PROCEEDINGS of the
24th International Congress on Acoustics

October 24 to 28, 2022 in Gyeongju, Korea

ABS-0471

# 2000 land-use regressions for road traffic noise predictions – how sample selection affects extrapolation weights

Jeroen STAAB [1,2,*], Arthur SCHADY [3], Kathrin WOLF [4], Sahar BEHZADI [4], Marco DALLAVALLE [4,5], Matthias WEIGAND [1,6], Tobia LAKES [2,7], Hannes TAUBENBÖCK [1,6]

[1] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Germany

[2] Humboldt-University Berlin, Geography Department, Germany

[3] German Aerospace Center (DLR), Institute of Atmospheric Physics (IPA), Germany

[4] Helmholtz Zentrum München GmbH - German Research Center for Environmental Health, Institute of Epidemiology, Germany

[5] Ludwig-Maximilians-Universität München, Institute for Medical Information Processing, Biometry and Epidemiology, Department of Epidemiology, Germany

[6] University of Würzburg, Institute of Geography and Geology, Department of Remote Sensing, Germany

[7] Integrative Research Institute on Transformations of Human-Environment Systems (IRI THESys), Germany

## ABSTRACT

The awareness that noise exposure is critical for human health is growing around the globe, and land-use regressions (LURs) are becoming a popular tool for producing noise exposure maps. One important factor for noise emissions is road traffic. The propagation in this regard is determined by the spatial layout of road infrastructure and the surrounding environment, respectively. LURs use geostatistical models and allow to extrapolate microphone measurements. In this study, we investigated whether models are prone to sampling artifacts. We used yearly averaged $L_{den}$ simulations, compliant to the European noise directive 2002/49/EG, as input for 2000 virtual field campaigns. We permuted different sampling schemes (*random*, *systematic*, *stratified*) and sizes (*n* = 50, 100, 200, 500 to 1000) 100 times. The overall model performances varied substantially between 0.61 – 0.95 for $R^2$, 1.94 – 7.46 dB(A) for mean absolute error and 2.47 – 10.03 dB(A) for root mean squared error. Comparing the eventual model terms using variance analyses (ANOVA), we found significant differences between the sampling schemes for traffic information and land cover (e.g. vegetated surfaces) features. Simultaneously, less than half of the LURs' weights differed significantly depending on the sampling size. Thus, our experiments give an in-depth view on the mechanics of LUR and their sensitivity with respect to sampled training data.

Keywords: Traffic noise, Exposure Assessment, Sensitivity Analysis

## 1 INTRODUCTION

Modern society beholds ongoing technological developments, increased wealth and mobility. With respect to road traffic noise, the WHO considers yearly averaged levels above 53dB(A) being unhealthy (1). To investigate the health impact of noise, two approaches exist (2): first, controlled lab experiments, where the biophysical reactions of human bodies to noise exposure can be investigated. It is obvious though, that this is impractical for large scale and long-term studies. Thus, second, scholars aim to extract information on the noise exposure by the addresses of study participants (3–5). To do so, microphones could be deployed in the field; however, this approach is limited due to costs and spatiotemporal limitations. In Europe, thanks to the European Noise Directive (6, also referred to as END), simulated $L_{den}$ levels are available for urban agglomerations with a population larger than 100,000 (6). This END compliant data can be utilized for health studies (e.g. 5,7). Other health studies investigate the impact of noise relative to noise exposures mapped using

[*] Jeroen.staab@dlr.de

land-use regressions (LUR) (e.g. 8).

LURs use geospatial information describing the surroundings; indicators such as distance to roads, built-up morphology and close by vegetated areas are used as predictors of a statistical model, e.g. least-squares (3,9–13). Simplified, noise $N$ is a function of noise emissions $E$ and subsequent interactions $I$ when propagating though the environment. After fitting the model, each predictor is supplemented with a weight $w$, such that the complete term may be denoted as $N = w_E E + w_I I$. In a previous study, we have investigated the impact of variances within the response variable and the respective spatial transferability (14). Thereby, we permuted sample sizes ($n$ = 50, 100, 200, 500 to 1000) and sampling schemes (*random*, *systematic*, *stratified*) 100 times and have shown respective sampling artifacts. However, the produced 2000 regressions terms themselves had not yet been investigated. It stands to reason, that with each respective sampling experiment, an individual configuration of weights was produced. We hypothesized to see distinct variations in the respective terms again. In particular, though, we assumed that the weights for some predictors are more prone to sampling artefacts than others.

As the maps produced with LURs are used for exposure assessments eventually, knowing their robustness is of high relevance with respect to their confidence. Also, having the most important predictors identified, computational costs could be eventually reduced for large-scale applications. Therefore, in this study, we investigated the regression terms thoroughly. After visualizing the variations, ANOVA was used to pin point the most concerned predictors.

# 2 DATA & METHODS

Large scale noise mapping is often limited by the respective data availability and computational resources (15). Mechanically, LURs models are, however, computationally comparable to conventional geostatistical modelling and, using satellite data, have been successfully deployed for air quality assessment on national scales (16–18). Analog, we strived for Earth observation data and other public geodata as input for our model. Methodically, this study continued upon (14); thus, the utilized input data and original experimental setup are only summarized. The focus of the study is put to the extension of the methodical framework - investigating the regressions' weights. In compliance to the END, all geospatial operations were conducted on a 10 x 10 Meter resolution.

Spatially, we focused our investigations on the German city Koblenz. We did so, as the city provided us with an area-wide, highly resolved noise map for training and validation. In 2022, Koblenz had 113,844 inhabitants living on 105 km² (19). The city has heterogenous types of built-up structures, ranging from a historic city center, industrial areas and low dense suburban formations. At the same time, with both, its rivers and forests, it features different natural environments embodied in a fluvial shaped topography.

## 2.1 Predictor Variables

Of all noise emitting sources, road traffic is affecting most people in Europe (20). Along this priority, information regarding its respective emissions is available in public datasets. The *OpenStreetMaps* (OSM) project holds information on the spatial layout of streets attributed with street type, number of lanes, speed limits and sometimes even road surface materials. Using ANOVA, we had previously shown that the road types *Motorway*, *Trunk*, *Primary*, *Secondary*, *Tertiary* and *Residential* have distinct mean noise levels on the roads themselves and can be used as proxy in LURs (14). After tunnels were excluded (as in 21), we integrated the road infrastructures in to our LUR in two fashions: First, as ushered by Harouvi et al. (12), the log-transformed distance to the next road of the respective type was used. Second, with respect to cumulated road exposures, the road length multiplied with their lanes within a given radius (9–11,13,21–23) was summarized. Based on preceding literature (3,9–11,13,21–23) and in relation to the logarithmic behavior of noise propagation (24), we used systematically scaled moving window radii ranging between 12.5 and 1600 Meters.

