

Modelling of human exposure to air pollution in the urban environment: a GPS-based approach

Daniela Dias · Oxana Tchepel

Received: 24 January 2013 / Accepted: 24 October 2013 / Published online: 24 November 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The main objective of this work was the development of a new modelling tool for quantification of human exposure to traffic-related air pollution within distinct micro-environments by using a novel approach for trajectory analysis of the individuals. For this purpose, mobile phones with Global Positioning System technology have been used to collect daily trajectories of the individuals with higher temporal resolution and a trajectory data mining, and geo-spatial analysis algorithm was developed and implemented within a Geographical Information System to obtain time–activity patterns. These data were combined with air pollutant concentrations estimated for several microenvironments. In addition to outdoor, pollutant concentrations in distinct indoor microenvironments are characterised using a probabilistic approach. An example of the application for PM_{2.5} is presented and discussed. The results obtained for daily average individual exposure correspond to a mean value of 10.6 and 6.0–16.4 $\mu\text{g m}^{-3}$ in terms of 5th–95th percentiles. Analysis of the results shows that the use of point air quality measurements for exposure assessment will not explain the intra- and inter-variability of individuals' exposure levels. The methodology developed and implemented in this work provides time-sequence of the exposure events thus making possible association of the exposure with the individual activities and delivers main statistics on individual's air pollution exposure with high spatio-temporal resolution.

Keywords Exposure assessment · Urban air pollution · PM_{2.5} · Traffic related · GIS · Trajectory data mining

Introduction

Exposure to air pollution is estimated to cause 1.3 million deaths worldwide/year in urban areas, and emissions from road traffic account for a significant share of this burden (WHO 2011). Therefore, an accurate assessment of human exposure is crucial for a correct determination of the association between the traffic-related air pollutants and the negative health outcomes (Hertel et al. 2008).

Exposure estimates to atmospheric pollutants can address individuals (personal exposure) or large population groups (population exposure) and can be based on direct (exposure monitoring) or indirect methods (exposure modelling) (Zou et al. 2009). In practice, monitoring of personal exposure is limited to studies with a small number of individuals because of the high costs associated with the measurements. Similarly, air quality time series provided by a monitoring network are frequently used as a good individual exposure indicator. Nevertheless, this estimate has been found to correlate poorly with personal exposures (Kousa et al. 2002; Koistinen et al. 2001; Oglesby et al. 2000) as it does not capture spatial heterogeneity in exposure to air pollution, time spent indoors and human mobility (Koistinen et al. 2001) thus leading to inaccuracies and underestimation of the health effects of air pollution (Szpiro et al. 2008; Peng and Bell 2010). In addition, the presence of individuals in direct vicinity to the emission sources may result in higher exposure concentrations than pollution levels registered at monitoring stations (Baklanov et al. 2007).

Consequently, exposure modelling technique arises as an alternative approach able to address the spatial and temporal variability of individual exposure concentrations and is recommended for exposure assessment (WHO 2002; Brunekreef

Responsible editor: Michael Matthies

D. Dias (✉)
Centre for Environmental and Marine Studies and Department of
Environment and Planning, University of Aveiro, 3810-193 Aveiro,
Portugal
e-mail: danieladias@ua.pt

O. Tchepel
Department of Civil Engineering, FCTUC, University of Coimbra,
3070-788 Coimbra, Portugal

and Holgate 2002; Brauer et al. 2002). Over the past 20 years, several exposure modelling methods have been developed with the aim of estimating exposure at the individual level. The major purpose of these models is to characterise air quality concentrations to be used as surrogate of personal exposure to air pollution and assumes that subjects within a demographic area (e.g. census units) are equally exposed to air pollution. Thus, over the past decade, air quality models have been integrated with geographic information system (GIS) in an attempt to reflect individual exposure by combining air pollutants concentration data with residence location (e.g. Gauderman et al. 2007; Hoek et al. 2008). To overcome some of these issues, several personal exposure models based on a microenvironment approach, such as AirPex (Freijer et al. 1998), SHEDS-PM (Burke et al. 2001), HAPEM (Özkaynak et al. 2008) and APEX (US Environmental Protection Agency 2009) are available. These models are designed to simulate the distribution of personal exposure in several microenvironments (e.g. outdoors, indoor-residential, public buildings, and workplaces) (Burke et al. 2001), by combining the estimated pollutant concentrations at every microenvironment and the time spent at visited microenvironments based on participant's diary or time–activity measurement databases (Kruize et al. 2003; Georgopoulos et al. 2005; Özkaynak et al. 2008; US Environmental Protection Agency 2006; Health Effects Institute 2010; Klepeis et al. 2001). However, by using this time–activity location data, the information on the actual “activity space” of individuals required for high-resolution exposure modelling is rarely available (Health Effects Institute 2010) and individual air pollution exposure context can be assumed as a series of independent microenvironment exposures (Ballesta et al. 2008).

Recently, there has been an increasing focus on using global positioning system (GPS) technology to collect the individual trajectory information to be used in combination with air pollution levels to estimate personal air pollution exposure levels in urban areas (Jensen 2006; Gerharz et al. 2009, 2013). However, such models are still strongly dependent on questionnaires' information to derive individual activity profile, providing exposure estimates only if the individual resides in a microenvironment which is specified in the model (Gerharz et al. 2013) and generally only indoor and in-vehicle microenvironments are identified, ignoring exposure during walking periods.

In this perspective, despite several efforts have been expended on characterising the spatial and temporal distributions of air pollution, much work remains in understanding the role of individual mobility in conditioning exposures in urban areas. Therefore, what has risen from literature reviews is that the time-sequence of exposure events is not preserved in exposure assessment, and the information to evaluate possible correlations in exposures to different pollutants because activities that are related in time are not conserved. The source–receptor relationship, especially for “hot-spots” peak exposure

is still insufficiently addressed and the contribution of traffic-related air pollution to the total exposure is not clear (Health Effects Institute 2010; Wang et al. 2009). In addition, the development of innovative models that reduce uncertainties in exposure characterisation is required (Lioy 2010).

Recent findings highlight that the population mobility is one of the factors that may affect significantly the exposure (Nethery et al. 2008; Beckx et al. 2009; Dons et al. 2011; Tchepel and Dias 2011). In this sense, the knowledge of where individuals spend time is essential for assessment of human exposure to air pollution and research on human behaviour or activities is a crucial component of modern and future exposure science (Lioy 2010). The predictability in human dynamics by studying the mobility patterns of individuals using mobile phones became an emerging field (Song et al. 2010), and GPS technology presents as a promising tool by monitoring real-time geographic positions.

In this perspective, the GPS-based exposure model to traffic-related air pollution (ExPOSITION) is developed in the present work for quantification of human exposure to traffic-related air pollutants within distinct microenvironments using air pollution modelling with high spatial–temporal resolution and a novel approach based on trajectory analysis of individuals. Enhanced resources, such as GIS, GPS and data mining techniques are considered by the model to analyse the human behaviour and individual's activities required for exposure assessment. For this purpose, information on detailed time–location data are collected for each individual at each moment by GPS-equipped mobile phones, offering many advantages over traditional time–location analysis, such as high temporal resolution and minimum reporting burden for participants (Rainham et al. 2010).

The GPS technology guarantees that there will be an increasing availability of large amounts of data affecting individual trajectories, at increasing localisation precision. However, significant uncertainties associated with the processing and classifying of raw GPS data is one of challenging issue for the exposure studies (Wu et al. 2010; Zheng and Zhou 2011). Thus, a trajectory data mining and geo-spatial analysis algorithm were developed and implemented within GIS to process the trajectories obtained with GPS-equipped mobile phones to automatically identify time–activity location in several microenvironments, such as commuting, indoor, and outdoor locations. The time–activity patterns of each individual are finally combined with information on pollutant concentrations that the individuals experiences at different microenvironments. For outdoor environment, atmospheric dispersion modelling and different modelling tools may be used to provide this external information on ambient concentrations for ExPOSITION model. Indoor concentrations are estimated considering not only outdoor/indoor infiltration factor but also the additional contribution of indoor emissions sources. In addition, a probabilistic approach are considered by the ExPOSITION

model to characterise the variability of the microenvironmental parameters providing an additional knowledge on the variation associated with microenvironmental concentrations and its contribution to the individual exposure estimates. The development of the ExPOSITION model is presented and described.

