Evaluation of the CALIOPE air quality forecasting system for epidemiological research: the example of NO_2 in the province of Girona (Spain)

Inmaculada Aguilera[a,b,c], Xavier Basagaña[a,b,c], María Teresa Pay[d], David Agis[a,b,c], Laura Bouso[a,b,c], Maria Foraster[a,b,c,e], Marcela Rivera[f], José María Baldasano[d,g], Nino Künzli[h,i]

[a] Centre for Research in Environmental Epidemiology, Barcelona, Spain
[b] Hospital del Mar Research Institute, Barcelona, Spain
[c] CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain
[d] Earth Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
[e] Pompeu Fabra University, Barcelona, Spain
[f] University of Montreal Hospital Research Centre, Montreal, Canada
[g] Environmental Modeling Laboratory, Technical University of Catalonia, Barcelona, Spain
[h] Swiss Tropical and Public Health Institute, Basel, Switzerland
[i] University of Basel, Basel, Switzerland

Corresponding author:
Inmaculada Aguilera, PhD
Centre for Research in Environmental Epidemiology (CREAL)
Doctor Aiguader 88
08003 Barcelona, Spain
Phone: +34 932147300 ; Fax: +34 932147302
E-mail: iaguilerajim@gmail.com

Abstract

Background: Air quality models are being increasingly used to estimate long-term individual exposures to air pollution in epidemiological studies. Most of them have been evaluated against measurements from a limited number of monitoring stations, which may not properly reflect the exposure characteristics of the study population.
Methods: We evaluated the performance of the high-resolution CALIOPE air quality forecasting system over a large sample of passive measurements of NO_2 conducted at 635 home outdoor locations of the Girona province (Spain) during several 4-week sampling campaigns over one year. Sampling sites were superposed over the 4 km x 4 km CALIOPE grid, and average NO_2 modeled concentrations were derived for all measurements conducted during the same sampling campaign at all the sampling sites located within the same grid cell. In addition, the ratio between measured and modeled concentrations for the whole study period at one fixed monitoring station was used to post-process the modeled values at the home outdoor locations.
Results: The correlation between measured and modeled concentrations for the entire study area (which includes urban settings, middle-size towns, and rural areas) was 0.78. After correcting the modeled concentrations by the measured to modeled ratio at the fixed station, they were very similar to the measured concentrations (27.7 $\mu g/m^3$ and 29.3 $\mu g/m^3$, respectively).However, the performance of the modeling system depends on the type of subarea and is affected by the sub-grid emission sources.
Conclusions: The evaluation over the heterogeneous Girona province showed that CALIOPE is able to reproduce the spatial variability of 4-week NO_2 concentrations at the small regional level, but not at the smaller within-city scale. CALIOPE output data is as a valuable tool to complement study-specific air pollution measurements by incorporating regional spatial variability and long-term temporal variability of background pollution in epidemiological research.

Keywords: air quality modeling; exposure assessment; model evaluation; NO_2

1. Introduction

Air pollution is the most important environmental factor affecting human health in Europe, contributing in a non-negligible manner to the total burden of mortality in urban areas (Boldo et al., 2006; Künzli et al, 2000). As a result, air quality management is a priority issue in European environmental policies. Modeling techniques have become very useful tools to study the dynamics and transport of air pollutants and to forecast air quality concentration for short term mitigation and public information and warnings, both applications for regulatory and scientific purposes (European Commission, 2008; EEA, 2011).

From the epidemiological research perspective, there is an increasing interest in using data from air quality models to estimate long-term individual exposures to air pollution in population-based studies. This interest lies in the ability of air quality models to reflect the spatial variability of air pollution both in the short- and long term, in contrast with other techniques such as land use regression (LUR) models, which usually better capture the spatial variability of air pollution at local scales but with a much lower temporal resolution (Marshall et al., 2008). In most cases, air quality models have not been evaluated in the particular study area or have been evaluated against measurements from a limited number of fixed monitoring stations (Bellander et al., 2001; Cyrys et al., 2005; Nyberg et al., 2000; Van den Hooven et al. 2012). Although air quality networks have the advantage of providing a good temporal coverage with routinely available regulatory air pollution data, data from a few stations provide poor spatial resolution to validate air quality models at the small regional or local scales. In addition, air quality networks are typically designed to assess air pollution levels for policy and regulatory purposes, not for health studies. Therefore, they may not properly reflect the exposure characteristics of the population living in a given study area, particularly for those pollutants that are more spatially heterogeneous (Brauer, 2010). For this reason, output data from air quality models should be preferably validated against observations in targeted locations selected to represent the real range of individual-level exposure within a study population. To our knowledge, only one study has been able to evaluate air pollution concentrations predicted by a dispersion model against NO_2 observations measured with passive samplers in a sample of residential outdoor locations from a Swiss population-based epidemiological study (Liu et al., 2007).

