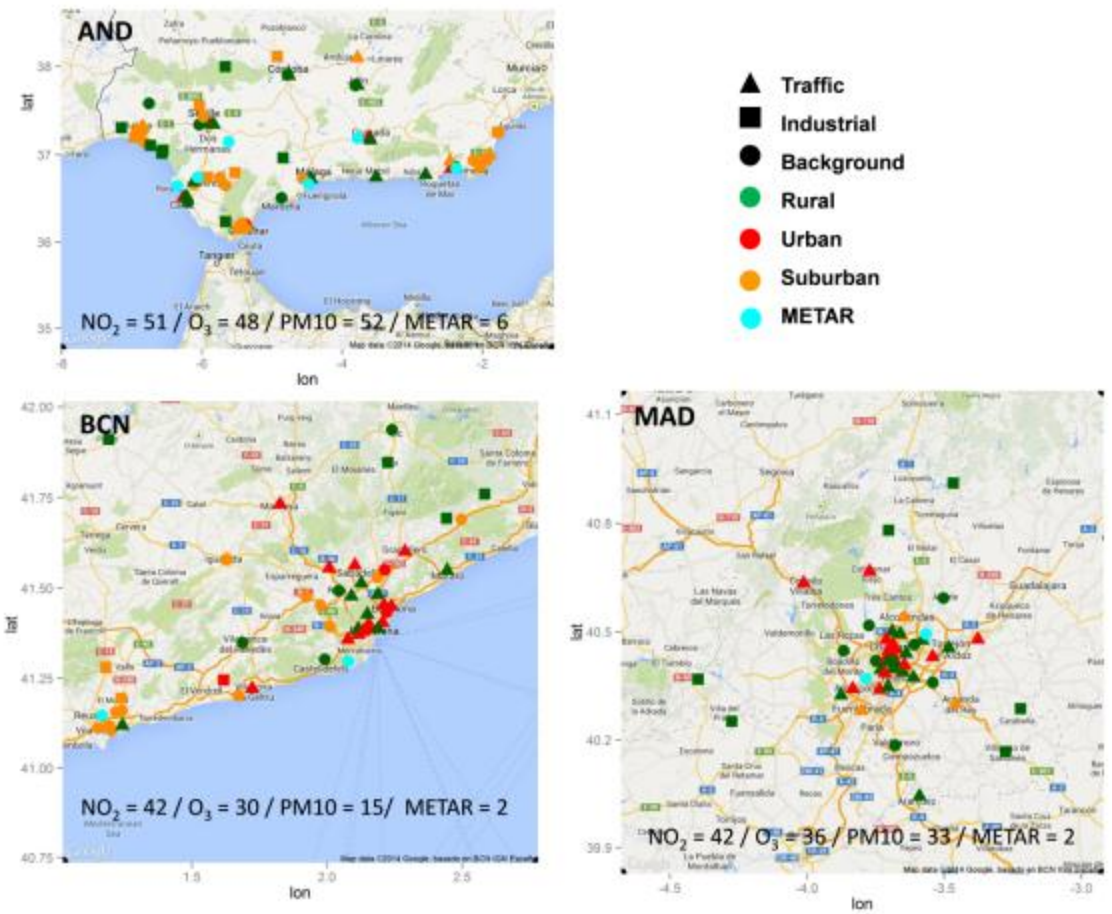



Figure 2. Air quality stations for $NO_2$, $O_3$ and $PM_{10}$ in the three domains under study (AND, BCN and MAD) in April 2013. Different types of stations are shown by symbols and color codes. The various symbols represent the major emission type affecting each station (Traffic: triangle; Industrial: square; and Background: circle), while the colors reflect the environment of each station (Urban: red; Suburban: green; and Rural: orange). Cyan dots represent METAR stations used in Sect. S1.