Methodology—human exposure modelling

The ExPOSITION model is developed to assess average short (e.g. daily) and long-term (e.g. annual) inhalation exposures of the individuals to traffic-related air pollutants over urban spatial scale with high spatial-temporal resolution. For this purpose, air pollution concentrations are estimated for different microenvironments (described in “Microenvironmental concentrations”) and combined with detailed time-activity patterns obtained from data collected by mobile phones with GPS technology (described in “Trajectory data mining” and “Time-activity patterns”). The ExPOSITION modelling system developed and applied in this study is schematically presented in Fig. 1 and described in the following sections.

Personal exposure is characterised by ExPOSITION model in terms of time-weighted average exposure concentration calculated from air pollutant concentration fields and time spent by individuals in different microenvironments (Eq. 1). It is important to highlight the distinction between air pollution “concentration” provided by dispersion models, and “exposure concentration” defined as amount of chemicals that comes into contact with the human body and take into account not only pollutant concentration fields but also the location of an individual and duration of the exposure. Thus, individual exposure is calculated by ExPOSITION as following:

$$\bar{E}_i = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} C(x, y, z, t)_i dt \quad (1)$$

where ($\mu\text{g m}^{-3}$) is the average exposure concentration for person i , $C_i(x, y, z, t)$ ($\mu\text{g m}^{-3}$) is the air pollutant concentration occurring at a particular point where the person i is located during the time t and spatial coordinate (x , y , and z) and t_1 and t_2 (hours) are the starting and ending times of the exposure event.

Exposure estimates are provided in $\mu\text{g m}^{-3}$ and can be determined for each individual as hourly, daily or annual average and resulting data can be exported for further analysis (e.g. epidemiological analysis and health impact assessment).

Microenvironmental concentrations

Specific microenvironments are distinguished in the exposure model including residence, other indoors, outdoors and in-

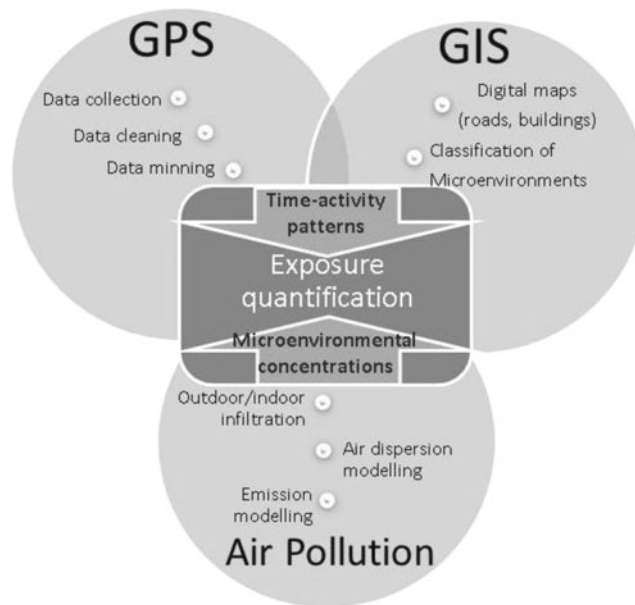


Fig. 1 Conceptual framework of the ExPOSITION modelling system

vehicle (Table 1). Two different approaches are considered to characterise pollution levels in these microenvironments. Thus, outdoor concentrations are estimated using atmospheric dispersion modelling and different modelling tools may be used to provide this external information for ExPOSITION as will be discussed in “Emission and air quality modelling”. For indoors and in-vehicle microenvironments a probabilistic approach was implemented as an integrated part of ExPOSITION algorithm. In this case, it is assumed that within a microenvironment the pollutants are homogeneously distributed and microenvironmental concentration $C(x, y, z, t)$ ($\mu\text{g m}^{-3}$) considered in Eq. 1 is calculated using a linear regression equation based on the outdoor/indoor infiltration factor α_j (dimensionless) and additional contribution of indoor pollution sources expressed as β_j ($\mu\text{g m}^{-3}$):

$$C(x, y, z, t) = \beta_j + \alpha_j \times C(x, y, z, t)_{\text{ambient}} \quad (2)$$

where $C(x, y, z, t)_{\text{ambient}}$ ($\mu\text{g m}^{-3}$) is the outdoor concentration that occurring in the immediate vicinity to the microenvironment j at time t and spatial coordinate (x, y, z).

Microenvironmental concentrations are estimated based on a probabilistic approach considered by the model that attempts to capture the variability in microenvironment parameters. In this sense, to calculate microenvironmental concentrations for each individual the ExPOSITION model randomly assigns the parameters β and α to each indoor location from empirical distributions taking into account the average and standard deviation obtained from literature review for each type of microenvironments (Table 1).

A single value is selected from the probabilistic distribution of each microenvironmental parameter α and β . These values

Table 1 Parameters used to determine PM2.5 concentrations in different microenvironments

Microenvironment	β		α		Data source
	Average	Standard deviation	Average	Standard deviation	
Residence	5.75	3.91	0.41	0.06	Hoek et al. (2008)
Vehicle (no smoking)	33.0	7.2	0.26	0.14	Burke et al. (2001)
Office (no smoking)	3.6	1.3	0.18	0.06	
School	6.8	1.4	0.60	0.09	
Public access	9.0	3.6	0.74	0.18	
Restaurant/bar	9.8	0.5	1.00	0.05	

are then used in the model to produce a single estimate of microenvironmental concentration. This process is repeated many times, with new values for each stochastic input parameter and probability distribution of exposure microenvironments is obtained.

Trajectory data mining

Trajectories of the individuals are required as one of the main inputs to the exposure modelling. The time–activity patterns are determined by the model based on an innovative approach developed for collection and analysis of data registered by mobile phones with GPS technology and thus providing the precise location and time spent at different microenvironments during the daily trajectories of individuals required for the exposure assessment.

Collection of time–location information using GPS technology provides continuous tracking of the individuals with high data resolution in time and in space. However, to overcome some of the significant uncertainties associated with the processing and classifying of raw GPS, automatic processing of GPS raw data using the trajectory data mining is implemented in ExPOSITION model.

To identify important patterns, several levels of GPS data processing are required (Fig. 2a). First, it is necessary to “clean” the GPS raw data to eliminate invalid entries. Subsequently, the places where the individual was stopped for a certain time period are distinguished from moving activities, like driving a vehicle. Finally, it is necessary to discover which of these points belong to the same activity/place. For this purpose, the data clustering process is implemented to distinguish significant places based on the analysis of spatial and temporal information of GPS points (Fig. 2a).

Thus, “significant places” are considered as those locations that play significant role in the activities of a person, carrying a particular semantic meaning such as the living and working places, the restaurant and shopping mall, etc., ignoring the transition between these places. Additionally, a “movement

activity” is a composition of movements with a frequent regularity of location change over time which can be aggregated by the purpose of the trip of an individual.

A preliminary processing of GPS data is implemented as a first step to “clean” the data and converts it into a standard format in preparation for the clustering approach. For this purpose, an error-checking algorithm was developed to remove invalid points. This algorithm considers a measurement as valid, if the GPS receiver is able to see at least four satellites and if the horizontal dilution of precision value is below 6 (Fig. 2b). Otherwise the measurement is considered invalid. In addition, the algorithm evaluates incorrect entries of the travel speed.

GPS datasets provide information on the locations in coordinate form (e.g. latitude and longitude) but contains no semantic meaning (Zhou et al. 2007a) like the address or characteristics of location, i.e. type of microenvironments. Therefore, it is necessary to extract and distinguish in the GPS data the locations where the individual stopped for a certain time period and these locations are designated as “stay points”. A stay point represents a geographic location in which the individual stays for a certain time period and in addition to a raw GPS point carries a particular semantic meaning.

The algorithm to extract stay points from GPS data is iterative and it is based on searching for locations where the user has spent a longer time period (Li et al. 2008). As presented in Fig. 2b, the extraction of a stay point S from a user’s GPS trajectory $P = \{p_1, p_2, \dots, p_K\}$, depends on two scale parameters: a distance threshold (D_{threh}) and a time threshold (T_{threh}). Thus, a single stay point S can be characterised by a group of consecutive GPS points p_i containing latitude (p_i .Lat), longitude (p_i .Long) and time (p_i .T):

$$S = \{p_i\}, \text{ where } m \leq i \leq n, \\ \text{Distance}(p_m, p_i) \leq D_{\text{threh}} \text{ and} \\ |p_n.T - p_m.T| \geq T_{\text{threh}}$$

A pre-processing of the GPS data and detection of the stay points is important to extract some important locations.

