Journal of Exposure Science and Environmental Epidemiology (2015) 25, 97–105
© 2015 Nature America, Inc. All rights reserved 1559-0631/15
www.nature.com/jes

ORIGINAL ARTICLE

Application of land use regression modelling to assess the spatial distribution of road traffic noise in three European cities

Inmaculada Aguilera[1,2], Maria Foraster[3,4], Xavier Basagaña[3,4], Elisabetta Corradi[1,2], Alexandre Deltell[5], Xavier Morelli[6,7], Harish C. Phuleria[1,2,8], Martina S. Ragettli[1,2], Marcela Rivera[9], Alexandre Thomasson[10], Rémy Slama[6,7] and Nino Künzli[1,2]

Noise prediction models and noise maps are used to estimate the exposure to road traffic noise, but their availability and the quality of the noise estimates is sometimes limited. This paper explores the application of land use regression (LUR) modelling to assess the long-term intraurban spatial variability of road traffic noise in three European cities. Short-term measurements of road traffic noise taken in Basel, Switzerland ($n = 60$), Girona, Spain ($n = 40$), and Grenoble, France ($n = 41$), were used to develop two LUR models: (a) a "GIS-only" model, which considered only predictor variables derived with Geographic Information Systems; and (b) a "Best" model, which in addition considered the variables collected while visiting the measurement sites. Both noise measurements and noise estimates from LUR models were compared with noise estimates from standard noise models developed for each city by the local authorities. Model performance (adjusted R^2) was 0.66–0.87 for "GIS-only" models, and 0.70–0.89 for "Best" models. Short-term noise measurements showed a high correlation ($r = 0.62–0.78$) with noise estimates from the standard noise models. LUR noise estimates did not show any systematic differences in the spatial patterns when compared with those from standard noise models. LUR modelling with accurate GIS source data can be a promising tool for noise exposure assessment with applications in epidemiological studies.

Journal of Exposure Science and Environmental Epidemiology (2015) **25,** 97–105; doi:10.1038/jes.2014.61; published online 17 September 2014

Keywords: traffic noise; noise propagation models; land use regression models; noise exposure

INTRODUCTION

Environmental noise is prevalent and ubiquitous, particularly that from traffic in urban areas.[1] Noise has been associated with annoyance, cardiovascular disease, and impaired cognitive performance, among other health outcomes.[2] The increasing concern about this environmental stressor warrants further research; however, this is sometimes prevented by the unavailability or low quality of the noise exposure assessment.

Noise prediction models within urban and metropolitan areas are useful tools to identify the priorities for action planning intended to manage noise and reduce noise pollution. Because assessing individual exposure to noise in large epidemiological studies is not feasible owing to logistical and budgetary reasons, noise prediction models and noise maps have been widely used for this purpose in epidemiological studies.[3–6] However, the availability of noise maps (or estimates from noise prediction models) depends on several factors, including the size of the area, the availability of input data, and the legal context. In the European Union (EU), the Environmental Noise Directive (2002/49/EC) requires EU Member States to determine the population exposure to noise through strategic noise maps, but this obligation does not apply for agglomerations with <100,000 inhabitants.[7] Even when available, a thorough case-by-case evaluation of these data should be done by an expert to ensure the quality. Maps are designed to identify noisy areas but not quieter areas[8] and the degree of refinement may vary across cities. The inaccessibility to data from noise prediction models is also an issue. As developing noise models requires specialized software and user expertise, it might not be possible for epidemiologists to obtain model predictions at given locations, and lower quality noise estimates may be obtained directly from existing digital noise maps. In other parts of the world, including North America, noise mapping efforts have been more variable and dependent on the awareness of the competent authorities concerning the integration of noise mapping into community planning.

Land use regression (LUR) modelling is currently one of the most used methods for assessing the exposure to air pollution in epidemiological studies.[9] This method uses least-squares regression modelling to predict air pollution levels based on the pollution monitoring data at a small number of locations and predictor variables collected mainly through Geographic Information Systems (GIS). LUR models are relatively inexpensive and their performance in detecting the small-scale variations of air pollution within urban areas is as good as other techniques such as dispersion models, which require more complex and costly inputs of data.[10] Besides, their empirical structure allows for the use of standardized approaches, which is a clear strength in multicentre studies. A recent example is the ESCAPE (European Study of

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland; [2]University of Basel, Basel, Switzerland; [3]Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; [4]CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; [5]Grup de Recerca en Enginyeria de Fluids, Energia i Medi Ambient (GREFEMA), University of Girona, Girona, Spain; [6]INSERM U823, Institut Albert Bonniot, Team of Environmental Epidemiology Applied to Reproduction and Respiratory Health, Grenoble, France; [7]Université Joseph Fourier (UJF), Grenoble, France; [8]Centre for Environmental Science and Engineering, Indian Institute of Technology Bombay, Mumbai, India; [9]University of Montreal Hospital Research Center (CRCHUM), Montreal, Canada and [10]Air Rhône-Alpes, Grenoble, France. Correspondence: Dr. Inmaculada Aguilera, Swiss Tropical and Public Health Institute, Socinstrasse 57, Basel 4051, Switzerland.
Tel.: +41 61284 8111. Fax: +41 61284 8101.
E-mail: inmaculada.aguilera@unibas.ch
Received 18 March 2014; revised 27 May 2014; accepted 12 July 2014; published online 17 September 2014

Cohorts for Air Pollution Effects) project, which has developed LUR models to estimate outdoor concentrations of several air pollutants at the home addresses of participants in over 30 cohort studies.[11–13]

In the noise exposure assessment field, LUR modelling has been very little explored. To our knowledge, only one study has applied this technique to assess the spatial variation of environmental noise.[14] It was conducted in an area of ~2,400 m^2 in northeast China, and used two subsets of sites: 101 for model development and 101 for model validation. The model explained 83.2% of the variability in environmental noise levels and showed a similar performance at three different spatial scales (from the "local" downtown to the "regional" urban sprawl area). Besides, the LUR model performed better than a spatial interpolation model (exponential Kriging) based on the same monitoring data.

The aim of the present study was to develop LUR models to explain the within-city variability of road traffic noise in three European cities, which could be of potential use in noise exposure assessment for epidemiological studies. These models have been developed using short-term measurements of road traffic noise levels at locations specifically selected for the present study. A second objective was to compare both the short-term noise measurements and the noise estimates from LUR models with road traffic noise levels estimated from the existing models developed by the local authorities.

MATERIALS AND METHODS

Study Areas

The Tri-Tabs (Tri-national traffic, air, noise, and health) project was conducted between 2009 and 2011 in three mid-sized European cities: Basel Switzerland (193,000 inhabitants), Girona Spain (96,700 inhabitants), and Grenoble France (156,000 inhabitants). Road traffic is the dominant inner-city source of both air pollution and noise in the three cities.

As the main objective of the project was to identify and compare the spatial characteristics of ultrafine particles and traffic-related noise, measurement sites were selected specifically to capture the spatial contrast of traffic-related air pollution and noise, based on variables such as distance to major roads, street density, building density, and population density. Sites directly affected by traffic flow were overrepresented as we anticipated more variation in air pollution and noise between them than between urban or regional background sites. In total, 60 measurement sites were selected in Basel, 40 in Girona, and 41 in Grenoble. Sites were geocoded using the Global Positioning System (GPS) in Basel and Grenoble, and manually corrected if needed using digital maps by someone who had visited the site personally. In Girona, sites were geocoded using a mapping application from the regional government and manually corrected to locate them at 2 m from the corresponding façade.