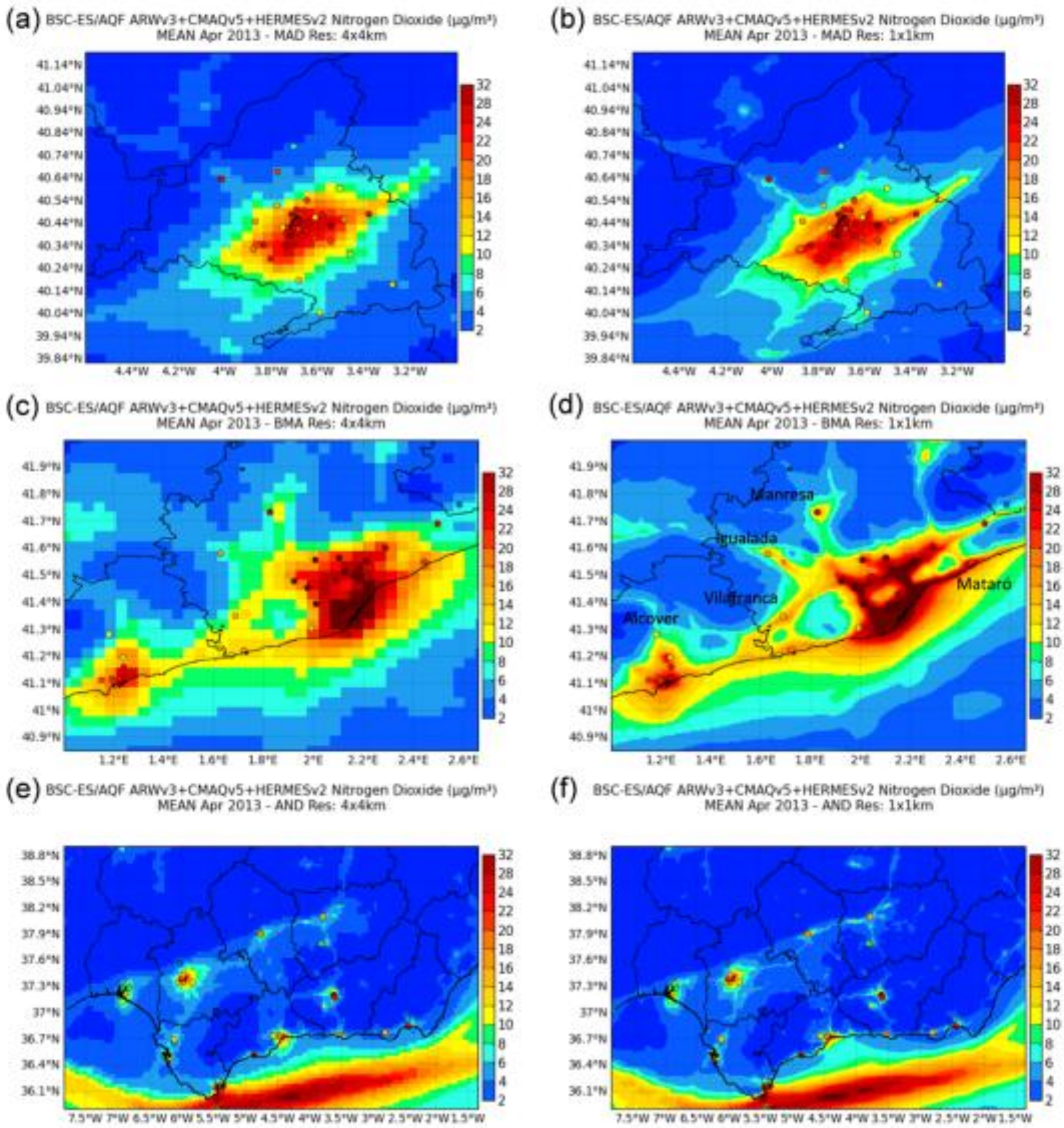



Figure 3. CALIOPE-AQFS mean $NO_2$ concentration ($\mu gm^{-3}$) in April 2013 over (a,b) MAD, (c,d) BCN, and (e,f) AND, as a function of horizontal resolution: 4 km (left column) and 1 km (right column). Dots indicate mean concentration at air quality stations.

