Environmental Research 199 (2021) 111231

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Land use regression modelling of community noise in São Paulo, Brazil

Michelle Raess [a,b], Alexandra Brentani [c], Bartolomeu Ledebur de Antas de Campos [a,b], Benjamin Flückiger [a,b], Kees de Hoogh [a,b], Günther Fink [a,b], Martin Röösli [a,b,*]

[a] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland
[b] University of Basel, Basel, Switzerland
[c] Department of Pediatrics at the Medical School of São Paulo University, São Paulo, Brazil

ARTICLE INFO

Keywords:
Noise measurement
Community noise
Land use regression
São paulo
Noise exposure

ABSTRACT

Noise pollution has negative health consequences, which becomes increasingly relevant with rapid urbanization. In low- and middle-income countries research on health effects of noise is hampered by scarce exposure data and noise maps. In this study, we developed land use regression (LUR) models to assess spatial variability of community noise in the Western Region of São Paulo, Brazil.We measured outdoor noise levels continuously at 42 homes once or twice for one week in the summer and the winter season. These measurements were integrated with various geographic information system variables to develop LUR models for predicting average A-weighted (dB(A)) day-evening-night equivalent sound levels (L_{den}) and night sound levels (L_{night}). A supervised mixed linear regression analysis was conducted to test potential noise predictors for various buffer sizes and distances between home and noise source. Noise exposure levels in the study area were high with a site average L_{den} of 69.3 dB(A) ranging from 60.3 to 82.3 dB(A), and a site average L_{night} of 59.9 dB(A) ranging from 50.7 to 76.6 dB (A). LUR models had a good fit with a R^2 of 0.56 for L_{den} and 0.63 for L_{night} in a leave-one-site-out cross validation. Main predictors of noise were the inverse distance to medium roads, count of educational facilities within a 400 m buffer, mean Normalized Difference Vegetation Index (NDVI) within a 100 m buffer, residential areas within a 50 m (L_{den}) or 25 m (L_{night}) buffer and slum areas within a 400 m buffer. Our study suggests that LUR modelling with geographic predictor data is a promising and efficient approach for noise exposure assessment in low- and middle-income countries, where noise maps are not available.

1. Introduction

Rapid urbanization and life style changes have led to increasing noise exposure in many low- and middle-income countries (LMIC). Growing evidence suggests that noise may have negative health consequences and should therefore be considered a major threat for public health (World Health Organization, 2011). Chronic exposure to noise has been linked to several negative health outcomes, such as annoyance (Brink et al., 2016), sleep disturbance (Frei et al., 2014), cardiovascular diseases (Héritier et al., 2017; Seidler et al., 2016; Sørensen et al., 2012; Vienneau et al., 2015), diabetes mellitus (Eze et al., 2017; Sørensen et al., 2013), respiratory diseases (Recio et al., 2016), depression (Seidler et al., 2017; Zijlema et al., 2016) and cognitive impairment (Schlittmeier et al., 2015; van Kempen et al., 2010). Community noise can have its origin from multiple sources, such as transportation (road traffic, railway and airport), industry, construction sites and social sources (Singh and Davar, 2004; Sheng and Tang, 2011). According to recent estimates of the European Environment Agency at least 20% of the European Union population live in areas where traffic noise levels are considered harmful to health. 21.8 million residents are annoyed by noise and 6.5 million sleep disturbed (European Environment Agency. Environmental noise in Europe -, 2020) These significant health impacts are most likely to be underestimated, with new WHO evidence suggesting health effects at lower levels than used for this burden of disease estimates (World Health Organization. Environmental Noise Guidelines for the European Region, 2018).

Abbreviations: GIS, geographic information systems; LUR, Land use regression; SP-ROC, São Paulo Western Region Birth Cohort; OSM, Open Street Map; NDVI, Normalized Difference Vegetation Index; LAeq, A-weighted equivalent sound pressure levels; dB(A), A-weighted decibel; L_{den}, A-weighted day-evening-night equivalent sound levels; L_{night}, A-weighted night sound levels; GPS, Global Positioning system; LOSOCV, leave-one-site-out cross validation; LMIC, low and middle income countries.

* Corresponding author. Socinstrasse 5959, 4051, Basel, Switzerland.
E-mail address: martin.roosli@swisstph.ch (M. Röösli).

https://doi.org/10.1016/j.envres.2021.111231
Received 8 January 2021; Received in revised form 20 April 2021; Accepted 22 April 2021
Available online 7 May 2021
0013-9351/

For guiding policy makers' decisions on noise thresholds and protective actions, noise prediction models are necessary, because measuring individual noise exposure is not feasible from a logistical and budgetary point of view and would need a lot of extra effort for accurate source apportionment. Gold standard for transportation noise exposure assessment are sound propagation models, which are based on the components influencing sound emissions and propagation in the environment (Björk et al., 2006; Tang and Wang, 2007). In recent years land use regression (LUR) modelling has been increasingly employed (Ragettli et al., 2016; Harouvi et al., 2018; Aguilera et al., 2015; Chang et al., 2019; Liu et al., 2020). This methodology, which was initially developed for air pollution studies (Morley et al., 2015; de Hoogh et al., 2014; Kloog et al., 2015), uses multiple regression between the noise measurements and land use predictors for producing high-resolution noise maps. It is a promising and relatively inexpensive method for assessing noise exposure, especially for areas where high quality numerical models and corresponding input data do not exist and for urban areas with a large spatial noise variability. It was demonstrated, that LUR modelling can potentially explain spatial variability and without systematic difference when comparing with standard sound propagation noise models (Aguilera et al., 2015). To the best of our knowledge, only two studies have used a LUR model to assess noise exposure in LMICs, where availability of suitable noise emission data is challenging and noise sources may be different (Sieber et al., 2017a; Xie et al., 2011).

With a population of over 20 million, São Paulo is one of the largest cities in the world and showed a rapid population growth in the past years. However, very scarce city-wide noise models are available in Brazil. The aim of this study was to develop a LUR model using one-week outdoor noise measurements collected in summer and winter including local GIS predictors to assess the spatial variability of community noise levels in São Paulo. The modeled noise exposure will ultimately be used to investigate effects of community noise exposure on children's development using data from São Paulo Western Region Birth Cohort (SP-ROC) study (Brentani et al., 2020).

2. Materials and methods

2.1. Study design and study areas

Outdoor noise levels were measured at 42 homes of SP-ROC participants once or twice during one week in the summer and the winter season in the Butanta-Jaguaré region of São Paulo. São Paulo has a population of over 20 million people and is the largest city in Brazil and South America. The Butanta-Jaguaré region, which is located in the western part of São Paulo, has an area of 62 km^2 and approximately 480,000 inhabitants (Secretaria Municipal da Saúde, 2018). The study area comprises of residential, industrial and commercial areas with a large and congested road network, extended build-up environment but also large green areas. Residential areas vary between middle-class and informal settlements (called "favelas"). Measurement locations were selected from addresses of SP-ROC study participants following the design used in the ESCAPE study (Beelen et al., 2013). The homes were selected to represent a broad range of potential low to high noise exposure situations among the SP-ROC population (Fig. 1). By visual inspection on Google Earth, we considered various situations in terms of settlement type, distance to street, railways, industry, surrounding green space, education etc. (Table 1) with an equal distribution across the study area.

