Accepted Manuscript

Development and performance evaluation of new AirGIS – A GIS based air pollution and human exposure modelling system

Jibran Khan, Konstantinos Kakosimos, Ole Raaschou-Nielsen, Jørgen Brandt, Steen Solvang Jensen, Thomas Ellermann, Matthias Ketzel

PII: S1352-2310(18)30726-X

DOI: <https://doi.org/10.1016/j.atmosenv.2018.10.036>

Reference: AEA 16333

To appear in: *Atmospheric Environment*

Received Date: 23 May 2018

Revised Date: 12 October 2018

Accepted Date: 21 October 2018

Please cite this article as: Khan, J., Kakosimos, K., Raaschou-Nielsen, O., Brandt, J., Jensen, S.S., Ellermann, T., Ketzel, M., Development and performance evaluation of new AirGIS – A GIS based air pollution and human exposure modelling system, *Atmospheric Environment* (2018), doi: <https://doi.org/10.1016/j.atmosenv.2018.10.036>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Development and Performance Evaluation of New AirGIS – A GIS Based Air Pollution and Human Exposure Modelling System

Jibran Khan^{1,2,*}, Konstantinos Kakosimos², Ole Raaschou-Nielsen^{1,3}, Jørgen Brandt¹, Steen Solvang Jensen¹, Thomas Ellermann¹, Matthias Ketzel^{1,4}

¹Department of Environmental Science, Aarhus University, Roskilde, Denmark

²Department of Chemical Engineering, Texas A&M University at Qatar, Doha, Qatar

³The Danish Cancer Society Research Centre, Copenhagen, Denmark

⁴Global Centre for Clean Air Research (GCARE), Department of Civil and Environmental Engineering, University of Surrey, Guildford, United Kingdom

Abstract

AirGIS, a Geographic Information Systems (GIS) based air pollution and human exposure modelling system, is routinely used in conjunction with the Operational Street Pollution Model (OSPM[®]), across the globe, to assess local- or street-scale air pollution. We developed a substantially revised version of AirGIS (hereafter, new AirGIS) as a new modelling system in open-source GIS i.e. PostgreSQL software with its spatial extension PostGIS to (1) optimize the model performance enabling model calculations for a large number of sites over a large geographical area, with limited computing resources (2) replace the outdated programming language Avenue (3) become independent of commercial GIS software. This paper, therefore, aims to describe the overall structure of new AirGIS modelling system together with its strengths and limitations. Furthermore, the new AirGIS has been evaluated against various measured datasets of ambient air pollution (NO_x, NO₂, PM₁₀ and PM_{2.5}). In terms of reproducing temporal variation (single location, time series of concentrations e.g. annual, daily etc.) of air pollution, the new model achieved correlations (R) in the range 0.45 – 0.96. While, in terms of reproducing the spatial variation (several locations, single time interval), the new AirGIS achieved correlations in the range 0.32 – 0.92. The new model, therefore, can be used for both short- and long-term air pollution exposure assessments to facilitate health related studies. However, the present evaluation of the new modelling system also revealed that the new AirGIS significantly overestimated the observed concentrations for two out of four datasets. The possible reasons for these errors and future directions to reducing the bias in the new model output have been discussed.

Keywords – Urban air pollution; human exposure modelling; model evaluation; GIS, PostgreSQL, AirGIS, OSPM

*Corresponding author at: Matthias Ketzel, Department of Environmental Science, Aarhus University, DK-4000, Roskilde, Denmark. E-mail address: mke@envs.au.dk, Phone: +45 871 58529.

Abbreviations and acronyms

AADT, annual average daily traffic; BC, black carbon; CO, carbon monoxide; CTM, chemistry-transport model; CoV, coefficient of variation; DEHM, Danish Eulerian Hemispheric Model; DSM, Danish Surface Model; DTM, Danish Terrain Model; ESCAPE, European Study of Cohorts for Air Pollution Effects; ESRI, Environmental Systems Research Institute; FAC2, factor of two statistic; GIS, Geographic Information Systems; GPS, Global Positioning System; LUR, land use regression; NO_x, nitrogen oxides ($\mu\text{g}/\text{m}^3$); NO₂, nitrogen dioxide ($\mu\text{g}/\text{m}^3$); NMB, normalized mean bias; NOVANA, Danish/National Monitoring and Assessment Programme for the Aquatic and Terrestrial Environment; O₃, ozone; OSM, OpenStreet map; OTM, OpenTransport map; OSPM, Operational Street Pollution Model; PM, particulate matter; PM₁₀, mass concentrations of particulate matter less than 10 μm in aerodynamic diameter ($\mu\text{g}/\text{m}^3$); PM_{2.5}, mass concentrations of particulate matter less than 2.5 μm in aerodynamic diameter ($\mu\text{g}/\text{m}^3$); RMSE, root-mean squared error; SOA, secondary organic aerosol; SPREAD, Spatial Distribution of Emissions to Air; SUB, Simple Urban Background methodology; UBM, Urban Background Model; WRF, Weather Research and Forecasting Model.

1. Introduction

Air pollution originating from anthropogenic and natural sources, has a great influence on the human health and environment. In particular, there has been extensive evidence associating serious health outcomes with air pollution exposures e.g. cardio-respiratory mortality (Hoek et al., 2013), myocardial infarction (Puett et al., 2014) and lung cancer (Raaschou-Nielsen et al., 2013). In large urban areas, variations in traffic flows and speeds, meteorology, land use, and terrain lead to the complex patterns of air pollution (Wilson and Zawar-Reza, 2006). Subsequently, relatively dense networks of routine air pollution monitoring are not sufficient to capture the spatial and temporal patterns of air pollution exposures in complex urban agglomerations. Also, direct measurement of exposures (e.g. via personal exposure measuring devices or biomarkers) is clearly not feasible for large study populations (Vienneau et al., 2009). Air pollution modelling is, therefore, indispensable. Within the epidemiology community, the use of so-called LUR models is well established to study population exposure in cohort studies (Beelen et al., 2013). Although, these models are adequate for exposure assessment but are often considered to be restricted to the time period and geographical area of the monitoring campaign used for their creation (de Hoogh et al., 2014). Also, LUR models cannot reliably be used for scenario calculations (Lefebvre et al., 2017).

Air quality dispersion models are effective tools to reproduce spatial and temporal variations of pollutants emitted into the atmosphere (Ketzel et al., 2011). Their use to implement air quality mapping and assessment, forecasting and management, at both national and local level, is also supported by the European Commission Directive 2008/50/EC, on Ambient Air Quality and Cleaner Air for Europe. In turn, air pollution dispersion models coupled with GIS are increasingly used to meet the needs of exposure assessment in health related studies (e.g. Morra et al., 2006; Beevers et al., 2013). Presently, there exist several air pollution dispersion models e.g. ADMS-Urban (Carruthers et al., 2000; CERC, 2017), DUSTAN (Allwine et al., 2006), AERMOD (Kesarkar et al., 2007; The United States EPA, 2017), STEMS-Air (Gulliver & Briggs, 2011), KCL-urban and CMAQ-urban (Beevers et al., 2013), IFDM (Lefebvre et al., 2011; 2013) based dispersion-kernel method (Lefebvre et al., 2017), PMSS (Duchenne and Armand, 2017) etc. capable to address air pollution mapping over a large geographical region with high spatial resolution and are adequate for air pollution exposure assessment studies.

For the assessment of local- or street-scale air pollution at many address locations, the Danish AirGIS system was developed by Jensen et al., (2001). It is a GIS-based model system for air pollution dispersion and human exposure modelling. The AirGIS system has been previously validated by Ketzel et al., (2011) for several compounds (NO_x, NO₂, CO, O₃) and time resolutions (annual, daily means etc.). The modelling system, in general, is capable to calculate ambient air quality and traffic-related human exposures at high temporal (hourly basis) and spatial resolution (individual address location close to the building façade) to facilitate air pollution epidemiology and urban air quality management and assessment (Ketzel et al., 2011). AirGIS has been routinely used for (i) the integrated Danish Air Quality Monitoring Programme under NOVANA (Ellermann et al., 2018; Hertel et al., 2007) (<http://dce.au.dk/en/monitoring/>), (ii) several health related studies based on short-term (e.g. Raaschou-Nielsen et al., 2000) and long-term (e.g. Andersen et al., 2016) exposure to ambient air pollution. It has also been used to develop a publicly available webGIS service with background concentration over Denmark at 1 km x 1 km resolution and street concentrations for 2.4 million address locations (Jensen et al., 2017).

The AirGIS system was implemented in Avenue scripting language as an extension in ArcView 3.x (Jensen et al., 2001) which is not supported by ESRI® anymore. Thus, it was not possible to update the former AirGIS at its present architecture. Moreover, it was also highly desired to optimize AirGIS model performance over a large study area with many address locations. In conjunction, recent advances in GIS tools allow for a more sophisticated handling of large spatial datasets. Thus, we developed a substantially revised version of the AirGIS system (new AirGIS) in open-source GIS. New AirGIS has been implemented in open source PostgreSQL¹ software (also known as Postgres) with its spatial extension PostGIS² in conjunction with R³ scripts for pre-processing and post-processing of the datasets. The main objectives of using open source software and geospatial programming have been to (1) optimize the model performance over a large geographical region with limited computing resources (2) comply with the latest developments as well as future adjustments of the model as per need/requirements (3) replace the outdated programming language Avenue and (4) become independent of commercial GIS software. In addition, in the new model system, the integration of inhomogeneous emissions scenario is planned as a future update once relevant GIS datasets are available. Subsequently, to facilitate future epidemiological studies, it was necessary to assess the performance of the new modelling system.

Thus, there are two main aims of this research work. First, to describe the development together with strengths, limitations and future prospects of the new AirGIS. Second, to evaluate the performance of new AirGIS modelling system against measured data, in terms of its feasibility for future air pollution dispersion modelling and exposure assessment studies, reflecting upon the key significance of this research work. In this study, new AirGIS has been evaluated against four measurements datasets (see chapter 2.4 for further details and references to the datasets): (1) several years of long-term measurements (1994 – 2015) from four permanent monitoring stations of the Danish air quality

¹ See: <https://www.postgresql.org/>

² See: <https://postgis.net/>

³ See: <https://www.r-project.org/>

monitoring network, (2) short-term measurements available to us as part of the ESCAPE project (2009-10), (3) another set of 5-week passive measurements along ten major streets in Copenhagen, Denmark (2011), (4) 1-month measurements campaign at the 204 addresses in Greater Copenhagen area, Denmark (1994-5). The new model's performance is assessed by taking into account different temporal (single location, annual, daily averages etc. of concentrations) (dataset 1, 2, 4) and spatial (several locations, single time interval) (dataset 2, 3, 4) variations of air pollution levels under varying urban settings representing a wide range of traffic patterns and street geometries. As compared to the previous validation study (Ketzler et al., 2011); this research work differs mainly in pollutants included, as also PM_{10} and $PM_{2.5}$ have been included in the new model evaluation.