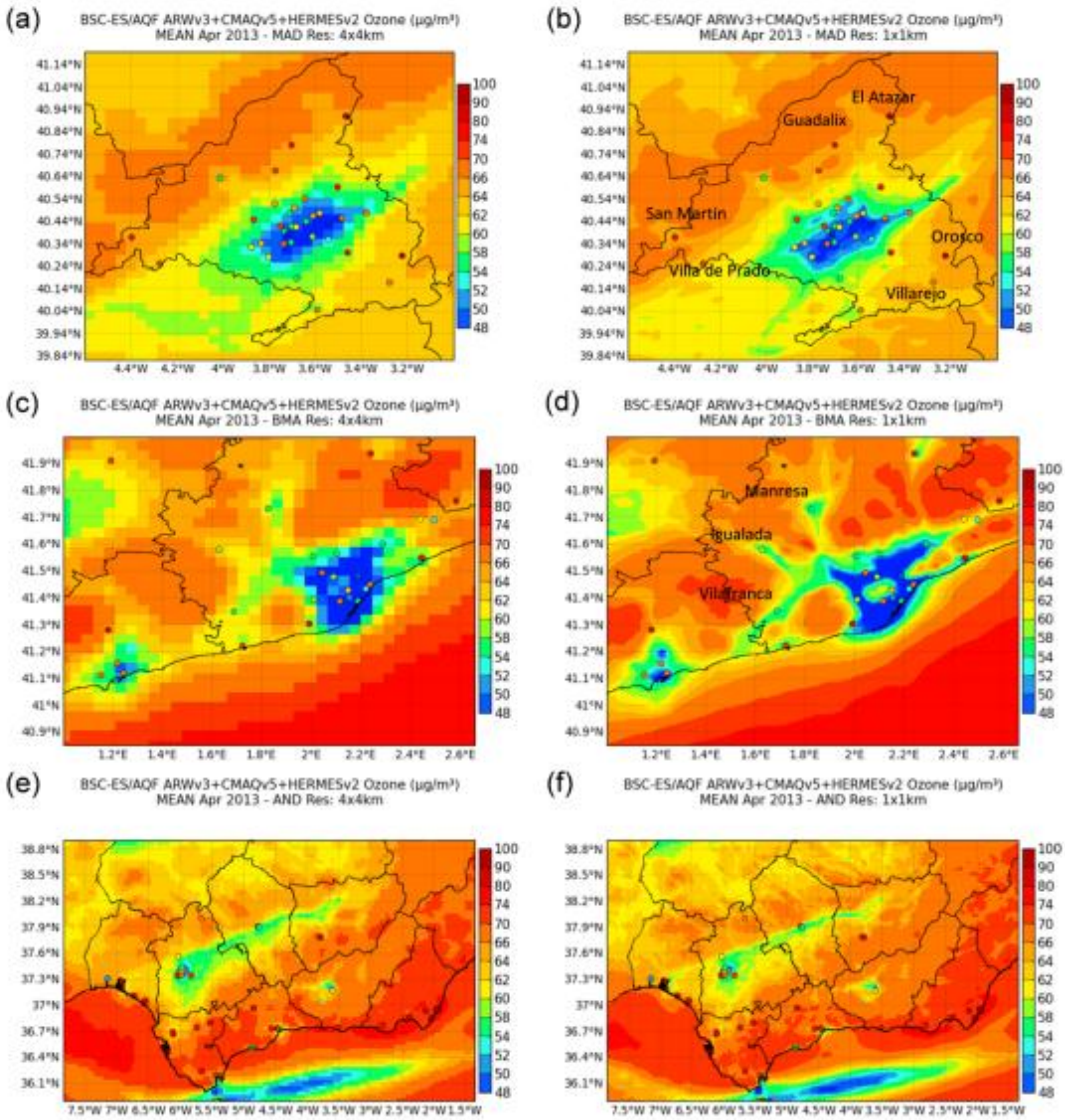



Figure 4. CALIOPE-AQFS mean $O_3$ concentration ($\mu gm^{-3}$) in April 2013 over (a,b) MAD, (c,d) BCN, and (e,f) AND, as a function of horizontal resolution: 4 km (left column) and 1 km (right column). Dots indicate mean concentration at air quality stations.

