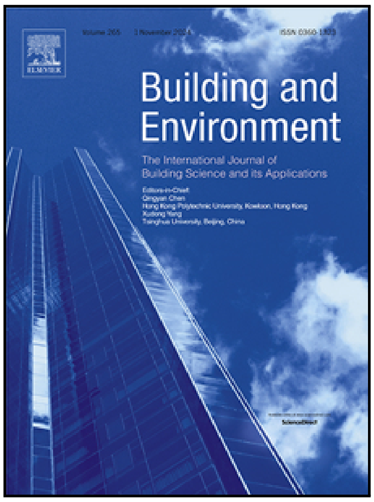
Journal Pre-proof



Machine Learning for Predicting Soundscape: From Individual-Level Deterministic Models to Group-Level Probabilistic Models

Huan Tong , Fan Xia , Andrew Mitchell , Francesco Aletta ,
Tin Oberman , Jian Kang

PII: S0360-1323(26)00002-8
DOI: https://doi.org/10.1016/j.buildenv.2026.114196
Reference: BAE 114196

To appear in: *Building and Environment*

Received date: 31 August 2025
Revised date: 29 December 2025
Accepted date: 31 December 2025

Please cite this article as: Huan Tong , Fan Xia , Andrew Mitchell , Francesco Aletta , Tin Oberman , Jian Kang , Machine Learning for Predicting Soundscape: From Individual-Level Deterministic Models to Group-Level Probabilistic Models, *Building and Environment* (2026), doi: https://doi.org/10.1016/j.buildenv.2026.114196

Highlights:

- Roughness, THD and vegetation are key predictors of the soundscape pleasantness.
- Lceq, Roughness, and RA are key features to perceive soundscape eventfulness.
- Gaussian process regression performs better at both individual and group levels.
- Group metrics perform better than individual metrics on capturing group perception.

Machine Learning for Predicting Soundscape: From Individual-Level Deterministic Models to Group-Level Probabilistic Models

Huan Tong [a], Fan Xia [a], Andrew Mitchell [b], Francesco Aletta [c], Tin Oberman [c], Jian Kang [c] *

[a] School of Architecture, Harbin Institute of Technology, Shenzhen, China

[b] The Bartlett School of Sustainable Construction, University College London, London, UK

[c] The Bartlett School of Environment, Energy and Resources, University College London, London, UK

* Corresponding Author

E-mail: j.kang@ucl.ac.uk

UCL Institute for Environmental Design and Engineering, The Bartlett, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, United Kingdom

Abstract:

Urban environments are used by a large number of diverse people but existing soundscape prediction model are focused on perceptual outcomes of an idealised average individual. With respect to developing group-level soundscape prediction models, it remains unclear which factors are important for predicting soundscapes and which types of models perform better for that task. Therefore, by relying on the International Soundscape Database this study aims at determining which factors can be used to predict soundscape and which model performs better at the group level. In this study, methods, such as correlation analysis, are used to select demographic, acoustic, visual, and geographic information factors that are significantly correlated with soundscapes. Subsequently, this study compares the performances of four models—linear regression, random forest, XGBoost, and Gaussian Process Regression (GPR)—in soundscape prediction tasks conducted at the individual and group levels. The results show that the equivalent sound pressure level ($|r| > 0.31$), roughness ($|r| > 0.34$), total harmonic distortion ($|r| > 0.31$), relative approach ($|r| > 0.30$) and vegetation ($|r| > 0.48$) are important to the soundscape prediction. The performance of the GPR model is better than the other three models at the individual level ($R^2_{ISOPleasant} = 0.36, MAE_{ISOPleasant} = 0.26, RMSE_{ISOPleasant} = 0.33, R^2_{ISOEventful} = 0.18, MAE_{ISOEventful} = 0.23, RMSE_{ISOEventful} = 0.29$). At the group level, the performance of the GPR model is also relatively high ($KL_{ISOPleasant} = 0.81, DME_{ISOPleasant} = 0.26, DME_{ISOEventful} = 0.38$). This study identifies the key acoustic and visual factors of soundscape perception and demonstrates the advantages of GPR. The introduction of a probability distribution–based framework is expected to predict soundscape at the group level and offer guidance for urban sound environment design.