2.2. Noise measurements

Two measurement campaigns were conducted in 2019. The first measurement campaign was done from 12 to 19 February 2019 (summer season) and the second one from 7 to August 14, 2019 (winter season). In total, data were collected from 42 different homes. Thereof, 26 homes were measured twice, eight homes exclusively in summer and eight homes in the winter season only, because some sites could not be measured again, for various reasons such as missing informed consent, participant not at home, new construction site. A-weighted (loudness of sounds in air as perceived by the human ear) equivalent sound pressure levels (LAeq) averaged over one-second (1-s) intervals were continuously measured during one week using a Noise Sentry RT type-II sound level meter data logger (Convergence Instruments, Sherbrooke, QC, Canada). The installation and retrieval of the measurement devices was performed by several field work teams in parallel, therefore starting and ending times of measurement were maximum 4 h apart. The process of installation and retrieval occurred on Tuesdays, in order to capture the weekend noise, considering the noise variation along working days and weekends. In order to receive reliable measurements and to avoid theft, vandalism or unwanted noise sources, the study team was given clear

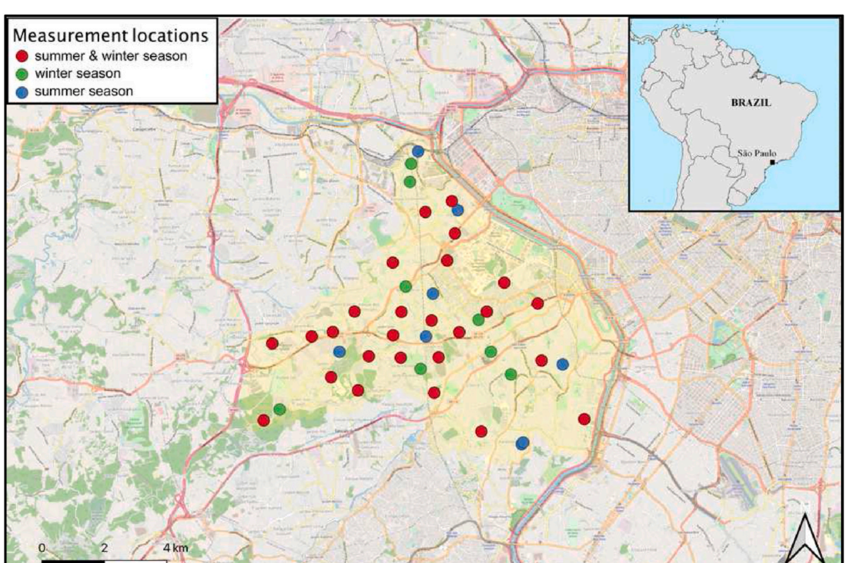

Fig. 1. Map of study area (yellow) and noise measurement locations. Out of a total of 42 locations, 26 locations were measured twice in summer and winter season (red), with additional 8 locations in summer season (blue) and winter season (green). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
List of LUR predictor variables.

Layer	Source	Year	Unit	Transformation	Expected effect
Land use (Green space, Industry, Train terminal, Residential, Waters)	OSM[a]/ESRI[b]	2019	Surface (ha)	25, 50, 100, 200, 400, 1000 buffer	+/−
Roads size S (motorway, trunk, primary, secondary, tertiary, residential roads) size M (motorway, trunk, primary, secondary, tertiary roads) size L (motorway, trunk, primary, secondary roads) size XL (motorway, trunk, primary roads) size XXL (motorway, trunk) separate road type: motorway/trunk/primary/secondary/tertiary/residential	OSM	2019	Distance (m)/ Length (m)	Distance and inverse distance to nearest road Road length within 25, 50, 100, 200, 400, 1000 buffer	+
NDVI	Landsat 8/ USGS[c]	2017	Mean index (−1 to 1)	25, 50, 100, 200, 400, 1000 buffer	−
Built-up environment	Landsat 8/ USGS	2017	Mean index (−1 to 1)	25, 50, 100, 200, 400, 1000 buffer	+
Traffic signals	OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest traffic signal Count within 25, 50, 100, 200, 400, 1000 buffer	+
Fuel stations	OSM[a]	2019	Distance (m)/Count	Distance/inverse distance to nearest fuel station Count within 25, 50, 100, 200, 400, 1000 buffer	+
Bus stations	OSM[a]	2019	Distance (m)/Count	Distance/inverse distance to nearest bus station Count within 25, 50, 100, 200, 400, 1000 buffer	+
Informal settlements („favelas“)	CEM[d]	2016	Surface (ha)	25, 50, 100, 200, 400, 1000 buffer	+
Shopping (Shopping Centers, Marketplaces)	Dados Abertos[e]/OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest shopping place Count within 25, 50, 100, 200, 400, 1000 buffer	+
Police stations	OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest police station Count within 25, 50, 100, 200, 400, 1000 buffer	+
Education (Schools, Kindergarten, College, (University))	OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest education place Count within 25, 50, 100, 200, 400, 1000 buffer	+
Gastronomy (Bars, Restaurants, Nightclubs, Fast food, Cafe, Cinema)	OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest gastronomy place Count within 25, 50, 100, 200, 400, 1000 buffer	+
Leisure (Playground, Sportcenters, Stadiums)	OSM[a]	2020	Distance (m)/Count	Distance/inverse distance to nearest leisure place Count within 25, 50, 100, 200, 400, 1000 buffer	+
Railway	OSM[a]	2020	Distance (m)	Distance/inverse distance to nearest railway station	+
Income	CEM[e]	2010	Mean (US$)	25, 50, 100, 200, 400, 1000 buffer	−
Household density	CEM[e]	2010	Density (n/ha)	25, 50, 100, 200, 400, 1000 buffer	+
Weekly NO2 passive sampling measurements	own data	2019	Concentration (ug/ m^3)	at measurement location	+
Hospital	OSM[a]	2019	Distance (m)	Distance/inverse distance to nearest hospital	+
Rain	Ecmwf[f]	2019	mm	at measurement location	+

Note: LUR, Land use regression; NDVI, Normalized Difference Vegetation Index; OSM, OpenStreetMap; CEM, Center for Metropolitan Studies; ha, hectare; m, meter; mm, milimeter

[a] OpenStreetMap contributors. OpenStreetMap [November 04, 2020]. Available from: https://www.openstreetmap.org.

[b] Environmental Systems Research Institute (Esri). [December 11, 2020]. Available from: https://www.esri.com.

[c] U.S. Geological Survey. Landsat 8 Mission [November 04, 2020]. Available from: https://www.usgs.gov.

[d] Centro de Estudos da Metrópole. Relatórios Favelas e Loteamentos - Estudo do CEM para Sehab/PMSP (English translation: Center for Metropolitan Studies (CEM). Report on Slums and Settlements - Study by CEM for the Municipal Housing Department (SEHAB)) [November 04, 2020]. Available from: http://centrodametropole.fflch.usp.br/pt-br/downloads-de-dados/relatorios-favelas-e-loteamentos-estudo-do-cem-para-sehabpmsp.