2. Materials and methods

2.1 The AirGIS system

This section summarizes the general working principle of the AirGIS modelling system.

AirGIS (<http://envs.au.dk/en/knowledge/air/models/airgis/>), is part of the multiscale integrated dispersion modelling system THOR⁴ covering coupled modelling of regional background concentration, urban background concentrations and street concentrations (Brandt et al., 2001). Based on national GIS datasets, AirGIS generates input files for the street pollution model OSPM[®] to estimate air pollution at address locations. Automatic generation of necessary input files (traffic and street geometry information etc.) for OSPM, is one of the key features of AirGIS that would otherwise be very tedious and time consuming to produce for a large number of addresses. This, consequently, enables estimation of air quality levels at a large number of addresses in an automatic and effective way. AirGIS is able to estimate air pollution at any address location in Denmark. In the case that the address is located along minor roads (≤ 500 veh/day) only the urban background concentration is assigned. In case the address is located near a road with significant traffic (> 500 veh/day) the street pollution model OSPM is applied additionally to the urban background. The air pollution concentration is modelled at a receptor point close to the building façade in a standard height of 2m, but it is possible to change the receptor height e.g. in case of available information about the floor number of addresses in a multi-floor apartment building.

The AirGIS system operates at three different levels of pollution (Figure 1). The working principle of these three dispersion models is summarized as follows: (1) The Danish Eulerian Hemispheric Model (DEHM) (Christensen, 1997) calculates the regional background concentrations in a 5.6 km x 5.6 km grid resolution. It is a three dimensional, offline, large-scale, Eulerian, nested grid (Denmark: 5.6 km x 5.6 km, Northern Europe: 17 km x 17 km etc.), atmospheric CTM model developed to study long-range transport of air pollution on the Northern Hemisphere. DEHM includes emissions from all sources outside the area including traffic, small-scale combustion, power plants, industrial units etc. using a comprehensive chemical scheme based on photochemistry and particles (Brandt et al., 2012) (2) The Urban Background Model (UBM) (Berkowicz, 2000b) calculates the urban background concentrations in a 1 km x 1 km grid

⁴ See also: <http://envs.au.dk/en/knowledge/air/models/thor/>

resolution. Being a multiple source model, it uses a Gaussian approach for horizontal dispersion and a linear approach for vertical dispersion up to the boundary layer. (3) The Operational Street Pollution Model (OSPM)⁵ (Berkowicz, 2000a; Kakosimos et al., 2010) calculates the street contributions where background concentrations from DEHM/UBM are included via AirGIS system. OSPM® uses a combination of a plume model for the direct contribution from the traffic source and a box model for the recirculating part of the pollutants inside the street canyon environment. While adding contributions from three pollution levels, NO_x-NO₂ non-linearity is taken into account in all models by a full atmospheric chemistry module in DEHM and simple NO-NO₂-O₃ chemistry modules in UBM and OSPM models.

The emissions database for Denmark has a high spatial resolution of 1 km x 1km and is based on SPREAD methodology (Plejdrup and Gyldenkerne, 2011) for the spatial distribution of national emissions. In the current model system, meteorological datasets (wind speed, wind direction, air temperature etc.) that are used as input to all models (DEHM, UBM, OSPM) are based on WRF model (NCAR, 2018). Moreover, regional background concentrations that are input to UBM are considered as spatially homogenous over the city and nearby surroundings. While treating background concentrations, it is important that a double counting of emissions should be avoided (Lefebvre et al., 2017). In our model chain, UBM takes DEHM concentrations 25 km upwind (hour by hour depending on wind direction) and models only emission of the closest 25 km – thereby avoiding double counting of emissions. Then, OSPM calculates the “street increment”, considering only the contribution of the closest street to the address, on top of the UBM background (1km x 1km horizontal resolution) not introducing double counting either.

Figure 1: An illustration of the spatial variation of three air pollution levels considered in the AirGIS system. DEHM: the Danish Eulerian Hemispheric Model; UBM: Urban Background Model; OSPM: Operation Street Pollution Model; *Light green* colour indicates the level of regional background concentrations; *Yellow* colour indicates the urban background concentrations; *Maroon* colour indicates the direct street contributions (traffic hotspots).

⁵ See also: <http://envs.au.dk/en/knowledge/air/models/ospm/>

The AirGIS modelling system, in general, performs calculations on an hourly basis and then concentrations are averaged over the time period corresponding to those used in the exposure studies. This allows a calculation of short-term and long-term averages of air pollution estimates which can be beneficial for health related studies based on exposure assessment at street level. Furthermore, the modelling system has been through continuous refinement i.e. in context of input datasets on traffic and street geometry, vehicle emission factors, chemistry etc. (e.g. Ketzel et al., 2012). Recently, the system has been extended to and validated for PM₁₀, PM_{2.5} and BC (Hvidtfeldt et al., 2018). At present, the complete modelling system allows for the estimation of various air pollutants including NO₂, NO_x, PM₁₀, PM_{2.5}, CO, O₃ and BC.

2.2 New AirGIS architecture

This section summarizes the new features and updates in the AirGIS system. In addition, similarities and differences in the two systems (old and new AirGIS) have also been highlighted.

2.2.1 New system overview

An overall overview of the new AirGIS is as follows. Unlike the former AirGIS, the new model system makes use of open-source GIS programming tools in PostgreSQL software, with GIS functions provided by its spatial extension i.e. PostGIS, in conjunction with R scripting language (The R Core Team, 2017) interface for pre- and post-processing of the datasets. The processing speed for spatial operations in PostGIS is greater than for other common desktop GIS applications due to its efficient use of spatial indexing (Gulliver et al., 2015). This is of particular importance for the application of new AirGIS that uses address points with exposure periods, road networks with traffic composition information and buildings footprints with building heights information as input for a large geographical region of interest. As such, PostGIS provides an effective environment in which to handle these large spatial data sets. R scripts (via R-Studio) provide a seamless interface to query data from PostgreSQL database as well as to execute PostGIS commands. The main aim of using R software has been (1) to strengthen the new system by its significant statistical computing capabilities (2) to perform preliminary data analyses as soon as PostGIS commands execution is completed (3) to provide a single flexible environment to perform all data handling from pre-processing via running GIS queries to post-processing and export to OSPM and UBM model runs. Following sub-sections summarize the working principle of the new model system as well as briefly compare it with the former AirGIS.

Figure 2 shows an overall structure and dataflow of the new AirGIS and is explained in the following.

Figure 2: An overall structure and dataflow of new AirGIS architecture. See also list of abbreviation at the beginning of the paper; Block **GIS data** represents the input GIS shapefiles (receptors points, road networks, building footprints); Block **AirGIS system** represents the new AirGIS processing workflow, numbered boxes indicate the key steps of workflow, see text; Block **QGIS interface** represents the visualization interface to view/assess air pollution exposure output.

Block **GIS data** (Figure 2) represents the spatial input data i.e. address points (with exposure periods), road networks (with traffic information) and building footprints (with building heights). Traffic information includes annual average daily traffic (AADT) flow for passenger cars, vans, lorries and buses and the travel speed for varying vehicle categories for each street. The input files are stored in a PostgreSQL/PostGIS database for the geometric processing.

Block **AirGIS System** (Figure 2) represents the central part of new AirGIS workflow. In this block, first, the closest/most busy street to each address location within a certain buffer distance (e.g. 30m – 50m, the empirical radii and easy to change by the user) are searched (via Postgres script) (Figure 2, “step 1”). During this search, the address points are divided on the basis of AADT value at the closest/most busy street into two categories i.e. urban background point (hereafter, background point) and street

concentrations point (hereafter, street point). That is, if AADT is less than 500 veh/day (also user defined and changeable), then the street contribution (calculated by OSPM) is usually very small and can be neglected. Therefore, such address points are assigned as background points (bypassing the OSPM calculations). Otherwise, the address points are assigned as street points.

Concerning the background points, they are processed only in connection to the urban background contributions. The coupled models chain works as follows: WRF provides meteorological input to DEHM model to estimating regional background concentrations that are input to the UBM model (Figure 2). The urban background concentrations are calculated with the UBM model and, subsequently, averaged according to the exposure periods and spatial grid, via R script (Figure 2, "step 2"). Thus, in the new system, transport and chemistry of pollutants are generally treated in the same way as the former system. Here, it should be noted that the previous versions of AirGIS, sometimes, also made use of (i) regional/background monitoring data as background concentrations (depending on the availability) (ii) simplified SUB method (Berkowicz, 2000b) to estimate urban background pollution levels. Presently, in the new system, background-monitoring data is only used to validate overall model output. Furthermore, only DEHM and UBM are used to estimate background concentrations.

After estimation of background concentrations (Figure 2, "step 2"), the new system processes the street points (AADT $\geq$ 500). That is, new AirGIS via an automatic Postgres script, generates for each address point an orthogonally projected point on the closest street centerline (referred as OSPM points) (Figure 2, "step 3"). Once OSPM points are generated, the new system produces the street configuration information (street width, lengths etc.) for each of these points. The process of generation of street configuration in the new system is described in detail in the sub-section 2.2.2.

Once street configuration is generated, the input files for OSPM calculations in the required format are produced. Hereafter, OSPM runs take place to calculate street contributions. Then, OSPM's output together with DEHM/UBM results is further processed through R-scripts (post-processing) (Figure 2, "step 4") to produce exposure output. This is one of the unique innovative features of the new AirGIS where pre- and post-processing is handled via R scripting interface enabling processing of large datasets. In the former AirGIS system, pre- and post-processing was handled in a number of manual steps.

Finally, block *QGIS⁶ interface* (Figure 2) (QGIS Development Team, 2018) represents the visualization interface during the whole modelling process, where both input and output files can be visualized readily in a direct link to the Postgres database.

2.2.2 Computation of street configuration

Here, in this section, the process of estimation of street configuration in the new system is summarized. We recall that for each OSPM point (residing on street center line), the new model system generates street configuration information by making use of scripts written in PostGIS programming environment and streets and buildings spatial data sets. The street configuration information, in general, represents

⁶ <https://qgis.org>

(i) the physical environment (e.g. street orientation, street width, height of buildings in different wind sectors) around the receptor points (ii) static data that will be produced only once for each address location.