Once emitted, a soundwave propagates through the environment. In an urban context, buildings and their respective volume therefore are essential (17,25,26). Heutschi et al. (27), for example, stress the unfavorable effects of dense street canyons. Opting for the highest data quality, building footprints and their respective height (Level-Of-Detail-1, LoD1) were retrieved from the Federal Agency for Cartography and Geodesy (28). The topographic position index was computed according to Weiss (29). Positive values depict superior locations with a lot of air volume available, while negative values occur when a pixel is lower than its surrounding. With respect to scope, the same moving window radii were chosen as defined above. Analog, the surrounding topography was integrated into the geostatistical model based on satellite derived height information. During data harmonization, the original EU-DEM (30) was rescaled to 10 x 10 Meter using bilinear interpolation.

Last but not least, propagating sound interacts with the ground surface. Plain and solid surfaces help noise traveling over long distances, while soft and porous material have sound absorbing effects (31–33). That said, we had previously produced a remote-sensing-derived land-cover classification (34), including seven land-cover classes expected to be relevant for noise mapping: These are *artificial land* such as paved roads and built-up areas, *water areas*, *open soil* as most commonly found on agrarian fields and opencast mines, as well as four classes differentiating vegetative biomass and its seasonality. Although this level of detail is not included in END compliant maps, we stress the proven correlation of biomass and sound attenuation (33,35). During data harmonization, we computed the respective fraction for each land-cover class using the moving window radii described above.

As computing multiple, systematically scaled radii introduced strong correlating covariates, we selected the most relevant scale *a priori* following the descriptions of Ragettli et al. (21) and Liu et al. (23). Eventually, 21 features were considered in our experimental setup (see Table 1).

Table 1 – Overview of investigated parameters, their source and properties

| Source | Feature | Attribute | Units | Min. | Max. |
|---|---|---|---|---|---|
| OSM | Motorway | Proximity | log(m) | 0 | 3.99 |
| | | Length800 | m | 0 | 17,609 |
| | Trunk | Proximity | log(m) | 0 | 3.744 |
| | | Length400 | m | 0 | 9,343 |
| | Primary | Proximity | log(m) | 0 | 3.67 |
| | | Length100 | m | 0 | 1,762 |
| | Secondary | Proximity | log(m) | 0 | 3.65 |
| | | Length1600 | m | 0 | 16,990 |
| | Tertiary | Proximity | log(m) | 0 | 3.48 |
| | | Length1600 | m | 0 | 28,002 |
| | Residential | Proximity | log(m) | 0 | 3.30 |
| | | Length800 | m | 0 | 21,673 |
| BKG | LoD1 | TPI800 | / | −1.77 | 44.20 |
| Copernicus | DEM | TPI1600 | / | −114.62 | 121.66 |
| Weigand et. al. (34) | Artificial land | Mean800 | % | 0 | 94.28 |
| | Open soil | Mean1600 | % | 0 | 1.32 |
| | High, seasonal veg. | Mean800 | % | 0.25 | 96.67 |
| | High, perennial veg. | Mean1600 | % | 0.01 | 16.96 |
| | Low, seasonal veg. | Mean1600 | % | 0.04 | 72.53 |
| | Low, perennial veg. | Mean800 | % | 0 | 66.01 |
| | Water areas | Mean400 | % | 0 | 3.99 |

## 2.2 Experimental Setup

### 2.2.1 Sampling Noise Simulation Data

Conventionally, LURs are used to extrapolate in-situ measurements (3,9–13,21–23). Reviewing the respective literature though showed sample sizes ranged from 40 (Girona in 10) to 1296 (Shanghai in 3) and sample location was more often stratified (3,9,10,13,21–23) than random (12). Stratification can be applied e.g. by using land-use classes as strata (13) or based on population (as in 3,22). Theoretically, also *systematic*, grid based, sampling could be conducted as well. As we want to assess in this study the uncertainties within the regression's terms introduced by varying sampling designs, we reproduced this range using an END compliant map as reference. Such maps are produced using highly accuracy engineering methods to simulate noise emission and its propagation using ray-tracing simulations (36,37). Also, the source specific map allows controlling the emitters in our experiments. For our study area, the road $L_{den}$ 2017 ranged from 12.8 to 88.3 dB(A).

Although, by using an existing map, no new exposure information is produced, subsampling it repeatedly allowed investigating the LUR models thoroughly. We varied the sample sizes from $n$ being 50, 100, 200, 500 or 1,000 and compared four sampling schemes (*random*, *systematic*, $stratified_{Lden}$, $stratified_{Urb.Atl.}$, see Figure 1a). As strata for the latter we used $L_{den}$ classes or 22 different LU/LC classes defined by the Urban Atlas (38) respectively. Each configuration was repeated 100 times such that in total 2,000 different sample sets, further referred to as virtual field campaigns, were conducted.

Looking at the overall Koblenz data set first, a mean $L_{den}$ value of 51.0 (standard deviation = 11.1) was computed by an engineering bureau. Compared thereto, Figure 1b depicts that both stratified approaches tend to have had higher mean values, $stratified_{Lden}$ in particular. True for all sampling schemes and also important for this study as well, small sample sizes tended to vary more over the 100 repetitions. Staab et al. (14) had consequently used a two-sided t-test to determine if the virtual field campaigns were representing the total population well. Almost all sample size and repetitions were significant ($p > 0.05$) for *systematic* sampled data sets, but most *random* and smaller $stratified_{Urb.Atl.}$ sample sets could be considered representative as well (Figure 1c).

### 2.2.2 Modeling

While some modern machine learning methods have more predictive power, others excel with their interpretability (39). That said, we chose linear least squares regression in order to investigate the models' weights depending on our various sampling experiments. At the same time, their low computational costs suited well to our experiments being repeated 2000 times. Most important for putting our findings into context though, linear regressions are commonly deployed in the noise mapping context (3,9–13). As the prerequired assumptions for least squares were met by most predictors, and keeping the focus on sampling artefacts, we selected a consistent feature set and did not consider forwards- (as in 3,10,11,13,21,40) or backwards-selecting implementations (9,12,22,23).

Eventually, each models' coefficient of determination ($R^2$), root mean squared error (RMSE) and mean absolute error (MAE) was computed (see summary in Table 2). Again, a high variance introduced by sampling artefacts was found. With regard to sampling scheme, we found that *random* and *systematic* sampling resulted in lower $R^2$ compared to stratified sampling approaches. Higher RMSE and MAE scores stand out at $stratified_{Lden}$ only. Overall, $R^2$ tends to decrease at larger sample sizes. It ranged between 0.61 at *random* Sampling, $n$ = 500 and seed 68 and 0.95 for $stratified_{Lden}$ sampling, $n$ = 50, seed = 50. Vice versa, RMSE and MAE both tend to be lowest at small sample sizes but increases steadily (for details see 14).