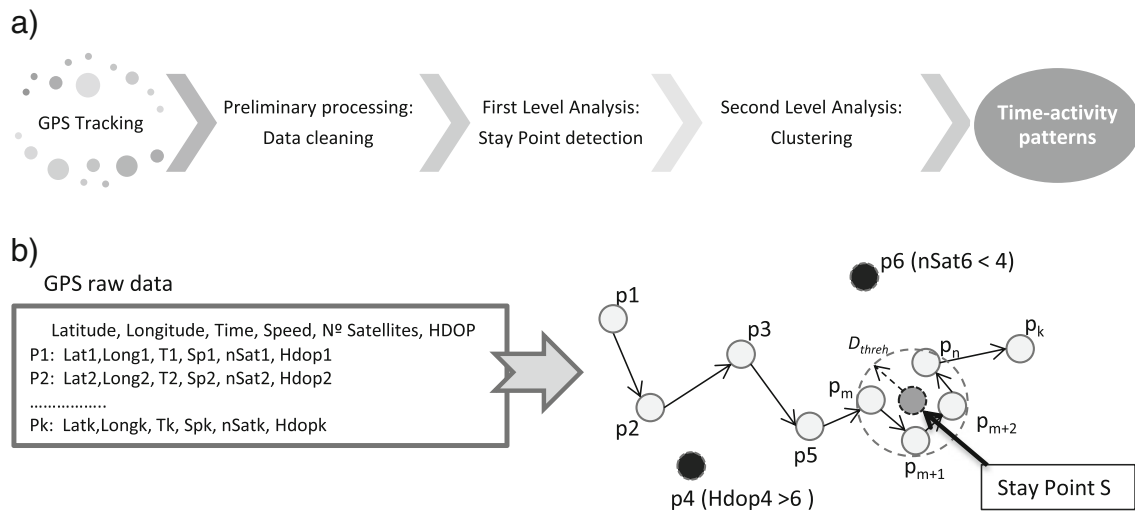


Fig. 2 a Schematic representation of the trajectory data mining analysis; b GPS raw data, GPS “clean” trajectory and stay points detection

However, the repetition of the same locations is not considered and each time that a location is discovered it is assumed as a new location. To overcome this problem, a second level analysis to group up different stay points with the same semantic meaning is implemented using cluster analysis.

Clustering is a data mining technique focused on detecting hidden groups, or clusters, among a set of objects (Bock 1996). In this study, to group the points belonging to the same premises, and thus define the personally significant places, a density-based clustering algorithm DJ-Cluster (Zhou et al. 2004, 2007a, b) was implemented. The DJ-Cluster algorithm is selected and applied in this study, as it is less vulnerable to noise and does not require the number of places as a parameter. However, the algorithm depends excessively on the density of the points and does not give importance to the time spent in each site, i.e. duration, which will be relevant for the exposure quantification.

In the clustering algorithm, the neighbourhoods within distance Eps are analysed for each point. If at least a minimum number ($MinPts$) of such neighbourhoods is found, the points are either grouped as a new cluster or joined with an existing cluster, and a significant place is created. Otherwise, the point is labelled as a moving activity (e.g. being in vehicle micro-environment) (Fig. 3). The following conditions define the density-based neighbourhood of a point and density-joinable relationships (Zhou et al. 2007a):

(a) Density-based neighbourhood of a point:

The density-based neighbourhood N of a point p , denoted by $N(p)$, is defined as:

$$N(p) = \{q \in Q | \text{dist}(p, q) \leq Eps\} \quad (3)$$

where Q is the set of all points, q is any point in the sample, Eps is the radius of a circle around p to defines the density. The following condition is also needs to be

satisfied for $N(p)$:

$$\#N(p) \geq MinPts \quad (4)$$

where $MinPts$ is the minimum number of points required in that circle.

(b) Density-joinable:

$N(p)$ is density-joinable to $N(q)$ denoted as $J(N(p), N(q))$, with respect to Eps and $MinPts$, if there is a point such that both $N(p)$ and $N(q)$ contain it.

For the objectives of this study, the sites are identified as personally significant places taking into account two variables: density and duration. In this perspective, DJ-Cluster algorithm was changed to implement additional condition based on duration of stay, as presented in Fig. 3. Thus, $N(p)$ defined in Eq. 3 needs to satisfy simultaneously two conditions:

$$\#N(p) \geq MinPts \wedge N(p) \cdot \text{duration} \geq MinDrt \quad (5)$$

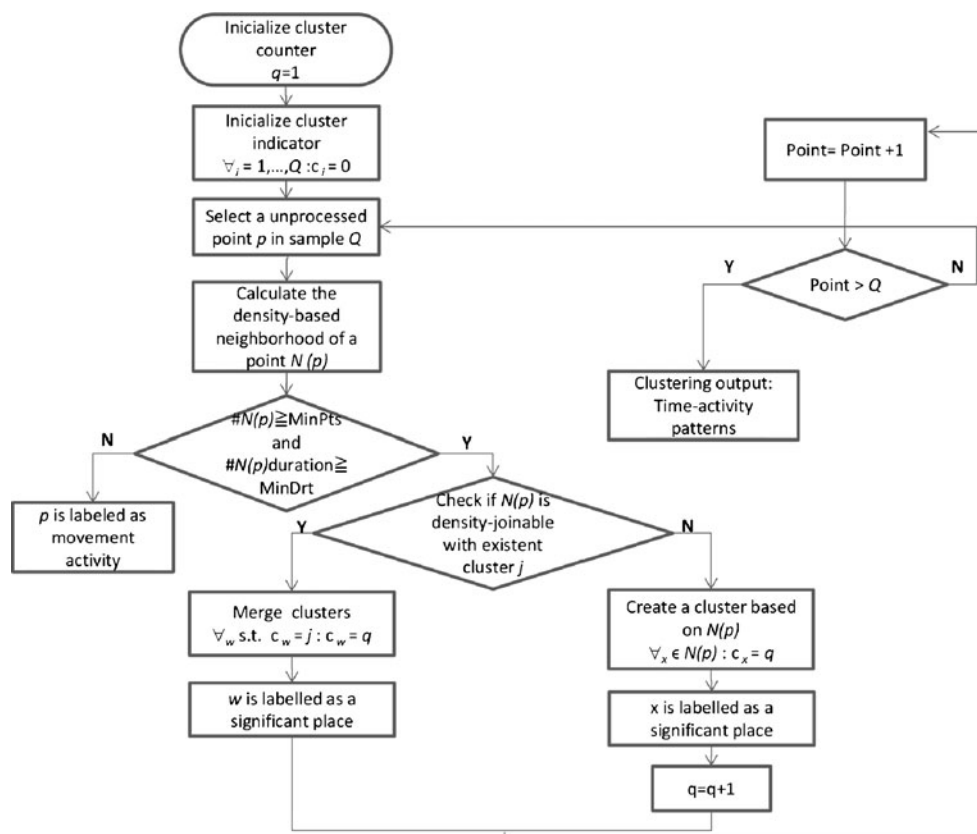
where $MinDrt$ is a parameter that represents the minimum duration at a location. Thus, p can be considered as a cluster or merged with an existent cluster in case that has a minimum number of points required $MinPts$ and a minimum duration $MinDrt$.

These data are further analysed within GIS environment for classification of microenvironments and to obtain information on time–activity patterns.

Time–activity patterns

Location of the individuals in space and in time is required to estimate individual exposure in a combination with pollutants concentration fields provided by air pollution dispersion model.

Fig. 3 Flowchart of the clustering process



To obtain information on time–activity patterns the significant places and movement activities extracted from the trajectory are further analysed within GIS environment to cross this information with other geo-spatial information. For this purpose, geoprocessing of GPS data is performed using ModelBuilder module provided by ArcGis 10. ModelBuilder can be thought of as a visual programming language for building workflows in which it is possible to create, edit, and manage geospatial analysis (Allen 2011).

The geoprocessing of GPS data is accomplished by considering analytical functions and several predefined criteria based on speed, time and spatial location register for the trajectory points to classify the significant places and movement activities to three activity categories: indoor, outdoor and in vehicle travel. The detailed GIS-maps are used to identify and to classify the microenvironments.

An indoor activity is distinguished from outdoor based on the time register. If the spending time in that point is equal or higher than 10 min, based on several tests conducted in this study and as presented by Ashbrook and Thad (2003), the significant place is identified as an indoor activity, and it is geographically located to the nearest indoor microenvironment, acquiring the entire attribute data associated to this microenvironment, such as microenvironment type (residence, workplace, restaurant, etc.). Additionally, the speed

value is analysed to distinguish outdoor activity from in vehicle travel. However, the higher speed values registered during driving a vehicle are not sufficient to identify a movement activity. In addition, activities like being static outdoor and in the traffic jam are difficult to distinguish based on speed criteria only. Thus, if the speed value is less than the speed of walking of 2 km h⁻¹ (Transportation Research Board 1994), the distance between the identified point and the nearest road will be analysed. If there is no intersection with the road network, the significant place is identified as an outdoor microenvironment, such as being in a park, sitting on a terrace, etc. Otherwise, vehicle microenvironment is identified.