Similar to the former AirGIS (Jensen et al., 2001), the new model system uses the concept of 2½ dimensional urban landscape model (Hansen et al., 1997) to estimate street configuration information for each OSPM point. The term “2½ dimensional” (also known as pseudo-3D) describes the 2D graphical projections to appear as three-dimensional (3D), when in fact they are not (Wikipedia, 2018). Thus, the estimated street configuration information (street orientation, street lengths etc.) of the both systems, due to the same concept, is essentially similar to each other with only difference that is, in the new system it is implemented (“re-programmed”) in a new language, here PostGIS.

The procedure of street configuration estimation is as follows. For each OSPM point (located at street center line), the new system first estimates the street orientation. The street orientation ($0^{\circ} - 180^{\circ}$) is usually computed clockwise according to the true north (Figure 3b) and determined by computing the direction of the street center line nearest to the receptor point. Based on this concept, the PostGIS script in the new system, first splits the road network shapefile (multilinestring) into individual line segments (linestring), and then the nearest line segment (edge) is used to estimate the street orientation.

Air pollution levels are influenced by the buildings layout in a street canyon environment (Vardoulakis et al., 2003; Xie et al., 2005; Shu et al., 2014). The new model system handles this phenomenon in the same way as the old system. The system creates 12 wind sectors, where each wind sector covers an angle of 30 degrees. Based on above criteria, new AirGIS generates wind sectors around OSPM point to estimate the height of the buildings in wind sectors (Figure 3a). First, the new system generates 12 wind sectors within 50m buffer (changeable) around each OSPM point (Step 1). Then, the first wind sector pie is aligned to the street orientation (Step 2) and subsequently it is used to locate the building as well as to identify the associated building height of that wind sector (Step 3). Furthermore, the new model also searches for the general building height i.e. the most prevalent height among wind sectors. If the prevalent height is zero, the general building height is estimated as zero, which is allowed and handled appropriately in OSPM.

Figure 3: A representative scenario of street configuration estimation in the new AirGIS via PostGIS scripts (visualization in QGIS software) (a) the generation of 12 wind sectors (with first wind sector aligned to the street orientation) to estimate the height of the buildings in wind sectors. Sectors 9 and 10 are the examples of “no buildings” case. Buildings heights (in meters) and examples of street lengths (length 1 and 2) have also been shown. The minimum and maximum allowable lengths are 50m and 200m, respectively (b) the example of street configuration parameters (width, orientation) in the new AirGIS system. Address point, OSPM point and receptor point have also been shown.

Next, the new system computes the street length, as distance from OSPM point to the nearest street intersections on both sides. Based on empirical values, the minimum allowed street length is 50m and the maximum allowed street length is 200m.

Among street configuration parameters, street width has a key importance. The new model system estimates the street width (Figure 3b) as follows. First, the new system finds the intersection of 50m buffer around the OSPM point and the nearest building polygons on each side of the street. Based on this intersection, the new model generates parallel lines on both sides of the street aligned with closest buildings. Then, distance from OSPM point to the parallel lines is calculated and summed as the estimated street width. This approach also works when there are buildings on only one side of the street. A brief comparison of former and new AirGIS systems (Figure S1, S2, S3 and Table S1), in terms of

their capability to generate street configuration information, is shown in *Appendix A, supplementary material*. The comparison shows a good agreement between estimated street configuration of the old and new systems. Furthermore, to summarize, Table 1 lists similarities and differences in the former and new AirGIS systems.

Table 1: List of similarities and differences in the former and new AirGIS systems.

System features	Former AirGIS	New AirGIS
Input (GIS shapefiles)	Addresses, road network, buildings	Addresses, road network, buildings
Scripting language	Avenue scripting language (ArcView 3.x)	PostGIS scripting in a Postgres database via R
Estimation of street configuration	GIS model concept by Hansen et al., 1997	GIS model concept by Hansen et al., 1997
Regional background contributions (long-range chemistry transport)	Measurements/ DEHM (5.6 km x 5.6 km for Denmark)	DEHM (5.6 km x 5.6 km for Denmark)
Urban background contribution	SUB method/Measurements/ UBM for parts of Denmark	UBM (1 km x 1 km grid) for all Denmark
Street contribution	OSPM [®]	OSPM [®]
Meteorological input	The WRF model	The WRF model
Pre- and post-processing	Manual steps	R
Visualization interface for exposure output	ArcView 3.x	QGIS

2.3 GIS base data

2.3.1 Road/traffic data

Road traffic data sets have a key importance for modelling traffic air pollution. In Denmark, road traffic information is based on a national traffic database in form of a GIS shapefile. The GIS road network is originally based on the TOP10DK road network of the National Survey and Cadastre from 1999 that was subsequently updated to KORT10 road network i.e. the nationwide object-oriented map, since 2007 (<https://kortforsyningen.dk/indhold/data>, Jensen et al., 2009a). The digital mapping of the roads is based on aerial photos (spatial resolution: 40 cm x 40 cm) and precision is high (about 1m) for the center line of the road (The Danish Geodata Agency, 2006). The traffic database, therefore, contains polyline shapes representing the road center line and a list of relevant traffic related attributes that are required for the geometric processing in new AirGIS, and later OSPM. The most relevant attributes of the GIS road network are road type (to specify street type, see Table 2), AADT, and travel speed (Jensen et al., 2009b).

The street type specifies the vehicle distribution and the diurnal variation in traffic, the so-called OSPM street types. Table 2 shows an overview of the street classification used in the AirGIS system. Each OSPM street type refers to a text file containing information about hourly traffic distribution for different days and vehicle types (passenger cars, vans, small and large trucks, buses). Days include Monday – Thursday, Friday, Saturday, Sunday and further divided on the month of July (summer

vacation month) and other months in the year. In this way, the system estimates the hourly traffic for each vehicle category and arbitrary hour of the year, required to calculate air pollution concentration. See Jensen et al., (2009a) for further details and an example of OSPM street “type B”.

Table 2: The street classification used in the AirGIS modelling system to estimate air pollution concentrations (using OSPM) at any address location in Denmark (source: Jensen et al., 2009a).

Street type	Description	Passenger cars	Vans	Small trucks (≤ 32 t)	Large trucks (> 32 t)	Buses
		%	%	%	%	%
B	Transit roads in larger cities	81.9	10.8	2.8	1.4	3.1
C	Distribution roads in residential areas	83.2	12.1	2.5	0.74	1.6
D	Roads with a mix of residential and business area	81.9	11.7	3.0	1.5	2.0
F	Access roads to larger cities	79.7	11.9	4.3	2.5	1.6

Furthermore, in addition to the above mentioned national/Danish datasets, several open-source GIS data of road links, buildings and address points are also available and useful when applying AirGIS outside Denmark, see Table S2 (*Appendix B, supplementary material*).

2.3.2 Building data

Information on building foot prints as polygon shapefile is also available for the whole of Denmark based on a national data set (Kort10 DK) obtained from the Danish Geodata Agency (<http://gst.dk>). Building height was estimated for each building based on the National Elevation Model that has a resolution of 1m x 1m calculated as the difference between DTM (Danish Terrain Model) and DSM (Danish Surface Model).

2.3.3 Address point data

Geocoded address locations (point data) are available for the whole Denmark via the Central Person Registry. For smooth processing of the new AirGIS algorithms, it is important that an address point should be inside building polygon. The address locations are assigned with exposure periods, which can be short-term and/or long-term as per population exposure study design. The receptor point is usually assumed to be at 2m height as standard and near the façade of the building closest to the address point or curbside of the street in case of no building. However, other heights can be modelled when information about height is available, e.g. floor number of an apartment.

2.4 Measured air pollution datasets

Model evaluation is indispensable for reliable air pollution exposure estimates. The new AirGIS modelling system has been evaluated against various available measurements datasets (Table 3). In addition, measurements of the Danish Air Quality Monitoring Programme (dataset 1) have also been used to calibrate the new modelling system in terms of modelled PM (see chapter 2.5).

Table 3: An overview of various measured air pollution datasets used for the performance evaluation of new AirGIS modelling system.

Dataset	Name	Pollutant	Measurement site	Measurement method	Time resolution	Location (city)	Time Period
1	Danish Air Quality Monitoring Programme (NOVANA)	NO _x , NO ₂ , PM ₁₀ , PM _{2.5}	4 permanent kerbside	Active/continuous	Hourly	Copenhagen, Odense, Aarhus, Aalborg	1994 - 2015
2	ESCAPE-EU Danish campaign	NO _x , NO ₂ , PM ₁₀ , PM _{2.5}	41 streets (20 streets for PM)	Passive, active	3 obs./ 14 days	Greater Copenhagen area	November 2009 – October 2010
3	Five weeks passive sampling campaign	NO ₂	10 major streets	Passive	1 obs./ 5 weeks	Copenhagen	October 24 – November 28, 2011
4	1-month campaign	NO ₂	204 addresses	Passive	6 obs./ 1 month	Greater Copenhagen area	1994 - 1995

Figure 4 shows the locations of measurement sites of datasets 2, 3 and 4 in Copenhagen, Denmark and Figure S4 (*Appendix C, supplementary material*) shows locations for dataset 1. The following subsections summarize various measured air pollution datasets.

2.4.1 Dataset 1 – The Danish Air Quality Monitoring Network

The urban part of the Danish Air Quality Monitoring Network consists of five permanent kerbside stations in four major cities of Denmark (see: <http://envs.au.dk/en/knowledge/air/monitoring/>, for more details; Ellermann et al., 2018). All stations monitors provide hourly measurements of various pollutants (NO_x, NO₂ etc.) Among five permanent stations, the street station in Odense, Denmark has been moved to the new location i.e. Grønløkkevej (ODGR), and has only been operational since 2015. However, measured data at the previous location i.e. Albanigade (ODAL) was available. Furthermore, at H.C. Andersens Boulevard (HCAB) monitoring station in Copenhagen, Denmark, there was a change in street layout in 2010 that moved traffic closer to the station (Ellermann et al., 2018). Consequently, this led to the lack of reliable historic measured data (e.g. year 2010) at HCAB. In turn, HCAB was excluded from the analyses. Thus, in this study, long-term series (22 years) of half-hourly measurements at four permanent street stations i.e. Jagtvej (JGTV) in Copenhagen, Albanigade (ODAL) in Odense (old location), Banegårdsgade (AARH) in Aarhus, and Vesterbro (AALB) in Aalborg, are used to evaluate the new model performance. For the new model evaluation, we used measured concentrations ($\mu\text{g}/\text{m}^3$) of NO_x, NO₂, PM₁₀ and PM_{2.5} in the years 1994-2015.