Short-Term Road Traffic Noise Measurements

Traffic-related noise was measured (together with UFP and road traffic flow) during 20 min in non-rush hours (defined separately for each city) and during weekdays only. Equivalent continuous A-weighted levels of diurnal noise were measured, producing LAeq averages (dB) corrected from abnormal peaks. Measurements were performed following the standard procedures recommended by the European Noise Directive 2002/49/EC. No measurements were performed in bad weather conditions, including raining and wind >4 m/s. To avoid noise reflections, the sound-level meter was deployed on a tripod at 1.5 m height and separated from noise barriers by at least 1.5 m. We used high-resolution sound-level meters, namely a CESVA SC30 in Girona and Grenoble and a Pulsar 30 in Basel, which were calibrated on a daily basis. Both were class 1 sound-level meters and had the same tolerance of ±0.7 dB. Thus, the use of two different devices was not expected to affect the results of city-specific LUR models.

After measurements, we followed a data cleaning and quality control protocol to identify and remove non-traffic-related events that produce noise, such as birds, trams, bus stops, car honking, people talking loud close to the microphone, car engine starting and leaving a nearby parking spot, construction noise, and so on. However, because of the very dense tram network in Basel, in some sites it was not possible to disentangle that source from the traffic-related noise.

In Basel, measurements were taken at all sites in three different seasons (spring, summer and winter 2011). In Girona and Grenoble, measurements were repeated in a subset of sites (25 in Girona and 26 in Grenoble) at close time intervals (during summer 2009 in Girona and during fall 2011 in Grenoble). Given that repeated noise measurements were very highly correlated over time in the three cities (intraclass correlation coefficient $r = 0.86–0.97$), we used the average of the repeated noise measurements as the best estimate of the site specific long-term noise levels.

Road Traffic Noise Estimates from Models Developed by Local Authorities

Long-term road traffic noise levels were derived from noise models developed by the local authorities in each city. Models were developed by the Universitat de Girona in 2005 (Girona), by the Federal Office for the Environment in 2010 (Basel), and by Acoucité in 2012 (Grenoble). The three models were similar in terms of quality of input data and noise modelling software, and were compliant with the European Noise Directive 2002/49/EC. Models included information on topography, height of buildings, noise barriers, reflections, ground effects, and Annual Average Daily Traffic (AADT). The traffic composition was based on either models or measurements. Low traffic densities were considered and modelled for Basel, and based on standard values for Girona and Grenoble. The resolution of the noise data was at least 1 dB(A). In addition, for Girona, the noise model was validated with 120 noise measurements carried out for 20 min, with a validation R^2 of 0.93 and a maximum difference of 3 dB(A) between the measured and modelled values.[15]

Noise levels at the three cities were calculated with CadnaA software. The local calculation method STL86+ was used for the Swiss model (SonBase), whereas NMPB-routes 96 and 2008 were applied in Girona and Grenoble, respectively.

Noise levels were estimated at receptor points close to the measurement sites within each of the road traffic noise models. We obtained long-term noise estimates during the day (L_{day}) and night (L_{night}) for Girona and Basel and also for the evening for Grenoble. In order to avoid noise reflections from building façades, as done for the measurements, noise predictions in Girona were estimated at 2 m from the façade of the site's postal address (with the exception of one site, which had a missing value owing to geocoding problems). Owing to the limitations in accessing the models, this procedure was not possible to be applied in Basel and Grenoble. Instead, noise predictions were estimated at the façade of the building closest to the measurement site in Basel (mean distance = 9.8 m, SD = 8.5) and at the exact location of the geocodes in Grenoble. The three road traffic noise models also had differences in other characteristics such as the surface considered for ground effects or the calculation height. These and other specific characteristics of the three models are summarized in Supplementary Table S1.

For comparison purposes, and given that short-term noise measurements were taken during non-rush hours, we also derived the A-weighted average level of the long-term noise predictions for 24 h ($L_{24\,h}$) as the time-weighted logarithmic mean of L_{day} and L_{night}:

$$L_{24h} = 10 \cdot \log_{10} \frac{1}{24}\left(16 \cdot 10^{\frac{L_{day}}{10}} + 8 \cdot 10^{\frac{L_{night}}{10}}\right)$$

Determinants of Spatial Distributions of Road Traffic Noise

Two types of predictor variables were used to explain the spatial distribution of our short-term road traffic noise measurements: variables collected using GIS and variables collected while visiting the measurement sites. The former are appealing as those are often available from publicly accessible data, whereas the latter require site visits and standard protocols.

Building and population density, digital road networks (with full coverage of all existing roads), bus networks, and land coverage were the GIS source data available for computing the GIS variables (defined in Table 1). The characteristics of the GIS source data are shown in Supplementary Table S2. Because digital road networks were obtained from different data sources, the definition of major roads was different in each city: roads up to class 2 (that is, at least 4 m wide with traffic in both directions) in Basel; roads from class 0 (motorways) to class 4 (local connecting roads) in Girona; and roads with AADT > 5,000 vehicles/day in Grenoble. For traffic-related variables (that is, those calculated using road and bus networks), we calculated buffers with radii of 50, 100, 200 and 500 m around each site. For building and population density and land use, we calculated buffers of 100, 200, 500 and 1,000 m. Considering the

© 2015 Nature America, Inc.

Table 1. Description of potential predictor variables collected from GIS data ($n=46$) and through measurements made during the noise measurement at Tri-Tabs sites ($n=17$).

Variable description (unit)	*Variable name*	*Buffer radii in m*	*Comments*
GIS variables (n = 46)			
Population density (number)	POP	100, 200, 500, 1000	
Building density (number)	BUILD	100, 200, 500, 1000	
Length of all roads (m)	ROADS	50, 100, 200, 500	
Length of major roads (m)	MAJORROADS	50, 100, 200, 500	
Total traffic load of all roads (veh·day/m)	TRAFFICLOAD	50, 100, 200, 500	Missing in Basel at some sites
Total traffic load of major roads (veh·day/m)	MAJORTRAFFICLOAD	50, 100, 200, 500	Missing in Basel at some sites
Average daily traffic at nearest road (veh/day)	TRAFNEAR	NA	Missing in Basel at some sites
Average daily traffic at nearest major road (veh/day)	TRAFMAJOR	NA	Missing in Basel at some sites
Distance to nearest road (m)	DISTNEAR	NA	Only used in Grenoble
Distance to nearest major road (m)	DISTMAJOR	NA	
Length of bus lines (m)	BUSLINES	50, 100, 200, 500	Only in Girona and Grenoble
Distance to nearest bus line (m)	DISTBUS	NA	Only in Girona
Distance to nearest bus stop (m)	DISTBSTOP	NA	Only in Girona
Residential land use (m^2)	LAND_RESID	100, 200, 500, 1000	
Industrial land use (m^2)	LAND_IND	100, 200, 500, 1000	
Agricultural, semi-natural, and forest land use (m^2)	LAND_AGRIC	100, 200, 500, 1000	