Finally, this detailed time–activity patterns for each individual will be linked with the pollutants concentration fields varying in space and in time provided by air pollution dispersion model described in the next section, allowing to produce exposure estimates within distinct microenvironments.

Emission and air quality modelling

Air quality modelling allows establishing the relationships between current emissions and current air quality at particular locations. Information on variability of air pollutant concentrations is essential for the exposure quantification and these data may be provided for ExPOSITION by any modelling

tools if it is compatible with their requirements in terms of spatial and temporal data resolution.

In this work, hourly traffic emissions required by the air quality model were estimated using the Transport Emission Model for Line Sources (TREM). The emission factors considered by TREM depend on average speed, fuel type, engine capacity and emission reduction technology. A new version Transport Emission Model for Hazardous Air Pollutants (TREM-HAP) prepared to calculate HAPs emissions (Tchepel et al. 2012) has been used to provide inputs for AUSTAL2000 dispersion model.

AUSTAL2000 is the official reference air dispersion model of the German Regulation on Air Quality Control for short-range applications (Janicke and Janicke 2002; Janicke 2004). The model is based on Lagrangian approach that simulates the dispersion of air pollutants by utilising a random walk process. Three-dimensional diagnostic wind fields is calculated based on a given initial wind profile and a given terrain profile and/or set of building shapes. Additionally, the vector of the turbulent velocity is randomly varied for every particle by using a Markov process (Janicke 2002; VDI 2000). The fundamental equation for the Lagrangian atmospheric dispersion of a single pollutant is given by Eq. 6.

$$C(x, t) = \int_0^t \int P(x, t|x', t') S(x', t') dx' dt' \quad (6)$$

where $C(x, t)$ is the average pollutant concentration in x at time t , $S(x', t')$ is the source term and $P(x, t|x', t')$ is the probability density function (PDF), that the hypothetical parcel moves from the point x' at time t' to the point x at time t . Therefore, if actual paths of the portions of air can be obtained, the simple calculation of the density of trajectories points provides an estimate of the concentration (Graff 2002).

The main objective of AUSTAL2000 application in the current study is the calculation of atmospheric dispersion of substances, including PM fractions (four different classes of the aerodynamic diameter) allowing to establish relationship between emissions and air quality and to provide hourly pollutants concentration fields. Additionally to input data on emissions, a continuous time series of meteorological parameters, including wind direction, wind speed and atmospheric stability are required by AUSTAL2000.

Currently, several studies using the AUSTAL2000 are available, as well comparative analyses with other dispersion models (Yau et al. 2010; Langner and Klemm 2011; Merbitz et al. 2012; Gerharz and Pebesma 2012).

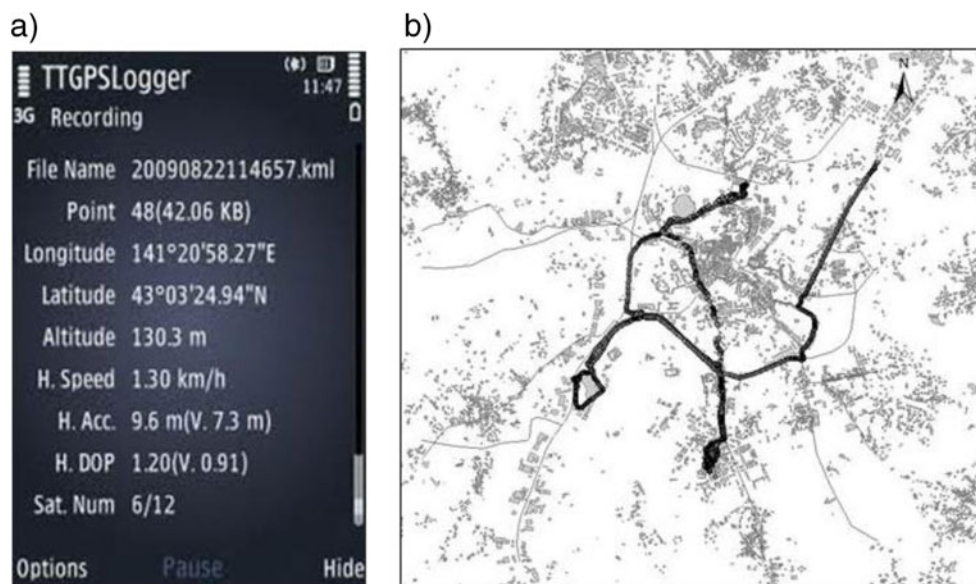
Model application

The Leiria urban area, situated in the central part of Portugal and covering eight sub-municipality units, was selected in this study for the model application. The study domain covering an area of 4.5×4.5 km² with 20-m grid resolution and a complex orography, containing about 5,000 buildings considered as obstacles for the air dispersion modelling. The Leiria urban area and road network considered in this study for the exposure quantification are presented in Fig. 4.

In this study, given the huge health impacts of PM even at levels of exposure currently being experienced by most urban populations (WHO 2006), the new exposure modelling tool was applied to provide a quantitative assessment of the personal exposure to PM_{2.5}.

Hourly PM_{2.5} emissions from road traffic were estimated by TREM based on the traffic volume for each road. For this purpose, data reported by Pinto et al. (2008) were used to

Fig. 4 **a** Data recording screen from mobile phone; **b** spatial visualisation of the GPS raw data recorded



characterise the number of vehicles for each road link. To estimate PM_{2.5} concentrations hourly simulations were conducted with AUSTAL2000 model taking into account hourly meteorological conditions, such as prevailing wind direction and wind speed and background concentrations measured at one monitoring station located in sub-urban area of the city with a temporal resolution of 15 min. In addition, the model system includes a diagnostic wind field model to account for terrain profile and buildings structures.

To characterise the variability in input parameters used to calculate microenvironmental concentrations (Eq. 2), a set of random inputs characterising the infiltration factor α and the contribution of indoor pollution β are generated for each microenvironment. The PDF for both parameters is determined using the information presented in Table 1. A combination of random values is used to create 625 independent inputs for each microenvironment to be considered by ExPOSITION for the exposure estimations.

TTGPSLogger tracking system (TTGPSLogger 2012) was used to collect GPS data providing trajectories of five individuals during a working day in November 2010. TTGPSLogger is a GPS logger software for Symbian S60 allowing to store detailed time–location information on geographic coordinates, speed and time at regular time intervals during its use over the daily activities of individuals. In addition, information on

the positioning accuracy of GPS receiver is provided (number of satellites, position dilution of precision, etc.). The GPS tracking log can be written in NMEA, GPX or KML format. For proper implementation of the trajectory data mining analysis, the GPS data was collected in one-second intervals.

Results and discussion

In this section, the results obtained with newly developed modelling tool for short-term PM_{2.5} exposure quantification are presented and discussed.

GPS data processing

TTGPSLogger tracking system installed on mobile phone is used to collect real-time latitude-longitude position of individuals, speed and time during their daily activities (Fig. 4a). This information was stored in a GPX file format that is compatible with GIS systems presenting very useful to analyse the spatial distribution of large amount of GPS raw data collected. Thus, during a typical working day of one of the individuals analysed in this study, 30,179 GPS raw points with a temporal resolution of 1 s are collected by TTGPSLogger tracking system. However, some of the collected GPS data points with