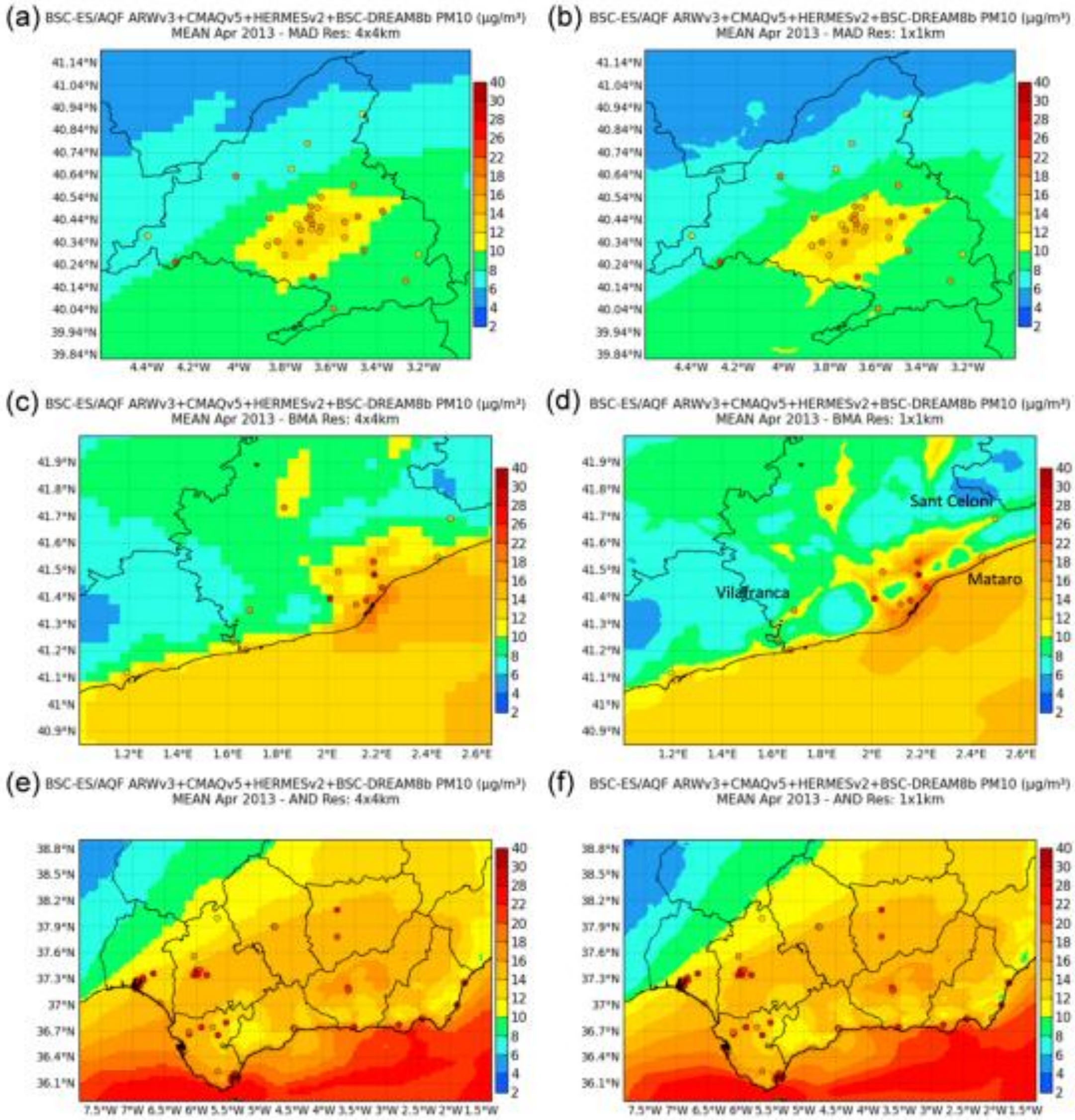
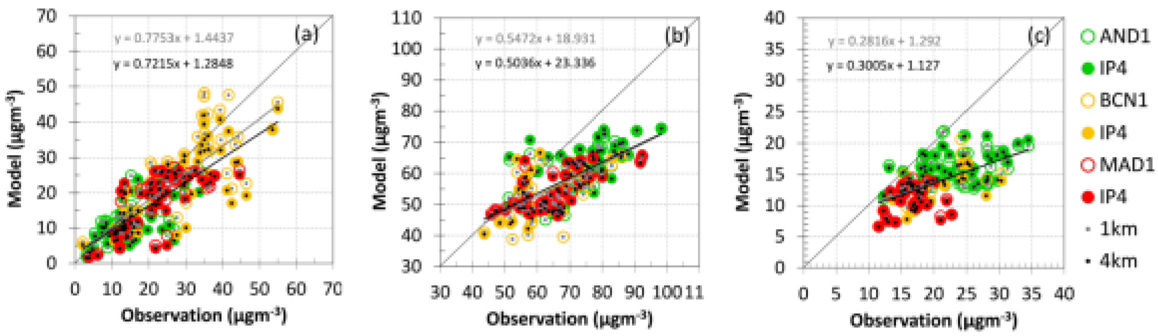