The present study aims to evaluate the CALIOPE air quality forecasting system as a tool for incorporating regional spatial variability and long-term temporal variability in air pollution exposure estimates for the participants of an epidemiological study (REGICOR) in the Girona province (northeast Spain). The CALIOPE project, funded by the Spanish Ministry of the Environment, has established a high-resolution air quality forecasting system for Spain (Baldasano et al., 2008b). The CALIOPE system combines four models to simulate air quality over Spain with a high spatial resolution (4 km x 4 km in the Iberian Peninsula domain), and a temporal resolution of 1 hour using the HERMES emission model specifically built up for Spain (Baldasano et al., 2008a). So far, CALIOPE is the only European system which includes the Saharan dust contribution on an hourly basis. Several evaluation studies (Baldasano et al., 2011; Pay et al., 2010; Pay et al., 2012) and near-real time evaluation (NRT) against air quality measurements on an hourly basis support the confidence on the system. Current air quality forecasts and NRT evaluations are available through http://www.bsc.es/caliope. The performance of the CALIOPE modeling system has been evaluated by comparing its estimates with measured concentrations of particulate matter less than 10 µm in aerodynamic diameter (PM_{10}) and gaseous pollutants (NO_2, SO_2, and O_3) at the Spanish network of air quality monitoring stations for a full year (Baldasano et al., 2011). Strong differences were found depending on the type of site (urban, suburban, and rural) and the main emission sources (traffic, industrial, and background). For NO_2, modeled concentrations showed good agreement with observations, particularly in rural areas influenced by background emissions, although levels were systematically underestimated in all background and many urban sites. A similar pattern was found for PM_{10} measurements. On the contrary, modeled O_3 levels performed better in urban settings, especially at traffic stations located in the largest cities (Madrid and Barcelona).

Given the importance of assessing the performance of air quality models in study-specific targeted locations for epidemiological research, the REGICOR study, which has conducted a large amount of NO_2 measurements at residential outdoor locations specifically selected to cover a broad range of traffic-related air pollution and urban settings, offers a unique opportunity to evaluate the performance of the CALIOPE modeling system for estimating NO_2 concentrations at a geographically-diverse province level.

2. Material and methods

2.1. CALIOPE modeling system description

The state-of-the-art CALIOPE modeling system has been described in detail elsewhere (Baldasano et al., 2008b; 2011). Briefly, it integrates a meteorological model (WRFv3.0.1.1), an emission model (HERMES 2004), a chemical transport model (CMAQv4.5), and a mineral dust atmospheric model (BSC-DREAM8b) together in an air quality modeling and forecasting system.

The Advanced Research Weather Research and Forecasting model (WRF-ARW) (Skamarock and Klemp, 2008) provides the meteorological fields required by the chemical transport model. The model top is defined at 50 hPa to resolve the troposphere-stratosphere exchanges properly. The High-Elective Resolution Modeling Emission System (HERMES) (Baldasano et al., 2008a) uses information and the state-of-the-art methodologies for emission estimations. It calculates emissions by sector-specific sources or by individual installation and stacks following a bottom-up approach for the reference year 2004. The Models-3 Community Multiscale Air Quality Modeling System (Models-3/CMAQ) (Byun and Schere, 2006) combines current knowledge in atmospheric science and air quality modeling with multi-processor computing techniques in an open-source framework to deliver concentration estimates of the main air pollutants (O_3, NO_2, SO_2, and PM_{10}, among others) and acid deposition. Finally, CALIOPE modeled particulate matter are achieved by adding the Saharan dust contribution from BSC-DREAM8bin (Pérez et al., 2006a; 2006b) model to the CMAQ outputs.

Hourly NO_2 concentrations from the CALIOPE system over one-year period (from 1st July 2007 to 31st June 2008) in a 4 km x 4 km grid covering the province of Girona were used for the present study.

2.2. NO_2 measurements in REGICOR

REGICOR (Girona Heart Registry) is a population-based study which aims at investigating the effect of chronic exposure to ambient air pollution on subclinical markers of atherosclerosis in the Girona province. Twelve of the original towns included in REGICOR that provided the maximum contrast in ambient air pollution levels in a pilot study, while keeping the travel distance between towns to a minimum, were selected for the present study. The final study area has approx. 4,550 km^2 and includes the average mid-size city of Girona and adjacent Salt (96,200 and 30,300 inhabitants, respectively), towns on the Mediterranean coast with maritime ports and tourist resorts, inland towns, and small villages located in valleys and mountains (Figure 1). In countries such as Spain, fixed site monitoring stations are rare in regions with small communities clustered in otherwise rather rural areas. The present study area has only one monitoring station, located at an urban background site in Girona city.

From the total number of 2,780 subjects who underwent a detailed cardiovascular health examination, we selected a subsample of 635 subjects for home outdoor NO_2 monitoring with passive Palmes tubes over 4 weeks (Palmes et al., 1976) sent to the participants. Homes were selected to cover a broad range of traffic characteristics and urban settings (e.g. low and high building-density areas, historical quarters, and peripheral residential areas), to be representative of the residential locations of the REGICOR participants, and to be well distributed across the 12 towns. In addition, 4-week NO_2 monitoring was performed every month at a reference site in each town for an entire year, in order to derive temporally-adjusted NO_2 concentrations in the sampling sites of the corresponding town. Results from our pilot study and from previous monitoring studies conducted by local authorities indicated that monthly mean NO_2 concentrations during spring and fall were similar to the annual mean. Thus, from the 23 sampling campaigns that were performed from June 2007 to July 2009, the largest ones were conducted during spring and fall. For the present study, NO_2 measurements performed between 1st July 2007 and 31st June 2008 were included.

2.3. CALIOPE NO_2 estimates for REGICOR sampling sites

Home addresses of REGICOR participants were geocoded at the front door level. Then, the geocoded addresses were superposed over the CALIOPE grid to identify the corresponding 4 km x 4 km cell for each address (Figure 2), and the average NO_2 modeled concentration was derived for each sampling site (home address) and for the corresponding 4-weeks sampling period, matching the period from the opening until the closing day of each passive sampler. In addition, daily NO_2 modeled concentrations were derived for the urban background station located in Girona city over the whole study period.

2.4. Statistical analysis

A first comparison was made between measured and modeled NO_2 concentrations at the fixed monitoring station in Girona city over the whole study period. Since the sampling period of Palmes tubes was 4 weeks, for each day of the study period we computed a 4-week average of measured and modeled concentrations, being the reference day the start date of the 4-weeks time window. The ratio between 4-weeks measured and modeled concentrations was derived for each day, and the average ratio for the whole study period was used to post-process the modeled values at the home outdoor sampling sites.