## 2.3 Regression Term Assessments

To deepen our knowledge on LUR artefacts introduced by different sampling settings, we took a specific look into the regression terms. With each virtual field campaign, an individual model with its respective term was produced. Following a deductive approach, first the regression terms itself are inspected, before an analysis of variance (ANOVA) confirms the observed trends.

Within the term, for each selected variable, an individual weight (also known as estimate) and p-value was computed. These weights were then compared across the sampling configurations, by aggregating them. For visualization, we used violin charts highlighting the most frequent estimates, further annotated with quartiles and mean values. We expected to see higher variances at smaller sample sizes, converging later. Making *a priori* assumptions concerning the effect of different sampling schemes though was difficult. Having had a look at the mean $L_{den}$ values of $stratified_{Lden}$, we could only assume the respective regression terms aligning to very loud levels such as emitted in close vicinity to road infrastructure.

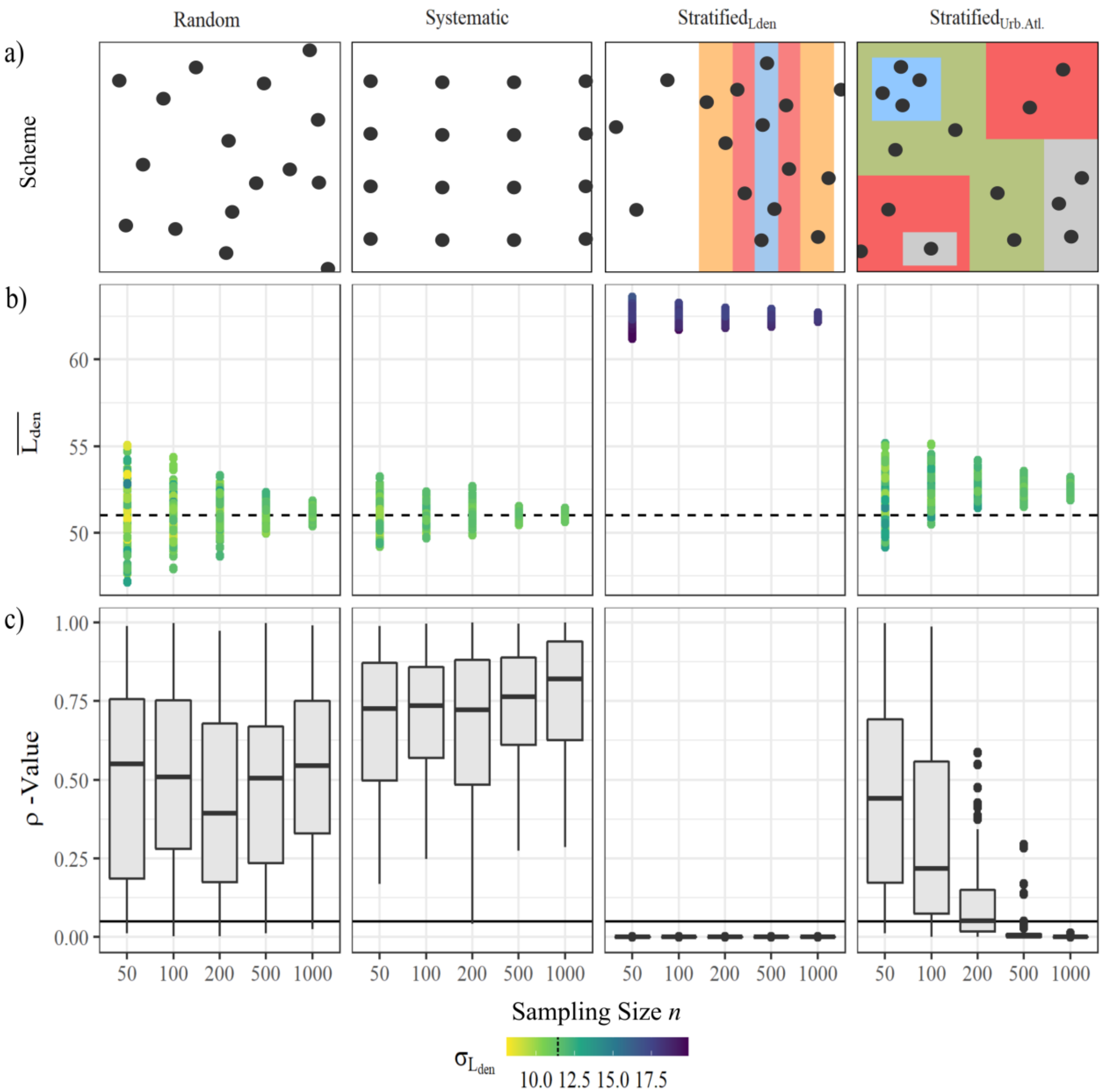


Figure 1 – Summary of different sampling schemes (a, where colored background refers to $L_{den}$ and Urban Atlas classes) and respective average $L_{den}$ per sample (b) compared to overall data set (dashed line) and its standard deviation (yellow to purple colors), using a two-sided t-test eventually (c).

Second, we particularly highlighted systematic differences based on the sampling design. We therefore utilized a multivariate analyses of variance (ANOVA) using sample size and sampling scheme as grouping variables. For each predictor it compared, whether the 100 guessed estimates per repetition had different means depending on the sampling design. A p-value below the confidence interval of 0.05 was interpreted as being significant. That means, that changing sample size or sampling scheme respectively would lead to a different estimate for this specific predictor. An ad-hoc Tukey test was computed where necessary, to pin point exceptional differences.

Table 2 – Statistical summary of accuracy metrics across all 2,000 experiments

| | Min | 1st Qu. | Median | Mean | 3rd Qu. | Max |
|---|---|---|---|---|---|---|
| $R^2$ | 0.6084 | 0.6994 | 0.7593 | 0.7618 | 0.8086 | 0.9517 |
| RMSE | 2.474 | 5.473 | 6.119 | 6.132 | 6.637 | 10.031 |
| MAE | 1.938 | 4.031 | 4.460 | 4.582 | 4.882 | 7.462 |

## 3 RESULTS

With this study, we aimed at deepening the previous experiments by Staab et al. (14) using END compliant maps for spatial extrapolations. We did so by investigating 2,000 LUR models produced to understand model variabilities as a function of sampling configurations.