Figure 5. CALIOPE-AQFS mean $PM_{10}$ concentration ($\mu g m^{-3}$) in April 2013 over (a,b) MAD, (c,d) BCN, and (e,f) AND, as a function of horizontal resolution: 4 km (left column) and 1 km (right column). Dots indicate mean concentrations at air quality stations.




| (e) | $r_{NO2}$ | |
|---|---|---|
| | 4km | 1km |
| All | 0.79 | 0.81 |
| AND | 0.62 | 0.71 |
| BCN | 0.72 | 0.73 |
| MAD | 0.70 | 0.74 |

| (f) | $r_{O3}$ | |
|---|---|---|
| | 4km | 1km |
| All | 0.69 | 0.73 |
| AND | 0.58 | 0.64 |
| BCN | 0.54 | 0.57 |
| MAD | 0.71 | 0.73 |

| (g) | $r_{PM10}$ | |
|---|---|---|
| | 4km | 1km |
| All | 0.61 | 0.58 |
| AND | 0.36 | 0.26 |
| BCN | 0.63 | 0.64 |
| MAD | 0.20 | 0.24 |

Figure 6. Monthly mean scatter plots for CALIOPE-AQFS (y-axis) and observed (x-axis) concentrations for the three study domains (AND in green, BCN in yellow, and MAD in red), as a function of horizontal resolution for (a) $NO_2$, (b) $O_3$ and (c) $PM_{10}$. Equations show the linear adjustment between models and observations at 1 km (light grey) and 4 km (dark grey). Spatial correlation coefficients as a function of resolution and domain are shown for (e) $NO_2$, (f) $O_3$, and (g) $PM_{10}$.