[e] Secretaria Municipal da Saúde 2018. Prefeitura da Cidade de São Paulo 2018 (English translation: Municipal Health Secretariat 2018. Prefecture of the City of São Paulo) [November 04, 2020]. Available from: http://www.prefeitura.sp.gov.br/cidade/secretarias/saude/.

[f] Instituto Brasileor de Geografia e Estatistica (English translation: Brazilian Institute of Geography and Statistics Foundation). Population Census 2010 [November 04, 2020]. Available from: https://www.ibge.gov.br.[7] European Centre for Medium-Range Weather Forecasts (ECMWF) [November 04, 2020]. Available from: https://www.ecmwf.int.

instructions on the device installations. Devices were mounted within the study participants' property, attached at external part of the home without direct local source affection such as air-conditioning, preferably at the most exposed façade. Special attention was also given to the height of microphone positioning (height of 2–3 m), since this was proven to be crucial (Montes González et al., 2020). Building structures varied substantially across areas, ranging from high rise apartment buildings in some areas to mostly small buildings in informal settlements (favelas). Each sampling site was geocoded with a Global Positioning System (GPS). Noise meters were calibrated before and after deployment in the field at 94 dB(A) with a Pulsar Acoustic Calibrator Model 105 Class 1. The devices have a noise floor (detection limit) at 31 dB(A), which is lower than the minimal levels measured during both measurement campaigns. However, two devices showed a higher lower detection limit of 37 and 49 dB(A). Measurements of these devices were checked with Robust regression on order statistics (ROS) method for censored measures. The maximum difference of various mean values (e.g. L_{night}, L_{den}, $L_{weekend}$ etc.) for these devices was found to be 0.04 dB(A). Therefore, no relevant bias was expected and these data were used without further adjustment.

For the analysis, measurements were restricted to the period from midnight after measurement begin to midnight before measurement end, in order to have the same times for each site, resulting in 6 successive days of continued noise monitoring per each site and measurement week. Virtually no missing data occurred (<0.1%). In order obtain robust mean values, upper outliers, defined as 1-s noise measurements exceeding the six-day mean by more than three standard deviations, were removed (3,953 out of 19,633,345 1-s measurements in summer season and 8,890 out of 20,271,184 in winter season). No lower outliers were removed, since quiet phases are expected to be always plausible. Additionally, measurements from three devices of summer season and from three devices of winter season were removed because of extreme recordings during a specific time period or suspicious, unexplainable patterns (see two examples in Supplementary Figure S1). L_{den} was calculated as an energetic mean over 6 consecutive days of the 1s-LAeq measurements at each site, by applying a penalty of 10 dB(A) for night time noise (23:00–7:00) and 5 dB(A) for evening noise (19:00–23:00). L_{night} refers to the energetic mean value of measurements from 23:00–7:00 and L_{Aeq24h} from 7:00–7:00 on the next day.

2.3. Predictor variables of noise exposure

GIS data related to factors relevant for noise exposure were collected for developing the LUR model and implemented in the program QGIS 3.4.15 (Table 1). Details on road geography and different facilities, such as schools, police stations, bus stations and gastronomic institutions were extracted from Open Street Map (OSM) (OpenStreetMap contributors, 2020) and were directly available in QGIS. The census of 2010 gave information about income and household density (Instituto Brasileor de Geografia e Estatistica. Population Census, 2010). Land cover information, such as the Normalized Difference Vegetation Index (NDVI, a substitute for green space) and built-up environment, were available from remote sensing data (based on Landsat 8 images acquired from the U.S. Geological Survey website (U.S. Geological Survey. Landsat 8 Mission [04.11, 2020)). Areas of informal settlements ("favelas") were defined by the Centro de Estudos da Metrópole (Centro de Estudos da Metrópole. Relatórios Favelas e Loteamentos - Estudo do CEM para Sehab/PMSP [04.11, 2020). The categories of land use were manually generated by applying aerial photographs from ESRI World Imagery (Environmental Systems Res, 2020) and data from OSM. The road system was categorized by OSM into motorway, trunk, primary, secondary, tertiary and residential roads. We made a further classification by combining these road types. We computed at each measurement site the distance to the GIS variable and/or the quantity of the GIS variable within circular buffers of different sizes (Table 1). We did not use predictors with more than 90% zero values which may happen for small buffer sizes.

In addition to GIS predictors, we also considered weekly NO2 levels, concurrently measured at the same sites like noise using passive gas samplers from Passam AG, Switzerland (Passam, 2021). Precipitation data obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) (European Centre for Mediu, 2020) was also used to check whether some heavy rains during the measurements in wet season could have impacted the noise measurements.

2.4. Development of LUR models

For developing the LUR model, we conducted supervised mixed linear regression analysis using measurement site as random intercept effect to account for multiple measurements from the same site. We used the noise metric L_{den}, because it best represents the average noise burden (World Health Organization. Environmental Noise Guidelines for the European Region, 2018), and L_{night}, because we also want to separately evaluate the effect of night time noise in subsequent epidemiological analysis. First, univariate regression with all predictor variables presented in Table 1 was performed. The variables with the highest adjusted explained variance (adjusted R^2) and with the expected direction of effect, which we defined previously, were then one by one added in the model with a stepwise forward selection. We continued this process until the adjusted R^2 did not further increase. To prevent from overfitting of the model, no more than 5 predictors were included, which represents less than 10% of the sample size. In a next step, we challenged each variable in the model with all their buffer sizes or distance type to check if any of them could further improve the model. We then removed variables with a p-value above 0.05 from the LUR model one by one. Additionally, we checked for heteroscedasticity and normality of residuals, collinearity of the predictor variables by calculating the Variance Inflation Factor (VIF), for influential observations by calculating the Cook's Distance and for spatial autocorrelation using Moran's I. In a last step, we performed a leave-one-site-out cross validation (LOSOCV) by means of the corresponding Stata function: measurement(s) from each location were sequentially left out from the model and noise exposure was predicted from a model without data from the corresponding site.

3. Results

In total after data cleaning, valid noise measurements were used over 6 consecutive days at 31 sites in summer season and 31 sites in winter season (Supplementary Figure S2) totaling to 62 valid measurements from 39 different locations for developing the LUR model. In Table 2 summary statistics for L_{day}, $L_{evening}$, L_{night}, L_{Aeq24h}, L_{den} of the cleaned noise measurements are presented. Expressed as L_{den} the arithmetic mean of all 1s-LAeq measurements was 73.1 dB(A) with a standard deviation of 9.9 dB(A). Highest and lowest site specific L_{den} were 82.3 and 60.3 dB(A). Fig. 2 presents the diurnal variability of noise for each day of week with highest noise levels in the evening. In particular on weekends (Friday-Sunday), noise levels in the evening and night hours were high. Noise levels of summer season (February) seem to be slightly higher in the later hours and lower in the earlier hours compared to winter season, which were measured in August. On average, L_{den} noise levels were similar in summer (69 dB(A)) and winter season (70 dB(A)) for the 23 sites with data from both seasons (L_{night}: 59 dB(A) and 60 dB (A)). Pearson and Spearman correlation between L_{den} and L_{night} measurements of summer and winter season were between 0.76 and 0.78 (Fig. 3).