Keywords: Soundscape prediction; Machine learning; Gaussian process regression; Group level evaluation

1. Introduction

Soundscape has been widely discussed in many fields, covering topics such as psychological restoration and implementation in urban planning[1,2]. A soundscape is defined by the International Organization for Standardization (ISO) as an "acoustic environment as perceived or experienced and/or understood by a person or people, in context"[3], and a simplification of soundscape perception is suggested to two-dimensional indicators, namely, ISOPleasant and ISOEventful[4]. As an important part of soundscape research, soundscape prediction holds application potential in the field of urban planning, which refers to the prediction of people's soundscape perceptions in unknown environments on the basis of existing soundscape data. Kang et al., noted that the importance of soundscape prediction is that it allows urban planners, architects and other designers to estimate how potential users will perceive the soundscape of a space they are designing, which they cannot otherwise assess[1,5]. As urban soundscape research and practice increasingly focus on holistic development, it is necessary to apply existing engineering tools and design methods to soundscape planning tasks in urban areas to promote the application of soundscapes in large-scale projects and a wide range of other fields[1,5,6]. Whether the goal is to determine the impacts of urban design or to integrate large-scale data at the community and urban levels, soundscape prediction models consisting of interacting factors constitute an important part of the implementation of soundscape[7]. To develop more practical soundscape prediction models, identifying the relevant factors and selecting appropriate modelling approaches remain central topics in research on soundscape.

Current research has identified factors that are correlated with soundscape perception, such as demographic, psychoacoustic, acoustic, visual, and contextual factors. These findings can help with selecting appropriate factors for soundscape prediction. First, demographic factors, such as age, gender, and education, are correlated with soundscape perception[7–9]. For instance, Fang et al., reported that females generally showed higher sensitivity and lower tolerance than males did towards several sounds and that higher education levels resulted in lower tolerance towards sounds. Guo et al., found that increased visit intensity could enhance the pleasantness perception[10]. Second, some studies have revealed that acoustic and psychoacoustic metrics are correlated with soundscape perception[7,8,11–16]. In terms of acoustic metrics, sound pressure level has been widely discussed. Regarding psychoacoustic metrics, loudness is the most widely considered factor, while other factors include sharpness, roughness, and the speech interference level[12,17–23]. For instance, Mitchell et al., reported that the A-weighted equivalent continuous sound pressure level (L_{Aeq}) is correlated with perceived pleasantness and eventfulness, with correlation coefficients of -0.34 and 0.37, respectively[24]. The research from Hong et al., finds that sound level played the most important role among all physical factors[25]. With respect to psychoacoustics factors, in the study by Mitchell et al., [24] loudness (N) is correlated with pleasantness (-0.37) and eventfulness (0.33), and roughness is correlated with pleasantness (-0.36) and eventfulness (0.32) [24]. Third, visual and contextual factors play important roles in soundscape research. Ricciardi reported that visual factors strongly influence the explained variance in soundscape pleasantness[13]. In terms of context, location is another important factor for conducting soundscape prediction, with location factors having higher influence on pleasantness, than on eventfulness, as reported by Mitchell et al., [24] and Erfanian et al[26]. Visual factors showed mediating effects on the soundscape restorativeness reported by Guo et al[27]. Furthermore, urban morphology and the percentages of natural and contextual factors are correlated with the perceptions of soundscapes[14,28,29]. Watts et al., reported that the greater the proportion of vegetation in the field of vision is, the more tranquillity people feel[15], whereas Puyana et al., reported that soundscapes are negatively correlated with areas possessing high building densities and positively correlated with the proportion of sea in the visual field[17]. Since different models have varying abilities to model the relationships between different variables, even though many factors have shown correlation with soundscape perception, it remains to be further discussed which factors can be used for soundscape prediction.