### 3.1 Mean LM weights

Due to the intuitive interpretability, we scrutinized the intercept of the models first. Spread along the Y Axis, Figure 2 shows the estimated intercept value in dB(A). With values ranging between -2.1 and 329.0, these were on average - 153.2 dB(A) -, relatively high. Starting from here, the predicted noise levels will be lower when adding the rest of the term. At a sample size of 50, these estimates were found to highly vary and one can barely see differences between the colored sampling schemes. If at all, *random* sampled field campaigns lead to on average higher intercept estimates (mean 157.1) compared to *systematic* (mean 148.4), $stratified_{Lden}$ (mean 153.8) and $stratified_{Urb.Atl.}$ (mean 142.5) experiments. With increased sample sizes though, the variance decreased and this observation solidified. *Id est*, at sample sizes of 1,000, the mean estimate is 7.7 dB(A) higher for $stratified_{Lden}$ and -5.3 dB(A) lower at $stratified_{Urb.Atl.}$ compared to *random* sampling (mean 152.9dB(A); 153.7 dB(A) for *systematic* sampling). With the related ANOVA, we only found significant differences in regards to sampling scheme. The Tukey test pointed out that these differences stand out for $stratified_{Urb.Atl.}$ in comparison to the other three approaches, as well as significant differences between estimates derived from $stratified_{Lden}$ and *systematic* field campaigns.

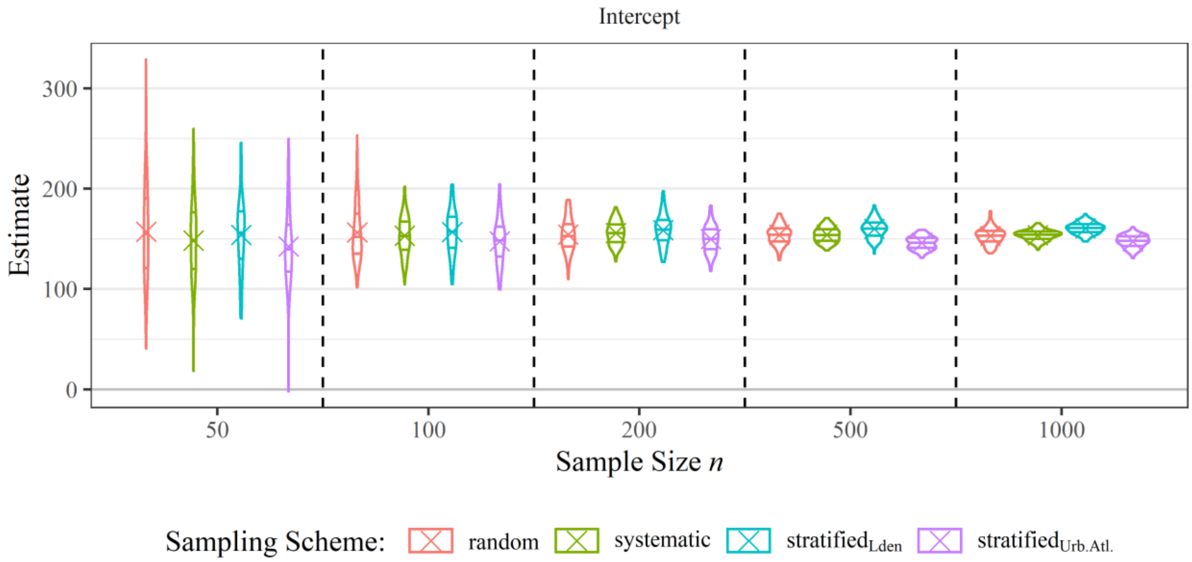


Figure 2 – Violin plots showing most frequent (width) estimated intercept value (Y axis) at different sample sizes (X axis) and sampling schemes (color). Vertical lines depict quartiles whereas cross (x) shows mean.

Looking at the estimated weights for our road variables (Figure 3), variance decreased at larger samples sizes, too. The distinct information was most visible at sample size of 1000. Thus, we focus in the following on describing these results in particular: Beginning with the log transformed distance to the closest road (Figure 3, left side), the mean estimate was negative for *motorways*, *trunks*, *primary*-, *secondary* and *residential* roads. This result is plausible as it would have predicted high $L_{den}$ values at close vicinity to these roads and the emitted sound levels decrease naturally over larger distances. In the following, we will refer to this effect as 'geometrical attenuation'. Physically, this effect is by nature constant and independent of the source. As part of a linear regression model though, the different estimates in the term rather correspond to respective emission levels of the source itself. A glimpse at the labeled Y-Axis of Figure 3 shows that the estimates for log-transformed distance to *motorways* were lowest. With respect to residential roads though, this rational was inverse and the estimates tended to even be positive. Looking at the mapped data, we could see this road type was most frequently appearing in residential areas. We thus comprehend a close vicinity to residential roads to proxy quiet areas and interpreted the estimate as confirming successful noise reduction