LUR models were developed for L_{den} and L_{night} and are described in Table 3a,3b. Five relevant predictors were identified: educational facilities within a 400 m buffer, inverse distance to the closest medium road (including motorway, trunk, primary, secondary and tertiary roads), proportion of informal settlements ("favelas") within a 400 m buffer, proportion of residential land use within a 50 m buffer (L_{den}) and 25 m buffer (L_{night}) and mean NDVI within a 100 m buffer. The same predictors were found for the L_{night} model, although a different buffer size for residential land use and differing coefficients. The most predictive variable was the inverse distance to the closest medium road explaining 26% of the measured spatial variability (Supplementary Table S1). Summary statistics of all relevant predictor variables can be found in Table 4. Fig. 4 describes the distribution of the measured and predicted noise values. The mean difference between predicted minus observed exposure was 0.12 dB(A) (standard error: 0.46 dB(A)) for L_{den} and 0.04 dB(A) (0.49 dB(A)) for L_{night} (Supplementary Figure S3). A substantial fraction of the measured spatial variability can be explained by the available GIS predictor variables with a R^2 of 0.61 for L_{den} and a

Table 2

Summary statistics (mean, standard deviation (SD), minimum (min), maximum (max) of the measured cleaned noise levels for five different metrics (L_{day}, $L_{evening}$, L_{night}, L_{Aeq24h}, and L_{den}) for all 1-s LAeq measurements and for site averages, in A-weighted decibels ([dB(A)]).

	all 1-s LAeq measurements[a]				site averages[b]			
Variable Name	Mean	SD	Min	Max	Mean	SD	Min	Max
Lday	69.5	9.2	32.4	105.4	66.4	5	58.1	78.3
Levening	69.5	9.8	33.5	109.3	66.4	5.2	55.7	77.0
Lnight	65.2	10.4	31.2	110.4	59.9	6.7	50.7	76.6
LAeq24h	68.5	10.7	31.2	110.4	65.4	5.2	56.5	76.4
Lden	73.1	9.9	32.4	120.4	69.3	5.7	60.3	82.3

[a] Energetic mean level.

[b] Mean values calculated in dB(A)-units.

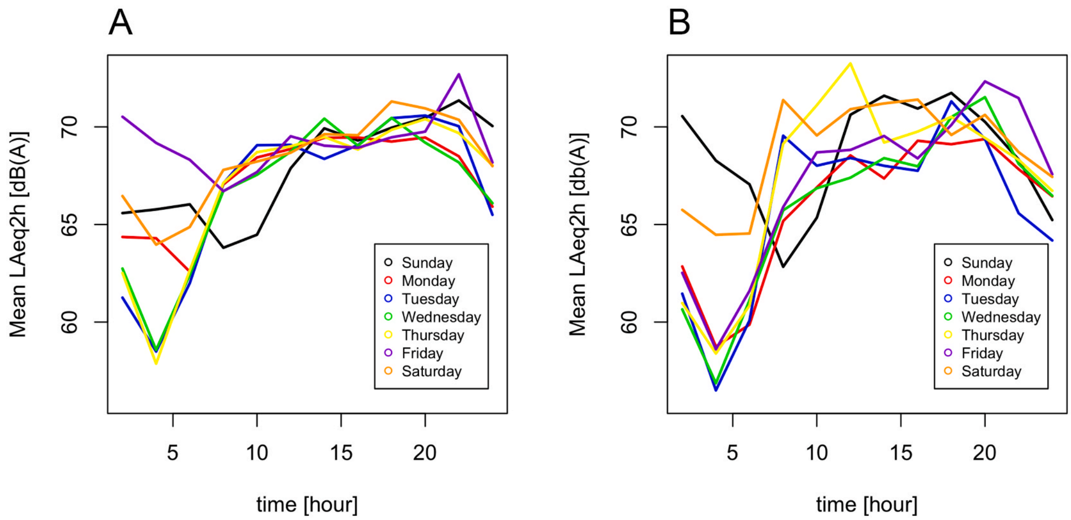

Fig. 2. Diurnal variation of 2-h average of L_{Aeq} [db(A)] by day of week represent by different colors for summer season (A) and winter season (B) . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

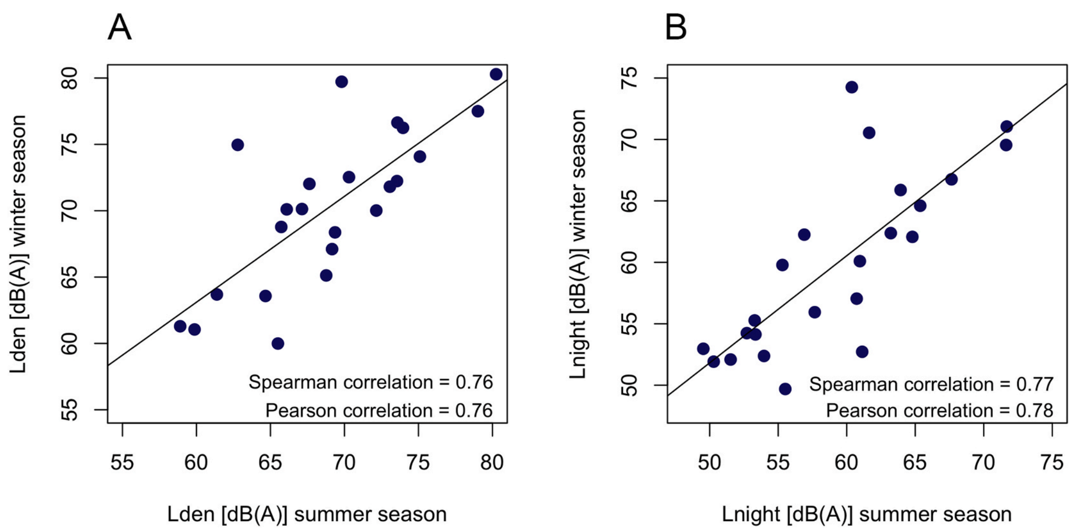

Fig. 3. Comparison of mean L_{den} (A) and L_{night} [db(A)] (B) measurements in summer and winter season at each location (n = 23).

Table 3a
Results of the LUR model explaining L_{den} measured at 39 sample sites ($R^2 = 0.612$). The coefficient (coef.) refers to L_{den} increase per unit of the predictor variable.

Variable Name	Buffer Radius (m)	Unit of the Coef. and the 95%-CI	Coef	(95% CI)	*p*-value
Residential land use	50	surface (ha)	−8.03	(-9.93, −6.14)	<0.001
Medium roads		inverse distance (m^{-1})	39.6	(33.6, 45.7)	<0.001
Educational facilities	400	count	1.14	(0.738, 1.54)	0.004
Informal settlements	400	surface (ha)	0.201	(0.126, 0.276)	0.007
NDVI	100	Mean index	−21.5	(-27.4, −15.6)	<0.001
Intercept		dB(A)	75.8	(73.5, 78.2)	<0.001

Note: NDVI, Normalized Difference Vegetation Index; ha, hectare; m, meter; dB (A), A-weighted decibels.