In order to compare modeled NO_2 concentrations with the passive tubes measurements, we averaged all measurements conducted during the same sampling campaign at all the sampling sites located within the same 4 km x 4 km cell. Because the number of measurements taken at each sampling site over the study period was heterogeneous, and the number of sampling sites located within a same grid cell differed across the study area, we additionally compared modeled and measured concentrations weighted by the total number of measurements used to obtain the average value at each grid cell for the same sampling campaign.

Descriptive statistics of both measured and modeled concentrations were calculated for the whole study area and for six different geographic clusters of adjacent towns, which reflect similarities in urban characteristics, topography, and traffic intensity levels. Correlation coefficients and root mean square errors (RMSE) were calculated to assess the performance of the CALIOPE modeling system over the sampling sites at the different groups of towns.

Analyses were performed using Stata 10.1 (StataCorp, College Station, TX) and R 2.12 (http://www.R-project.org).

3. Results

The correlation between daily measured and modeled NO_2 concentrations at the urban background monitoring station of Girona city was moderate for the whole study period (Pearson's r=0.35) but low for the extreme seasons, i.e. winter and summer (Pearson's r<0.10). Modeled NO_2 concentrations were persistently underestimated (mean=8.8 $\mu g/m^3$, SD=3.3 $\mu g/m^3$) as compared to measured concentrations (mean=35.7 $\mu g/m^3$, SD=10.4 $\mu g/m^3$). The average ratio of 4-week modeled to measured concentrations for the whole study period at the fixed monitoring station was 0.25. This ratio was used to post-process the modeled NO_2 concentrations in all the sampling sites of the study area.

A scatter plot of 4-week measured and modeled (corrected by the average ratio) NO_2 concentrations at the fixed monitoring station of Girona is shown in Figure 3. The highest modeled concentrations (>70 µg/m^3), mainly corresponding to the winter season, are overestimated by the model after post-processing, which contributes to reduce the correlation found for this period.

Table 1 shows the descriptive statistics for measured and modeled 4-week NO_2 concentrations in the whole REGICOR area and by groups of towns. After averaging the number of passive measurements conducted during the same sampling campaign and located within the same grid cell, a total number of 165 measurements were used to assess the performance of the CALIOPE modeling system. The adjusted mean modeled NO_2 concentration for the whole study area was very similar to the mean measured concentration (27.7 and 29.3 µg/m^3, respectively). The Spearman correlation between modeled and measured concentrations for the entire study area was high (r=0.78). When the six groups of towns were considered independently, correlations were moderate to high (r=0.40 – 0.76) and did not follow a clear geographic pattern. On the contrary, an urban-rural pattern was shown by the RMSE, with higher values in the urban area of Girona and Salt (11.3 µg/m^3), lower values in rural areas (7.7 µg/m^3 in Banyoles-Porqueres and 6.7 µg/m^3 in Llagostera-Santa Cristina) and intermediate values in the remaining middle-size inland and coastal towns.

The linear relationship between modeled and measured NO_2 concentrations can be visualized in Figure 4. The scatter plots show the difference in terms of absolute concentrations between the urban area of Girona-Salt and the rest of the REGICOR area, as well as the lack of clear outliers, neither in Girona-Salt nor in the rest of the study area. When the weight of the number of tubes for each measurement site is incorporated into the scatter plot (Figure 4B), the higher contribution of the reference sites and the larger concentration of sampling sites over the 4 km x 4 km cells in Girona-Salt are clearly highlighted, and the R^2 decreased from 0.61 to 0.33.

4. Discussion

In the present study, we have found a good correlation between NO_2 concentrations measured with Palmes tubes at a large sample of residential locations and modeled NO_2 concentrations with the CALIOPE modeling system over one year across the geographically-diverse province of Girona, Spain. Despite the high overall correlation, modeled concentrations were systematically underestimated in the whole study area, which includes urban settings, inland and coastal towns, and rural areas. CALIOPE predictions were successfully corrected for this underestimation using continuous data over the study period from the fixed monitoring station representative of the urban background concentrations in the city of Girona. Modeled concentrations showed a substantially higher correlation with measurements at residential locations than with measurements at the fixed monitoring station. These results indicate that relying on daily measurements from a small sample of monitoring sites is insufficient to properly assess the performance of an air quality modeling system such as CALIOPE at the province (i.e. small regional) level over long time periods.

The performance of the CALIOPE modeling system has been previously evaluated by using hourly air pollution data from the 68 monitoring stations of the Spanish air quality network over the year 2004. The correlation coefficient between measured and modeled NO_2 concentrations was 0.53 for the whole sample of monitoring stations and decreased when the authors split the sample into urban stations (r=0.47), suburban stations (r=0.39), and rural stations (r=0.51) (Baldasano et al., 2011). Although the differences between this previous study and the present one in terms of study area (Spain vs. Girona province) and type of NO_2 measurements (automatic monitoring of hourly concentrations vs. passive monitoring of monthly concentrations) make direct comparisons difficult, the better performance of CALIOPE found in our study supports the argument that, for epidemiological studies on air pollution and health, a comprehensive evaluation of output data from air quality models should be made against measurements in locations specifically selected to represent the exposure distribution among the study population.