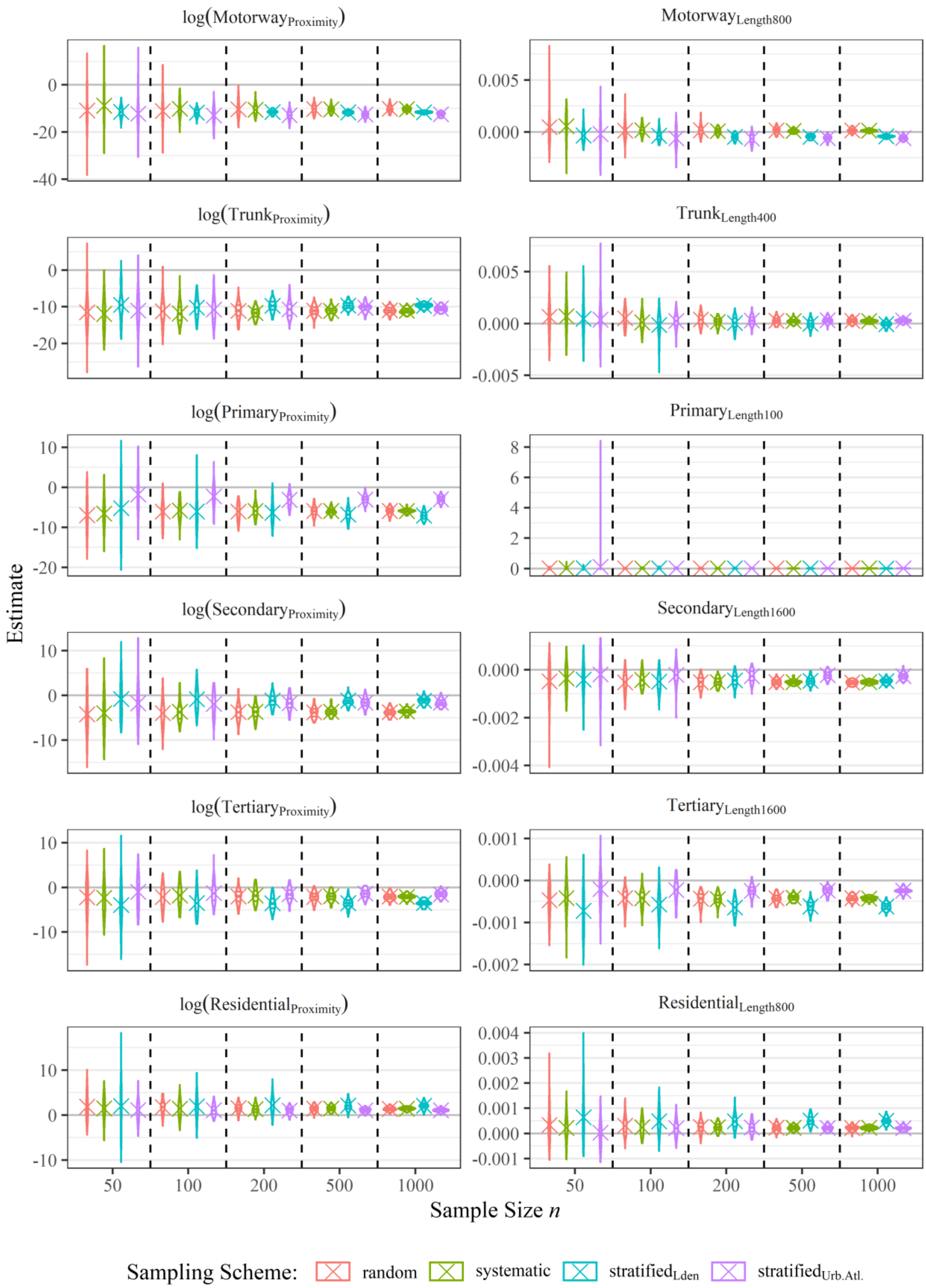


Figure 3 – Violin plots showing most frequent (width) estimated values (Y axis) for road variables (log transformed distance to closest road left and summed road length right) at different sample sizes (X axis) and sampling schemes (color). Vertical line

management (e.g. speed limits), whereas increased noise levels were found at larger distances to residential areas only. For a better readability, we will reference hereto as the ‘residential area effect’. Comparing the different sampling schemes next, we again saw conformity between *random* and *stratified* sampling, but distinct deviations for the two stratified schemes. Looking at $stratified_{Lden}$ first, the geometric attenuation effect was stronger compared to *random* sampling for *motorways*, *primary*- and *tertiary* roads and less with respect to *trunks* and *secondary* roads. Also, the residential area effect was most significant here. Estimates derived from samples using the $stratified_{Urb.Atl.}$ scheme showed a comparably mellow residential area effect and less geometric attenuation for all road types but *motorways*. Generally spoken, the variance analyses and the subsequent Tukey test confirmed significant different estimates for most road types and each potential comparison but *random* and *stratified* sampling. The only exception hereto was found for *primary* roads, where the estimates of $stratified_{Lden}$ were similar to the two.

Furthermore, observable trends for cumulated road lengths solidified at larger sample sizes. It is notable though, that the variances for $stratified_{Lden}$ experiments was generally bigger compared to the other 1,400. With respect to the Y Axis of Figure 3, right side, readers need to aware of the predictor units (Meters) relative to radius and the respective value range (c.f. Table 1). Significant differences can be seen for the estimates derived regarding the cumulated *motorway* length within an 800 Meter radius. Systematic and random sampling both tended to have slightly higher values compared to the two stratified approaches. Looking at $Trunk_{Length400}$ next, $stratified_{Lden}$’s mean was -0.00006 at $n$ = 1000, while positive for the others (*random* = 0.00027, *systematic* = 0.00024, $stratified_{Urb.Atl.=}$ 0.00028). Having investigated the mean estimates with respect to *primary* roads next, the values at $stratified_{Urb.Atl}$ were significantly higher compared to the ones derived using the other schemes. The mean estimates for $Secondary_{Length1600}$ was highest at $stratified_{Urb.Atl.}$ sampling (mean = -0.00027 at $n$ = 1000), while the Tukey test depicted the others to be compared thereto significantly lower. For tertiary roads, the mean estimates at $n$ = 1000 were -0.00042 and -0.00043 at *random* and *stratified* sampling, but as low as -0.00062 $stratified_{Lden}$ and only -0.00024 using $stratified_{Urb.Atl.}$. Last but not least, with respect to the sum of *residential* roads length within an 800 Meter radius, $stratified_{Lden}$ stood out. Here, the mean estimate at $n$ = 1000 was 0.0005, while only about half that much at the other three sampling configurations.

Reviewing the estimated weights for the nine environmental predictors (Figure 4), variance was decreasing at larger samples sizes again. It is interesting to note though, that the variance tended to be higher for $stratified_{Lden}$ in general though. The estimates regarding the topographical position index derived from the built-up model ($LoD1_{TPI800}$) tended to be rather negative. This corresponds to sound pressure levels being predicted to be higher in street canyons. Vice versa though, the TPI derived from the digital elevation model at a 1,600 Meter radius ($DEM_{TPI1600}$) has slight positive values for *random* and *systematic* sampling. The comparably larger positive values for $stratified_{Lden}$ would have led to higher predicated noise levels in superior locations. Interestingly, only $statified_{Urb.Atl.}$ showed a distinct tendency for negative estimates, such that noise levels were higher in valleys. The mean estimates with respect to the surrounding landcover fractions of *artificial land* were very close to zero. With respect to these fractions ranging between 0 and 100 percent, a weight of -0.044, such as found at $n$ = 1000 for both stratified approaches equally, led to a reduction of -4.4 dB(A) in the extreme cases of fully impervious surfaces (e.g. in the city center). An increased amount of *open soil* within the range of 1,600 Meters led to increased noise levels. Although physically, such materials have a sound absorbing effect, comparing it to the mapped data, we could see that this land cover class was most commonly located in disperse industrial and agricultural areas such as found along the major roads. *High, seasonal* and *low, perennial vegetation* both tended to decrease noise predictions, while the estimates were inconsistent for *high, perennial vegetation* and gravitated to be positive for *low, seasonal vegetation*. It was previously discussed though (14), that vegetation does only play a minor role in END compliant noise mapping. Thus, drawing conclusions from these estimates is difficult. Most probably, high vegetation factions rather indicated periphery. The estimates with respect to *water areas* within a neighboring scope of 400 Meters inclined to be positive for all schemes but $statified_{Urb.Atl.}$ sampling. This is logical, as water surfaces in general allow sound travel further distances.