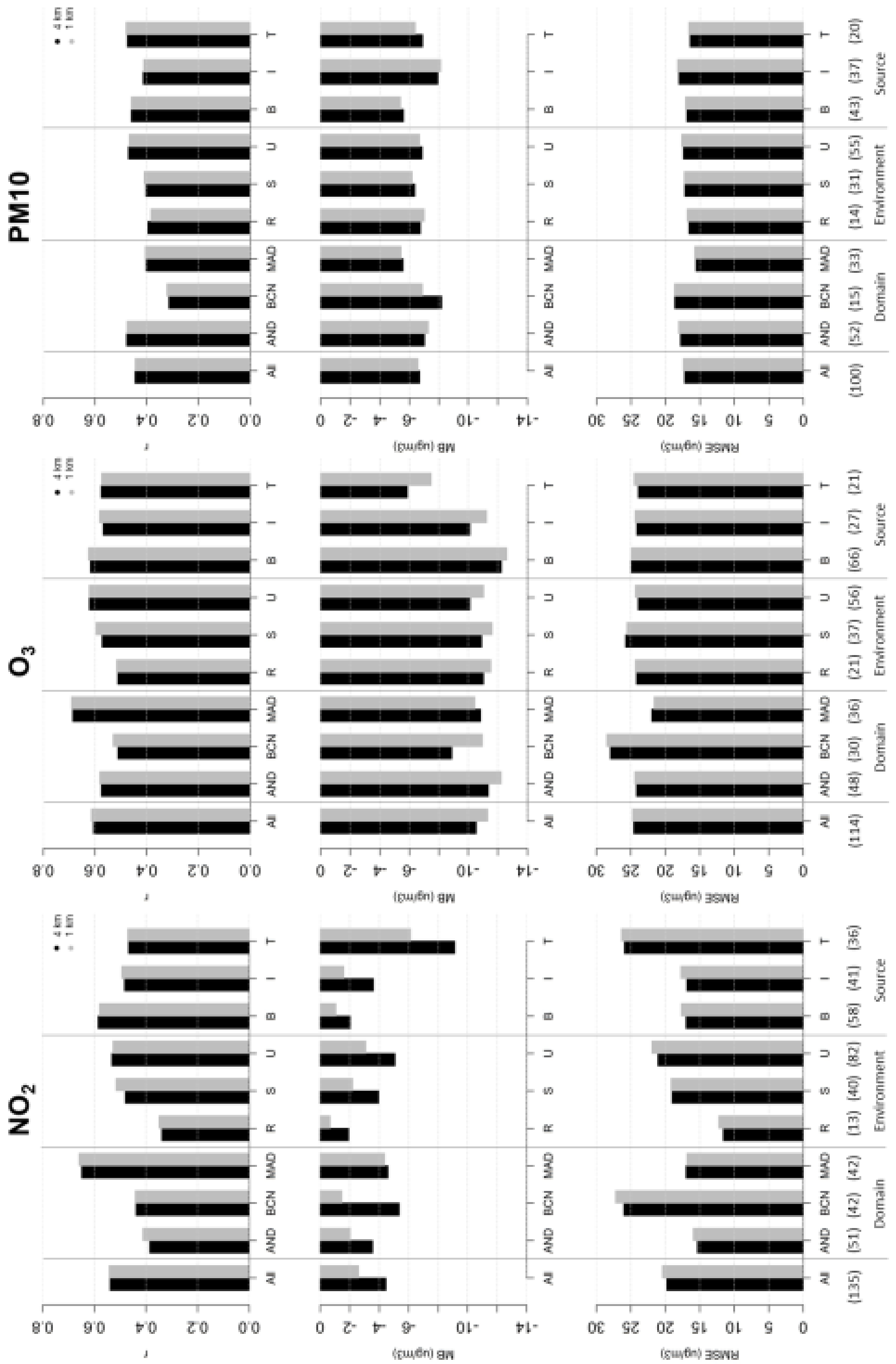



Figure 7. Statistics (r, MB, and RMSE in rows) for each pollutant ($NO_2$, $O_3$, and $PM_{10}$ in columns) on an hourly basis as a function of horizontal resolution: 4 km (black) and 1 km (grey). Four categories are considered: all stations (all), domain (AND, BCN and MAD), station environment (R, S, and U), and main sources (B, I, and T).

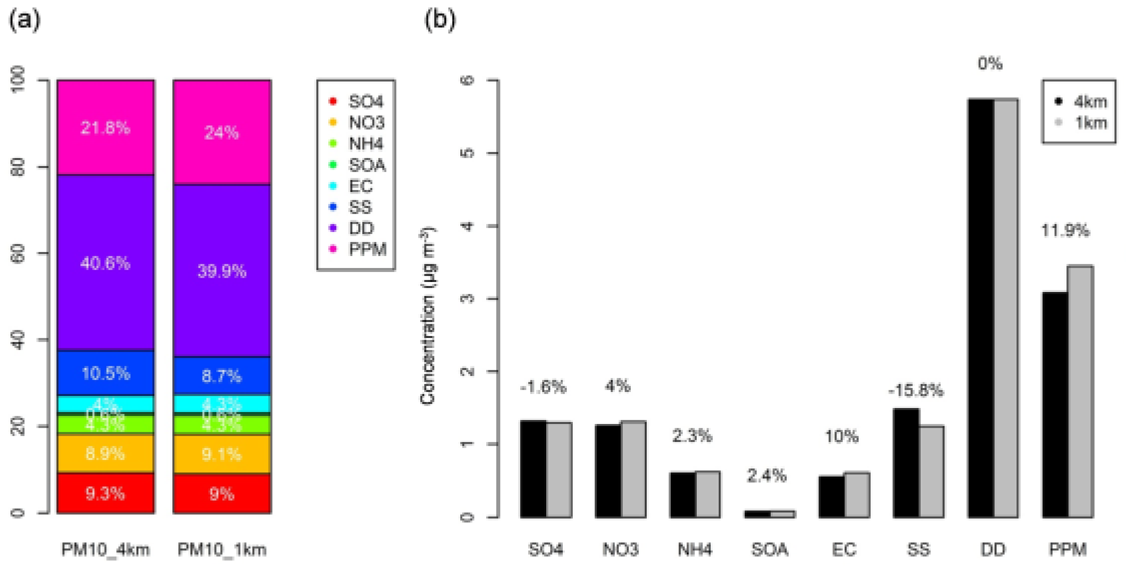



Figure 8. Resolution effect on $PM_{10}$ components in April 2013. (a) Percentage of $PM_{10}$ components: sulfate (SO4), nitrate (NO3), ammonium (NH4), secondary organic aerosol (SOA), elemental carbon (EC), sea salt (SS), desert dust (DD), and primary particulate matter (PPM). (b) $PM_{10}$ component concentrations in the 1-km simulation (black) and 4-km simulation (grey). Numbers over bars indicate the % of increase when increasing resolution.