Many researches have attempted to predict soundscape based on different models. For instance, Mitchell et al., used linear multilevel models to predict soundscape perception during the covid pandemic[24] and Kang et al., used linear regression to predict soundscape based on the data collected from individual responses[30]. With the development of artificial intelligence, many machine learning models have been used for soundscape prediction. For instance, random forest (RF), eXtreme Gradient Boosting (XGBoost), and support vector machine (SVM) are widely used for soundscape prediction. Versümer et al., used LR, RF, XGBoost, and SVM to investigate the soundscape model while considering the fixed and mixed effects on different soundscape datasets[23]. An experiment conducted by Fan and Giannakopoulos et al., demonstrated the potential of the SVM model in soundscape prediction tasks[31,32], and Lunden et al., used an SVM to predict soundscape perceptions and reported that it is possible to predict the responses of human to acoustic environments on the basis of data derived from acoustic signals[33]. Zhao et al., used a gradient-boosted regression tree (GBRT) model to predict soundscapes, achieving an R^2 of 0.48 [29]. Meanwhile, with the advancement of artificial neural networks (ANNs), Yu and Kang constructed an ANN model to predict soundscape perceptions[34]. Their experiment demonstrates that the ANN model performs effectively and reliably in soundscape prediction tasks. Machine learning shows considerable potential for application in soundscape prediction tasks. However, above studies build models by leveraging existing datasets of individual perceptual responses to predict the perception of each record. Specifically, these studies select deterministic models, such as RFs, XGBoost and other AI models[23,34,35]. These models operate by identifying specific functions within a high-dimensional vector space to fit the individual-level soundscape perception data obtained from questionnaires. Moreover, these studies rely individual-level (pointwise or per-data) evaluation criteria for their models, such as R^2, the $RMSE$, etc. These evaluation standards describe the accuracy of models in terms of predicting individual soundscape perceptions. However, urban environments are intended sometimes for specific groups and communities, and sometimes also for unknown users and groups of users, however rarely for idealised individuals, so paying more attention to soundscape perception at the group level is necessary so to address public space users in an inclusive and sustainable way. Soundscape perception at the group level can be described as a set of soundscape perceptions perceived by a group of people in the same content. Individual-level prediction requires the accurate determination of the specific rating assigned by each person. In contrast, soundscape prediction at the group level involves predicting the proportion of different soundscape evaluation of individuals. Most current research focuses on the individual level, while discussions concerning soundscape prediction at the group level remains limited. In this case, the goal of soundscape prediction can change from accurately predicting individual perceptions to predicting the aggregate soundscape perception at the group level—in other words, predicting not individual soundscape perceptions but the soundscape perception distribution of a group of people.

Therefore, this study aims to explore the potential of machine learning models for predicting soundscape perceptions at the group level and evaluate the model performance. In this context, this study attempts to answer the following research questions.

(1). Considering the soundscape prediction task at the group level, which factors are correlated with soundscape perception (ISOPleasant and ISOEventful) and thus can be used for soundscape prediction?

(2). Among the different models that are available for soundscape prediction, what metrics can be used to evaluate their performance at the group level, and which model performs best at the individual and group levels?

In this study, by relying on the International Soundscape Database, Linear Regression (LR), Random Forest (RF), XGBoost (XGB), and Gaussian process regression (GPR) models are selected for soundscape prediction. Pointwise and groupwise accuracy metrics are used to assess the models at both the individual and group levels.

2. Methodology

This study is designed to answer each research question using corresponding methods (Figure 1). Data were sourced from the International Soundscape Database (ISD) and OpenStreetMap (OSM). Through classification and extraction processes, we obtained survey data (including soundscape perceptions and demographics), acoustic data, visual data from videos, and geographic information. Fivefold cross-validation was applied to the dataset. The employed feature extraction methods were tailored to each model to identify the factors that were relevant to soundscape perception. The trained models were tested on test sets, and the performance attained across all folds was aggregated to evaluate the predictive effectiveness of each model at both the individual and group levels.