Variable description	*Variable name*	*Unit/Category*	*Comments*
Variables collected while visiting the noise measurement sites (n = 17)			
Type of site	SITE_TYPE	Traffic Background	
Type of neighbourhood	NEIGHBOURHOOD	Seven categories based on the type of buildings and building density	Only in Basel and Girona
Canyon-like street	CANYON	Yes No	Only in Grenoble
Street width (including sidewalks)	ST_WIDTH	Meters	
Speed of cars	SPEED	Slow Normal	
Inclination of the street	INCLINATION	Degrees	
Number of parking lanes	P_LANES	Number	
Number of running lanes	R_LANES	Number	
Traffic direction	T_DIRECTION	One way Both ways	
Height of the building at measurement site	HEIGHT_SITE	Meters	
Height of the building on the opposite side of measurement site	HEIGHT_OPP_SITE	Meters	
Mean height of buildings on the side of measurement site	MEAN_HEIGHT_SITE	Meters	
Mean height of buildings on the opposite side of measurement site	MEAN_HEIGHT_OPP_SITE	Meters	
Count of light vehicles in 20 min	AUTOVEHFLOW	Number	
Count of heavy vehicles in 20 min	TRUCKFLOW	Number	
Count of motorbikes in 20 min	MOTOFLOW	Number	
Sum of all vehicles and motorbikes in 20 min	TOTVEHFLOW	Number	

relatively small size of the study areas and the characteristics of existing noise propagation models, we did not use buffers over 1,000 m.

A second group of variables was obtained while performing the field work, to characterize each measurement site in detail (Table 1). None of these variables were available in GIS format when the study was performed. They related to street characteristics (street width, number of running and parking lanes, canyon-like street, speed of vehicles, street inclination, and traffic direction), building configuration in the street (height of buildings at both street sides) and surroundings (type of neighbourhood), and site characteristics (type of site depending on traffic intensity). Short-term measurements of road traffic flow (light vehicles, heavy vehicles, and motorbikes), conducted as part of the Tri-Tabs study objectives, were also included in this group. We averaged the repeated measurements of traffic flow because of their very high temporal correlation (intraclass correlation coefficient $r=0.93$–0.94). A detailed description of the collection of these variables is reported in the Supplementary Material.

Development of LUR Models

Linear regression models were developed for each city, following the standardized approach previously applied within the ESCAPE project in a large number of European study areas.[11,13] The supervised forward stepwise selection procedure started with univariate regression analyses with all the predictor variables. Some traffic-related variables had been previously log-transformed because they showed an exponential relationship with noise measurements. The predictor giving the highest explained variance (adjusted R^2) and a plausible direction of effect (defined *a priori*) was included in the model. Subsequent variables were included in the model provided that, with each inclusion, the absolute increase in adjusted R^2 was $>1\%$, the coefficient of the new variable had the predefined direction of effect, and the direction of effect for variables already in the model did not change. As a final step, variables with *P*-value >0.1 were sequentially removed.

After developing the models, regression diagnostics were conducted to identify influential values and multicollinearity among predictors, and residuals were tested for heteroscedasticity. Leverage *vs* residual squared plots and Cook's *D* statistics were used to detect the influential values. A maximum Variance inflation factor (VIF) of 3 was chosen as cutoff value for indicating multicollinearity (common rule of thumb for multicollinearity is VIF >5). Models' residuals were tested for spatial autocorrelation using Moran's *I*.

Model performance was evaluated by leave-one-out crossvalidation (LOOCV), that is, one of the sites was removed, the model was re-fitted,

© 2015 Nature America, Inc.

Table 2. Descriptive statistics[a] of the main characteristics of the Tri-Tabs sites, short-term noise measurements and long-term noise estimates ($L_{24\,h}$) derived from the standard models developed by the local authorities.

	Basel (n = 60)		*Girona* (n = 40)		*Grenoble* (n = 41)	
	Mean (SD)	*P10–P50–P90*	*Mean (SD)*	*P10–P50–P90*	*Mean (SD)*	*P10–P50–P90*
Distance to nearest major road	56.3 (72.6)	2.2–21.7–184.6	168.3 (188.9)	11.6–120.3–477.8	54.4 (69.4)	1.1–14.4–171.4
Length of all roads within 50 m	171.2 (57.8)	98.4–179.8–239.4	222 (71.7)	134.6–219.3–318.4	97.8 (65.9)	1–98.8–167.2
Population density within 100 m	875.4 (778.5)	2–673.5–2102.5	277.3 (263.2)	31.1–194.7–606.1	819.2 (612.2)	156.1–550.2–1757.6
Height of the building at measurement site	12.3 (5.4)	6.6–11.2–18.2	11.9 (5.3)	6–12.2–18.7	12.7 (8.9)	7–13–25
Total vehicles count, 20 min	76.8 (71.5)	8.8–65.3–169.2	111.5 (134.9)	6.3–46.5–325.8	215.1 (150.1)	10.5–207–445.5
Measured traffic noise, 20 min (dB(A))[b]	61.1 (6.4)	52.9–62.4–68.8	63.9 (6.2)	54.9–64.6–72.3	64.5 (6.7)	52.8–66.9–71.0
Modelled traffic noise, $L_{24\,h}$ (dB(A))	52.3 (9.4)	39.4–54.2–64.5	64.8 (5.7)	56.3–64.8–72.6	62.9 (9.8)	47.6–67.2–72.1

[a]Mean, standard deviation, and percentiles 10, 50, and 90. [b]Measured during daytime.

and the noise level at the omitted site was estimated. After repeating this procedure for all the sampling sites, the LOOCV R^2 was calculated as a measure of the model performance and the prediction error was expressed as the root mean squared error (LOOCV RMSE).

Two LUR models were developed for each city. The first model considered only predictor variables derived with GIS ("GIS-only" model). This model could be used for noise mapping and could be applied to predict noise levels at any location within the study areas. The second model was developed using all the potential predictor variables available at the Tri-Tabs sites (that is, both GIS variables and variables collected while visiting the measurement sites) in order to obtain the best possible model ("Best" model) in terms of explained variance, but with a limited application to generally estimate noise levels and maps as the extended set of variables was available only at the measurement sites.

Comparative Analyses

We performed two comparative analyses. First, we calculated Spearman's correlation coefficients to assess the correlation between short-term road traffic noise measurements in Tri-Tabs sites and long-term estimates from noise models developed by the local authorities. We also fitted a linear regression between measures and estimates, and plotted the residuals *vs* the fitted values.

The second analysis compared the long-term estimates from road traffic noise models developed by the local authorities with the noise estimates derived from the LUR models specifically developed for this study, possibly to use as estimates of long-term noise exposure as well. A linear regression was fitted using the estimates from road noise models developed by the local authorities as the dependent variable. The residuals were tested for heteroscedasticity and spatial autocorrelation, and plotted in maps in order to identify potential geographical patterns.

Statistical analyses were performed using STATA 12.1 (Stata Corp., College Station, TX, USA).