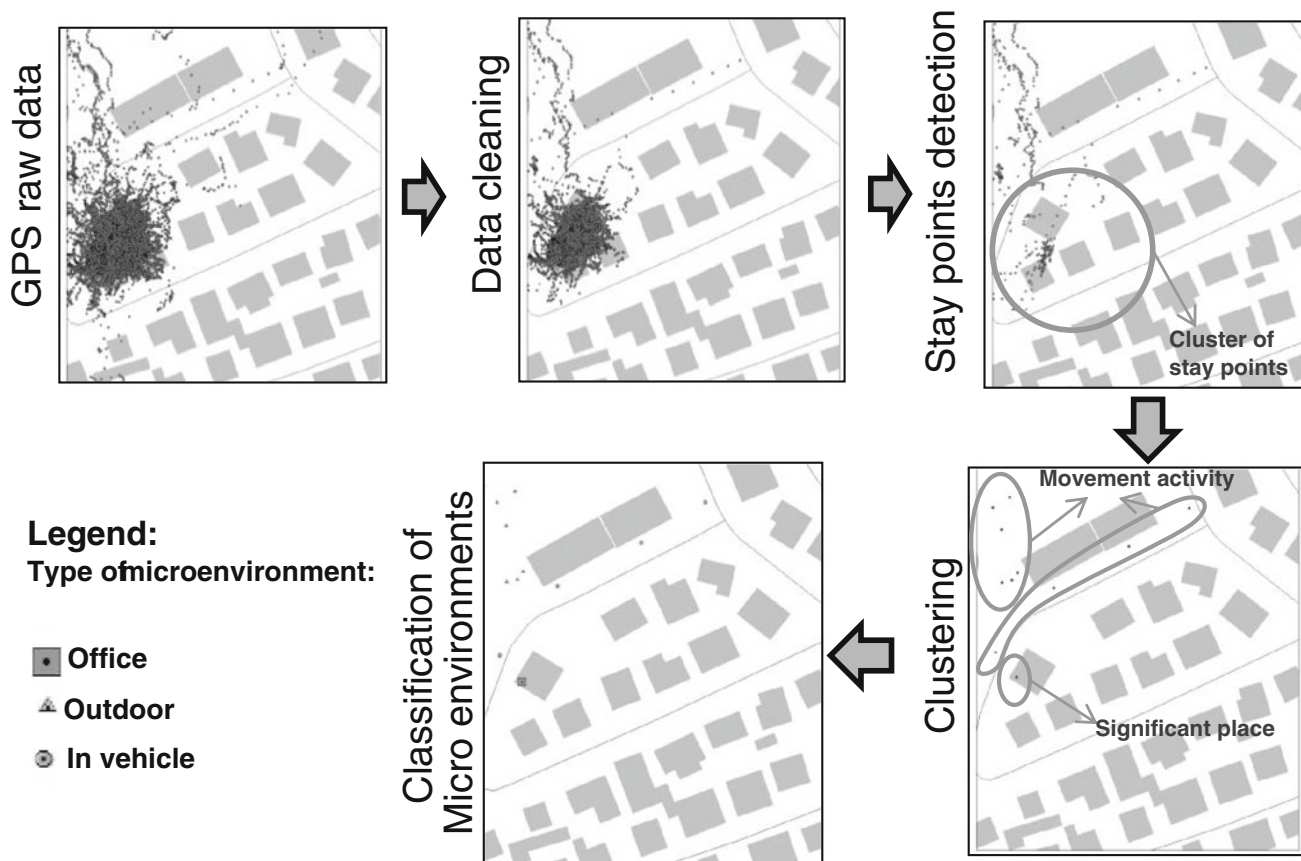


Fig. 5 Example illustrating the data processing applied to GPS raw data

invalid information, such as incorrect entries of speed values achieving maximum of 650 km h^{-1} , are identified.

Most of the invalid measurements observed in this study are from areas where the individual has stayed indoors due to the obstruction of the GPS signal inside of buildings. Furthermore, there are some situations where the GPS receiver located inside buildings does not lose the signal but the data collected are affected by significant errors achieving about 60 m of distance from the actual position. Another limitation observed is a gap of GPS information during some periods (from 15 s to 10 min) depending on the GPS status.

Taking into account the limitations detected during the analysis of GPS raw data, cleaning of the data and their processing are required to predict the time–activity patterns (Fig. 5).

In Fig. 5, a sequence of images with a zoom to the workplace of the individual 4 is presented to illustrate the formation of three clusters detected from the data. The first image corresponds to the raw GPS points recorded. Subsequently, the data cleaning allows to remove invalid points from collected GPS raw points. However, this approach is only a pre-processing to detect errors and inconsistencies in data. The stay point detection algorithm reduces significantly the number of GPS points that are consequently used for the clustering. In the example presented in Fig. 5, one significant place and two movement activity clusters are identified from the set of stay points.

The locations resulting from the clustering algorithm are further analysed within GIS environment to cross this information with other geo-spatial information and to obtain detailed time–activity patterns classified by different types of microenvironments. Thus, in the case of individual 4, 30,179 collected GPS raw points resulted in 15,978 stay points, originating 295 locations that are linked with the pollutants concentration in distinct microenvironments to assess its individual exposure.

Emission, air quality and meteorological data

To estimate human exposure to PM_{2.5}, hourly traffic emissions and air pollutants concentrations were estimated. Figure 6a illustrates the spatial variations in hourly traffic-related emissions across the study area obtained by linking TREM-HAP outputs to GIS maps. As could be seen in the figure, higher emission values are observed for main city entrances.

Apart from road transport, other emission sources that play an important role for the study area are characterised in the current study by considering the background concentrations of PM_{2.5}. As mentioned in “Model application” the background concentrations of PM_{2.5} and meteorological data were monitored in a sub-urban location at one fixed monitoring station. It should be mentioned that the study period was analysed and selected for this work as the PM_{2.5} concentrations observed for this sampling period were not affected by long-range transport.

The meteorological data obtained during the sampling period are presented in the Fig. 7a. As could be seen, the higher wind intensities are achieved with winds blowing from the North, which is also the predominant wind direction, although there is a significant contribution of the East direction. As regards the variation throughout the day, generally the wind speed gradually increases, reaching maximum values at 6 p.m. and minimum values during the night. Therefore, this information on prevailing wind direction and wind speed for each hour, terrain profile and building shapes of the city are explicitly considered by the meteorological module and reflected in the modelled PM_{2.5} concentrations.

The time series of hourly background concentrations of PM_{2.5} measured from fixed monitoring station is presented in the Fig. 7b evidencing a gradually decrease for PM_{2.5}

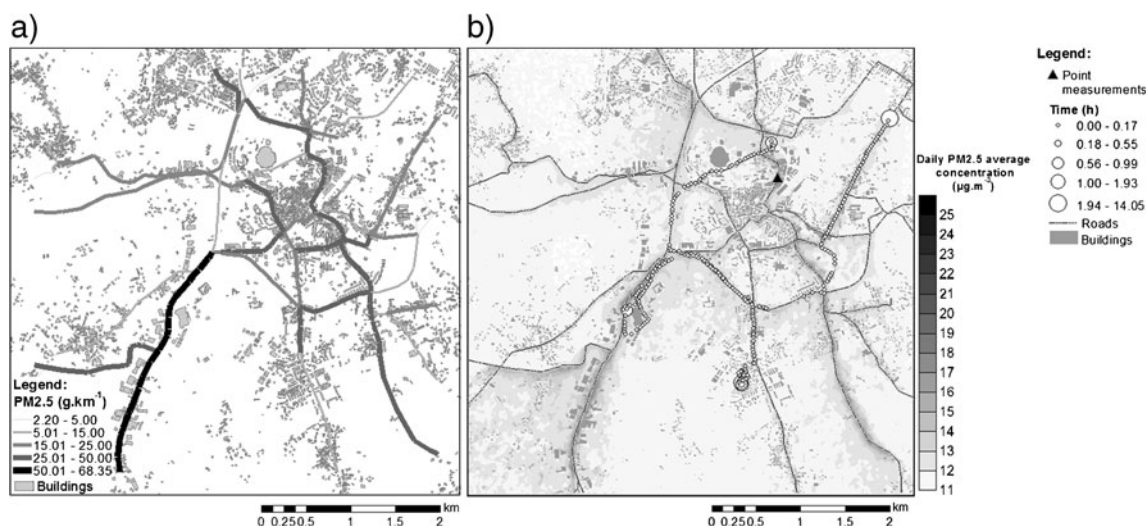


Fig. 6 Spatial distribution of **a** hourly PM_{2.5} emissions (g.km^{-1}), **b** daily average PM_{2.5} concentration ($\mu\text{g m}^{-3}$) and time spent by the individual in each microenvironment