In most of the epidemiological studies, however, model evaluation has been often limited to a few monitoring sites, usually those from the air quality networks (Bellander et al., 2001; Cyrys et al., 2005; Nyberg et al., 2000; Van den Hooven et al., 2012), which are unable to identify potential differences in model performance across different subareas. The REGICOR study provided a unique opportunity to identify such geographical variations in performance, given the intensive NO_2 passive monitoring over one year at a large number of sampling sites, specifically selected to cover a broad range of geographic characteristics, to reflect the highest contrast in air pollution levels, and to represent the spatial distribution of the study participants. The comparison of CALIOPE estimates with NO_2 measurements revealed a good performance at the small regional level of a 4 km x 4 km grid but the inability to reflect the spatial gradients at the smaller locale scale.

The underestimation of modeled NO_2 concentrations at the fixed monitoring station is not surprising, since urban stations are more likely to be influenced by very local emission sources which are more difficult to be captured by the model. Regarding the low correlation between measured and modeled concentrations at this station, the previous assessment of the CALIOPE modeling system over the whole Spanish air quality monitoring network already identified that measurements at urban stations from small and medium-sized cities generally showed poorer agreement with modeled estimations, whereas modeled concentrations were more accurate at urban stations located in larger and better characterized cities such as Barcelona or Madrid, for which a higher spatial resolution (1x1 km) was available (Baldasano et al., 2011).

When the performance of CALIOPE was assessed against the measurements of Palmes tubes across the REGICOR study area, results showed that not only urban but also suburban and rural areas were systematically underestimated. The methodology used to estimate traffic emissions at medium- and small-sized cities may explain this result. Besides, biomass burning and natural NO_x production such as lightning are not currently treated in the model, and biogenic emissions in HERMES may need further revision. To correct this underestimation in the CALIOPE forecasting system, an updated version of the inventory is being developed (Guevara et al., 2012).

Uncertainties in the simulated air pollution concentrations are common in modeling techniques. Misrepresentation of atmospheric dynamics and chemistry, geographic differences in the quality of input data, and the choice of modeling domain and grid structure are among the causes associated with model uncertainties (Borrego et al., 2008; Sicardi et al., 2012). Several post-processing techniques have been applied to correct the modeled concentrations based on observed data, such as Bayesian-based techniques, data assimilation, or the Kalman filter technique (Napelenok et al., 2008). The latter has been successfully applied within CALIOPE to improve the prediction of tropospheric ozone (Sicardi et al., 2012), and further works in Portugal have extended the method to other pollutants including NO_2 (Borrego et al., 2011).

Although the correlation between modeled and measured concentrations is relatively high in the cities of Girona and Salt, the highest mean RMSE value for this urban area is indicative of a poorer performance of the model in comparison with suburban and rural areas. This can be explained by two factors. First, the HERMES emission model incorporates emissions data from on-road transport, but in cities with less than 500,000 inhabitants the availability of traffic emissions is restricted to major roads only (Baldasano et al., 2008a). Thus, the influence of high and very local on-road traffic is not well characterized in middle-size cities. Second, the large number of measurements with Palmes tubes taken in Girona and Salt provides a wide range of NO_2 concentrations which are considerably diluted when the sampling sites are superposed on a grid of 4 km x 4 km to obtain the predicted values, given the small area of these two cities (45.7 km^2).

The problem of comparing grid model predictions with point measurements is well known, namely the within-grid cell variability in emission sources due to different land uses, topography, traffic activities, and other characteristics that typically vary at finer scales. This problem is partially solved, but never eliminated entirely, by using finer scale grid sizes or by applying within-grid model treatments for the major point sources (Ching et al., 2006). This is particularly relevant for NO_2, which usually shows a heterogeneous spatial distribution at the intraurban level, with relatively large contrast within short distances (Lewné et al., 2004). In fact, a study conducted in several urban areas of Switzerland demonstrated that even a dispersion model with a fine spatial resolution of 200 m x 200 m could not accurately predict NO_2 at traffic locations in urban areas (Liu et al., 2007). It is important to highlight, however, that assessing the very local variability of air pollution in any Spanish city is beyond the scope of the CALIOPE modeling system.

Estimating long-term exposures to air pollution in population-based studies is always challenging and requires considerable efforts in terms of time and resources. In order to minimize the exposure misclassification resultant from the use of central monitoring sites to assign community-level exposures, several exposure assessment methodologies are being increasingly applied in cohort studies, such as interpolation methods, LUR models, or dispersion models, each of them providing a different degree of spatiotemporal resolution (Jerrett et al., 2005). LUR modeling is probably the most used approach in the last years because it provides a detailed spatial resolution at a relatively low cost, but its poor temporal resolution poses a problem when attempting to estimate chronic exposures to air pollution over long periods of time. As a result, a few recent studies have combined LUR models with output data from dispersion models, satellite-based data, or meteorological data in order to incorporate long-term temporal variation (Kloog et al., 2011; Liu et al., 2012; Molter et al., 2010). Despite the limited capability of the CALIOPE modeling system for assigning individual exposures in epidemiological studies at the intraurban level, it is able to capture the spatiotemporal pattern in air pollution concentrations at the regional scale and therefore shows great potential as a complementary tool for assigning long-term exposures to air pollution in epidemiologic studies at this scale. The ability to estimate concentrations for any time length makes it particularly appealing for the investigation of the contribution of health effects on different shorter and intermediate time scales. In particular, in pregnancy cohorts one could easily derive exposure terms for any time period.

5. Conclusions

The present study has assessed the performance of the CALIOPE air quality forecasting system over a large sample of 4-week NO_2 measurements taken with passive samplers in the province of Girona within the framework of an epidemiological study. The evaluation over the study area showed that the model is able to reproduce the spatial variability of 4-week NO_2 concentrations, with a correlation coefficient of 0.78 and a RMSE of 9.8 $\mu g/m^3$. Results also showed that the performance of the modeling system depends on the type of subarea, and is affected by the sub-grid emission sources, particularly in urban settings. Despite its limitations in simulating within-city contrasts in air pollution concentrations, this study strongly supports the use of CALIOPE output data as a valuable and complementary tool to incorporate regional spatial variability and long-term temporal variability in epidemiologic studies.