RESULTS

Summary statistics of the main characteristics of the Tri-Tabs sites, short-term noise measurements and long-term noise estimates from the standard models developed by the local authorities are reported in Table 2. In Girona, short-term measurements and long-term estimates had a very similar distribution and were highly correlated ($r = 0.75$). The correlation was similar in Grenoble ($r = 0.78$) but the distribution showed lower noise estimates as compared with short-term measurements in the lower percentile. The correlation in Basel was lower than in the other two cities ($r = 0.62$).

In Basel, short-term measurements generally showed higher noise levels as compared with noise estimates developed by the local SonBase model. This underestimation of the local noise model can be visualized in Figure 1. In the residuals *vs* fitted plots, the distribution of the residuals is wider for higher fitted values in Basel, which correspond mainly to traffic sites. In Girona and especially in Grenoble the opposite situation occurred, that is, the distribution of the residuals was wider among background sites.

Table 3 summarizes the LUR models for each city. Models considering only GIS variables as predictors ("GIS-only" models) explained 66%, 87% and 73% of the spatial variability in noise levels in Basel, Girona and Grenoble, respectively. In Basel, the site with the highest value for MAJORROADS_50 had a Cook's *D* of 0.9. This predictor variable included segments of a highway that run underneath the city at this location, which may interfere in the relationship between the predictor and noise levels. For this reason the site was removed from the analysis, although the exclusion did not substantially affect either the model parameters or its performance (data not shown). Only variables derived using road networks and bus networks entered in the three models, including simple distance metrics such as DISTMAJOR (which explained by itself 61% of the spatial variability in Basel and 53% in Girona and Grenoble). Despite being all traffic-related variables, they had acceptable multicollinearity as indicated by the VIF values. The scatter plots of the modelled *vs* measured noise levels (Figures 2 and 3) showed a cluster of sites around a modelled noise level of 65 dB(A) in Basel. This is due to the distribution of the predictor variable MAJORROADS_50, which has a value between 98 and 100 m for ~25% of the sites. For Grenoble, the residuals *vs* fitted plots in Figure 2 show some heteroscedasticity, but it is not severe.

The "Best" models were derived considering both GIS variables and additional variables collected while visiting the measurement sites. In Basel and Girona, the improvement in terms of adjusted R^2 was very moderate (Table 3). In Basel, the "Best" model included the short-term measurements of heavy vehicles flow, which probably would have not been necessary if GIS variables with traffic intensity data (see Table 1) would have been available at all the measurement sites. In Girona, the "Best" model included the height of the site, which was missing at two sites. In Grenoble, none of the additional variables collected in the field work entered in the "Best" model, therefore the "Best" model and the "GIS-only" model are the same.

The comparison between noise levels estimated with the LUR models developed with GIS variables and estimated noise levels from standard models developed by the local authorities is shown in Figure 4. LUR estimates of noise in Basel were higher than those from the SonBase model in almost all the sampling sites, as expected, given that LUR models were built using the short-term measurements as dependent variable. Similarly to what was found in Figures 2 and 3, some sites were clustered around a LUR estimate of 65 dB(A), for the reasons mentioned above. In Girona there was a high correlation between the noise estimates from the two models, with very similar absolute values and with differences that did not strongly depend on the noise levels. Grenoble

© 2015 Nature America, Inc.

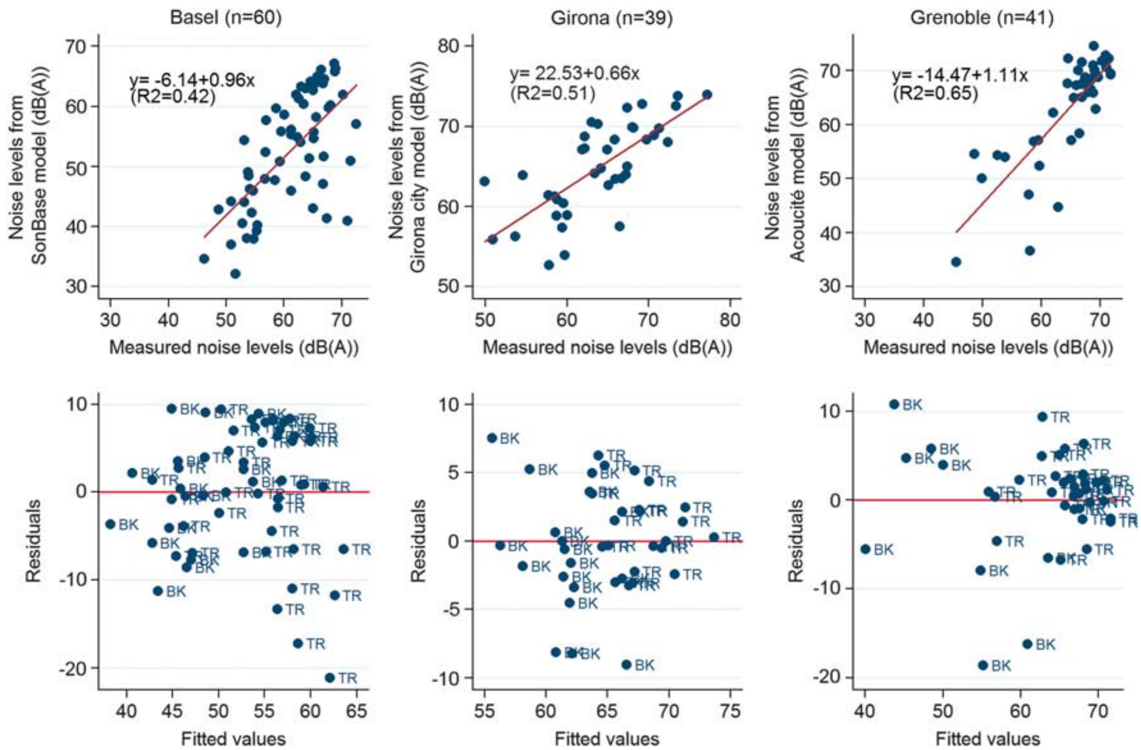

Figure 1. Comparison between short-term noise measurements in the three cities and long-term noise estimates from standard models developed by the local authorities ($L_{24\,h}$ in dB(A)). In the residuals *vs* fitted plots, traffic and background sites are labelled as TR and BK, respectively.

Table 3. Land use regression models for predicting noise levels measured at Tri-Tabs sites, using (a) only GIS variables ("GIS-only" model), and (b) both GIS variables and variables collected while visiting the Tri-Tabs sites ("Best" model).