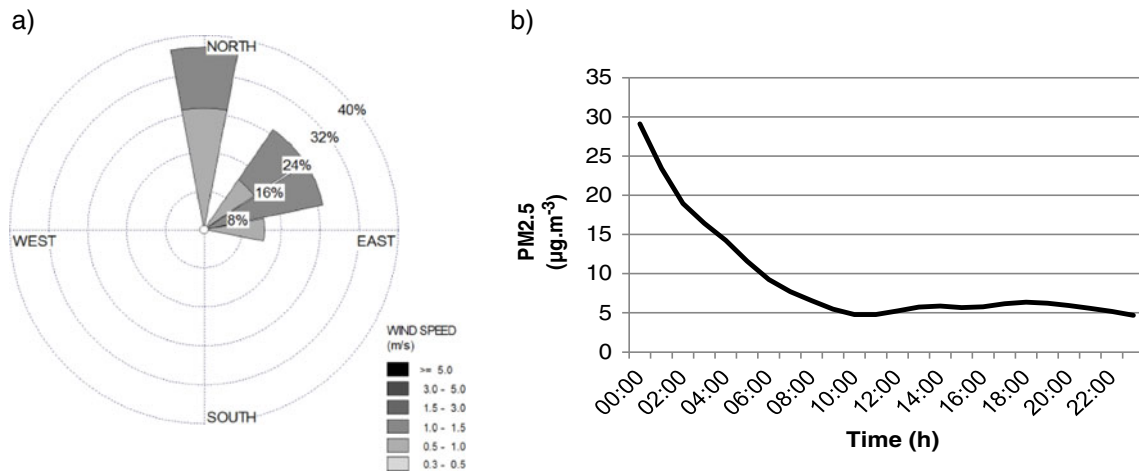


Fig. 7 **a** Hourly wind speed obtained from measurements as a function of wind direction; **b** temporal variation of hourly background PM2.5 concentrations

concentrations during the night-time period. The lower concentrations of PM2.5 are observed for 11 a.m., affected by the lower emissions that were observed during the same time period, thus influencing consequently PM2.5 concentrations. Conversely, the hourly maximum value of 29.1 µg m⁻³ are reached at 12 a.m.

The spatial distribution of the air pollutants concentration obtained by the AUSTAL2000 is presented in Fig. 6b showing that distribution of pollution levels within the study domain is not homogeneous. In addition, time–activity patterns obtained for one of the individuals are presented in the figure as an example. The analysis of results examines the PM2.5 concentration variation in space and in time provided by air pollution dispersion model and the influence of time spent in each microenvironment type. Thus, these findings enhance the importance of taking into account the high spatial and temporal variations in outdoor concentrations, the “microenvironmental” variations imposed by a variety of indoor and outdoor locations and the time spent indoors to obtain accurate personal exposure estimates to air pollution.

Personal exposure modelling

The individual exposure assessment performed by the ExPOSITION model is presented in this section. For better

understanding of the contribution of different microenvironments to the daily average PM2.5 exposure in the study area at a typical working day, several statistical parameters, including average individual exposure, 5th and 95th percentiles and extreme values were analysed (Table 2).

As could be seen in Table 2, the largest variability in the exposure concentration is identified for outdoor and residence microenvironment. Exposure concentration calculated for in vehicle are characterised by smaller variability range but higher absolute values in comparison with the other types of microenvironments. In addition, it is possible to verify that the variability in the PM2.5 exposure concentration in each microenvironment type is significant showing the importance to consider this variability in individual exposure modelling.

As expected, the indoor microenvironments represent a great relevance for the exposure of individuals (Fig. 8). Conversely, it is possible to verify that being outdoors represents a very low contribution to the exposure because corresponds only about 2 % of the time spent by individuals during their daily activities, which suggests that outdoor concentrations measurements should be used carefully for human exposure quantification. However, outdoor concentrations represent an important part of the pollution levels estimated for indoor microenvironments because of outdoor/indoor infiltration.

Table 2 Exposure concentration for PM2.5 (µg m⁻³) in different microenvironments

Microenvironment	Average	Percentile 5	Percentile 95	Minimum	Maximum
Residence	10.2	7.7	17.8	7.0	18.1
Workplace	8.7	4.5	11.7	4.5	16.1
Public access	14.5	13.5	16.3	13.5	26.1
Bar/restaurant	16.7	15.2	17.9	15.0	18.0
Vehicle	35.2	34.6	37.4	20.2	44.6
Outdoor	7.5	5.2	12.7	4.8	41.6

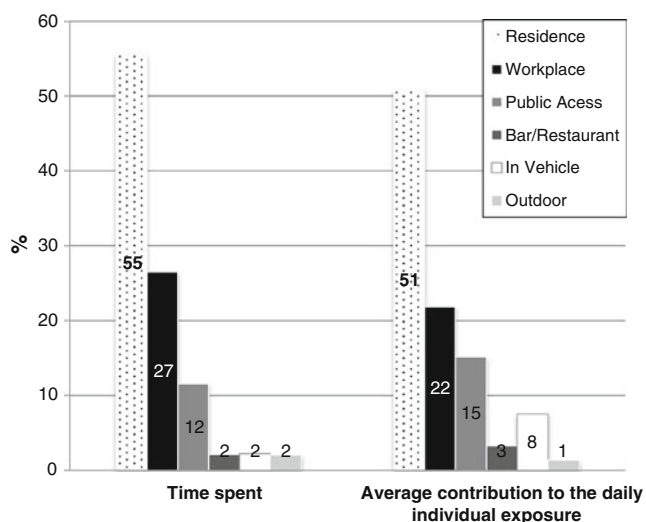


Fig. 8 Distribution of time spent by individuals and average contribution of different microenvironments to daily individual exposure

To better understand the individual exposure obtained during the simulation period, a temporal variation of the exposure concentration was analysed as presented in Fig. 9. Several statistical parameters, including hourly average exposure concentration, 5th and 95th percentiles are analysed for each individual.

The results show that the five individuals are exposed to different PM_{2.5} concentrations during their daily activities, and a significant variability in PM_{2.5} exposures across the individuals is evident in Fig. 9. Analysing the individual exposure concentrations during night time (between 9 p.m. and 7 a.m., approximately), when the people stay in residence, the hourly exposure concentrations presents a similar trend with the outdoor concentrations but different magnitude. However, throughout the day and depending on the daily activity of the individuals the hourly average exposure concentrations tend to be more variable. The highest exposure

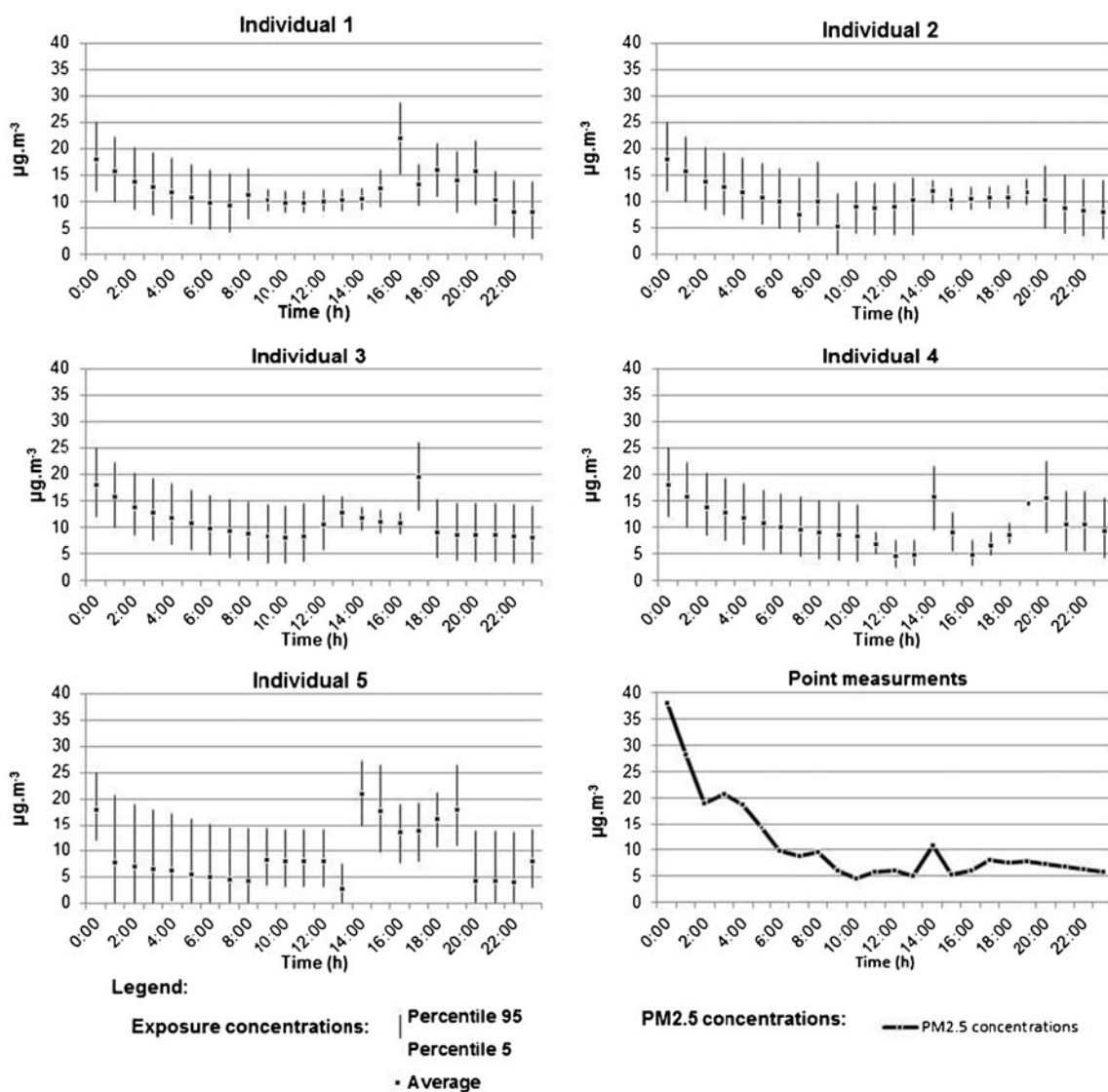


Fig. 9 Temporal variation of individual exposure concentrations and outdoor concentrations of PM_{2.5}