Funding: CALIOPE air quality simulations were performed on the MareNostrum supercomputer hosted by the Barcelona Supercomputing Center – Centro Nacional de Supercomputación (BSC-CNS). Other data used in this project were funded by: Fondo de Investigación Sanitaria, FIS PI060258; Fundacio La Marató de TV3, grant number: 081632; the French Agency ADEME (TRI-TABS - 2009/1/121) and CREAL Pilot Project Funds 2009. Dr. Rivera received funding from DIUE GC/FSE (2009FI_EX 00041). Ms. Foraster received funding from Instituto de Salud Carlos III.

References

Baldasano, J.M., Güereca, L.P., López, E., Gasso, S., Jiménez-Guerrero, P., 2008a. Development of a high-resolution (1 km x 1 km, 1 h) emission model for Spain: the High-Elective Resolution Modelling Emission System (HERMES). Atmospheric Environment 42, 7215-7233.

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

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

Bellander, T., Berglind, N., Gustavsson, P., Jonson, T., Nyberg, F., Pershagen, G., Jarup, L. , 2001. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. Environmental Health Perspectives 109, 633-639.

Boldo, E., Medina, S., Le Tertre, A., Hurley, F., Mücke HG, Ballester, F., Aguilera, I., Eilstein, D., 2006. APHEIS: Health Impact Assessment of long-term exposure to PM2.5 in 23 European cities. European Journal of Epidemiology 21, 449-458.

Borrego, C., Monteiro, A., Ferreira, J., Miranda, A.I., Costa, A.M., Carvalho, A.C., Lopes, M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environment International 34, 613-620.

Borrego, C., Monteiro, A., Pay, M.T., Ribeiro, I., Miranda, A., Basart, S., Baldasano, J.M., 2011. How bias-correction can improve air quality forecast over Portugal. Atmospheric Environment 45, 6629-6641.

Brauer, M., 2010. How much, how long, what, and where: air pollution exposure assessment for epidemiologic studies of respiratory disease. Proceedings of the American Thoracic Society 7, 111-115.

Byun, D.W., Schere, K.L., 2006. Review of the governing equations, computational algorithms and other components of the models-3 community multiscale air quality (CMAQ) modeling system. Applied Mechanics Review 59, 51-77.

Ching, J., Herwehe, J., Swall, J., 2006. On joint deterministic grid modeling and sub-grid variability conceptual framework for model evaluation. Atmospheric Environment 40, 4935-4945.

Cyrys, J., Hochadel, M., Gehring, U., Hoek, G., Diegmann, V., Brunekreef, B., Heinrich, J., 2005. GIS-based estimation of exposure to particulate matter and NO2 in an urban area: stochastic versus dispersion modeling. Environmental Health Perspectives 113, 987-992.

European Commission, 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. Off. J. Eur. Comm.

EEA, 2011. The application of models under the European Union's Air Quality Directive: A technical reference guide. Technical report 10/2011 (available through http://www.eea.europa.eu/publications/fairmode).

Guevara, M., Arévalo, G., Gassó, S., Martínez, S., Soret, A., Ferrer, G., Baldasano, J.M., 2012. Updating and Improvement of the High-resolution (1km x 1km, 1h) Emission Model for Spain: HERMES v.2.0. ACCENT-IGAC-GEIA Conference, Toulouse (France) 11-13 June 2012.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Analysis and Environmental Epidemiology 15, 185-204.

Kloog, I., Koutrakis, P., Coull, B.A., Lee, H.J., Schwartz, J., 2011. Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. Atmospheric Environment 45, 6267-6275.

Künzli, N., Kaiser, R., Medina, S., Studnicka, M., Chanel, O., Filliger, P., Herry, M., Horak, F.Jr., Puybonnieux-Texier, V., Quénel, P., Schneider, J., Seethaler, R., Vergnaud, J.C., Sommer, H., 2000. Public-health impact of outdoor and traffic-related air pollution: a European assessment. Lancet 356, 795-801.

Lewné, M., Cyrys, J., Meliefste, K., Hoek, G., Brauer, M., Fischer, P., Gehring, U., Heinrich, J., Brunekreef, B., Bellander, T., 2004. Spatial variation in nitrogen dioxide in three European areas. Science of the Total Environment 332, 217-230.

Liu, L.J., Curjuric, I., Keidel, D., Heldstab, J., Künzli, N., Bayer-Oglesby, L., Ackermann-Liebrich, U., Schindler, C., 2007. Characterization of source-specific air pollution exposure for a large population-based Swiss cohort (SAPALDIA). Environmental Health Perspectives 115, 1638-1645.

Liu, L.J., Tsai, M.Y., Keidel, D., Gemperli, A., Ineichen, A., Hazenkamp-von Arx, M., Bayer-Oglesby, L., Rochat, T., Künzli, N., Ackermann-Liebrich, U., Straehl, P., Schwartz, J., Schindler, C., 2012. Long-term exposure models for traffic related NO2 across geographically diverse areas over separate years. Atmospheric Environment 46, 460-471.

Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: Comparison of estimation methods. Atmospheric Environment 42, 1359-1369.

Molter, A., Lindley, S., de Vocht, F., Simpson, A., Agius, R., 2010. Modelling air pollution for epidemiologic research - Part I: A novel approach combining land use regression and air dispersion. Science of the Total Environment 408, 5862-5869.