City	*N*	*Determinants*	*Range*[a]	*Coefficient*	*VIF*[b]	*Seq* R^2 [c]	*LOOCV* R^2 [d]	*LOOCV RMSE*[e]
"GIS-only" model								
Basel	59	DISTMAJOR	0.3; 301.2	−0.043	2.25	0.60	0.66	3.81
		MAJORROADS_50	1; 225	0.045		**0.66**		
Girona	40	TRAFNEAR (log-transformed)	4.2; 10.4	1.915	2.52	0.72	0.84	2.47
		MAJORROADS_500	1; 4730	0.001	1.20	0.80		
		TRAFLOAD_50	38441; 3952260	1.82×10^{-6}	2.34	0.85		
		DISTBUS (log-transformed)	0.3; 5.5	−0.657	1.37	**0.87**		
Grenoble	41	DISTNEAR	0.1; 180.9	−0.078	1.73	0.62	0.69	3.76
		MAJORROADS_50 (log-transformed)	0; 5.3	1.305		**0.73**		
"Best" model								
Basel	60	DISTMAJOR	0.3; 301.2	−0.041	2.10	0.61	0.68	3.65
		TRUCKFLOW (log-transformed)	−2.3; 3.0	1.485	1.92	0.68		
		MAJORROADS_50	1; 316.5	0.023	1.43	**0.70**		
Girona	38	TRAFNEAR (log-transformed)	4.2; 10.4	1.863	2.45	0.72	0.88	2.25
		MAJORROADS_500	1; 4730	0.002	1.16	0.80		
		TRAFLOAD_50	38441; 3952260	2.11×10^{-6}	2.25	0.85		
		HEIGHT_SITE	0; 25.4	0.20	1.13	**0.89**		
Grenoble	41	DISTNEAR	0.1; 180.9	−0.078	1.73	0.62	0.69	3.76
		MAJORROADS_50 (log-transformed)	0; 5.3	1.305		**0.73**		

[a]Range of values of each predictor variable in the sample of noise measurement sites. [b]VIF, variance inflation factor . [c]Sequential R^2. Bold font indicates final R^2 for each model . [d]LOOCV R^2, leave-one-out crossvalidation R^2. [e]LOOCV RMSE, leave-one-out crossvalidation root mean square error.

showed the largest correlation between the two noise estimates ($R^2 = 0.76$), but the differences between them depended on the noise levels: The LUR estimates tended to be higher than estimates obtained from the Acoucité model at lower noise levels.

The residuals of the relationship between noise estimates from LUR models and noise estimates from standard models developed by the local authorities were plotted in maps of the three study areas to visualize their spatial distribution (Supplementary Figure

© 2015 Nature America, Inc.

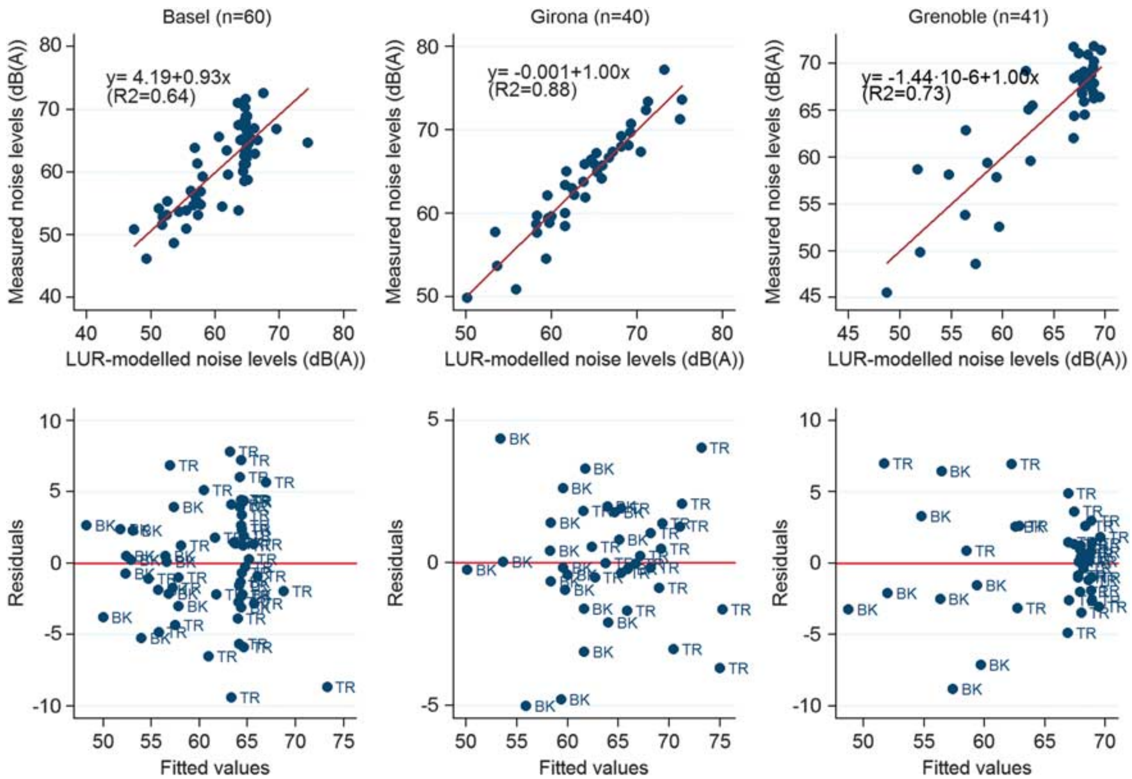

Figure 2. Scatter plots of noise estimates from Tri-Tabs LUR models (using the "GIS-only" model) *vs* short-term noise measurements at the Tri-Tabs sites. In the residuals *vs* fitted plots, traffic and background sites are labelled as TR and BK, respectively.