Table 3b
Results of the LUR model explaining L_{night} measured at 39 sample sites ($R^2 = 0.677$). The coefficient (coef.) refers to L_{night} increase per unit of the predictor variable.

Variable Name	Buffer Radius (m)	Unit of the Coef. and the 95%-CI	Coef	(95% CI)	*p*-value
Residential land use	25	surface (ha)	−39.3	(-46.9, −31.7)	<0.001
Medium roads		inverse distance (m^{-1})	49.9	(43.5, 56.3)	<0.001
Educational facilities	400	count	1.30	(0.883, 1.72)	0.002
Informal settlements	400	surface (ha)	0.295	(0.217, 0.373)	<0.001
NDVI	100	Mean index	−24.1	(-30.3, −17.8)	<0.001
Intercept		dB(A)	67.0	(64.5, 69.5)	<0.001

Note: NDVI, Normalized Difference Vegetation Index; ha, hectare; m, meter; dB (A), A-weighted decibels.

Table 4
Summary statistics of the GIS predictor variables used in the LUR model to explain L_{den} and L_{night}.

Variable Name	Buffer Radius (m)	Unit	Mean	SD	Min	Max
Residential land use	25 (L_{night})	surface (ha)	0.148	0.076	0	0.196
	50 (L_{den})	surface (ha)	0.573	0.292	0	0.785
Medium roads		inverse distance (m^{-1})	0.047	0.089	0.001	0.451
Educational facilities	400	count	0.887	1.380	0	7
Informal settlements	400 (L_{night})	surface (ha)	4.480	7.233	0	40.760
NDVI	100	Mean index	0.271	0.092	0.161	0.554

Note: NDVI, Normalized Difference Vegetation Index; ha, hectare; m, meter; dB (A), A-weighted decibels.

R^2 of 0.68 for L_{night} (Table 5). For both models, the differences between the model R^2 and the LOSOCV R^2 was less than 10%, which indicates stable models. No signs for heteroscedasticity were found and the residuals were approximately normally distributed (Supplementary Figure S3). There is neither concern about influential observations with Cook's Distance values of maximum 0.49 nor for collinearity of the predictors with the Variance Inflation Factor (VIF) having a maximum value of 1.19. Additionally, no sign for spatial autocorrelation was found (p-value for Moran's I 0.65 for L_{den} and 0.56 for L_{night}). From Fig. 5, which shows a map of the study area with a resolution of 20 × 20 m with predicted values of L_{den} and L_{night}, we can see that the model captures well high noise levels near main roads and informal settlements. More than half of predicted values in the study area lie over 65 dB(A) for L_{den} and over 55 dB(A) for L_{night} (Table 6).

We also tested other LUR models, such as for L_{day}, different weekdays and using the average value of both seasons instead of separate measurements, as well as separate models for summer and winter season.

Table 5
Summary of LUR model validation.

	native Model			LOSOCV[a]		
Variable Name	R^2	adj R^2	RSME	R^2	adj R^2	RSME
L_{den}	0.61	0.58	3.0	0.56	0.52	3.8
L_{night}	0.68	0.65	3.4	0.63	0.60	4.1

[a] Leave-one-site-out cross validation.

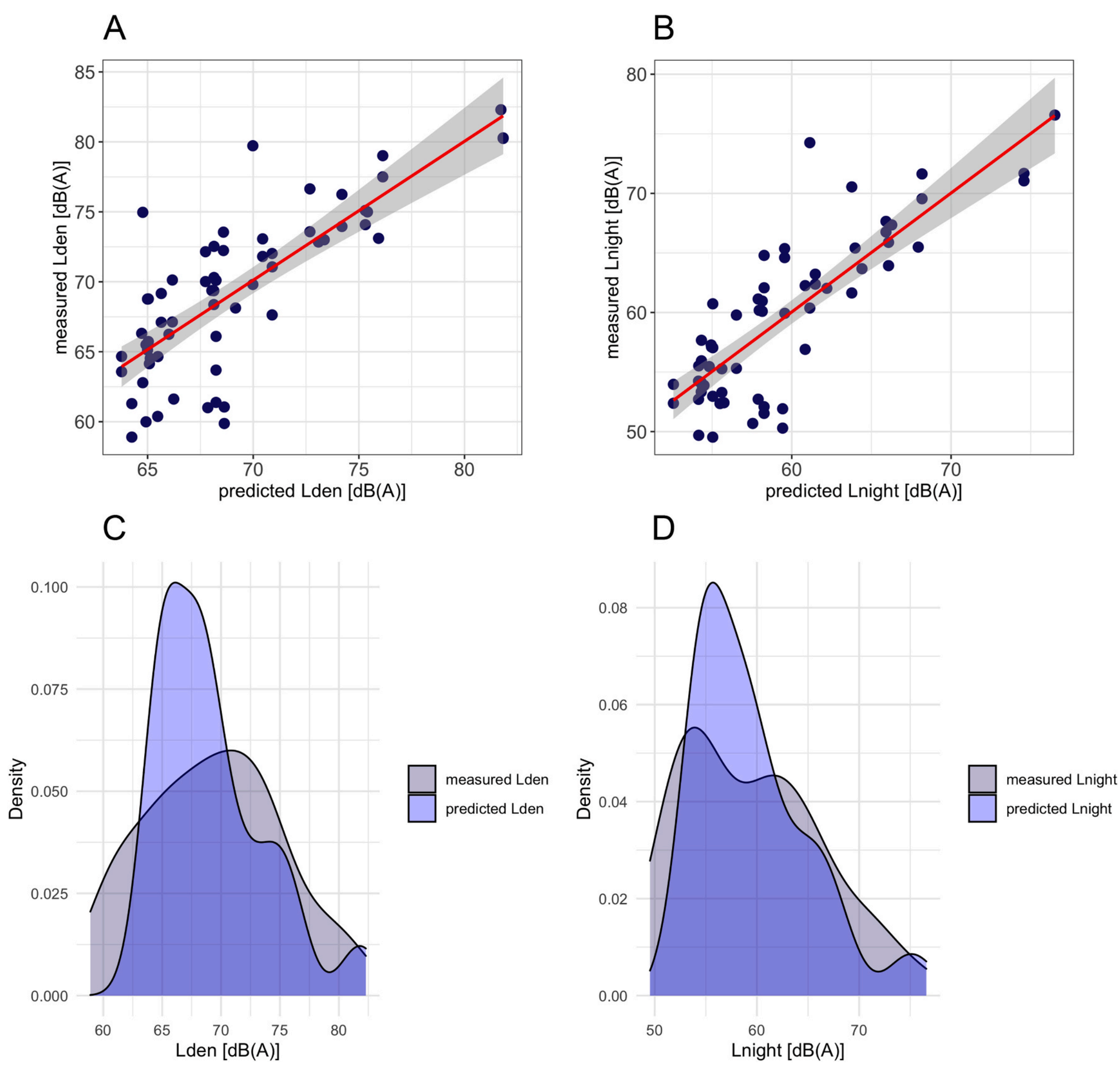

Fig. 4. Scatter plot and distribution of predicted noise against measured noise (A and C: L_{den} and B and D: L_{night}) (n = 62). In the scatter plot, the fitted value line (red) and the 95% confidence interval (grey zone) are also displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