Napalenok, S.L., Pinder, R.W., Gilliland, A.B., Martin, R.V., 2008. A method for evaluating spatially-resolved NOx emmissions using Kalman filter inversion, direct sensitivities, and space-based NO2 observations. Atmospheric Chemistry and Physics 8, 5603-5614.

Nyberg, F., Gustavsson, P., Jarup, L., Bellander, T., Berglind, N., Jakobsson, R., Pershagen, G., 2000. Urban air pollution and lung cancer in Stockholm. Epidemiology 11, 487-495.

Palmes, E.D., Gunnison, A.F., DiMattio, J., Tomczyk, C., 1976. Personal sampler for nitrogen dioxide. American Industrial Hygiene Association Journal 37, 570-577.

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

Pay, M.T., Piot, M., Jorba, O., Gassó, S., Gonçalves, M., Basart, S., Dabdubd, D., Jiménez-Guerrero, P., Baldasano, J.M., 2010. A full year evaluation of the CALIOPE-EU air quality modeling system over Europe for 2004. Atmospheric Environment 44, 3322-3342.

Pérez, C., Nickovic, N., Pejanovic, G., Baldasano, J.M., Özsoy, E., 2006a. Interactive dust-radiation modeling: a step to improve weather forecast. Journal of Geophysical Research 111, D16206.

Pérez, C., Nickovic, N., Baldasano, J.M., Sicard, M., Rocadenbosh, F., Cachorro, V.E., 2006b. A long Saharan dust event over the Western Mediterranean: lidar, sun photometer observation and regional dust modeling. Journal of Geophysical Research 111, D15214.

Sicardi, V., Ortiz, J., Rincon, A., Jorba, O., Pay, M.T., Gasso, S., Baldasano, J.M., 2012. Assessment of Kalman filter bias-adjustment technique to improve the simulation of ground-level ozone over Spain. Science of the Total Environment 416, 329-342.

Skamarock, W.C., Klemp, J.B., 2008. A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. Journal of Computational Physics 227, 3465-3485.

Van den Hooven, E.H., Pierik, F.H., van Ratingen, S.W., Zandveld, P.Y., Meijer, E.W., Hofman, A., Miedema, H.M., Jaddoe, V.W., de Kluizenaar, Y., 2012. Air pollution exposure estimation using dispersion modelling and continuous monitoring data in a prospective birth cohort study in the Netherlands. Environmental Health 11, 9.

Table 1. Descriptive statistics of measured and modeled NO_2 concentrations in the whole REGICOR study area and by groups of towns. Means, standard deviations (SD), and root mean square errors (RMSE) are shown in $\mu g/m^3$

Area	**No. of tubes**	**No. of measurements**[a]	**Measured mean (SD)**	**Modeled mean (SD)**	**Modeled mean after correction**[b] **(SD)**	**r**[c]	**RMSE**[d]
All REGICOR towns	884	165	29.3 (15.0)	6.8 (3.5)	27.7 (14.3)	0.78	9.8
Girona-Salt	534	47	46.0 (12.9)	10.6 (2.8)	43.1 (11.3)	0.58	11.3
Banyoles-Porqueres	47	26	20.8 (7.1)	3.7 (0.9)	15.1 (3.8)	0.68	7.7
Blanes	141	20	28.2 (10.7)	7.8 (2.0)	31.7 (8.1)	0.40	10.8
La Bisbal-Palafrugell-Palamós	110	29	25.7 (12.8)	6.4 (3.0)	26.2 (12.2)	0.63	10.5
Olot-Sant Joan	34	26	21.0 (7.2)	3.6 (1.1)	14.9 (4.5)	0.45	8.9
Llagostera-Sta. Cristina	18	17	16.2 (4.8)	5.4 (1.8)	21.9 (7.4)	0.76	6.7

[a] After averaging tubes located within the same 4 km x 4 km cell during the same sampling campaign

[b] After adjusting the modeled mean by the annual modeled to measured ratio at the fixed monitoring station

[c] Pearson correlation coefficient

[d] Computed with the corrected modeled concentrations

Figure captions

Figure 1. Map of the REGICOR study area including the location of NO_2 sampling sites and study participants

Figure 2. Annual average concentrations of NO_2 (µg/m^3) for the study period simulated by CALIOPE at a 4 km x 4 km spatial resolution

Figure 3. Four-week modeled vs. measured concentrations of NO_2 (µg/m^3) at the fixed urban background monitoring station of Girona city over the 1-year study period. y=16.0 + 0.6x (R^2=0.19). The scatter plot includes the 1:1 reference line. Modeled concentrations are corrected by the modeled to measured ratio obtained at this station for the whole period.

Figure 4. Modeled vs. measured concentrations of NO_2 (µg/m^3) at the REGICOR sampling sites during a 4-weeks period. Measurements conducted during the same sampling campaign at all the sites located within the same grid cell are averaged. In (B), the regression line is weighted by the number of measurements used for computing the mean value at each sampling site, and the size of the dots and triangles is proportional to this number. y=5.8 + 0.7x (R^2=0.61) in (A), and y=17.5 + 0.5x (R^2=0.33) in (B). The scatter plots include the 1:1 reference line. Modeled concentrations are corrected by the annual modeled to measured ratio at the fixed monitoring station.