2.4.2 Dataset 2 – Danish measurement campaign within ESCAPE project

Another dataset of short-term measurements was available as part of ESCAPE (<http://www.escapeproject.eu/>, Eeftens et al., 2012), in which 20 study areas across Europe were included to investigate the spatial variation of particles and nitrogen oxide (NO_x). The Danish measurement campaigns (Figure 4a) were conducted from November 2009 to October 2010 and covered 41 sites near Copenhagen for the measurements of NO_x, NO₂ and 20 sites for the measurements of PM₁₀ and PM_{2.5}. A 14-days measurement campaign was conducted three times for each site (N=41 sites for NO_x, NO₂; N=20 for PM). Failed measurements were repeated later to have three valid observed values. Further details about measurement campaign and samplers used are provided in Eeftens et al., (2012).

2.4.3 Dataset 3 – NO₂ 2011 measurement campaign

In order to evaluate the performance of AirGIS/OSPM for more locations in addition to permanent sites (Dataset 1), a 5-weeks passive measurement campaign was performed to measure NO₂ concentrations (in µg/m³) along ten busy roads (Figure 4b) in Copenhagen, Denmark from October 24th, 2011 to November 28th, 2011 (The FORCE Technology, 2018). The measurements were conducted using passive samplers by IVL, Gothenburg, Sweden (Ferm & Svanberg, 1998) that were mounted at lamp posts, traffic signs or building façade in about 2m height. This dataset was also used previously to validate former AirGIS/OSPM (Ketzel et al., 2012).

2.4.4 Dataset 4 – NO₂ measurement campaign at the 204 addresses

Another dataset from a comprehensive measurement campaign at the 204 addresses was available within the Childhood Cancer project (see Raaschou-Nielsen et al., 2000 for more details). Measurement campaigns were conducted at 103 street locations in the central Copenhagen, Denmark, and 101 locations in the Greater Copenhagen Area (20 – 50 km outside). Seven measurement campaigns covering 30 locations took place in October – November 1994 and in April – June 1995. At each measurement site, monthly mean concentrations of NO₂ were measured for six consecutive months (N=1224). Passive samplers were placed about 0.5m from the building façade and 4m above street level.

Figure 4: Locations of various measured datasets used to evaluate the performance of New AirGIS modelling system (a) Dataset 2 – the ESCAPE project measurement sites (NO_x, NO₂: 41 sites, PM: 20 sites) in the Greater Copenhagen Area during November 2009 – October 2010 (b) Dataset 3 – locations of the 10 measuring points (NO₂) along major roads in Copenhagen using passive samplers during October 24th, 2011 to November 28th, 2011 (c) Dataset 4 – NO₂ measurement sites at the 204 addresses in the Greater Copenhagen Area during October – November 1994 and April – June 1995.

2.5 PM calibration and model evaluation statistics

The modelling of PM concentrations is still under development within the AirGIS system and new components of the PM mass have recently been added e.g. secondary organic aerosol (SOA) in DEHM. However a comparison of modelled PM₁₀ and PM_{2.5} concentrations against measurements using the EU-reference method to determine PM at the permanent stations of the Danish Air Quality Monitoring

Programme (Ellermann et al., 2018) reveals that the model still underestimates the measured PM. The reason for this is most likely a remaining underestimation for some of the PM constituents in DEHM such as primary organic PM, secondary aerosols or water content in PM, and underestimation in OSPM of non-exhaust from traffic (road wear, tyre wear and brake wear). In order to compensate for the underestimation by the model we applied correction factors of 1.46 and 1.26 to all our modelled concentrations of PM₁₀ and PM_{2.5} values in this study. However, no calibration factors were applied to the modelled NO_x and NO₂.

Model performance is often evaluated on the mixture of spatial (several locations in space, only one time average) and temporal (one single location, timer series of many measurements e.g. annual, daily etc.) variation of predicted values. In our modelling system, the temporal variation of the modelled values evaluates mainly the whole chain of air pollution dispersion models (DEHM/UBM/OSPM) but to less extent the GIS part. While the spatial variation of modelled values reflects the performance of the whole AirGIS modelling system and the correctness of the input data, e.g. building foot prints, generation of street configurations or traffic data in the data base.

The performance of the new model system was primarily evaluated using spatial and temporal correlations (Pearson's correlation coefficient "r"). In addition, various other model evaluation statistics i.e. the coefficient of variation (CoV), root mean-squared error (RMSE), normalized mean bias (NMB), factor of two statistic (FAC2) (Carslaw, 2015) were also used (see *Appendix D, supplementary material* for the definitions). All statistical analyses were performed in R version 3.4.0 (<https://www.r-project.org/>) using "OpenAir" R package (Carslaw and Ropkins, 2012).

3. Results and discussions

3.1 Model evaluation against measured dataset 1

3.1.1 Annual averages

This section presents the results of new AirGIS evaluation against measured dataset 1 with a set of annual averages for various pollutants and for a large number of years (1994 – 2015). Thus, this part of the model validation is mainly on temporal validation of the long-term trends. These trends in the model are dominated by the trends in the emission data superimposed by year-to-year changes in the meteorology i.e. variable average wind speed.

Results for several years of new model runs at four measurement stations i.e. JGTV, ODAL, AARH and AARH are given as 22 years annual averages in Figure 5. For each station, trends for both NO_x and NO₂ are shown for modelled and observed street level concentrations (in µg/m³). A significant decrease in NO_x concentrations levels at all stations can clearly be observed, mainly caused by the changes in traffic emissions (Ketzel et al., 2011). This trend is, in general, well reproduced by the new AirGIS over a significant period of time. The changes in traffic emissions depend on traffic volume, vehicle distribution (light and heavy duty vehicles, diesel versus gasoline) and vehicle specific emissions factors. Similar to the former AirGIS system, long-term variation of emission factors is handled by the COPERT-IV model (<http://emisia.com/products/copert>). Moreover, detailed information on changes in traffic volume or

pattern is required to assess traffic emissions behaviour at a particular location. However, lack of such information might result in significant deviation between modelled and measured concentrations considering single years. In this connection, more accurate traffic datasets based on systematic traffic counts, have only been available since 2007 (Ellermann et al., 2018).

At all stations there seems to be a good match between measured and modelled values of NO_x while sometimes new model overestimated modelled values (e.g. for 1996 at ODAL) and underestimated significantly at AARH and AALB stations (2009 – 2015). These discrepancies may be related to uncertainties in the historic traffic and emissions datasets. As stated above, for the most recent years, more reliable traffic and emissions datasets are available, as compared to more uncertain historic traffic and emissions data. Therefore, historic traffic counts should be integrated into the new model system when they are available. Correlation coefficients between 22 years annual average modelled and measured values at these stations are found to be 0.95, 0.93, 0.96 and 0.87 respectively. These high correlations indicate that new AirGIS reproduced well the long-term temporal trends of modelled NO_x.

For NO₂ the observed concentrations have been decreasing on all stations, over the past two decades (Figure 5). In general, the long-term temporal NO₂ trends are also well reproduced by the new modelling system over a significant period of time (1994 – 2015). The new model, however, sometimes overestimates and underestimates observed NO₂ concentrations. In particular, new AirGIS overestimates the observed NO₂ values at JGTV, ODAL and AALB stations for the years 1995 – 2005. Especially for the streets in Odense (ODAL) and Aalborg (AALB) the new model output seems to be shifted by 10 – 13 µg/m³ and 14 – 18 µg/m³ respectively towards higher values. This significant discrepancy, likewise NO_x, may also be due to the uncertainties in historic traffic and emissions inventory data. Furthermore, uncertainties in our model parameters, (e.g. the NO-NO₂ reaction rate chemistry), the fraction of direct NO₂ emissions may also be the possible reasons for over-predicted values. While, for the years 2010 – 2015, there is a good match between measured and observed NO₂ values at all stations. In addition, a high correlation was found between 22 years annual average modelled and measured NO₂ concentrations at JGTV, ODAL and AARH stations with correlation coefficients of 0.83, 0.89 and 0.96. While, for AALB station, the correlation coefficient was found to be moderate i.e. 0.45, caused by the discrepancy before 2005. After 2005, the agreement at AALB is very good.

481
482 Figure 5: Annual averages trends for observed and modelled NO_x (left column) and NO₂ (right column) at four
483 urban street stations (measurements are part of the Danish air quality monitoring network; Ellermann et al., 2017),
484 dataset 1.

485 As noted above, in the previous validation study (Ketzel et al., 2011), the AirGIS system was not
486 evaluated for PM₁₀ and PM_{2.5}. The performance of the new modelling system in reproducing long-term
487 trends (annual averages) of PM₁₀ and PM_{2.5} is shown in Figure 6, however only for JGTV due to data
488 availability. The new model slightly under-predicted observed PM₁₀ values for all the years with some

exceptions. Possible reasons for these PM_{10} under-predictions can be uncertainties in traffic and emissions inventory data and inaccurate street configuration information. While, in terms of $PM_{2.5}$ concentrations, the new model under-predicted observed value for all years except 2012 and 2013. Despite these underestimations, the correlation coefficients between measured and modelled PM_{10} and $PM_{2.5}$ values were found to be 0.86 and 0.91, respectively. Thus, new AirGIS reproduced long-term trends of PM_{10} and $PM_{2.5}$ in a good agreement with the observed values at JGTV station.

Figure 6: Annual averages trends for the observed and modelled PM_{10} (left column) and $PM_{2.5}$ (right column, limited comparison due to the lack of measured data) at Jagtvej (JGTV) street station, dataset 1. Model results include calibration see chapter 2.5.

3.1.2 Daily averages

For epidemiological research investigating health effects due to the short-term air pollution exposure (e.g. Chen et al., 2018), the daily pollutants concentrations are used. Therefore the following subsections demonstrate the performance of the new model in reproducing the short-term temporal trends (daily averages) of various air pollutants. As example, time periods for a recent year, 2015 and more historic years 2005 and 2006 were selected.

Year 2015

This section presents the results of new AirGIS evaluation against measured dataset 1 with a set of 4-months daily average air pollution levels during 1 January – 30 April 2015 for NO_x and NO_2 . In terms of PM_{10} and $PM_{2.5}$, however, the new model evaluation is presented during 1 April – 31 July 2015 due to data availability.

Table 4 presents the validation statistics on the daily averages (year 2015) of measured and predicted air pollution. Figure 7 (left panel) shows the day-to-day variation of modelled and observed NO_x ($\mu g/m^3$) at JGTV, ODAL and AARH. Data at AALB was not available. At all stations, in general, there seems to be a very good match between daily averages of modelled and measured NO_x (Figure 7, left panel). This is reflected by high positive correlation coefficients i.e. 0.88 (at JGTV), 0.87 (at ODAL) and 0.79 (at AARH), respectively (Table 4). The new modelling system, however, sometimes over- and under-estimated the observed NO_x levels. At JGTV and AARH stations, for example, the under-estimations were in the range 10% – 27% (RMSE: 26.3 – 27.5 $\mu g/m^3$) (Table 4).