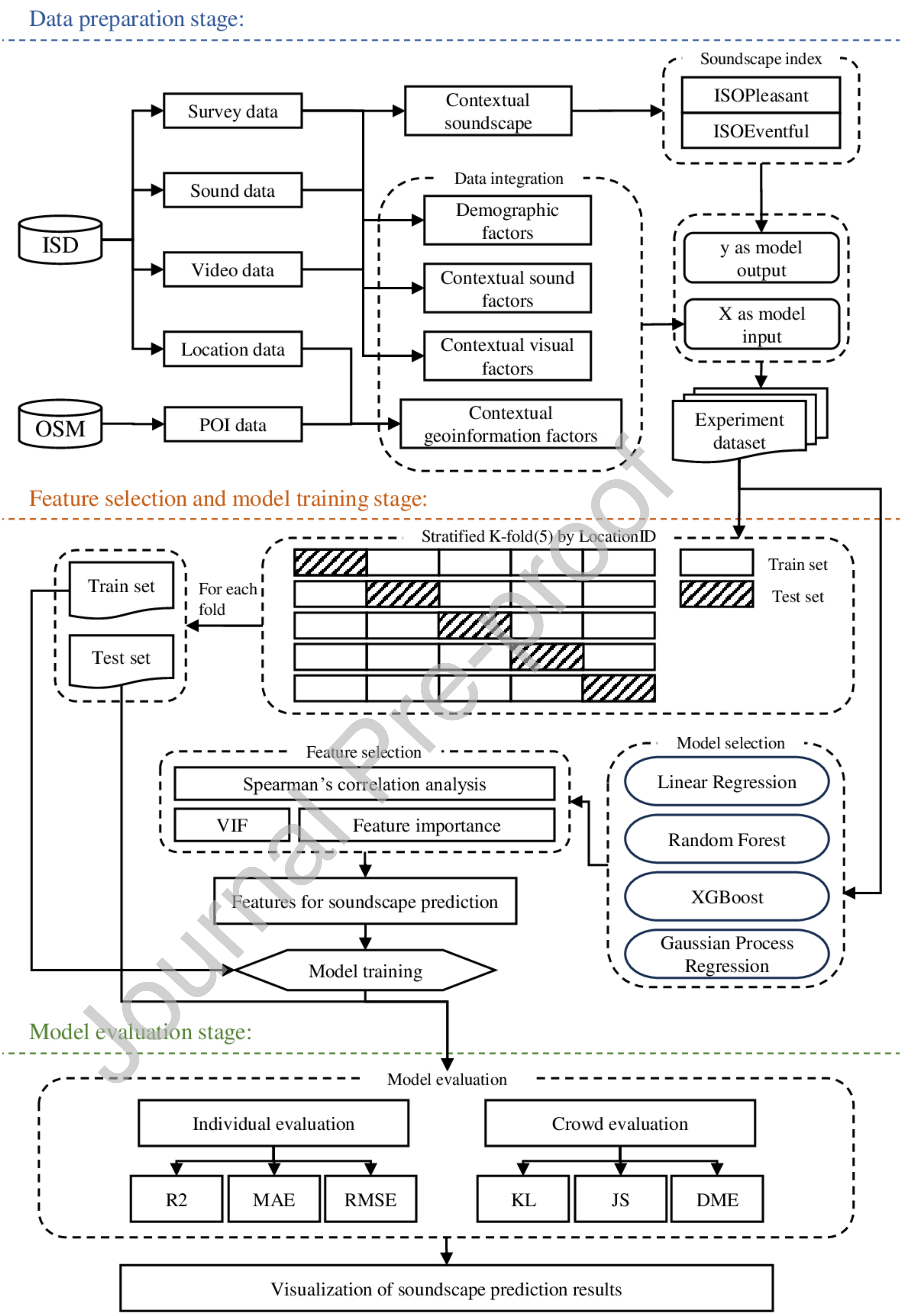


Figure 1 Overall experimental flow and framework of this study.

Specifically, Section 2.1 describes the soundscape method and introduces the source and description of the soundscape database. Section 2.2 provides a detailed introduction to the classification and filtering methods employed for the soundscape prediction-related factors, and Section 2.3 introduces the models used in this study and the reasons for choosing these models. It also explains how these models were evaluated and the reasons for choosing specific evaluation methods.