Figure 1
Click here to download high resolution image

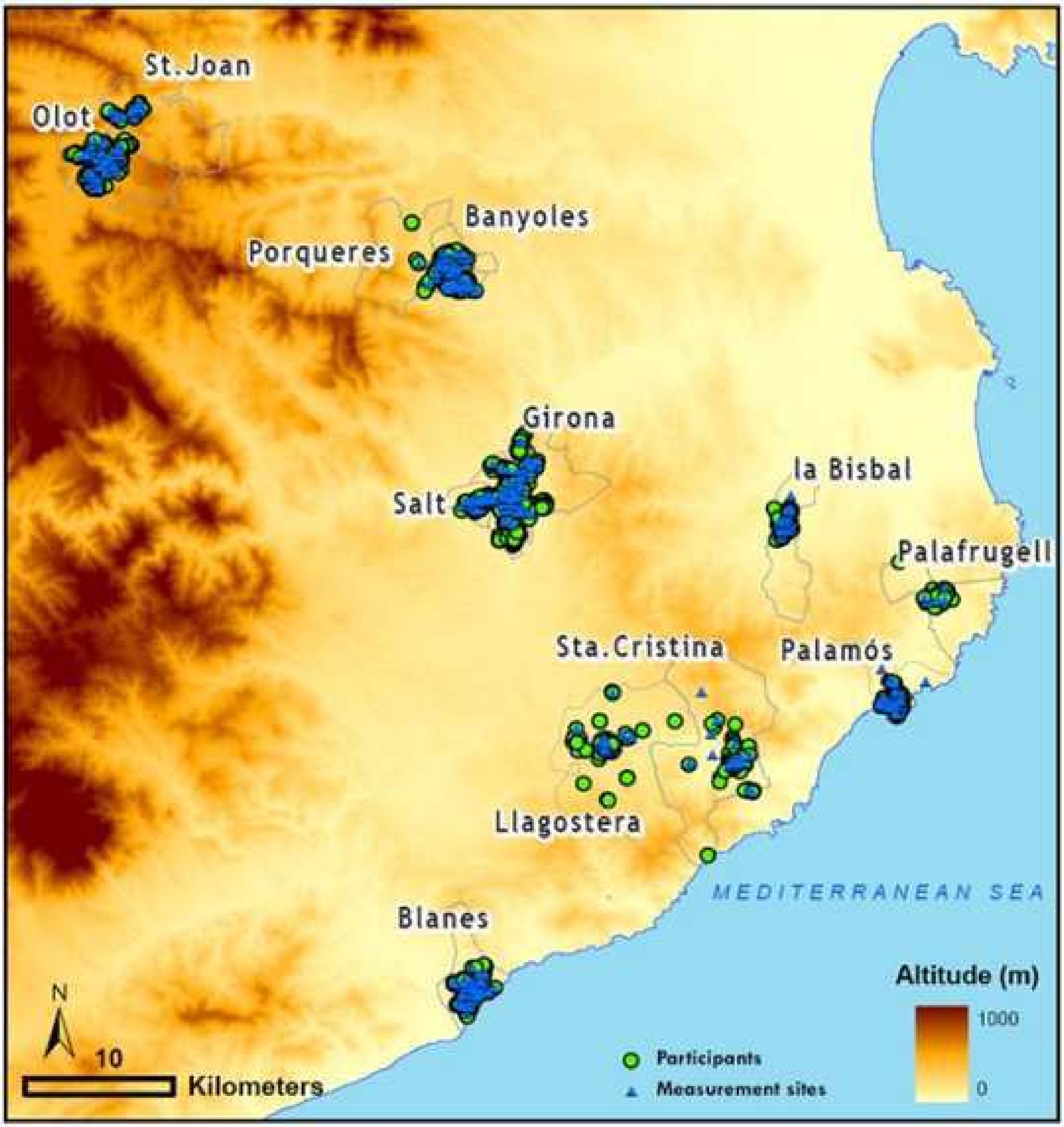

Figure 2
Click here to download high resolution image