In terms of NO₂ (Figure 7, right panel), a similar kind of good agreement between observed and modelled values (µg/m³) was observed. The correlation coefficients were found to be 0.84 (at JGTV), 0.81 (at ODAL) and 0.80 (at AARH), respectively (Table 4). In general, the new AirGIS reproduced well the daily averages trends of modelled NO₂. Nevertheless, in a few cases under- and over-estimations of observed NO₂ could also be seen. At all stations, RMSE was in the range 7.6 – 9.3 µg/m³ (Table 4). These deviations are most likely related to uncertainty in the modelled day-to-day variation of local meteorology (wind speed and direction, humidity, temperature etc.). In addition, uncertainties in the representativeness of the urban background contributions may also be related to these discrepancies in the new model output.

Figure 7: Daily averages trends for the observed and modelled NO_x (left column) and NO₂ (right column) at four urban street stations, dataset 1, during 1 January – 30 April 2015.

The data relating to the measured and modelled PM₁₀ and PM_{2.5} are shown in Figure 8 for the street station in Copenhagen (JGTV). Clearly, a very good agreement between daily averages (µg/m³) of modelled and measured PM₁₀ as well as PM_{2.5} can be seen (Figure 8). This is depicted by high positive correlation coefficients i.e. 0.84 (PM₁₀) and 0.88 (PM_{2.5}), respectively (RMSE: 2.9 – 4.2 µg/m³) (Table 4).

Figure 8: Daily averages trends (during 1 April – 31 July 2015) for the observed and modelled PM₁₀ (left column) and PM_{2.5} (right column) at Jagtvej (JGTV) street station, dataset 1. Model results include calibration see chapter 2.5.

Table 4: Validation statistics for observed (obs.) versus modelled (mod.) concentrations (daily averages) of various air pollutants (NO_x, NO₂, PM_{2.5}, PM₁₀) for dataset 1 in the year 2015. Av = average (µg/m³), FAC2= factor of two statistic, NMB = normalized mean bias, RMSE = root mean squared error (µg/m³), R = Pearson's correlation coefficient.

Oxides of Nitrogen											
Station	Method	NO _x					NO ₂				
		Av	NMB	FAC2	RMSE	R	Av	NMB	FAC2	RMSE	R
JGTV	obs.	70.3					31.9				
	mod.	52	-0.27	0.97	27.5	0.88	30.3	-0.05	0.99	7.6	0.84
ODAL	obs.	31					17.9				
	mod.	33.2	0.08	0.87	13.7	0.87	19.3	0.08	0.90	7.9	0.81
AARH	obs.	67.4					30.8				
	mod.	60	-0.10	0.94	26.3	0.79	30.3	-0.05	0.98	9.3	0.80
Particulate matter											
Station	Method	PM ₁₀					PM _{2.5}				
		Av	NMB	FAC2	RMSE	R	Av	NMB	FAC2	RMSE	R
JGTV	obs.	20.1					11.4				
	mod.	19.5	-0.04	0.99	4.2	0.84	11.1	-0.02	0.98	2.9	0.88

Years 2005 and 2006

This section presents the results of new AirGIS evaluation against measured dataset 1 with a set of 4-months daily average air pollution levels during 1 January – 30 April 2006 (NO_x, NO₂), 1 April – 31 July 2005 (PM₁₀) and 1 April – 31 July 2006 (PM_{2.5}).

Table 5 presents the validation statistics on the daily averages (years 2005 and 2006) of measured and predicted air pollution. Figure 9 (left panel) shows the observed and modelled NO_x values (µg/m³) as daily averages, in the period 1 January – 30 April 2006 at JGTV, ODAL, AARH and AALB. The agreement between measured and predicted NO_x (µg/m³) seems to be quite good at street stations in Copenhagen

(JGTV) and Odense (ODAL) with correlation coefficients of 0.76 and 0.87, respectively (Table 5). While, for the streets in Aalborg (AALB) and Aarhus (AARH), moderate (0.53) to slightly higher correlation (0.67) (Table 5) was observed. Discrepancies in the new model output, in terms of over- and under-predictions, could also be observed particularly at JGTV. For example, new AirGIS systematically under-predicted the observed NO_x at JGTV street station during 15 – 21 February and 20 – 27 April 2006. Moreover, significant over-estimations (13%) of observed NO_x values were found in the new model output at AARH station (see NMB in Table 5). At AALB, there seems to be a significant discrepancy in the modelled NO_x (FAC2 = 0.64), and the new system under-estimated the observed NO_x by 7%. One of the possible reasons of these under- and over-estimations may be related to the uncertainty in the predicted meteorology. Furthermore, it was noted above (see section 3.1.1) that more precise traffic datasets based on systematic traffic counts, have only been available since 2007 (Ellermann et al., 2018). Thus, uncertainty in traffic emissions can also be related to these significant errors in new modelling system output. Similar uncertainties apply for the further pollutants and will not be repeated.

For the case of NO₂ (Figure 9, right panel), correlation between modelled and measured values ($\mu\text{g}/\text{m}^3$) seems to be good (0.83) at street station in Odense (ODAL) (Table 5). For the street stations in Copenhagen (JGTV) and Aarhus (AARH), correlation was found to be slightly higher i.e. 0.67 and 0.63. While, moderate correlation (0.59) between the measured and modelled NO₂ was observed at AALB street station. RMSE, at all stations, was in the range 10.4 – 26.3 $\mu\text{g}/\text{m}^3$ (Table 5).

Figure 9: Daily averages trends for observed and modelled NO_x (left column) and NO₂ (right column) at four urban street stations, dataset 1, during 1 January – 30 April 2006.

Figure 10 shows the observed and modelled daily averages trends of PM₁₀ (during 1 April – 31 July 2005) and PM_{2.5} (during 1 April – 31 July 2006) at Jagtvej (JGTV) street station. In general, there is a good

agreement between the measured and modelled PM_{10} values ($\mu\text{g}/\text{m}^3$), which is reflected by high positive correlation coefficient $r=0.79$ (Table 5). In terms of $\text{PM}_{2.5}$, a moderate correlation i.e. $r=0.53$ was found between the measured and modelled values ($\mu\text{g}/\text{m}^3$). Significant over- and under-estimations can clearly be seen (Figure 10) (Table 5) especially for the case of $\text{PM}_{2.5}$, where NMB was found to be 0.31. These discrepancies may be related to the uncertainties in the predicted meteorology and other factors, as highlighted above. In general, these underlying uncertainties in the model (Figures 9, 10 and Table 5) are a bit higher for older data (before ~2010) compared to more recent years. This might lead to slightly higher uncertainty in the estimated historic air pollution exposure.

Figure 10: Daily averages trends for the observed and modelled PM_{10} (left column) (during 1 April – 31 July 2005) and $\text{PM}_{2.5}$ (right column) (during 1 April – 31 July 2006) at Jagtvej (JGTV) street station, dataset 1. Model results include calibration see chapter 2.5.

Table 5: Validation statistics for observed (obs.) versus modelled (mod.) concentrations (daily averages) of various air pollutants (NO_x , NO_2 , $\text{PM}_{2.5}$, PM_{10}) for dataset 1 in the years 2005 and 2006. Av = average ($\mu\text{g}/\text{m}^3$), FAC2= factor of two statistic, NMB = normalized mean bias, RMSE = root mean squared error ($\mu\text{g}/\text{m}^3$), R = Pearson's correlation coefficient.

Oxides of Nitrogen											
Station	Method	NO_x					NO_2				
		Av	NMB	FAC2	RMSE	R	Av	NMB	FAC2	RMSE	R
JGTV	obs.	129.9					57.1				
	mod.	104.9	-0.19	0.98	41.6	0.76	51.7	-0.09	0.99	14.1	0.67
ODAL	obs.	87.4					36.9				
	mod.	75.5	-0.13	0.91	30.2	0.87	37.2	0.02	0.98	10.4	0.83
AARH	obs.	120.3					49.6				
	mod.	134.8	0.13	0.91	52.8	0.67	56.8	0.16	0.96	17.9	0.63
AALB	obs.	134.6					47.5				
	mod.	125	-0.07	0.64	117.9	0.53	49.5	0.05	0.90	26.3	0.59
Particulate matter											
Station	Method	PM_{10}					$\text{PM}_{2.5}$				
		Av	NMB	FAC2	RMSE	R	Av	NMB	FAC2	RMSE	R
JGTV	obs.	22.4					13.5				
	mod.	26.6	0.19	0.99	7.8	0.79	17.8	0.31	0.94	7	0.53

3.2 Model evaluation against measured dataset 2

This section presents the results of new AirGIS evaluation against the short-term measured dataset 2 as part of the Danish measurements campaign in Copenhagen, Denmark (ESCAPE-EU project) conducted from November 2009 to October 2010 for various pollutants (NO_x, NO₂, PM₁₀, PM_{2.5}). We present the performance evaluation of the new modelling system in terms of reproducing both temporal and spatial variation of the observed values.

In Table 6, we present the descriptive statistics on observed and modelled values of NO_x, NO₂, PM₁₀ and PM_{2.5}. Figure 11 (a, b) shows the scatterplots of modelled NO_x and NO₂ (µg/m³) in terms of reproducing temporal & spatial variation (3 sets of 14 days averages, N=123) of the observed values. There is a considerable scatter in the modelled values of both NO_x (Figure 11a) and NO₂ (Figure 11b) as compared to the measured ones. Furthermore, there seems to be a better agreement between measured and modelled values of NO₂ as compared to NO_x. The correlation coefficients between modelled and measured NO_x and NO₂ were found to be 0.74 and 0.73 (Table 6). Figure 12 (a, b) shows the scatterplots of modelled NO_x and NO₂ (in µg/m³) in terms of spatial variation (average concentrations per site, N=41) of the same observed values. Clearly, some scatter in the modelled values of NO_x (Figure 12a) and NO₂ (Figure 12b) and the deviation from one-to-one line can also be observed in this case. The correlation coefficients in terms of reproducing spatial variation of observed NO_x and NO₂ were 0.79 and 0.81, slightly higher than for the temporal & spatial variation.