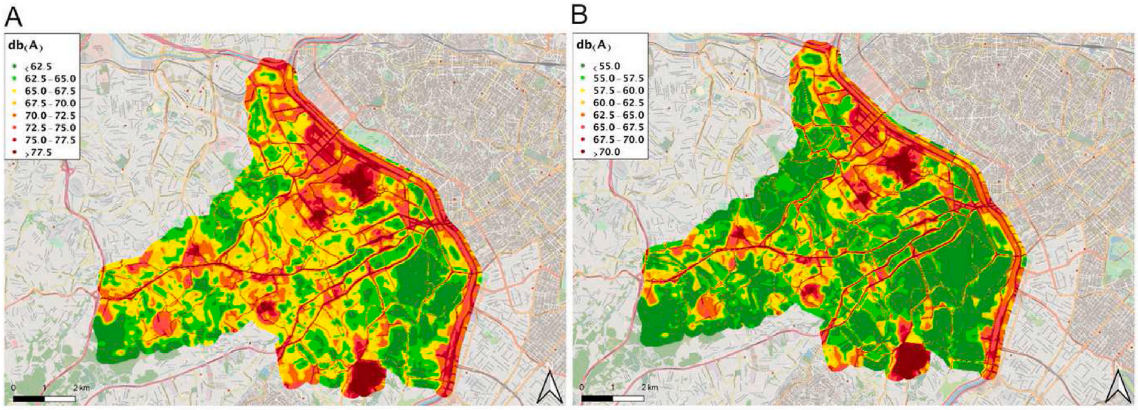

Fig. 5. Map of study area with modeled L_{den} (A) and L_{night} (B) with a 20 × 20 m resolution.

Table 6
Predicted cumulative noise distribution for L_{den} and L_{night} in all 20 × 20 m grid points of the study area.

Threshold [dB(A)]	Cumulative proportion L_{den} threshold [%]	Cumulative proportion L_{night} threshold [%]
55	100.0	65.0
60	97.8	36.0
65	68.6	16.3
70	33.2	6.8
75	11.7	3.8

These models achieved similar outcomes with comparable GIS predictors, R^2 and validation results, underlining the accuracy of our developed LUR models (see Supplementary Table S2-8, Figures S4-7).

4. Discussion

We developed LUR models for L_{den} and L_{night} derived from continuous six-day noise measurements at 39 homes to assess the spatial variability of community noise levels in São Paulo, Brazil. We obtained a moderate to good model fit (R^2 of 0.56 for L_{den} and 0.63 for L_{night} using LOSOCV).

To our knowledge, this is the first LUR modelling study of South America and only the third one in a low- and middle-income country. The first study in a low- and middle-income country, which was conducted in China by Xie et al. (2011), could explain 70% of the observed variance. However, they only considered roads as predictor variables. The LUR model of the second study, conducted by Sieber et al. (2017a) in informal settings of the Western Cape, South Africa, explained only 13% of the noise exposure variability. Their explanation for the low model performance was that GIS predictor data was inaccurate and constantly changing in informal settlements. Although several informal settlements are situated within our study area and most GIS predictor data were obtained from Open Street Map, which is an open-source tool and therefore prone to errors and missing data, we achieved a LUR model performance comparable to previous models from high-income countries, which achieved R^2 values between 0.47 and 0.89 (Ragettli et al., 2016; Harouvi et al., 2018; Aguilera et al., 2015; Chang et al., 2019; Liu et al., 2020).

In our models of L_{den} as well as L_{night}, the inverse distance to the nearest medium sized road were the most predictive variable for noise levels according to explained variance. Transport related predictor variables have already previously been proven to have a good correlation with noise measurement (Ragettli et al., 2016; Harouvi et al., 2018; Aguilera et al., 2015). Additionally, the negative association of NDVI and residential land use with noise levels were demonstrated in a previous study of Ragettli et al. (2016). However, our study is the first to find educational facilities and informal settlements ("favelas") as relevant predictor variables for increased noise levels. Several studies previously measured elevated noise levels in schoolyards, as a result of screaming, loud playing children (Kapetanaki et al., 2018; Sarantopoulos et al., 2014). Informal settlements have not only high population densities (Paraisópolis, one of the largest informal settlement in São Paulo, with 1.000 inhabitants per hectare (Théry, 2020)) but also increased outdoor activities due to limited indoor space such as outdoor weekend parties. Therefore, identifying informal settlements as substantial noise predictor is crucial, given that more than 11% of people in São Paulo living in informal settlements with an upward trend (Instituto Brasileor de Geografia e Estatistica. Population Census, 2010).

In the guidelines of the World Health Organization (WHO) for community noise, L_{den} of less than 54 dB(A) is recommended for road traffic noise and less than 45 dB(A) for L_{night} (World Health Organization. Environmental Noise Guidelines for the European Region, 2018). Our community noise measurements were above these guidelines at all sites and our mean L_{den} in São Paulo of 69 dB(A) was also very high compared to other countries. In Western Cape of South Africa, a mean L_{den} of 63 dB(A) was measured (Sieber et al., 2017b). Modelled road traffic noise was 54 dB(A) in Switzerland (Cantuaria et al., 2018) and between 44 and 52 dB(A) in Munich (Tiesler et al., 2013). Another study in São Paulo (Moura-de-Sousa and Alves Cardoso, 2002), which measured noise levels in a different area, presented mean L_{Aeq24h} values between 61 and 75 dB(A) which fits very well the exposure range of our maps of the study area. One reason for higher noise levels, compared to results from other studies of mainly high-income countries, can be louder traffic noise because of less modern and therefore louder cars and pavement. Another factor explaining higher noise levels might include climate differences, with warmer and therefore more comfortable temperatures for outside activities, which can potentially lead to noise sources, such as street markets or parties. Last but not least, the population density in our study area is higher than the study areas of most other cited studies (average 79 inhabitants per hectare, but highly variable depending on area) (Secretaria Municipal da Saúde, 2018). The association of high population densities and noise levels has been shown in several studies (European Environment Agency, 2020; Stewart et al., 1999).

In addition to the high outdoor noise exposure values, it has to be considered for health impact assessment that buildings in this climate zone are usually not well sound insulated and windows are kept open, resulting in higher transmission of outdoor noise into the homes (De et al., 2005). Epidemiological research on noise is thus warranted for such areas, although still quite rare (Paiva-Vianna and Cardoso, 2016; Barbaresco et al., 2019).