2.1 Soundscape description and data sources

A soundscape is an acoustic environment as perceived by humans in context. In general, there are many different ways to describe soundscapes. To simplify the process of quantifying soundscapes and aid subsequent researchers in terms of comparing their experimental results, this study employed the widely used soundscape description standard published by the International Organization for Standardization (ISO) in 2014[3]. As the most widely used method for describing soundscapes, this standard involves describing soundscape perceptions based on eight perceptual attributes (PAs)[36]. The PAs are derived from a Likert-based questionnaire containing 8 scales, including pleasant, vibrant (or exciting), eventful, chaotic, annoying, monotonous, uneventful, and calm (Figure 2). During the questionnaire procedure, these PAs are assessed independently of each other; however, they are conceptually considered to form a two-dimensional circumplex with pleasantness and eventfulness on the x- and y-axes, respectively, where all regions of the space are equally likely to accommodate a given soundscape assessment[37]. To facilitate analysis of PA responses, Part 3 of ISO 12913 provides a coordinate transformation into the two primary dimensions on the basis of the $45°$ relationship between the diagonal axes and the pleasant and eventful axes[3]. In theory, this coordinate pair then encapsulates information from all 8 PA dimensions into two dimensions that are more easily understandable and analysable. The ISO coordinates and the PAs are arranged around the circumplex, as shown in Figure 2. The $cos45°$ term operates to project the diagonal terms down onto the x- and y-axes, and $\frac{1}{4+\sqrt{32}}$ scales the resulting coordinates to the range of $(-1, 1)$. The results of this transformation are shown in Figure 2. In this approach, soundscape perception has a standard description and modelling method[38]. In summary, this study employed ISOPleasant and ISOEventful to describe the two dimensions of soundscape perceptions, which were also the prediction targets of the models.

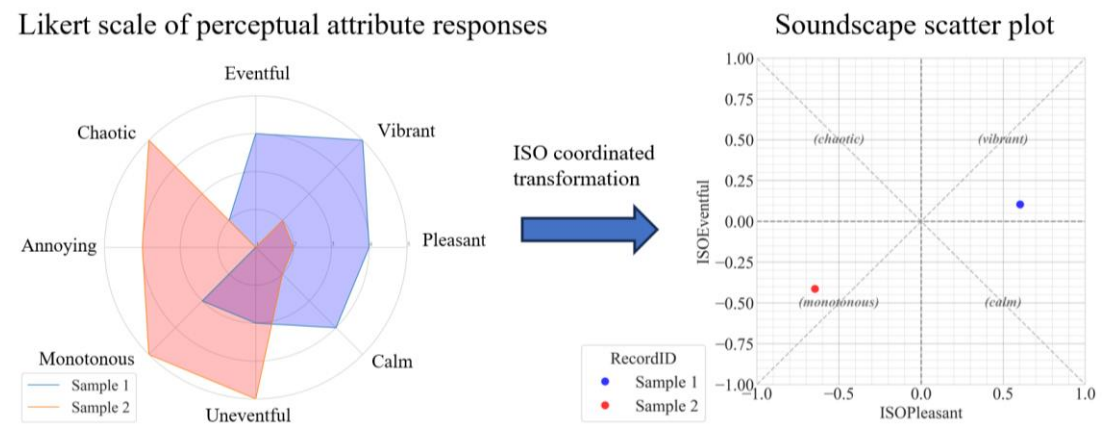


Figure 2 Example representations of two soundscape assessments. Left: Radar plot of two example perceptual attribute (PA) ratings on the Likert scale (1–5). Right: Scatter plot of the same assessments on the soundscape circumplex, transformed according to ISO 12913 Part 3.

On the basis of a literature review, this study selected soundscape perceptions (ISOPleasant and ISOEventful) as the prediction targets for the model. With respect to factor selection, the factors used in this study were categorized into four groups: demographic, acoustic, visual, and geographic factors. The soundscape perception data, demographic data, acoustic factors data, and visual factors data were provided by the International Soundscape Database (ISD), while the geographic factor data were provided by the Python API of OSM.