levels are related with both the magnitude of pollutant concentrations and the time spent in specific microenvironments as, for example, could be seen in Fig. 9 for the individual 1 at 4 p.m.

Overall, the daily average exposure to PM_{2.5} predicted by the ExPOSITION model correspond to $10.6 \mu\text{g m}^{-3}$ in terms of the mean value for all individuals and $6.0\text{--}16.4 \mu\text{g m}^{-3}$ in terms of 5th–95th percentiles. Comparing the mean value obtained by the model and estimated from air quality measurements at a fixed point ($11 \mu\text{g m}^{-3}$) represented in Fig. 6b, an agreement between the approaches is evidenced. However, the ExPOSITION model reveals additional inter- and intra-variability of individuals' exposure levels, suggesting limited representativeness of air quality concentrations obtained from point measurements to characterise individual exposure to urban air pollution.

The results obtained from the ExPOSITION model are in good agreement with the daily average exposure reported for other European cities such as Helsinki ($9.9 \mu\text{g m}^{-3}$) (Koistinen et al. 2001) and Amsterdam ($14.5 \mu\text{g m}^{-3}$) (Janssen et al. 2005). The current study shows that high PM_{2.5} exposure is mainly attributed to indoor microenvironments rather than outdoor, as also presented by Georgopoulos (2005). In this context, individual time activities patterns and time spent at different microenvironments during the day should be of prime concern in addition to the variability in the pollution levels, as presented by Burke et al. (2001).

Conclusions

Under this work, a GPS-based personal exposure model based on an innovative approach for trajectory analysis was developed. For this purpose, a time–activity pattern discovery sequence, based on trajectory data mining and geo-spatial analysis within GIS, was developed to extract useful time–location information from GPS raw data collected by a mobile phone with a GPS tracking system carried by the user during their daily activities. Taking into account the limitations detected during the analysis of GPS raw data, the results obtained during the several levels of GPS data analyses indicate that this approach could be used to analyse the human behaviours and activities required for exposure assessment.

Time series of individual exposure concentrations to PM_{2.5} are presented for the entire study area characterising a person's contact with a given pollution levels at different microenvironments. The results show a significant contribution of indoor microenvironments to the total exposure values thus stressing that individual exposure depends not only on the exposure pollution levels but also on the time spent in the microenvironment during the day. In addition, the low contribution of outdoor environment to the daily individual

exposure suggests that the concentration peaks of air pollution are not co-located in time and space with the time period that the individual spend outdoors.

The uncertainty of the exposure estimates are not addressed in the current application of the ExPOSITION model. Uncertainties in indoor and outdoor concentrations could be both relevant and should be taken into account in ExPOSITION model as they contributed to the total exposure uncertainties with varying strengths depending on the time people spent in different environments. However, as mentioned before, the uncertainty related to the human mobility during the exposure assessment period is overcome by the exposure modelling approach developed in this study.

Overall, by comparing the personal exposure modelling results with fixed-point measurements, the ExPOSITION results clearly outperforms the traditional approach of using urban background measurements as proxy for the personal exposure.

In contrast to exposure model approaches currently available (STEMS (Gulliver and Briggs 2005); Beckx et al. 2009; Gerharz et al. 2009, 2013), our model approach focuses not only on the indoor, outdoor or journey-time exposure to outdoor concentration but provides exposure estimates for the whole activity profile including the contribution of indoor emission sources and the variability of microenvironmental parameters. Additionally, current exposure models lack the temporal and spatial resolution of the presented approach that can therefore be considered for application areas different from those of the existing models, focussing on the exposure at the individual level.

The methodology developed and applied in this study allows to estimate and analyse the magnitude, frequency and the inter- and intra-variability of personal exposure levels, as well the contribution of different microenvironments, clearly addressing the time-sequence of the exposure events and source–receptor relationship, which is essential for health impact assessment and epidemiological studies. The ExPOSITION model contributes to better understanding of individual exposure in urban areas, by providing information on individual exposure taking into account where individuals actually spend their time and the high spatial and temporal variations of the “microenvironmental” concentrations imposed by a variety of indoor and outdoor locations, essential in the selection of strategies to reduce exposure to urban air pollution and related health effects.

An application of the ExPOSITION model developed in this study to benzene and validation of the modelling approach against individual exposure measurements will be presented in the future work.

Acknowledgements The authors thank the Portuguese Foundation for Science and Technology for the Ph.D. grant of D. Dias (SFRH/BD/47578/2008).