A strength of our study is the short sampling interval of one second allowing for the fact that community noise shows a high temporal variability. Previous studies used intervals between 2 min and 20 min

(Ragettli et al., 2016; Harouvi et al., 2018; Aguilera et al., 2015; Chang et al., 2019). One limitation of this study is that no data on traffic intensity was available, which has been identified as relevant noise predictor in previous studies (Aguilera et al., 2015; Chang et al., 2019). However, with the different road type variables, this aspect is considered to some extent. Besides traffic intensity, also other temporal predictors would have been interesting, in order to be able to predict on a daily/weekly/monthly scale. However, we checked for differences in LUR models using different days and times, which did not lead to relevant differences in model results. Furthermore, the number of measurement locations in our study area is relatively small. The study of Basagana et al. (Basagaña et al., 2012) demonstrated, that LUR models with small sample sizes and high number of potential predictors offered, tend to give higher and more inflated R^2 and LOSOCV R^2 values. We therefore tried to limit the set of predictors in the final model to five. Finally, noise measurements were conducted for two weeks only and thus may not have captured all relevant differences between seasons and weather conditions. However, the good correlation between summer and winter measurements indicates that seasonal variations are not substantial. By conducting continuous measurements over a week, we collected data from a variety of common weather conditions. The two measurement campaigns represent dry and wet season including some heavy rains during the measurements in wet season. We checked for rain as a predictor variable, however it did not have relevant spatial resolution.

5. Conclusions

Our study demonstrates the feasibility of LUR models to estimate community noise exposures for epidemiological research in areas where no public noise maps are available. Week-to-week variation of noise is relatively small compared to air pollution (Allen et al., 2009), as also demonstrated in these data. Thus, short measurement campaigns are suitable to estimate long-term spatial distribution of noise levels. Such estimates are also useful for epidemiological research and to assess the public health impact of community noise.

Funding

Data collection was supported by the Eckenstein-Geigy Professorship.

Ethical approval

Approval was obtained from the Swiss ethics committee (AO_2020–00025) and Brazilian ethics committee (01604312.1.0000.0065). This article does not contain any studies involving human participants performed by any of the authors.

Author contribution

Michelle Raess conducted the analysis, interpreted the results, and wrote the first draft of the manuscript. Benjamin Flückiger was involved in the fieldwork and data collection and contributed to the analysis of the results. Bartolomeu Ledebur de Antas de Campos and Alexandra Brentani were involved in the fieldwork, data collection and the design of the study. Kees de Hoogh supervised the GIS analysis, Günther Fink and Martin Röösli contributed to the design of the study, analysis and interpretation of the results. All authors contributed to the manuscript writing and approved the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2021.111231.

References

Aguilera, I., Foraster, M., Basagaña, X., Corradi, E., Deltell, A., Morelli, X., et al., 2015. Application of land use regression modelling to assess the spatial distribution of road traffic noise in three European cities. J. Expo. Sci. Environ. Epidemiol. 25 (1), 97–105.

Allen, R.W., Davies, H., Cohen, M.A., Mallach, G., Kaufman, J.D., Adar, S.D., 2009. The spatial relationship between traffic-generated air pollution and noise in 2 US cities. Environ. Res. 109 (3), 334–342.

Barbaresco, G.Q., Reis, A.V.P., Lopes, G.D.R., Boaventura, L.P., Castro, A.F., Vilanova, T.C.F., et al., 2019. Effects of environmental noise pollution on perceived stress and cortisol levels in street vendors. J. Toxicol. Environ. Health, Part A 82 (5), 331–337.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., et al., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., et al., 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – the ESCAPE project. Atmos. Environ. 72, 10–23.

Björk, J., Ardö, J., Stroh, E., Lövkvist, H., Ostergren, P.O., Albin, M., 2006. Road traffic noise in southern Sweden and its relation to annoyance, disturbance of daily activities and health. Scand. J. Work. Environ. Health 32 (5), 392–401.

Brentani, A., Scoleze Ferrer, A.P., Brentani, H., Liu, C.H., Grisi, S., Valente, M.H., et al., Cohort profile: São Paulo Western Region Birth Cohort (ROC), 2020. Int. J. Epidemiol. 49 (5), 1438, 1438g.

Brink, M., Schreckenberg, D., Vienneau, D., Cajochen, C., Wunderli, J.M., Probst-Hensch, N., et al., 2016. Effects of scale, question location, order of response alternatives, and season on self-reported noise annoyance using ICBEN scales: a field experiment. Int. J. Environ. Res. Publ. Health 13.

Cantuaria, M.L., Usemann, J., Proietti, E., Blanes-Vidal, V., Dick, B., Flück, C.E., et al., 2018. Glucocorticoid metabolites in newborns: a marker for traffic noise related stress? Environ. Int. 117, 319–326.

Centro de Estudos da Metrópole Relatórios Favelas e Loteamentos - Estudo do CEM para Sehab/PMSP [04.11.2020], 2020. http://centrodametropole.fflch.usp.br/pt-br/downloads-de-dados/relatorios-favelas-e-loteamentos-estudo-do-cem-para-sehabpmsp.

Chang, T.-Y., Liang, C.-H., Wu, C.-F., Chang, L.-T., 2019. Application of land-use regression models to estimate sound pressure levels and frequency components of road traffic noise in Taichung, Taiwan. Environ. Int. 131, 104959.

De, E., Duarte, A., Viveiros, E., 2005. Sound insulation performance of Brazilian dwellings: from colonial ages to contemporary architecture. Environmental Noise Control 2.

de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M.J., et al., 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. Environ. Int. 73, 382–392.

Environmental Systems Research Institute (ESRI), 11.12.2020. https://www.esri.com.

European Centre for Medium-Range Weather Forecasts (ECMWF), 2020. Precipitation data [04.11.2020]. Available from. https://www.ecmwf.int.

European Environment Agency, 2016. Quiet Areas in Europe. The Environment Unaffected by Noise Pollution (Report No 14/2016) [18.11.2020]. Available from. http://www.eea.europa.eu/highlights/mapping-europe2019s-quiet-areas.

European Environment Agency, 2020. Environmental noise in Europe - 2020 (EEA Report No 22/2019) [01.12.2020]. Available from. https://www.eea.europa.eu/publications/environmental-noise-in-europe.

Eze, I.C., Foraster, M., Schaffner, E., Vienneau, D., Heritier, H., Rudzik, F., et al., 2017. Long-term exposure to transportation noise and air pollution in relation to incident diabetes in the SAPALDIA study. Int. J. Epidemiol. 46 (4), 1115–1125.

Frei, P., Mohler, E., Röösli, M., 2014. Effect of nocturnal road traffic noise exposure and annoyance on objective and subjective sleep quality. Int. J. Hyg Environ. Health 217 (2–3), 188–195.

Harouvi, O., Ben-Elia, E., Factor, R., de Hoogh, K., Kloog, I., 2018. Noise estimation model development using high-resolution transportation and land use regression. J. Expo. Sci. Environ. Epidemiol. 28 (6), 559–567.

Héritier, H., Vienneau, D., Foraster, M., Eze, I.C., Schaffner, E., Thiesse, L., et al., Transportation noise exposure and cardiovascular mortality: a nationwide cohort study from Switzerland, 2017. Eur. J. Epidemiol. 32 (4), 307–315.

Hervé Théry, 20.11.2020. Picturing Inequalities [20.11.2020]. Available from. https://halshs.archives-ouvertes.fr/halshs-01718713.

Instituto Brasileor de Geografia e Estatistica, 2010. Instituto Brasileor de Geografia e Estatistica, 2010. Population Census 2010 [04.11.2020]. Available from. https://www.ibge.gov.br.