Similarly, Figure 13 (a, b) shows the scatterplots of modelled PM₁₀ and PM_{2.5} (µg/m³) in terms of reproducing temporal & spatial variation (3 sets of 14 days averages, N=60) of the measured values. Whereas, Figure 14 (a, b) shows the scatterplots of same pollutants in terms of reproducing spatial variation (total average per site, N=20). The correlation coefficients between modelled and measured PM₁₀ and PM_{2.5} (µg/m³) (temporal & spatial variation) were found to be 0.74 and 0.80, respectively. In terms of reproducing the spatial variation of observed PM₁₀, PM_{2.5} values much lower correlations were observed for PM₁₀ (r=0.62) and PM_{2.5} (r=0.32). However, these correlations in terms of assessing spatial variation of modelled PM are based on only 20 data points and therefore relatively uncertain and very sensitive to single outliers.

For all compounds, the new model overestimates the concentrations compared to the observed values by 25% to 71% (see NMB in Table 6). For NO₂ and NO_x the overestimation is quite large which is in contrast to the overall comparison with dataset 1, where the agreement was much better. Cyrus et al., (2012) in their work relating to the results of ESCAPE study compared Harvard Impactors samplers (NO_x, NO₂) with the chemiluminescence method being the EU reference method. The authors reported generally no to modest underestimation. Another companion paper (Eeftens et al., 2012) focusing on PM observations revealed that the same kind of limited numbers of samplers were used to measure PM during ESCAPE study. Interestingly, in Eeftens's work we noticed no such comparison to the EU reference method in terms of the measured PM. However, Hvidtfeldt et al., (2018) in their work, compared the Harvard Impactors used in the ESCAPE study with a high-quality fixed site monitoring instrument, SM200 (OPSIS Sweden, 2018), comparable to the EU reference method (Ellermann et al., 2018). The comparison

revealed that the Harvard Impactors systematically measured lower values of PM (PM_{2.5}: 31% lower on average, PM₁₀: 25% lower on average).

In a sensitivity analysis (not included in the manuscript), we observed that the model performance does not depend on the type of ESCAPE sites (rural, urban, street) and AirGIS performs similar for all sites

Figure 11: Evaluation of the new AirGIS modelling system against measured dataset 2 (a) Modelled average concentrations of NOx (in $\mu\text{g}/\text{m}^3$) plotted against observed NOx concentrations ($N=123$) (b) Same plot for NO₂ (in $\mu\text{g}/\text{m}^3$) ($N=123$).

Figure 12: Evaluation of the new AirGIS modelling system against measured dataset 2 (a) Modelled average concentrations of NOx (in $\mu\text{g}/\text{m}^3$) (average per site) plotted against corresponding observed NOx concentrations ($N=41$) (b) Same plot for NO₂ ($\mu\text{g}/\text{m}^3$) ($N=41$).

Figure 13: Evaluation of new AirGIS against dataset 2 (a) Modelled average concentrations of PM_{10} (in $\mu\text{g}/\text{m}^3$) plotted against observed PM_{10} concentrations ($N=60$) (b) Same plot for $\text{PM}_{2.5}$ (in $\mu\text{g}/\text{m}^3$) ($N=60$).

Figure 14: Evaluation of new AirGIS against dataset 2 (a) Modelled average concentrations of PM_{10} (in $\mu\text{g}/\text{m}^3$) (average per site) plotted against observed PM_{10} concentrations ($N=20$) (b) Same plot for $\text{PM}_{2.5}$ (in $\mu\text{g}/\text{m}^3$) ($N=20$).

Table 6: Validation statistics of observed (obs.) versus modelled (mod.) values of air pollutants: NO_x , NO_2 ($N=41$, 123) and $\text{PM}_{2.5}$, PM_{10} ($N=20$, 60) for dataset 2. $N=123$, 60 represents the mix of spatial and temporal variation while $N=41$, 20 shows the spatial variation of modelled air pollution. Av = average ($\mu\text{g}/\text{m}^3$), Min = minimum ($\mu\text{g}/\text{m}^3$), Max = maximum ($\mu\text{g}/\text{m}^3$), MB = mean bias ($\mu\text{g}/\text{m}^3$), FAC2= factor of two statistic, NMB = normalized mean bias, RMSE = root mean squared error ($\mu\text{g}/\text{m}^3$), r = Pearson's correlation coefficient.

Pollutant	Method	Both					N = 123, 60			N = 41, 20		
		Av	Min	Max	MB	NMB	RMSE	FAC2	r	RMSE	FAC2	r
NO_x	obs.	28.1	3.1	143.8								
	mod.	48.3	11.7	152.2	19.7	0.71	29.1	0.60	0.74	28.1	0.61	0.79
NO_2	obs.	16.4	1.7	54.8								
	mod.	28.1	9.5	58.4	11.3	0.70	14.1	0.62	0.73	13.1	0.58	0.81
PM_{10}	obs.	17.8	9.9	29.7								
	mod.	22.4	12.3	37.9	4.6	0.25	6.3	0.99	0.74	5.9	0.99	0.62
$\text{PM}_{2.5}$	obs.	11.3	4.0	21.3								
	mod.	14.8	7.8	23.6	3.6	0.31	4.6	0.93	0.80	4.2	0.99	0.32

3.3 Model evaluation against measured dataset 3

This section presents the results of new AirGIS evaluation (spatial variation) against measured dataset 3 (The FORCE Technology, 2018) i.e. 5-weeks passive NO₂ measurements along ten busy streets in Copenhagen, Denmark in the period October 24th to November 28th, 2011. It should be noted that a similar kind of evaluation was performed by Ketzel et al., (2012) by making use of former AirGIS system thus a direct comparison between modelled NO₂ concentrations of old and the new AirGIS system and measured values is given and summarized below.

The comparison of modelled NO₂ concentrations of former and new AirGIS system has been shown with the measured values in Figure 15. The comparison for the measurements at one of the permanent monitoring stations i.e. H.C. Andersens Boulevard (1) and at Ågade shows very good agreement between the measured and modelled NO₂ concentrations of former AirGIS system. For the same streets i.e. at H.C. Andersens Boulevard (1) new model system, however, slightly over-predicted the observed value. Also, the predicted NO₂ concentration of new AirGIS at H.C. Andersens Boulevard (1) was found to be higher than the former AirGIS predicted value. It should be noted that due to a change in street layout traffic has moved close to the station H.C. Andersens Boulevard (1). Thus, based on parallel measurements this rearrangement is estimated to have led to a jump of 8 µg/m³ in the concentration of NO₂ (Ellermann et al., 2016). AirGIS calculations are more representative of the measurements without the jump as they reflect pollution levels close the building façade. If the measurements are corrected for the jump, then there is a somewhat higher overestimation for H.C. Andersens Boulevard (1).

Figure 15: Evaluation of the new AirGIS modelling system against dataset 3: NO₂ modelled concentrations (old and new system) plotted against corresponding measured values. NO₂ concentrations are at selected streets in the period October 24, 2011 to November 28, 2011. Since no measurements exist for Nørre Søgade, only modelled NO₂ value has been shown.

At Ågade, a large discrepancy was observed in the predicted NO₂ concentration by the new modelling system. The predicted NO₂ concentration of new AirGIS (59.2 µg/m³) at Ågade was significantly higher than both former AirGIS predicted concentration (48.5 µg/m³) and the observed value (48 µg/m³). In some cases i.e. Sydhavnsgade and Fredensgade new AirGIS overestimated NO₂ concentrations significantly as compared to the measured values. While, the former AirGIS system under-predicted NO₂ concentration for both streets i.e., Sydhavnsgade and Fredensgade. At H.C. Andersens Boulevard (3), both modelling systems over-predicted NO₂ concentrations. Likewise, for the case of Jagtvej (1) new modelling system under-predicted NO₂ concentrations while there is a good match between the measured NO₂ concentration and new model predicted NO₂ value at Nordre Fasanvej (5). In particular, the largest discrepancy between measurements and new model results was found for Lyngbyvej (2). Interestingly, the largest discrepancy in former AirGIS output was also found at the same street section i.e. Lyngbyvej (2). Thus, it may be inferred that the two systems more or less behaves the same way in estimating modelled concentrations. The street sections Sydhavnsgade, Ågade and Fredensgade have one-sided building facades. Thus, most likely, the new model performs as good at these locations as on “full” street canyons. The presence of one-sided building facades might be one of the reasons of the discrepancy in the new model’s output. The comparison of street configuration generated by the old and new systems also revealed few differences in the estimated street-width. That is, the new system while generating street configuration information, found different street to estimate its width (see *Appendix A, supplementary material* for further details). In addition to this, there might exist some inaccuracies in our input datasets i.e. traffic volume and vehicle composition etc. Furthermore, unaccounted changes in traffic patterns in emissions and traffic datasets as well as new UBM/DEHM model runs may also be concluded to the explanations why new modelling system could not reproduce the observed NO₂ concentrations in 2011. Thus, re-evaluation of the model input, and street geometry of these street sections should be performed in future and may potentially help to improve new model estimates.

In Table 7, we present descriptive statistics (NMB, RMSE, FAC2 etc.) on the modelled and measured concentrations of NO₂ (µg/m³). The Pearson’s correlation coefficient (*r*) between modelled NO₂ (new AirGIS) and measured concentrations was found to be 0.76. While, mean error (ME), RMSE, NMB and FAC2 were computed as 7.5 µg/m³, 9.2 µg/m³, 0.13 and 0.99, respectively. In terms of former AirGIS and measured values, correlation coefficient was found to be 0.78. While, there was lower bias in the modelled NO₂ as compared to the new system i.e. mean error (ME) = 4.6 µg/m³ and RMSE = 6.7 µg/m³. In terms of the comparison of old and new system (modelled NO₂), correlation coefficient was found to be 0.80. It should be noted that the computed correlations are based upon *n*=10 observations and thus of the limited value. However, based on above, it may be established that the new model system reproduce similar kind of air pollution estimates as compared to the former AirGIS. In sum, new AirGIS showed good performance in reproducing the spatial variation of NO₂ values against measured dataset 3 at selected streets in Copenhagen, Denmark though the absolute concentrations calculated with the new version were not quite as good as the old version.

Table 7: Descriptive statistics on observed and modelled NO₂ values (in µg/m³) (old and new AirGIS modelling systems) for dataset 3, NO₂ 2011 measurement campaign along ten busy streets in Copenhagen, Denmark. *r* =

Pearson's correlation coefficient, FAC2 = factor of two statistic (FAC2), NMB = normalized mean bias, ME = mean error ($\mu\text{g}/\text{m}^3$), RMSE = root mean-squared error ($\mu\text{g}/\text{m}^3$).