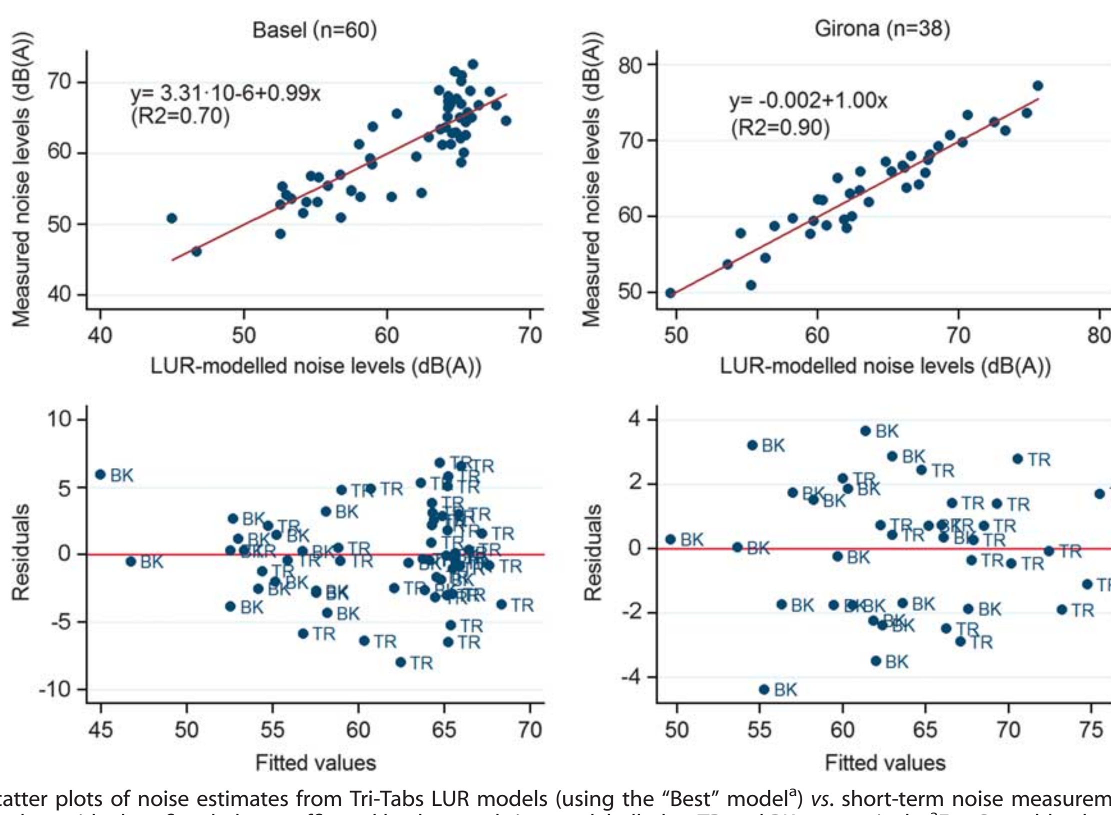

Figure 3. Scatter plots of noise estimates from Tri-Tabs LUR models (using the "Best" model[a]) *vs*. short-term noise measurements at the Tri-Tabs sites. In the residuals *vs* fitted plots, traffic and background sites are labelled as TR and BK, respectively. [a]For Grenoble, the "Best" model is equal to the "GIS-only" model.

© 2015 Nature America, Inc.

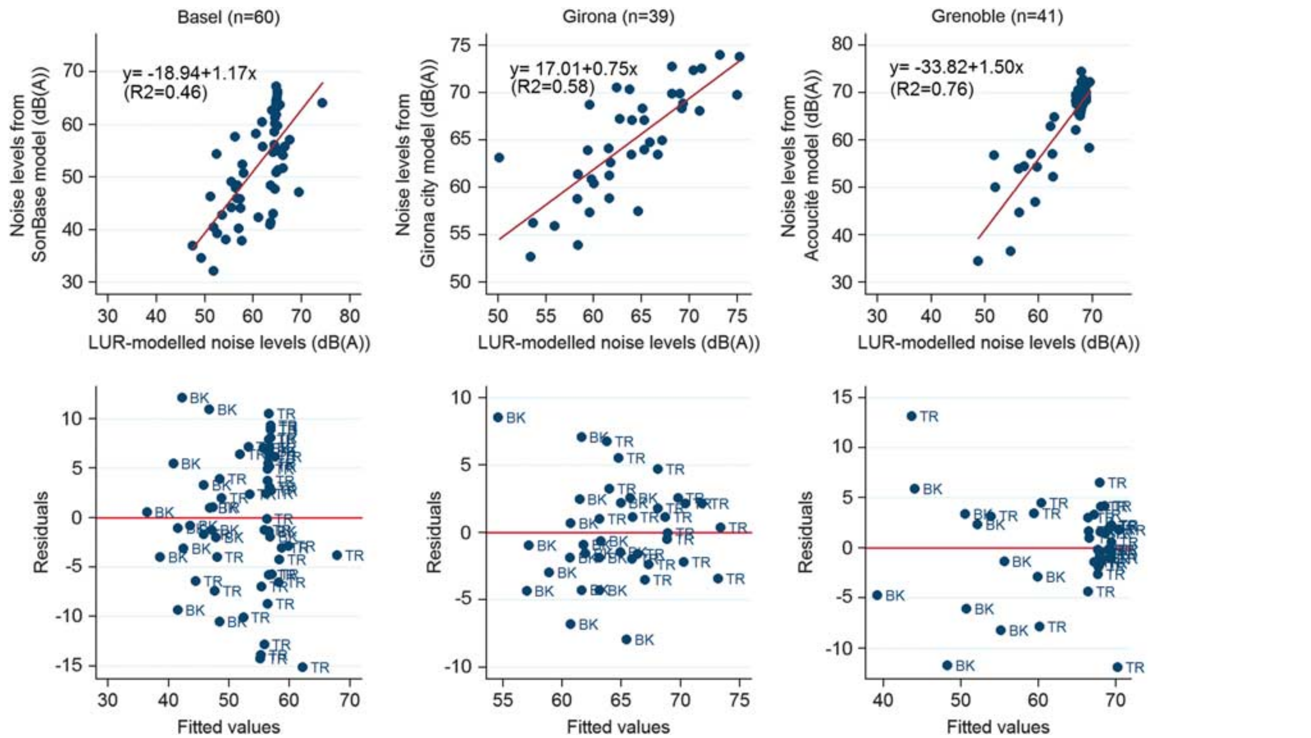

Figure 4. Comparison between the noise estimates from Tri-Tabs LUR models (using the "GIS-only" model) and long-term noise estimates derived from standard models developed by local authorities. In the residuals *vs* fitted plots, traffic and background sites are labelled as TR and BK, respectively.

S1). We did not observe any clear geographical pattern, and the Moran's *I* statistic indicated that there was no spatial autocorrelation of the residuals in any of the three cities.

Finally, as previously mentioned, the "GIS-only" models can also be used for noise mapping. As an example, a map of the noise levels in Girona is shown in Figure 5. It was created by deriving the noise estimates for the centroids of 50-m grid cells over the study area, and then converting the centroids into a 50×50 m raster map. The size of the grid was chosen by taking into account the smallest buffer size used for the LUR modelling.

DISCUSSION

We explored for the first time in Europe the application of LUR modelling (widely used for air pollution exposure assessment) to assess the long-term spatial variability of road traffic noise in three European cities. "GIS-only" models (that is, those using only predictor variables derived with GIS) explained between 66% and 87% of the spatial variability in the measured short-term noise levels and included traffic-related variables only. These percentages of explained variation are similar to those reported for traffic-related air pollution in other European studies that developed LUR models from a similar number of sampling sites.[11,16–18] The increase in R^2 values obtained in the "Best" models (i.e. those using GIS variables and other variables collected visiting the measurement sites) was very slight as compared to the "GIS-only" models. Given that data collection during field work is always an investment of time and resources, our results indicate that it is possible to develop LUR models for explaining intraurban noise variability without the need to take additional information at the measurement sites. Besides, a LUR model with only GIS variables can be used for noise mapping as illustrated in Figure 5.

Short-term noise measurements made at the Tri-Tabs sites were highly correlated with the noise estimates from the standard models developed by the local authorities ($r = 0.62–0.78$). Although noise estimates from standard models represent long-term noise levels, there is evidence that short-term noise measurements (taken during non-rush hours) are reasonable surrogates of longer-term measures.[19–20] A similar correlation coefficient ($r = 0.62$) was found in a study in the metropolitan area of Vancouver where 5-min measurements at 103 sites were compared with noise estimates from a noise model.[21]

Because noise measurements were performed during non-rush hours, we assumed they were comparable with $L_{24\,h}$ estimates from standard models, following the same rationale as with traffic counts.[22] Although $L_{24\,h}$ estimates are on average 1.4 dB(A) lower than L_{day}, using L_{day} instead of $L_{24\,h}$ would have shown the same results in terms of correlation coefficients and spatial variation of the noise estimates, as the correlation between L_{day} and $L_{24\,h}$ was 1 in the three cities. This is the case because nighttime levels are commonly derived using a day/night ratio in traffic flows.[8]

Because noise prediction models and noise maps based on the European Directive are designed to identify noisy areas, they may show a poorer performance at locations with lower noise levels,[8] which would explain the wider distribution of the residuals at background sites in Girona and Grenoble shown in Figure 1. In Basel, however, we found the opposite situation (that is, the distribution of the residuals was wider for higher fitted values), as well as a systematic underestimation of the Swiss SonBase model developed by the local authorities as compared with our short-term noise measurements. These two findings can be due to the fact that, of the Tri-Tabs cities, only Basel has trams running on many streets, providing an independent source of noise in the noise measurements that was not possible to fully remove during the data cleaning process. This source was not integrated in the SonBase model as it is specific for road traffic. Other characteristics of the SonBasel model, such as the use of a grassland surface for ground effects, can additionally explain this underestimation.