### 3.2 LM Terms ANOVA

Summarizing the conducted ANOVA, differences in sampling scheme and sample size were pin pointed (Figure 5). Where $p$ is below 0.05, a significant difference was found to at least one group. Looking at the estimated *intercept* for example, p < 0.001 stresses a highly significant difference

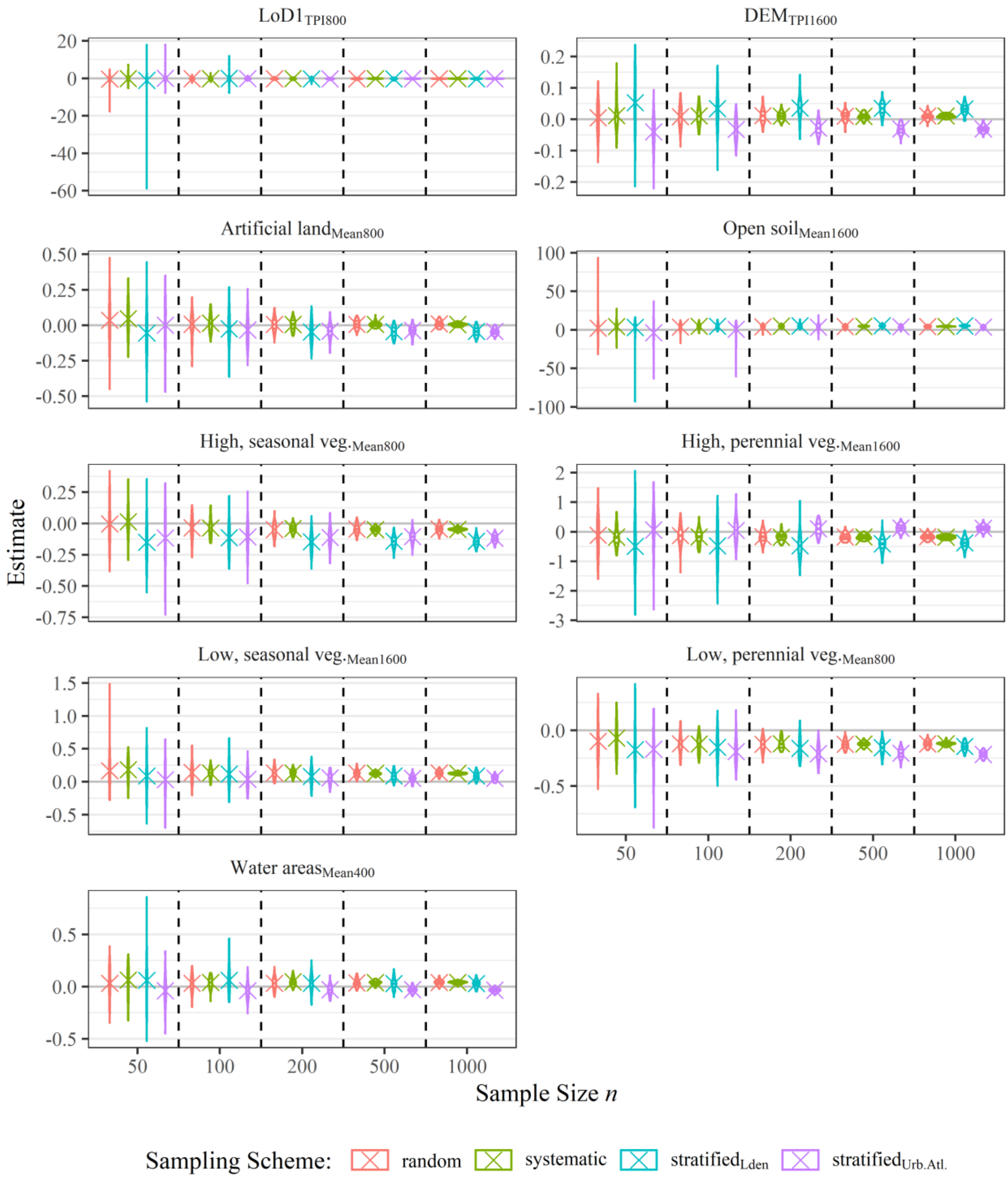


Figure 4 – Violin plots showing most frequent (width) estimated intercept value (Y axis) for environmental predictors at different sample sizes (X axis) and sampling schemes (color). Vertical lines depict quartiles whereas cross (x) shows mean.

between the sample schemes, but not with respect to sample size ($p$ = 0.101) and the ad-hoc Tukey-test has highlighted, where such differences occurred. In summary, it is shown that almost every estimate was influenced by the chosen sample scheme. Nevertheless, with respect to sample size, this was true for approximately every second estimate only.

## 4 DISCUSSION

In this study, we utilized 2,000 virtual field campaigns to investigate sampling effects on the eventual regression terms in LURs. Using ANOVA, we proved significant biases by sampling scheme