NO_2 ($\mu\text{g}/\text{m}^3$)	R	R-squared	FAC2	NMB	ME	RMSE
New system (mod.) and Observed	0.76	0.58	0.99	0.13	7.5	9.2
Old system (mod.) and Observed	0.78	0.61	0.99	0.06	4.6	6.7
Old and new system (mod.)	0.80	0.65	0.99	0.04	5.7	6.6

3.4 Model evaluation against measured dataset 4

This section presents the results of new AirGIS evaluation against measured dataset 4 i.e. one-month NO_2 measurement campaign at the 204 addresses in Copenhagen and its outskirts, Denmark in October – November 1994 and in April – June 1995 (the Childhood Cancer project, Raaschou-Nielsen et al., 2000). Similar to section 3.2, the performance of new modelling system in terms of reproducing temporal and spatial variation of the observed pollutant has been assessed, here NO_2 ($\mu\text{g}/\text{m}^3$).

Figure 16 shows the scatterplot of 1224 modelled NO_2 concentrations ($\mu\text{g}/\text{m}^3$) against their measured ones (six observed values for 204 measurement sites – temporal and spatial variation). Although, a very high positive correlation i.e. 0.89 shows a very good agreement between measured and modelled NO_2 values, there is a significant overestimation by the new model system – especially for the lower concentrations. Similar kind of overestimations can be observed in Figure 17 where modelled and measured NO_2 concentrations ($\mu\text{g}/\text{m}^3$) have been averaged per measurement site (spatial variation). The correlation coefficient, in this case (Figure 17), was found to be 0.92. Clearly, the modelled concentrations of NO_2 seem to be considerably scattered in both cases (Figures 16, 17) i.e. reproducing temporal and spatial variation of observed concentrations. In addition, there seems to be a large number of cases where the model gave large deviations from the measured values (points residing far away from the 1:1 line). For $N=1224$, the root-mean squared error (RMSE) and factor of two statistic (FAC2) were computed as $16.5 \mu\text{g}/\text{m}^3$ and 0.58, respectively. For the case of $N=204$, RMSE and FAC2 were $15.9 \mu\text{g}/\text{m}^3$ and 0.60. Moreover, for both cases ($N=1224$ and 204), the mean bias (MB) was found to be same i.e. $14.9 \mu\text{g}/\text{m}^3$. All these point towards a large bias and existing uncertainties i.e. over- and under-estimations in the new model output, which needs to be further investigated. One of the possible reasons for this large bias in new AirGIS output can be uncertainties in our traffic emissions database or an issue with too low mixing heights (H_{mix}) in the UBM model in regions close to the water as it is the case for many points especially in dataset 4. The UBM model is currently under development to address this issue. While evaluating former AirGIS system against the same measured dataset, Ketzel et al., (2011) pointed out towards possible differences in selection of receptor points' locations near street intersections. This issue should also be studied for the new model.

746
 747 Figure 16: Evaluation of new AirGIS against measured dataset 4: NO₂ modelled concentrations ($\mu\text{g}/\text{m}^3$) plotted
 748 against the measured ones ($\mu\text{g}/\text{m}^3$) at the 204 addresses in Copenhagen and its surroundings. At each location, 6
 749 monthly periods were available (N=1224).

750
 751 Figure 17: Evaluation of new AirGIS against measured dataset 4: NO₂ modelled concentrations ($\mu\text{g}/\text{m}^3$) (average
 752 per site) plotted against the measured ones ($\mu\text{g}/\text{m}^3$) (average per site) at the 204 addresses in Copenhagen and its
 753 surroundings (N=204).

754 3.5 Summary for all four datasets

In summary, the above analyses revealed that, the new model's performance varies across the four datasets. The new AirGIS performed better against measured dataset 1 (street stations, national monitoring network) as compared to datasets 2 – 4 (showing sub-city level variations). This might be due to the underlying uncertainties in the measured datasets 2 – 4, all using substantially simpler instruments e.g. passive sampling. The Danish air quality monitoring network that dates back to 1980s and is operated according to the EU directives on air quality (Ellermann et al., 2018). Thus, more reliable measurements have been available to evaluate the model performance, and might explain the better performance of new AirGIS against dataset 1. Nevertheless, possible uncertainties in the new model system in terms of inaccurate in traffic flow, speed, composition and emission factors or errors in representativeness of urban background etc. can also be related to the varying performance of the new model across the datasets.

4. Strengths, limitations and future work

The new AirGIS modelling system has a number of strengths. First, the implementation of the new system in PostgreSQL/PostGIS offers excellent efficiency, in terms of processing time for a large spatial dataset with limited computing resources, as compared to the former AirGIS system. For example, as a random test, the former AirGIS system took 1962 seconds (32.7 minutes) to process 1000 address points on a 2.0 GHz, 8 GB RAM, 64 bit, Intel Xeon(R) based machine. While, new AirGIS took only 24.2 seconds to process the same 1000 address points at the same machine. In another test, the former AirGIS system took 419 seconds (approx. 7 minutes) to process 87 address points while new AirGIS took only 4.1 seconds. Thus, data processing capability of new AirGIS has been optimized and the new system is about 100 times faster as compared to the former AirGIS modelling system. It should be noted that the optimized processing speed of the new system merely reflects upon the time to generate input files for OSPM® as the computing time is the same for DEHM/UBM/OSPM®. Nevertheless, this is, still, very important when a health related air pollution exposure study is based on millions of address points in a region of interest. For such case, the new modelling system should be able to produce exposure estimates in a significantly reduced time due to highly optimized processing time of new AirGIS and batch processing capability of OSPM®. Secondly, the use of open-source GIS programming based environment in conjunction with R-scripts allows for easy model adjustments for future requirements as well as for further developments. Thirdly, new AirGIS has been developed so that it is transferable to other locations as long as there is sufficient detailed information on traffic volume, vehicle distribution and speeds with building footprints data as well as other required input data.

In addition to the over- and under-predictions of the observed values, our new model – however – does have some other limitations. First, the new model system – similar to the old system – only models traffic as a local source of air pollution with high spatial resolution while other sources e.g. wood burning, chimneys (point sources) - even when known with spatially correct resolution - are taken into account only in the background contributions in a resolution of 1km x 1km. Second, as noted in *Appendix A, supplementary material*, the new system, sometimes, finds different street to estimate its width. This, consequently, leads to the incorrect estimation of street width when compared to the old system. Also, new AirGIS sometimes overestimates and underestimates street width specifically when buildings are not parallel to the street (case of asymmetric geometric features) which was, however,

also the case of the former version. Third, in the moment it was not possible to implement inhomogeneous emissions for OSPM® as developed by Ottosen et al. (2016) in the new model due to the lack of information about how traffic volume is distributed on different lanes of streets.

Furthermore, the limitations of this study should also be considered. The new system's performance in reproducing the long-term trends of PM₁₀ and PM_{2.5} has so far only been assessed for one street in Copenhagen.

More work has to be done in the future (i) to evaluate the new modelling system for other pollutants (e.g. CO, O₃ etc.) (ii) to further test and improve the street configuration generation (e.g. street width estimation) of the new system (iii) to implement inhomogeneous emissions scenario once relevant datasets are available..

5. Conclusions

The GIS-based air pollution and human exposure modelling system, AirGIS, has been substantially revised and updated as a new modelling system for air pollution exposure assessment in Denmark. The new model is implemented in open-source GIS (Postgres/PostGIS) via R-scripts interface. Being in its evaluation phase, the performance of the new model system has been evaluated against four measured datasets of ambient air pollution (NO_x, NO₂, PM₁₀ and PM_{2.5}). In terms of reproducing temporal variation of measured air pollution levels (dataset 1, 2, 4), new AirGIS achieved correlations in the range 0.45 – 0.96. While, in terms of reproducing the spatial variation of measured air pollution levels (measured dataset 2, 3, 4), new AirGIS achieved correlations in the range 0.32 – 0.92. Thus, in sum, the new model showed good performance in terms of correlation coefficients against measured values and gave a good description of the spatial and temporal variation of ambient air pollution within a region of interest. The data processing capability of new AirGIS has also been optimized and new model is about 100 times faster as compared to the former AirGIS modelling system in terms of generation of street configuration data. Therefore, the new AirGIS is well suited for air pollution exposure modelling e.g. in connection to health related studies. However, the bias between measurements and model can occasionally be large (especially with dataset 2 and 4 where there is a significant overestimation by the new model) probably caused by measurement uncertainty combined with possible errors in the model input.

Acknowledgements

This publication was funded by the Graduate School of Science and Technology (GSST) (Project # 21315), Aarhus University, Denmark and a NPRP award (NPRP # 7-649-2-241) from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

- (NCAR), T. N. C. f. A. R. (2018). WEATHER RESEARCH AND FORECASTING MODEL. Retrieved from <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>
- Agency, T. D. G. (2006). *Kort10 - en del af den geografiske infrastruktur*. Retrieved from <https://gst.dk/media/gst/64666/kort10.pdf>

- Allwine, K. J., Rutz, F. C., Shaw, W. J., Rishel, J. P., Fritz, B. G., Chapman, E. G., . . . Seiple, T. E. (2006). *DUSTRAN 1.0 user's guide: A GIS-based atmospheric dust dispersion modeling system*. Retrieved from
- Andersen, Z. J., Ravnskjaer, L., Andersen, K. K., Loft, S., Brandt, J., Becker, T., . . . Brauner, E. V. (2016). Long-Term Exposure to Fine Particulate Matter and Breast Cancer Incidence in the Danish Nurse Cohort Study. *Cancer Epidemiology and Prevention Biomarkers*, cebp. 0578.2016.
- Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., . . . Marcon, A. (2013). Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe—the ESCAPE project. *Atmospheric Environment*, 72, 10-23.
- Beevers, S. D., Kitwiroon, N., Williams, M. L., Kelly, F. J., Anderson, H. R., & Carslaw, D. C. (2013). Air pollution dispersion models for human exposure predictions in London. *Journal of Exposure Science and Environmental Epidemiology*, 23(6), 647-653.
- Berkowicz, R. (2000). A simple model for urban background pollution. *Environmental Monitoring and Assessment*, 65(1), 259-267.
- Berkowicz, R. (2000a). OSPM—a parameterised street pollution model *Urban Air Quality: Measurement, Modelling and Management* (pp. 323-331): Springer.
- Brandt, J., Christensen, J. H., Frohn, L. M., Palmgren, F., Berkowicz, R., & Zlatev, Z. (2001). Operational air pollution forecasts from European to local scale. *Atmospheric Environment*, 35, S91-S98. doi:[http://dx.doi.org/10.1016/S1352-2310\(00\)00415-5](http://dx.doi.org/10.1016/S1352-2310(00)00415-5)
- Brandt, J., Silver, J. D., Frohn, L., Geels, C., Gross, A., Hansen, A. B., . . . Villadsen, H. (2012). An integrated model study for Europe and North America using the Danish Eulerian Hemispheric Model with focus on intercontinental transport of air pollution. *Atmospheric Environment*, 53, 156-176.
- Carruthers, D., Edmunds, H., Lester, A., McHugh, C., & Singles, R. (2000). Use and validation of ADMS-Urban in contrasting urban and industrial locations. *International Journal of Environment and Pollution*, 14(1-6), 364-374.
- Carslaw, D. (2015). The openair manual: open-source tools for analysing air pollution data. Retrieved from http://www.openair-project.org/PDF/OpenAir_Manual.pdf
- Carslaw, D. C., & Ropkins, K. (2012). openair — An R package for air quality data analysis. *Environmental Modelling & Software*, 27-28, 52-61. doi:<https://doi.org/10.1016/j.envsoft.2011.09.008>
- CERC. (2017). ADMS model. Retrieved from Cambridge Environmental Research Consultants, <http://www.cerc.co.uk/environmental-software/ADMS-model.html>
- Chen, C., Liu, C., Chen, R., Wang, W., Li, W., Kan, H., & Fu, C. (2018). Ambient air pollution and daily hospital admissions for mental disorders in Shanghai, China. *Science of the Total Environment*, 613, 324-330.
- Cyrus, J., Eeftens, M., Heinrich, J., Ampe, C., Armengaud, A., Beelen, R., . . . Hoek, G. (2012). Variation of NO₂ and NO_x concentrations between and within 36 European study areas: Results from the ESCAPE study. *Atmospheric Environment*, 62, 374-390. doi:<https://doi.org/10.1016/j.atmosenv.2012.07.080>
- de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M. J., . . . Cesaroni, G. (2014). Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environment international*, 73, 382-392.
- Duchenne, C., & Armand, P. (2017). *Development of a generic modelling chain for weather prediction and potential atmospheric dispersion from the meso-scale to local scale*. Paper presented at the 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO18), CNRS, Bologna, Italy. http://www.harmo.org/conferences/Proceedings/Bologna/Bologna_proceedings.asp
- Eeftens, M., Tsai, M.-Y., Ampe, C., Anwander, B., Beelen, R., Bellander, T., . . . Hoek, G. (2012). Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within