BAr

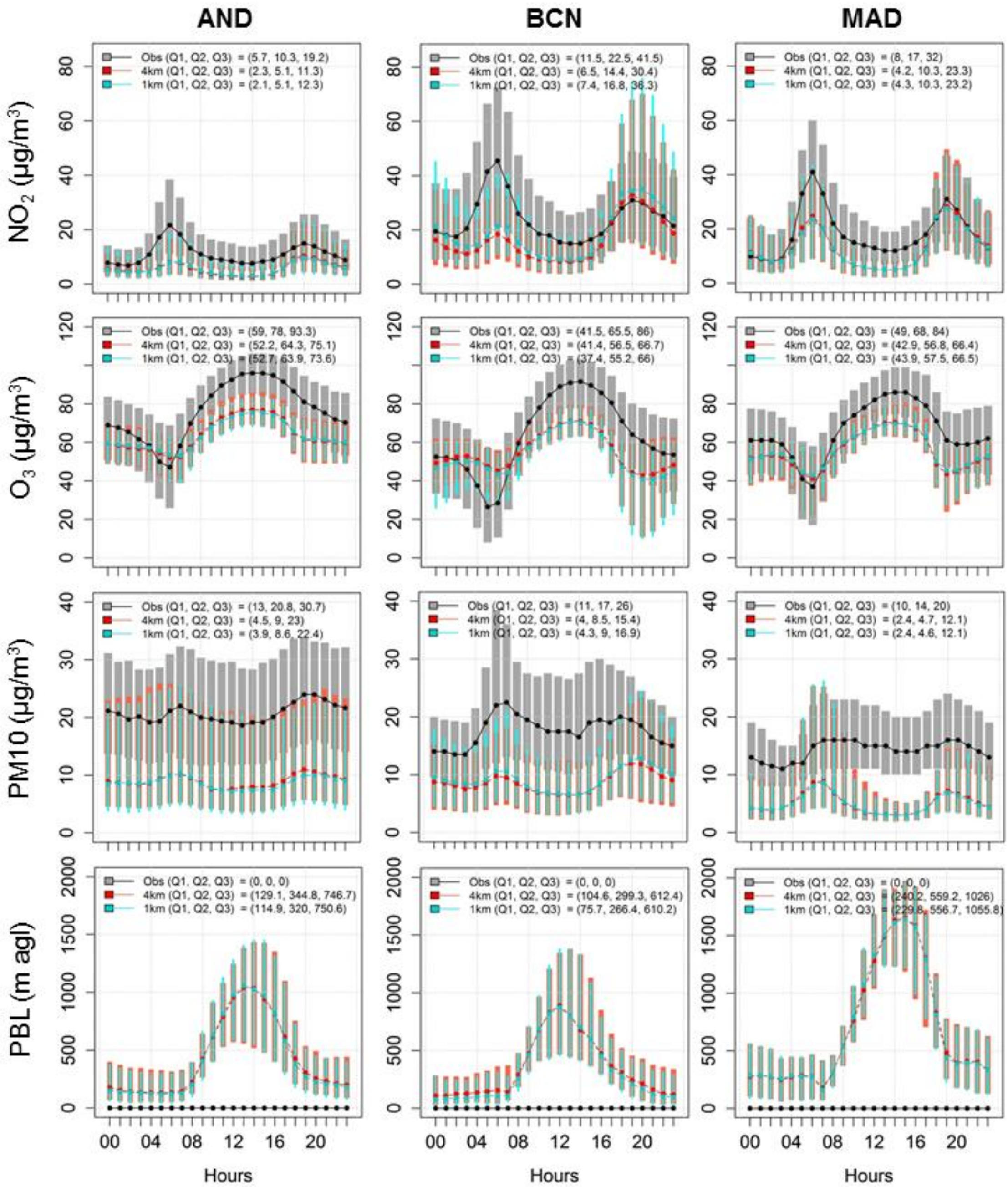



Figure 9. Daily cycles for $NO_2$, $O_3$ and $PM_{10}$ for each study domain at available stations as a function of resolution. No observations of PBL are available. Q1, Q2 and Q3 indicate quartiles for the daily cycle. Bars show Q1 and Q3 at each hour.