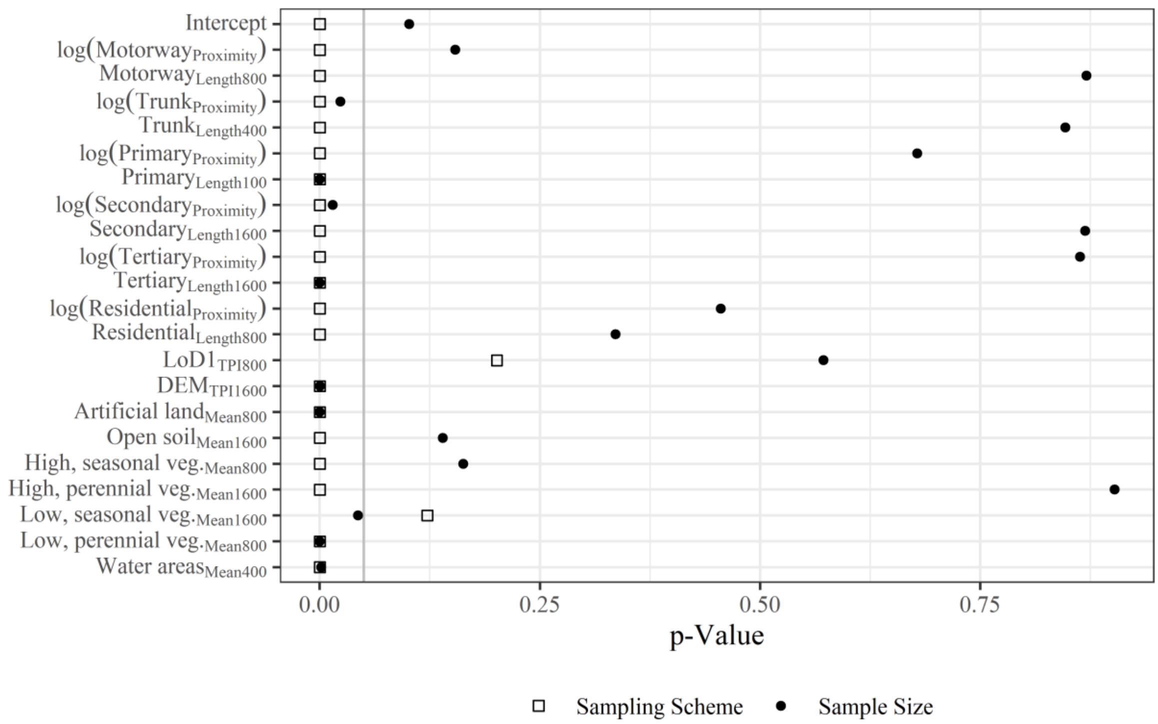


Figure 5 – Plot showing p-value (X axis) of ANOVA investigating estimates of predictors (Y axis) depending on the grouping variables sampling scheme and sample size (shape). Gray vertical line denoting 0.05 threshold for variable significance.

for almost all predictors and the same was true for nine of 22 predictors with respect to sample size. However, the interpretation of these results also requires a discussion of the methodological background.

First of all, the use of END compliant road traffic noise simulations from the city of Koblenz as training data kept the emitting sources constant and allowed for a rigorous repetition of drawing samples. We are aware though, that these samples differentiate from in-situ microphone measurements in regards to sematic content, observation period and expenditures when collecting very large sample sizes They can, for example, also measures soundscapes (c.f. 41). Also, the physical reachability of sample locations may be blocked due to access restrictions or the natural landscape. With respect to the investigated sampling schemes, we did only compare the three basic types identified. Population-weighted location-allocation, as conducted by Ragettli et al. (21), may be allocated into the group of stratified sampling, but was not reproduced here to limit the extent of conducted experiments. Nevertheless, we want to stress its semantic relevance for epidemiological studies.

Second, the 105 km² test site, Koblenz, is relatively small. So far, we argued, that with its heterogeneous urban morphology and landscape structure, it is an interesting test site for our experiments. Looking at the high variances of small sample sizes (i.e. $n = 50$), we assume the observed effects were stressed by this heterogeneity. While we thus overall promote sample sizes should be as large as possible, a limited scope of sample might be sufficient for small and homogenous research areas. This can particularly be true, if an appropriate sampling scheme was chosen, representing the population well.

Third, a comprehensive data set was compiled spanning twelve road traffic noise emission variables and nine predictors depicting the natural and built-up environment. However, the various scales selected using univariate regression models (as suggested by Ragettli et al. (21) and Liu et al. (23)), may have been biased by multifactorial relationships (14). Furthermore, no additional feature selection, as in forwards- or backwards selective regression, has been conducted, which may have preserved some covariances within the experiment. While we are aware these are decreasing the interpretability of the terms, a reduction was difficult to place within the setup as the selected predictors would possibly change between each iteration. Another approach would be switching to advanced machine learning models as presented by Liu et al. (23), but computing 2,000 random forests would have bound large computation resources. That said, the insights of this study allow the sampled data set constant in future studies and focus on such developments.

## 5 CONCLUSION

Lately, LURs are a trending topic in noise mapping for large-scale areas. With our previous study investigating spatial transferability (14) and the methodical addition on the regression terms themselves investigated in this study, we begin to understand the inside mechanics of LUR. By comparing 2,000 sample data sets, we have seen significant changes in the eventual regression terms. Some studies though use supervised LUR (e.g. 3), defining a terms direction *a priori*, e.g. log transformed distance to road should have a negative estimate and is excluded from the model elsewise. Rationally, this approach makes sense. With our experiments, we found regression weights varying across the zero line at smaller sample sizes particularly. Additionally, particular considering the surrounding environment at an extended radius may introduce proxy biases as well. Most important though, our experiments showed that the chosen sampling scheme can bias the prediction. Only *systematic* sampling was found to retrieve robust sample sets unbiased by the sample size. We therefore conclude that previously published studies have produced valuable noise maps with respect to their research aims. However, based on the findings of our study we want to stress that the interpretation of the respective terms and cross comparisons in general, requires to take the effects of the sampling schemes into account. We conclude that only a systematic analysis will allow, the estimated weights to highlight factors relevant in urban planning.

## ACKNOWLEDGEMENTS

This study was mainly funded by the German Federal Environmental Foundation (DBU), the project "Noise2NAKO$^{AI}$" (funding code ZT-I-PF-5-42) granted by the Helmholtz Association and the "Digitaler Atlas 2.0." project funded by the German Aerospace Center. The funding sponsors had no role in the design of the study, the interpretation, the process of writing of the manuscript, or the decision to publish the study. We want to furthermore thank the authorities of Koblenz for sharing the simulated noise levels. Further, both, the EU-DEM and the Urban Atlas were produced with funding by the European Union (https://land.copernicus.eu). The LoD1 model was provided by the German Federal Agency for Cartography and Geodesy (BKG, 2015). The administrative borders of the Rhine valley were provided by the Netherlands-palatinates Office for Land Surveying and Geoinformation (www.lvmergeo.rlp.de). Last but not least, we acknowledge Nils Karges for his support to this work.

## REFERENCES

1. World Health Organization. Regional Office for Europe (edit.). Environmental noise guidelines for the European Region. 2018. Available from: http://www.euro.who.int/en/publications/abstracts/environmental-noise-guidelines-for-the-european-region-2018
2. Öhrström E, Rylander R, Björkman M. Effects of night time road traffic noise—an overview of laboratory and field studies on noise dose and subjective noise sensitivity. J Sound Vib. 1988 Dec 22;127(3):441–8.
3. Xu X, Ge Y, Wang W, Lei X, Kan H, Cai J. Application of land use regression to map environmental noise in Shanghai, China. Environ Int. 2022 Mar;161:107111.
4. Cerletti P, Eze IC, Schaffner E, Foraster M, Viennau D, Cajochen C, et al. The independent association of source-specific transportation noise exposure, noise annoyance and noise sensitivity with health-related quality of life. Environ Int. 2020 Oct 1;143:105960.
5. Riedel N, Scheiner J, Müller G, Köckler H. Assessing the relationship between objective and subjective indicators of residential exposure to road traffic noise in the context of environmental justice. J Environ Plan Manag. 2014 Sep 2;57(9):1398–421.
6. 2002/49/EG. Assessment and management of environmental noise. L 189/12. 2002. Available from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:189:0012:0025:EN:PDF
7. Havard S, Reich BJ, Bean K, Chaix B. Social inequalities in residential exposure to road traffic noise: An environmental justice analysis based on the RECORD Cohort Study. Occup Environ Med. 2011 May 1;68(5):366–74.
8. Raess M, Valeria Maria Brentani A, Flückiger B, Ledebur de Antas de Campos B, Fink G, Röösli M. Association between community noise and children's cognitive and behavioral development: A prospective cohort study. Environ Int. 2022 Jan;158:106961.