References

- Allen D (2011) Getting to Know ArcGIS ModelBuilder. ESRI PR. 362 pp
- Ashbrook D, Thad S (2003) Using GPS to learn significant locations and predict movement across multiple users. *Pers Ubiquit Comput* 7: 275–286
- Baklanov A, Hänninen O, Slordal LH, Kukkonen J, Bjergene N, Fay B, Finardi S, Hoe SC, Jantunen M, Karppinen A, Rasmussen A, Skouloudis A, Sokhi RS, Sorensen JH, Odegaard V (2007) Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmos Chem Phys* 7:855–874
- Ballesta PP, Field RA, Fernandez-Patier R, Galan-Madruga D, Connolly R, Caracena AB, De Saegera E (2008) An approach for the evaluation of exposure patterns of urban populations to air pollution. *Atmos Environ* 42:5350–5364
- Beckx C, Int Panis L, Arentze T, Janssens D, Torfs R, Broekx S, Wets G (2009) A dynamic activity-based population modelling approach to evaluate exposure to air pollution: methods and application to a Dutch urban area. *Environ Impact Assess Rev* 29:179–185
- Bock HH (1996) Probabilistic models in cluster analysis. *Comput Stat Data Anal* 23:5–28
- Brauer M, Hoek G, Van Vliet P, Meliefste K, Fisher P, Wijga A, Koopman LP, Neijens HJ, Gerritsen J, Kerkhof M, Heinrich J, Bellander T, Brunekreef B (2002) Air pollution from traffic and the development of respiratory infections and asthmatic and allergic symptoms in children. *Am J Respir Crit Care Med* 166:1092–1098
- Brunekreef B, Holgate ST (2002) Air pollution and health. *Lancet* 360: 1233–1242
- Burke JM, Zufall MJ, Ozkaynak H (2001) A population exposure model for particulate matter: case study results for PM_{2.5} in Philadelphia, PA. *J Expo Anal Environ Epidemiol* 11:470–489
- Dons E, Int Panis L, Van Poppel M, Theunis J, Willems H, Torfs R, Wets G (2011) Impact of time–activity patterns on personal exposure to black carbon. *Atmos Environ* 45:3594–3602
- Freijer JI, Bloemen HJT, Loos S, Marra M, Rombout PJA, Steentjes GM, van Veen MP (1998) Modelling exposure of the Dutch population to air pollution. *J Hazard Mater* 61:107–114
- Gauderman WJ, Vora H, McConnell R, Berhane K, Gilliland F, Thomas D, Lumann F, Avol E, Künzli N, Jerrett M, Peters J (2007) Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study. *Lancet* 369:571–577
- Georgopoulos PG, Wang SW, Vyas VM, Sun Q, Burke J, Vedantham R, McCurdy T, Ozkaynak H (2005) A source-to-dose assessment of population exposures to fine PM and ozone in Philadelphia, PA, during a summer 1999 episode. *J Expo Anal Environ Epidemiol* 15: 439–457
- Gerharz L, Pebesma E (2012) Using geostatistical simulation to disaggregate air quality model results for individual exposure estimation on GPS tracks. *Stoch Environ Res Risk Assess* 27:223–234
- Gerharz LE, Krüger A, Klemm O (2009) Applying indoor and outdoor modeling techniques to estimate individual exposure to PM_{2.5} from personal GPS profiles and diaries: a pilot study. *Sci Total Environ* 407:5184–5193
- Gerharz LE, Klemm O, Broich AV, Pebesma E (2013) Spatio-temporal modelling of individual exposure to air pollution and its uncertainty. *Atmos Environ* 64:56–65
- Graff A (2002) The new German regulatory model—a Lagrangian particle dispersion model. In: 8th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
- Gulliver J, Briggs DJ (2005) Time-space modeling of journey-time exposure to traffic-related air pollution using GIS. *Environ Res* 97: 10–25
- Health Effects Institute (2010) Traffic-related air pollution: a critical review of the literature on emissions, exposure, and health effects, HEI Special Report 17. Health Effects Institute, Boston
- Hertel O, Jensen S, Hvidberg M, Ketzel M, Berkowicz R, Palmgren F, Wählin P, Glasius M, Loft S, and Vinzents P et al (2008) Assessing the Impacts of Traffic Air Pollution on Human Exposure and Health. In: Fischer M et al (eds) Road pricing, the economy and the environment, pp. 277–299
- Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Environ* 42:7561–7578
- Janicke L (2002) Lagrangian dispersion modeling. *Part Matter Agric* 235: 37–4, 3-933140-58-7
- Janicke I (2004) AUSTAL2000. Programbeschreibung zu Version 2.1. Stand 2004-12-23. Ingenieurbüro Janicke
- Janicke L, Janicke U (2002) A modelling system for licensing industrial facilities, UFOPLAN 200 43 256. German Federal Environmental Agency UBA, German
- Janssen NAH, Lanki T, Hoek G, Vallius M, de Hartog JJ, Van Grieken R et al (2005) Associations between ambient, personal, and indoor exposure to fine particulate matter constituents in Dutch and Finnish panels of cardiovascular patients. *Occup Environ Med* 62:868–877
- Jensen SS (2006) A GIS-GPS modeling system for personal exposure to traffic air pollution. *Epidemiology* 17:S38–S38
- Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hem SC (2001) The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expo Anal Environ Epidemiol* 11:231–252
- Koistinen KJ, Hänninen OO, Rotko T, Edwards RD, Moschandreas D, Jantunen MJ (2001) Behavioral and environmental determinants of personal exposures to PM_{2.5} in EXPOLIS-Helsinki, Finland. *Atmos Environ* 35:2473–2481
- Kousa A, Oglesby L, Koistinen K, Kunzli N, Jantunen M (2002) Exposure chain of urban air PM_{2.5}—associations between ambient fixed site, residential outdoor, indoor, workplace and personal exposures in four European cities in the EXPOLIS-study. *Atmos Environ* 36:3031–3039
- Kruize H, Hänninen O, Breugelmans O, Lebreit E, Jantunen M (2003) Description and demonstration of the EXPOLIS simulation model: two examples of modeling population exposure to particulate matter. *J Expo Anal Environ Epidemiol* 13:87–99
- Langner C, Klemm O (2011) A comparison of model performance between AERMOD and AUSTAL2000. *J Air Waste Manag Assoc* 61:640–646
- Li Q, Zheng Y, Xie X, Chen Y, Liu, W, Ma W-Y (2008) Mining user similarity based on location history. In: Proceedings of the 16th ACM SIGSPATIAL international conference on advances in geographic information systems, GIS '08, pp. 34:1–34:10
- Lioy PJ (2010) Exposure science: a view of the past and milestones for the future. *Environ Health Perspect* 118:1081–1090
- Merbitz H, Fritz S, Schneider C (2012) Mobile measurements and regression modeling of the spatial particulate matter. *Sci Total Environ* 438:389–403
- Nethery E, Leckie SE, Teschke K, Brauer M (2008) From measures to models: an evaluation of air pollution exposure assessment for epidemiological studies of pregnant women. *Occup Environ Med* 65:579–586
- Oglesby L, Künzli N, Röösli M, Braun-Fahrlander C, Mathys P, Stern W, Jantunen M, Kousa A (2000) Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution. *J Air Waste Manag Assoc* 50:1251–1261
- Özkaynak H, Palma T, Touma J, Thurman J (2008) Modeling population exposures to outdoor sources of hazardous air pollutants. *J Expo Sci Environ Epidemiol* 18:45–58
- Peng RD, Bell ML (2010) Spatial misalignment in time series studies of air pollution and health data. *Biostatistics* 11:720–740

- Pinto N, Silva JP, Pereira PM (2008) Projeto Mobilidade Sustentável para o Município de Leiria, Relatório 1- Diagnóstico e Princípios Orientadores de Intervenção, Laboratório de Planeamento, Transportes e Sistemas de Informação Geográfica. Instituto Politécnico de Leiria, Portugal
- Rainham D, McDowell I, Krewski D, Sawada M (2010) Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. *Soc Sci Med* 70:668–676
- Song C, Qu Z, Blumm N, Barabási A (2010) Limits of predictability in human mobility. *Science* 327:1018–1021
- Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman J (2008) Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal interactions. Working paper 337, UW Biostatistics Working Paper Series
- Tchepel O, Dias D (2011) Quantification of health benefits related with reduction of atmospheric PM10 levels: implementation of population mobility approach. *Int J Environ Health Res* 21:189–200
- Tchepel O, Dias D, Ferreira J, Tavares R, Miranda AI, Borrego C (2012) Emission modelling of hazardous air pollutants from road transport at urban scale. *Transport* 27(8):299–306
- Transportation Research Board (1994) Special Report 209. Highway Capacity Manual, Washington
- TTGPSLogger (2012). <http://code.google.com/p/ttgpslogger/>. Accessed November 2012
- US Environmental Protection Agency (2006) Total risk integrated methodology (TRIM) air pollutants exposure model documentation (TRIM.Expo/APEX, Version 4)–volume i: user’s guide. U.S. Environmental Protection Agency, Research Triangle Park, NC
- US Environmental Protection Agency (2009) Human exposure modeling air pollutants exposure model. Available from (http://www.epa.gov/ttn/fera/human_apex.html) Accessed 28th April 2013
- VDI (2000) Guideline 3945, Part 3. Environmental meteorology-atmospheric dispersion model—particle model. Guideline
- Wang S-W, Tang X, Fan ZH, Liou PJ, Georgopoulos PG (2009) Modelling personal exposures to ambient air toxics in Camden, New Jersey: an evaluation study. *J Air Waste Manag* 59:733–746
- WHO and EC (2002) Guidelines for concentration and exposure-response measurement of fine and ultra-fine particulate matter for use in epidemiological studies, EUR 20238 EN 2002
- WHO (2006) Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide-global update 2005. Summary of Risk Assessment, Geneva
- WHO (2011) Air quality and health. Fact sheet No. 313. Updated September 2011. Available from (<http://www.who.int/mediacentre/factsheets/fs313/en/>). Accessed 1 January 2013
- Wu J, Jiang C, Liu Z, Houston D, Jaimes G, McConnell R (2010) Performances of different global positioning system devices for time–location tracking in air pollution epidemiological studies. *Environ Health Insights* 4:93–108
- Yau KH, Macdonald RW, The JL (2010) Inter-comparison of the AUSTAL2000 and CALPUFF dispersion models against the Kincaid data set. *Int J Environ Pollut* 40:267–279
- Zheng Y, Zhou X (2011) Computing with spatial trajectories. Springer, New York
- Zhou C, Frankowski D, Ludford P, Shekhar S, Terveen L (2004) Discovering personal gazetteers: an interactive clustering approach. In: *Proc ACMGIS 266–273*
- Zhou C, Bhatnagar N, Shekhar S, Terveen L (2007a) Mining personally important places from GPS tracks. In: *ICDEW ’07: Proceedings of the 2007 I.E. 23rd International Conference on Data Engineering Workshop*, Washington, DC, USA. IEEE Computer Society. pp. 517–526
- Zhou C, Frankowski D, Ludford P, Shekhar S, Terveen L (2007b) Discovering personally meaningful places: an interactive clustering approach. *ACM Trans Inform Syst* 25:3
- Zou B, Wilson JG, Zhan FB, Zeng Y (2009) Air pollution exposure assessment methods utilized in epidemiological studies. *J Environ Monit* 11:475–490