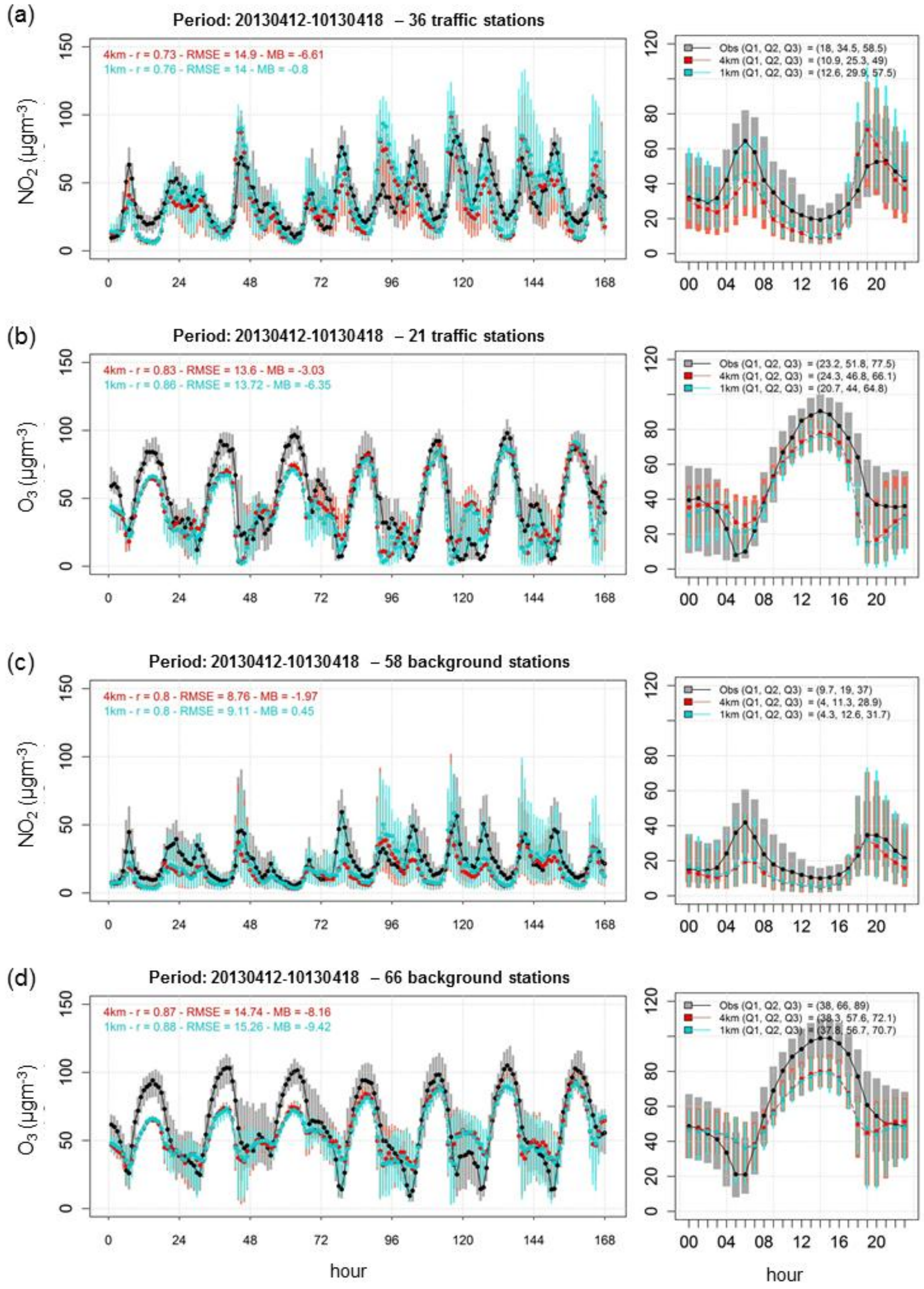



Figure 10. Temporal series and daily cycles for $NO_2$ and $O_3$ at background (a and b, respectively) and traffic stations (c and d, respectively) for the episode of 12-18th April, 2013. Q1, Q2 and Q3 indicate quartiles for the daily cycle. Bars show Q1 and Q3 at each hour.