9. Xie D, Liu Y, Chen J. Mapping Urban Environmental Noise: A Land Use Regression Method. Environ Sci Technol. 2011 Sep 1;45(17):7358–64.
10. Aguilera I, Foraster M, Basagaña X, Corradi E, Deltell A, Morelli X, et al. Application of land use regression modelling to assess the spatial distribution of road traffic noise in three European cities. J Expo Sci Environ Epidemiol. 2015;25(1):97–105.
11. Sieber C, Ragettli MS, Brink M, Toyib O, Baatjies R, Saucy A, et al. Land Use Regression Modeling of Outdoor Noise Exposure in Informal Settlements in Western Cape, South Africa. Int J Environ Res Public Health. 2017 Oct;14(10):1262.
12. Harouvi O, Ben-Elia E, Factor R, Hoogh K de, Kloog I. Noise estimation model development using high-resolution transportation and land use regression. J Expo Sci Environ Epidemiol. 2018 Nov;28(6):559–67.
13. Chang TY, Liang CH, Wu CF, Chang LT. Application of land-use regression models to estimate sound pressure levels and frequency components of road traffic noise in Taichung, Taiwan. Environ Int. 2019 Oct 1;131:104959.
14. Staab J, Schady A, Weigand M, Lakes T, Taubenböck H. Predicting traffic noise using land-use regression—a scalable approach. J Expo Sci Environ Epidemiol. 2022;1–12.
15. Staab J, Weigand M, Schady A, Wurm M, Lakes T, Taubenböck H. Europäische Lärmkarten – Methodik und Bewertung im Kontext überregionaler Umweltgerechtigkeitsstudien. In Bremen; 2020. (17; vol. IPP-Schriften).
16. Novotny EV, Bechle MJ, Millet DB, Marshall JD. National Satellite-Based Land-Use Regression: NO2 in the United States. Environ Sci Technol. 2011 May 15;45(10):4407–14.
17. Xu H, Bechle MJ, Wang M, Szpiro AA, Vedal S, Bai Y, et al. National PM2.5 and NO2 exposure models for China based on land use regression, satellite measurements, and universal kriging. Sci Total Environ. 2019 Mar 10;655:423–33.
18. Knibbs LD, Hewson MG, Bechle MJ, Marshall JD, Barnett AG. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ Res. 2014 Nov 1;135:204–11.
19. Federal Statistical Office Germany. GENESIS-online. Available from: https://www-genesis.destatis.de/genesis/online.
20. European Environment Agency. Environmental noise in Europe - 2020. 2020;(22):104.
21. Ragettli MS, Goudreau S, Plante C, Fournier M, Hatzopoulou M, Perron S, et al. 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. 2016 Nov;26(6):597–605.
22. Goudreau S, Plante C, Fournier M, Brand A, Roche Y, Smargiassi A. Estimation of Spatial Variations in Urban Noise Levels with a Land Use Regression Model. Environ Pollut. 2014 Sep 24;3(4):p48.
23. Liu Y, Goudreau S, Oiamo T, Rainham D, Hatzopoulou M, Chen H, et al. Comparison of land use regression and random forests models on estimating noise levels in five Canadian cities. Environ Pollut. 2020 Jan 1;256:113367.
24. Ostashev VE, Wilson DK. Acoustics in moving inhomogeneous media. CRC Press; 2015.
25. Schady A, Heimann D. Einfluss der Siedlungsform auf die Schallimmission an Fassaden – numerische Simulationen. Lärmbekämpfung. 2018;13(6):204–10.
26. Guedes ICM, Bertoli SR, Zannin PHT. Influence of urban shapes on environmental noise: A case study in Aracaju — Brazil. Sci Total Environ. 2011 Dec;412–413:66–76.
27. Heutschi K. A simple method to evaluate the increase of traffic noise emission level due to buildings, for a long straight street. Appl Acoust. 1995;44(3):259–74.
28. Federal Agency for Cartography and Geodesy (BKG). LoD1-DE. Available from: https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/sonstige-geodaten/3d-gebaudemodelle-lod1-deutschland-lod1-de.html
29. Weiss A. Topographic position and landforms analysis. In 2001.
30. European Environment Agency. EU-DEM v1.1 — Copernicus Land Monitoring Service. 2016. Available from: https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1
31. Attenborough K. Sound Propagation in the Atmosphere. In: Rossing TD, editor. Springer Handbook of Acoustics. New York, NY: Springer; 2007. p. 113–47.
32. Aylor D. Noise Reduction by Vegetation and Ground. J Acoust Soc Am. 1972;51(1B):197–205.
33. Margaritis E, Kang J, Filipan K, Botteldooren D. The influence of vegetation and surrounding traffic noise parameters on the sound environment of urban parks. Appl Geogr. 2018 May;94:199–212.
34. Weigand M, Staab J, Wurm M, Taubenböck H. Spatial and semantic effects of LUCAS samples on fully

automated land use/land cover classification in high-resolution Sentinel-2 data. Int J Appl Earth Obs Geoinformation. 2020 Jun 1;88:102065.
35. Reethof G. Effect of Plantings on Radiation of Highway Noise. J Air Pollut Control Assoc. 1973 Mar 1;23(3):185–9.
36. Garg N, Maji S. A critical review of principal traffic noise models: Strategies and implications. Environ Impact Assess Rev. 2014;46:68–81.
37. Steele C. A critical review of some traffic noise prediction models. Appl Acoust. 2001;62(3):271–87.
38. European Environment Agency. Urban Atlas 2018 — Copernicus Land Monitoring Service. 2020. Available from: https://land.copernicus.eu/local/urban-atlas/urban-atlas-2018
39. Carvalho DV, Pereira EM, Cardoso JS. Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics. 2019 Jul 26;8(8):832.
40. Raess M, Brentani A, Ledebur de Antas de Campos B, Flückiger B, de Hoogh K, Fink G, et al. Land use regression modelling of community noise in São Paulo, Brazil. Environ Res. 2021;199:111231.
41. Karges N, Staab J, Rauh J, Wegmann M, Taubenbock H. Soundscapes on edge - The real-time machine learning approach for measuring Soundscapes on resource-constrained devices. PROCEEDINGS of the 24th International Congress on Acoustics; 2022 28.10; Gyeongju, Korea.