To our knowledge, this is the first study introducing LUR modelling for noise exposure assessment in Europe. Our results in

© 2015 Nature America, Inc.

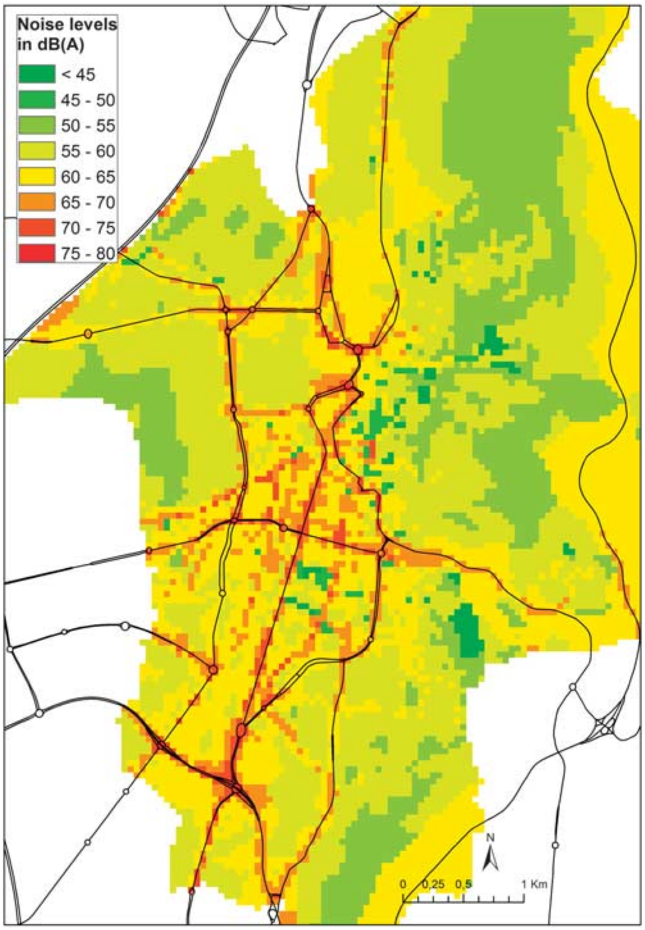

Figure 5. Map of the estimated noise levels using LUR modelling in Girona.

terms of explained variance of noise levels at the within-city scale are comparable to those reported by the only study on noise LUR modelling published so far, conducted in the Chinese municipality of Dalian and at three different spatial scales, intraurban being one of them.[14] But the comparison of the model performance between the two studies is difficult, given the differences in the selection of sampling sites (mainly routine monitoring sites in Dalian *vs* purpose-designed sites in Tri-Tabs cities), size of the sampling data set (101 sites in Dalian *vs* a range of 39–60 sites in Tri-Tabs cities), and model validation method (validation data set in Dalian *vs* LOOCV in Tri-Tabs cities). For smaller sample sizes, it has been shown that the adjusted R^2 and the LOOCV estimate (as compared with an independent validation data set) tend to give higher and more inflated values, particularly if the number of potential predictors offered is high.[23] The size of our sampling data set was too small to have an independent validation data set. However, the number of predictor variables offered for developing our "GIS-only" models ($n=46$) was low if compared with those used in the recently published air pollution LUR models from the ESCAPE project.[11,13]

The accuracy of the measurement sites' geocodes is very important when using models with fine-scale spatial resolution to estimate environmental exposures. In the air pollution exposure field, it has been shown that the choice of the geocoding technique used to estimate individual exposures at the residential addresses may have an impact on the health effect estimates, even if the differences in exposure estimates are very small.[24] This issue is even more important when applied to noise exposure assessment, because the decrease of noise levels with distance is considerable: 3 dB(A) and 6 dB(A) per doubling of distance from a line (that is, moving) source and a point (that is, static) source, respectively.[25] The distance of the measurement sites from the building façades is also an important factor to avoid the influence of noise reflections. A subanalysis performed in Girona using the GIS variables derived from the original geocodes provided by the mapping application (that is, before relocating them at 2 m from the façade of their postal address) showed a change in the predictors' coefficients of the "GIS-only" model between 5% and 73%, resulting in a difference in the noise estimates from −0.5 to 1.3 dB(A). Also, the lower correlations observed in Basel between long-term noise predictions from SonBase model and both short-term noise measurements (Figure 1) and noise estimates from the Tri-Tabs LUR model (Figure 4) could be partly explained by the method used to obtain noise estimates from SonBase model. As noise predictions were estimated at the façade of the building closest to the measurement site, their accuracy was directly affected by the distance from the measurement site to the corresponding building.

The importance of the geocodes accuracy also applies to the precision of the input data. Similar to what happens with air pollution, LUR models may be unable to represent the extremely local variations in noise levels (over distances of tens of metres) that may occur near line sources, particularly if the input data have not enough quality and spatial resolution.[9] In this context, the buffer sizes of the predictor variables (particularly the traffic-related ones) is also an issue to explore more in detail in future studies.

Although LUR models should not be considered as a substitute of high-quality noise prediction models and more efforts must be done to promote and use the latter, the application of LUR models to estimate noise exposures in epidemiological research may have advantages in some specific situations. In the EU, despite the effort that has been already done in establishing common noise assessment methods to produce strategic noise maps, access to noise estimates for epidemiological studies may be complicated depending on the specific data needs. Besides, strategic noise maps are not required for smaller agglomerations (that is, <100,000 inhabitants). It may also happen that noise estimates from existing noise models can be obtained only through spatial linkage with the digital noise maps, resulting in estimates of poorer quality. In all these cases, the use of LUR models could potentially solve some of these problems. The development of LUR models may also be of interest in urban areas from other countries where noise is clearly a public health problem but there are no noise maps available from competent authorities, or they are not sufficiently precise to be used in epidemiological studies. In addition, LUR models can be developed for time windows not covered by the standard noise maps.

Because noise levels tend to be quite stable over time, and short-term measurements can provide a reliable estimate of longer-term levels,[19–20] another advantage of developing LUR models for noise exposure assessment in epidemiological studies is that, unlikely for air pollution, noise LUR models do not require the performance of measurements during periods of several days or weeks and several campaigns over the year.