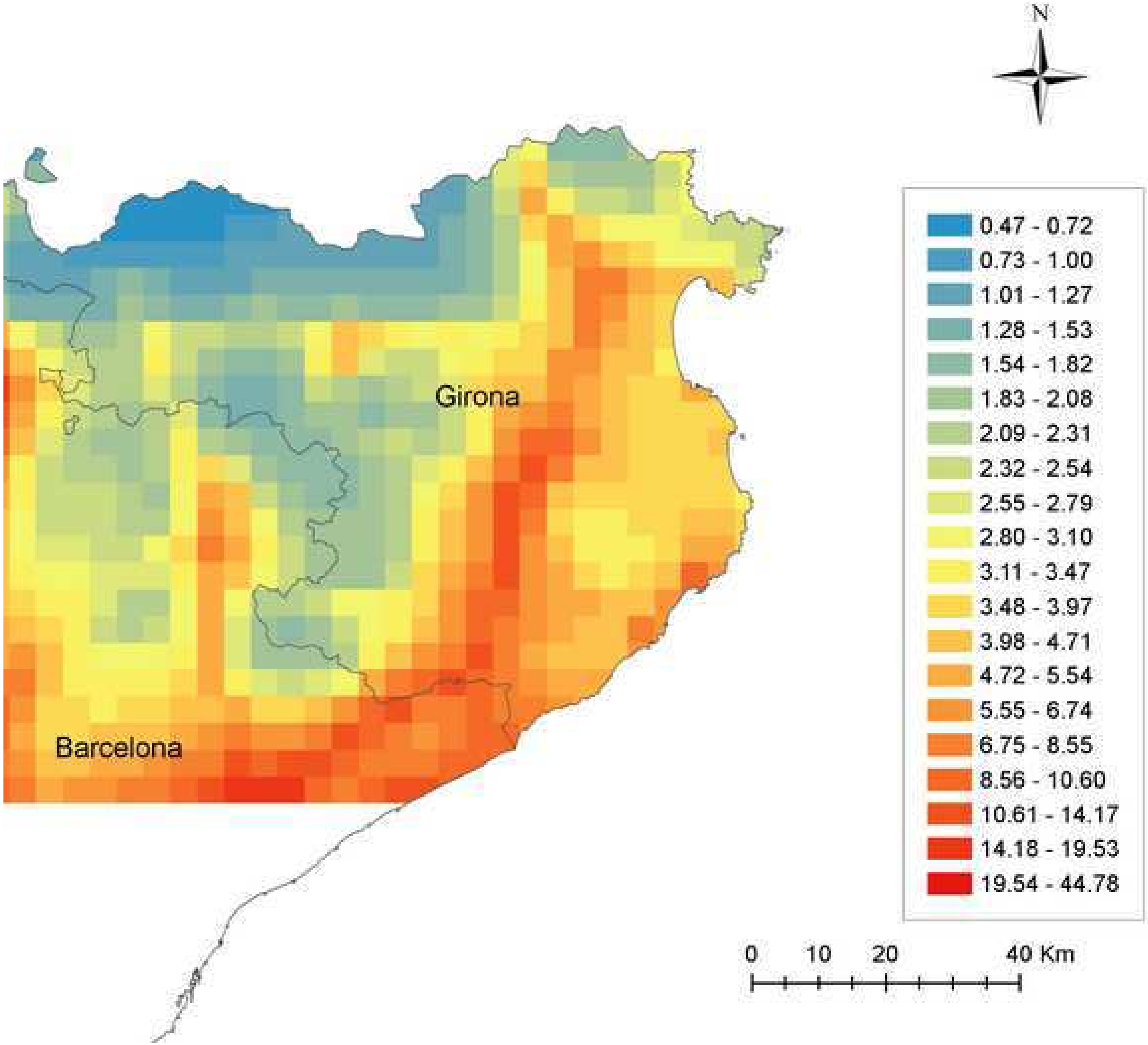

Figure 3
Click here to download high resolution image

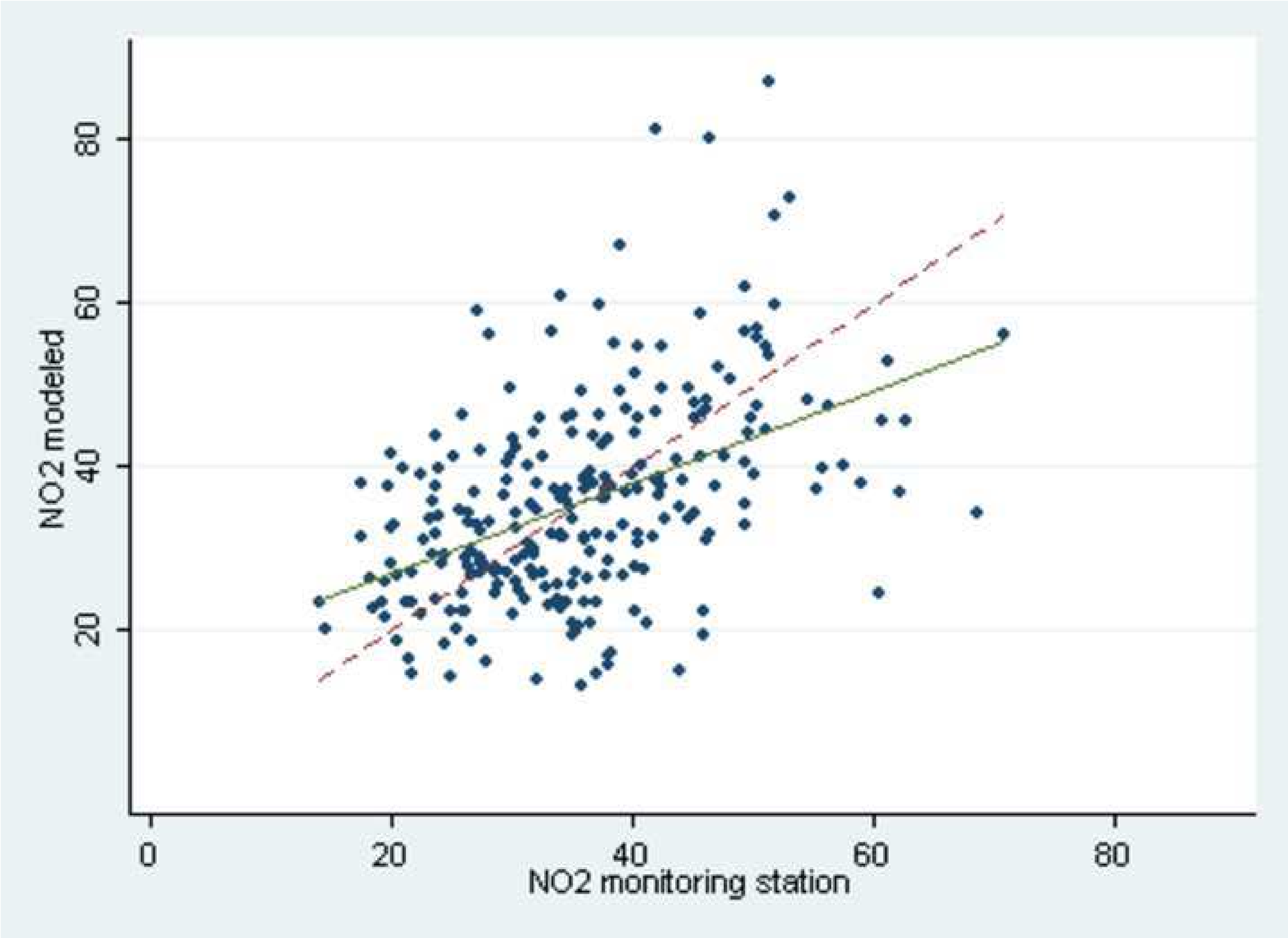

Figure 4
Click here to download high resolution image