Kapetanaki, S., Konstantopoulou, S., Linos, A., 2018. Noise Measurements and Sources of Noise in Primary Schools Located in Vulnerable Areas in Greece: A Cross-Sectional Study. Occupational Medicine & Health Affairs, p. 6.

Kloog, I., Sorek-Hamer, M., Lyapustin, A., Coull, B., Wang, Y., Just, A.C., et al., 2015. Estimating daily PM(2.5) and PM(10) across the complex geo-climate region of Israel using MAIAC satellite-based AOD data. Atmospheric environment (Oxford, England. 1994) 122, 409–416.

Liu, Y., Goudreau, S., Oiamo, T., Rainham, D., Hatzopoulou, M., Chen, H., et al., 2020. Comparison of land use regression and random forests models on estimating noise levels in five Canadian cities. Environ. Pollut. 256, 113367.
Montes González, D., Barrigón Morillas, J.M., Rey Gozalo, G., Godinho, L., 2020. Evaluation of exposure to road traffic noise: effects of microphone height and urban configuration. Environ. Res. 191, 110055.
Morley, D.W., de Hoogh, K., Fecht, D., Fabbri, F., Bell, M., Goodman, P.S., et al., 2015. International scale implementation of the CNOSSOS-EU road traffic noise prediction model for epidemiological studies. Environ. Pollut. 206, 332–341.
Moura-de-Sousa, C., Alves Cardoso, M., 2002. Urban noise in the city of sao Paulo. In: Noise and Health, vol. 4. An important problem of public health, Brazil, pp. 57–63.
OpenStreetMap contributors, 04.11.2020. OpenStreetMap. https://www.openstreetmap.org.
Paiva-Vianna, K.M., Cardoso, M.R., 2016. Noise Pollution, Acoustic Maps and Health Surveys: the Case of the Municipality of São Paulo. Noise Pollution in Urban and Industrial Environments: Measurements and Noise Mapping, Brazil, pp. 123–139.
Passam, A.G., 2021. Laboratory for environmental analysis and air pollution [16.04.2021]. Available from: https://www.passam.ch/products.
Ragettli, M.S., Goudreau, S., Plante, C., Fournier, M., Hatzopoulou, M., Perron, S., et al., 2016. Statistical modeling of the spatial variability of environmental noise levels in Montreal, Canada, using noise measurements and land use characteristics. J. Expo. Sci. Environ. Epidemiol. 26 (6), 597–605.
Recio, A., Linares, C., Banegas, J.R., Diaz, J., 2016. The short-term association of road traffic noise with cardiovascular, respiratory, and diabetes-related mortality. Environ. Res. 150, 383–390.
Sarantopoulos, G., Lykoudis, S., Kassomenos, P., 2014. Noise levels in primary schools of medium sized city in Greece. Sci. Total Environ. 482–483, 493–500.
Schlittmeier, S.J., Feil, A., Liebl, A., Hellbr Ck, J.R., 2015. The impact of road traffic noise on cognitive performance in attention-based tasks depends on noise level even within moderate-level ranges. Noise Health 17 (76), 148–157.
Secretaria Municipal da Saúde, 2018. Prefeitura da Cidade de São Paulo 2018 [04.11.2020]. Available from. http://www.prefeitura.sp.gov.br/cidade/secretarias/saude/.
Seidler, A., Wagner, M., Schubert, M., Droge, P., Romer, K., Pons-Kuhnemann, J., et al., 2016. Aircraft, road and railway traffic noise as risk factors for heart failure and hypertensive heart disease-A case-control study based on secondary data. Int. J. Hyg Environ. Health 219 (8), 749–758.
Seidler, A., Hegewald, J., Seidler, A.L., Schubert, M., Wagner, M., Droge, P., et al., 2017. Association between aircraft, road and railway traffic noise and depression in a large case-control study based on secondary data. Environ. Res. 152, 263–271.
Sheng, N., Tang, U.W., 2011. Spatial analysis of urban form and pedestrian exposure to traffic noise. Int. J. Environ. Res. Publ. Health 8 (6), 1977–1990.
Sieber, C., Ragettli, M.S., Brink, M., Toyib, O., Baatjies, R., Saucy, A., et al., 2017a. Land use regression modeling of outdoor noise exposure in informal settlements in western Cape, South Africa. Int. J. Environ. Res. Publ. Health 14.
Sieber, C., Ragettli, M.S., Brink, M., Toyib, O., Baatjies, R., Saucy, A., et al., 2017b. Land use regression modeling of outdoor noise exposure in informal settlements in western Cape, South Africa. Int. J. Environ. Res. Publ. Health 14 (10), 1262.
Singh, N., Davar, S.C., 2004. Noise pollution-sources, effects and control. J. Hum. Ecol. 16 (3), 181–187.
Sørensen, M., Andersen, Z.J., Nordsborg, R.B., Jensen, S.S., Lillelund, K.G., Beelen, R., et al., Road traffic noise and incident myocardial infarction: a prospective cohort study, 2012. PloS One 7 (6), e39283.
Sørensen, M., Andersen, Z.J., Nordsborg, R.B., Becker, T., Tjønneland, A., Overvad, K., et al., Long-term exposure to road traffic noise and incident diabetes: a cohort study, 2013. Environ. Health Perspect. 121 (2), 217–222.
Stewart, C.M., Russell, W.A., Luz, G.A., 1999. Can population density be used to determine ambient noise levels? J. Acoust. Soc. Am. 105 (2).
Tang, U.W., Wang, Z.S., 2007. Influences of urban forms on traffic-induced noise and air pollution: results from a modelling system. Environ. Model. Software 22 (12), 1750–1764.
Tiesler, C.M., Birk, M., Thiering, E., Kohlböck, G., Koletzko, S., Bauer, C.P., et al., 2013. Exposure to road traffic noise and children's behavioural problems and sleep disturbance: results from the GINIplus and LISAplus studies. Environ. Res. 123, 1–8.
U.S. Geological Survey, 2020. Landsat 8 Mission [04.11.2020]. Available from. https://www.usgs.gov.
van Kempen, E., van Kamp, I., Lebret, E., Lammers, J., Emmen, H., Stansfeld, S., 2010. Neurobehavioral effects of transportation noise in primary schoolchildren: a cross-sectional study. Environ. Health 9, 25.
Vienneau, D., Schindler, C., Perez, L., Probst-Hensch, N., Röösli, M., 2015. The relationship between transportation noise exposure and ischemic heart disease: a meta-analysis. Environ. Res. 138, 372–380.
World Health Organization, 2011. Burden of Disease from Environmental Noise:. Quantification of Healthy Life Years Lost in Europe. Copenhagen.
World Health Organization, 2018. Environmental Noise Guidelines for the European Region (2018). Geneva, Switzerland.
Xie, D., Liu, Y., Chen, J., 2011. Mapping urban environmental noise: a land use regression method. Environ. Sci. Technol. 45 (17), 7358–7364.
Zijlema, W.L., Wolf, K., Emeny, R., Ladwig, K.H., Peters, A., Kongsgård, H., et al., The association of air pollution and depressed mood in 70,928 individuals from four European cohorts, 2016. Int. J. Hyg Environ. Health 219 (2), 212–219.