Unlike noise prediction models, LUR models are built from real noise measurements taken at a given sample of sites. Because it is not possible to document that the association remains linear outside the range of predictor variables observed at the sampling sites, the application of LUR models to estimate noise levels at locations where the predictor variables fall outside the range of observed values is limited. Therefore, when possible the measurement sites should be carefully selected in order to be representative for the purpose of application of the LUR model.[26]

Because we did not have an independent data set of noise measurements, it was not possible to formally compare the performance of the LUR models developed in the present study against

© 2015 Nature America, Inc.

the standard noise models developed by the local authorities. Results indicate a good correlation between the two estimates, but with differences depending on the noise levels, particularly in Basel and Grenoble. In Girona, both models provided very similar absolute values of noise, despite having been developed using different sources of traffic intensity data (as shown in Supplementary Tables S1 and S2). As the comparison was performed at the noise measurement sites used for developing the LUR models, the correlation would probably have been lower if performed in other locations, such as in a data set of study participants' addresses. Further research is needed to assess the performance of LUR models in comparison with current noise prediction models for epidemiological analyses.

In conclusion, LUR modelling with accurate GIS source data can be a promising tool for noise exposure assessment with potential applications in epidemiological research, particularly in areas where noise prediction models or noise maps from competent authorities are not available.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The study was supported by ADEME (French Environment and Energy Management Agency; TRI-TABS study contract 0966C0304). Inmaculada Aguilera holds a postdoctoral fellowship awarded from the Catalan Agency for Management of University and Research Grants (BP-DGR-2011-00214). We thank Alex Ineichen (SwissTPH), Marta Cirach (CREAL), and Cécile Fontagné (Air Rhône-Alpes) for their assistance in deriving the GIS variables; n-Sphere (Zürich, Switzerland) and Acoucité Noise Observatory (Lyon, France) for providing the modelled noise estimates in Basel and Grenoble, respectively; and La Métro city council (Grenoble) for providing access to local data.

REFERENCES

1 European Environment Agency. Transport at a crossroads TERM 2008: indicators tracking transport and environment in the European Union. EEA report. Office for Official Publications of the European Communities: Luxembourg, 2009. Available at <http://www.eea.europa.eu/publications/transport-at-a-crossroads> (last accessed 5 May 2013).
2 Basner M, Babisch W, Davis A, Brink M, Clark C, Janssen S *et al.* Auditory and non-auditory effects of noise on health. *Lancet* 2014; **383**: 1325–1332.
3 Babisch W. Transportation noise and cardiovascular risk: updated review and synthesis of epidemiological studies indicate that the evidence has increased. *Noise & Health* 2006; **8**: 1–29.
4 Floud S, Vigna-Taglianti F, Hansell A, Blangiardo M, Houthuijs D, Breugelmans O *et al.* Medication use in relation to noise from aircraft and road traffic in six European countries: results of the HYENA study. *Occup Environ Med* 2011; **68**: 518–524.
5 Stansfeld SA, Berglund B, Clark C, Lopez-Barrio I, Fischer P, Ohrström E *et al.* Aircraft and road traffic noise and children's cognition and health: a cross-national study. *Lancet* 2005; **365**: 1942–1949.
6 Van Kempen E, Babisch W. The quantitative relationship between road traffic noise and hypertension: a meta-analysis. *J Hypertens* 2012; **30**: 1075–1086.
7 European Commission. Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 relating to the assessment and management o environmental noise. Off J Eur Commun L189, 12–25.
8 European Commission's Working Group - Assessment of Exposure to Noise (WG-AEN). Position paper: Good practice guide for strategic noise mapping and the production of associated data on noise exposure. 2003. Available at <http://ec.europa.eu/environment/noise/pdf/best_practice_guide.pdf> (last accessed 18 March 2013).
9 Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P *et al.* A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos Env* 2008; **42**: 7561–7578.
10 Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T *et al.* A review and evaluation of intraurban air pollution exposure models. *J Expo Anal Environ Epidemiol* 2004; **15**: 185–204.
11 Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X *et al.* Development of NO_2 and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe: The ESCAPE project. *Atmos Env* 2013; **72**: 10–23.
12 de Hoogh K, Wang M, Adam M, Badaloni C, Beelen R, Birk M *et al.* Development of land use regression models for particle composition in twenty study areas in Europe. *Environ Sci Technol* 2013; **47**: 5778–5786.
13 Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, Cirach M *et al.* Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM coarse in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol* 2012: **46**: 11195–111205.
14 Xie D, Liu Y, Chen J. Mapping urban environmental noise: a land use regression method. *Environ Sci Technol* 2011: **45**: 7358–7364.
15 Deltell A. *Elaboració del mapa acústic de trànsit de la ciutat de Girona. Degree Project Dissertation [in Catalan]*. Universitat de Girona: Spain, 2005.
16 Aguilera I, Sunyer J, Fernández-Patier R, Hoek G, Aguirre-Alfaro A, Meliefste K *et al.* Estimation of outdoor NO_x, NO_2, and BTEX exposure in a cohort of pregnant women using land use regression modeling. *Environ Sci Technol* 2008; **42**: 815–821.
17 Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer P, Gehring U *et al.* Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* 2003; **14**: 228–239.
18 Morgenstern V, Zutavern A, Cyrys J, Brockow I, Gehring U, Koletzko S *et al.* Respiratory health and individual estimated exposure to traffic-related air pollutants in a cohort of young children. *Occup Environ Med* 2007; **64**: 8–16.
19 Allen RW, Davies H, Cohen MA, Mallach G, Kaufman JD, Adar SD. The spatial relationship between traffic-generated air pollution and noise in 2 US cities. *Environ Res* 2009; **109**: 334–342.
20 Davies HW, Vlaanderen JJ, Henderson SB, Brauer M. Correlation between co-exposures to noise and air pollution from traffic sources. *Occup Environ Med* 2009; **66**: 347–350.
21 Gan WQ, McLean K, Brauer M, Chiarello SA, Davies HW. Modeling population exposure to community noise and air pollution in a large metropolitan area. *Environ Res* 2012; **116**: 11–16.
22 Rivera M, Basagaña X, Aguilera I, Agis D, Bouso L, Medina-Ramón M *et al.* Spatial distribution of ultrafine particles in Girona, Spain: a land-use regression model. *Atmos Environ* 2012; **54**: 657–666.
23 Basagaña X, Rivera M, Aguilera I, Agis D, Bouso L, Elosua R *et al.* Effect of the number of measurement sites on land-use regression models in estimating local air pollution. *Atmos Environ* 2012; **54**: 634–642.
24 Jacquemin B, Lepeule J, Boudier A, Arnould C, Benmerad M, Chappaz C *et al.* Impact of geocoding methods on associations between long-term exposure to urban air pollution and lung function. *Environ Health Perspect* 2013; **121**: 1054–1060.
25 Lee CSY, Fleming GG, Measurement of highway-related noise. Report No. FHWA-PD-96-046 and DOT-VNTSC-FHWA-96-5. John A. Volpe National Transportation Systems Center, Acoustics Facility, Cambridge, MA: May 1996.
26 Wang M, Beelen R, Eeftens M, Meliefste K, Hoek G, Brunekreef B. Systematic evaluation of land use regression models for NO_2. *Environ Sci Technol* 2012; **46**: 4481–4489.

Supplementary Information accompanies the paper on the Journal of Exposure Science and Environmental Epidemiology website (http://www.nature.com/jes)

© 2015 Nature America, Inc.