- 20 European study areas and the relationship with NO₂ – Results of the ESCAPE project. *Atmospheric Environment*, 62, 303-317. doi:<http://dx.doi.org/10.1016/j.atmosenv.2012.08.038>
- Ellermann, T., Nygaard, J., Nøjgaard, J.K., Nordstrøm, C., Brandt, J., Christensen, J., Ketzel, M., Massling, A. & Jensen, S.S. (2018). *The Danish Air Quality Monitoring Programme. Annual Summary for 2017*. Retrieved from <http://dce2.au.dk/pub/SR281.pdf>
- EPA, U. (2017). Air Quality Dispersion Modeling - Preferred and Recommended Models: AERMOD Modeling System. Retrieved from <https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models#aermod>
- Ferm, M., & Svanberg, P.-A. (1998). Cost-efficient techniques for urban- and background measurements of SO₂ and NO₂. *Atmospheric Environment*, 32(8), 1377-1381. doi:[http://dx.doi.org/10.1016/S1352-2310\(97\)00170-2](http://dx.doi.org/10.1016/S1352-2310(97)00170-2)
- Gaines Wilson, J., & Zawar-Reza, P. (2006). Intraurban-scale dispersion modelling of particulate matter concentrations: Applications for exposure estimates in cohort studies. *Atmospheric Environment*, 40(6), 1053-1063. doi:<http://dx.doi.org/10.1016/j.atmosenv.2005.11.026>
- Gulliver, J., & Briggs, D. (2011). STEMS-Air: A simple GIS-based air pollution dispersion model for city-wide exposure assessment. *Science of the Total Environment*, 409(12), 2419-2429. doi:<http://dx.doi.org/10.1016/j.scitotenv.2011.03.004>
- Gulliver, J., Morley, D., Vienneau, D., Fabbri, F., Bell, M., Goodman, P., . . . Fecht, D. (2015). Development of an open-source road traffic noise model for exposure assessment. *Environmental Modelling & Software*, 74, 183-193. doi:<http://dx.doi.org/10.1016/j.envsoft.2014.12.022>
- Hansen, H., Jensen, S., & Berkowicz, R. (1997). *Estimating Street Air Quality Using a 2½ Dimensional Urban Landscape Model*. Paper presented at the Den Store Nordiske Gis-konference.
- Hertel, O., Ellermann, T., Palmgren, F., Berkowicz, R., Løfstrøm, P., Frohn, L. M., . . . Christensen, J. (2007). Integrated air-quality monitoring—combined use of measurements and models in monitoring programmes. *Environmental Chemistry*, 4(2), 65-74.
- Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: a review. *Environmental Health*, 12(1), 43.
- Jensen, S. S., Berkowicz, R., Hansen, H. S., & Hertel, O. (2001). A Danish decision-support GIS tool for management of urban air quality and human exposures. *Transportation Research Part D: Transport and Environment*, 6(4), 229-241.
- Jensen SS, H. M., Pedersen J, Storm L, Stausgaard L, Becker T, Hertel O. (2009a). *GIS-baseret national vej- og trafikdatabase 1960-2005 - Faglig rapport fra DMU. Roskilde, Denmark: Afdelingen for Atmosfærisk Miljø, Danmarks Miljøundersøgelser*. Retrieved from <http://www2.dmu.dk/Pub/FR678.pdf>
- Jensen, S. S., Larson, T., Deepti, K. C., & Kaufman, J. D. (2009b). Modeling traffic air pollution in street canyons in New York City for intra-urban exposure assessment in the US Multi-Ethnic Study of atherosclerosis and air pollution. *Atmospheric Environment*, 43(30), 4544-4556. doi:<http://doi.org/10.1016/j.atmosenv.2009.06.042>
- Kakosimos, K. E., Hertel, O., Ketzel, M., & Berkowicz, R. (2010). Operational Street Pollution Model (OSPM)—a review of performed application and validation studies, and future prospects. *Environmental Chemistry*, 7(6), 485-503.
- Kesarkar, A. P., Dalvi, M., Kagainkar, A., & Ojha, A. (2007). Coupling of the Weather Research and Forecasting Model with AERMOD for pollutant dispersion modeling. A case study for PM₁₀ dispersion over Pune, India. *Atmospheric Environment*, 41(9), 1976-1988.
- Ketzel, M., Berkowicz, R., Hvidberg, M., Jensen, S. S., & Raaschou-Nielsen, O. (2011). Evaluation of AirGIS: a GIS-based air pollution and human exposure modelling system. *International Journal of Environment and Pollution*, 47(1-4), 226-238.

- Ketzel, M., Jensen, S. S., Brandt, J., Ellermann, T., Olesen, H. R., Berkowicz, R., & Hertel, O. (2012). Evaluation of the street pollution model OSPM for measurements at 12 streets stations using a newly developed and freely available evaluation tool. *Journal of Civil and Environmental Engineering*.
- Lefebvre, W., Maiheu, B., Hooyberghs, H., Janssen, S. & Blyth, L. . (2017). *Developing a method for high-resolution annual concentration assessment and scenario analysis for NO2 for the whole of europe*. Paper presented at the 18th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO18), CNRS, Bologna, Italy. <http://www.harmo.org/conferences/Proceedings/Bologna/publishedSections/H18-119-Lefebvre.pdf>
- Lefebvre, W., Van Poppel, M., Maiheu, B., Janssen, S., & Dons, E. (2013). Evaluation of the RIO-IFDM-street canyon model chain. *Atmospheric Environment*, 77, 325-337. doi:<https://doi.org/10.1016/j.atmosenv.2013.05.026>
- Lefebvre, W., Vercauteren, J., Schrooten, L., Janssen, S., Degraeuwe, B., Maenhaut, W., . . . Lefebvre, F. (2011). Validation of the MIMOSA-AURORA-IFDM model chain for policy support: Modeling concentrations of elemental carbon in Flanders. *Atmospheric Environment*, 45(37), 6705-6713. doi:<https://doi.org/10.1016/j.atmosenv.2011.08.033>
- Morra, P., Bagli, S., & Spadoni, G. (2006). The analysis of human health risk with a detailed procedure operating in a GIS environment. *Environment international*, 32(4), 444-454. doi:<http://dx.doi.org/10.1016/j.envint.2005.10.003>
- OP SIS AB (2018). Gas monitoring. Retrieved from <https://www.opsis.se/en/About-us>
- Plejdstrup, M. S. G., S. (2011). *Spatial distribution of emissions to air – the SPREAD model*. Retrieved from <http://www.dmu.dk/Pub/FR823.pdf>
- Puett, R. C., Hart, J. E., Yanosky, J. D., Spiegelman, D., Wang, M., Fisher, J. A., . . . Laden, F. (2014). Particulate matter air pollution exposure, distance to road, and incident lung cancer in the Nurses' Health Study cohort. *Environmental health perspectives*, 122(9), 926.
- Raaschou-Nielsen, O., Hertel, O., Vignati, E., Berkowicz, R., Jensen, S. S., Larsen, V. B., . . . OLSEN, J. H. (2000). An air pollution model for use in epidemiological studies: evaluation with measured levels of nitrogen dioxide and benzene. *Journal of Exposure Science and Environmental Epidemiology*, 10(1), 4.
- Shu, S., Yang, P., & Zhu, Y. (2014). Correlation of noise levels and particulate matter concentrations near two major freeways in Los Angeles, California. *Environmental Pollution*, 193, 130-137.
- Team, Q. D. (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from <https://qgis.org/en/site/>
- Team, T. R. C. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Vardoulakis, S., Fisher, B. E. A., Pericleous, K., & Gonzalez-Flesca, N. (2003). Modelling air quality in street canyons: a review. *Atmospheric Environment*, 37(2), 155-182. doi:[http://doi.org/10.1016/S1352-2310\(02\)00857-9](http://doi.org/10.1016/S1352-2310(02)00857-9)
- Vienneau, D., de Hoogh, K., & Briggs, D. (2009). A GIS-based method for modelling air pollution exposures across Europe. *Science of the Total Environment*, 408(2), 255-266. doi:<https://doi.org/10.1016/j.scitotenv.2009.09.048>
- Wikipedia. (2018). 2.5D.
- Xie, X., Huang, Z., & Wang, J.-s. (2005). Impact of building configuration on air quality in street canyon. *Atmospheric Environment*, 39(25), 4519-4530. doi:<http://dx.doi.org/10.1016/j.atmosenv.2005.03.043>

Research Highlights

- Development of an updated version of AirGIS air pollution and human exposure modelling system
- Overview of the new system's architecture and workflow
- Evaluation of the new model system against various measurements datasets
- Overview of the new system's strengths, limitations and future outlook