The ISD was utilized as the primary data source in this study[39]. This database comprises results derived from a series of soundscape assessment campaigns conducted across Europe, following a data collection protocol [40] designed to provide survey, environmental and acoustic data collected in urban public spaces with the goal of creating a unified dataset for the development of a predictive soundscape model, as well as a soundscape index. The protocol contained two-stages: a “Recording Stage,” during which approximately 15-minute 360° videos, corresponding spatial audio recordings and sound level measurements were captured across different locations, and a subsequent “Questionnaire Stage”. In the latter stage, on-site soundscape evaluations were collected from the public using the questionnaire based on ISO/TS 12913-2: 2018[41], yielding both demographic information and perceptual assessments. Simultaneously, acoustic measurements were continuously recorded using sound level metres, while 30-second binaural audio recordings and 360° photographs were obtained for the participants taking questionnaires at that moment (usually one or two individuals at a time) to document their immediate perceptual environments. These recordings were subsequently analysed to derive corresponding acoustic and psychoacoustic metrics. This research used perceptual and demographic data obtained from the questionnaire stage of the ISD, consisting of 1,147 valid records obtained from London stored in the comma-separated value (CSV) format. The acoustic measures were derived from binaural recordings collected during the same questionnaire stage. A computational analysis of these recordings produced the required acoustic and psychoacoustic factors, which were also provided in the CSV format and maintained a direct correspondence with the perceptual and demographic data. The visual data were obtained from the 360° videos captured during the ISD recording stage. For each location, frames were systematically sampled from the video recordings at predetermined intervals. These frames were then processed using a semantic segmentation algorithm based on the Cityscapes dataset to quantify the relative proportions of different visual elements (e.g., buildings, sky, roads, vegetation, and vehicles) [42]. The resulting metrics, categorized by scene type, were maintained in CSV files.

Table 1 Acoustic and psychoacoustic factors included in the International Soundscape Database (ISD), along with their symbols, units, and standards.

Factors	Symbol	Unit	Standard
L_{Zeq}	L_{Zeq}	dB	IEC 61672
L_{Ceq}	L_{Ceq}	dB	IEC 61672
L_{Aeq}	L_{Aeq}	dB	IEC 61672
Loudness	N	sones	ISO 532-1
Sharpness	S	acum	ISO 532-1
Roughness	R	asper	Hearing Model
Tonality	T	tuHM	ECMA-74 (17th)
Fluctuation Strength	FS	vacil	Hearing Model
(Total) Harmonic Distortion	THD		
Impulsiveness	I		Hearing Model
Speech Interference Level	SIL4	dB	ISO 9921:2003
Relative Approach 2D	RA	cPA	Hearing Model

Regarding geographic factors, this study used OSM as the data source. Specifically, on the basis of the latitude and longitude coordinates provided in the ISD, a 100-m buffer was established. The number of POIs within this buffer was considered a geographic factor. This is because the number of POIs can indirectly reflect the land use type of an area, which in turn affects people's sound perceptions. For instance, a high number of bars indicates that the surrounding area is likely a business region, which often indicates that the area is likely to be more eventful. Because OSM provides a wide variety of POI types, each with similarities, this study consolidated these POI types. Similarly, because the classifications of the urban visual elements contained in the Cityscapes dataset also overlap, the visual factor data were also consolidated.

2.2 Factor selection

In this study, different factor extraction methods were used for different models to obtain characteristic factors that could be used for soundscape prediction. The advantage of this approach is that it allowed us to assess the contributions of various factors to soundscape prediction from multiple perspectives. Because different models utilize different prediction methods, different factors contribute differently to their prediction performance. For instance, a linear model selects only factors with linear relationships for soundscape prediction purposes, whereas nonlinear features are not recognized by the model.

For the LR model, this study chose to calculate the correlation coefficient and variance inflation factor (VIF) to extract the characteristic factors. Specifically, factors with absolute correlation values greater than 0.3 and VIFs less than 10 were selected as model input factors. With respect to the RF and XGB, feature importance values were used in this study to gradually screen for factors. Specifically, first, an importance threshold was set, then all factors were used for modelling, factors with importance levels greater than the threshold were screened out, these factors were then remodelled, and the factors with importance values greater than the threshold were screened out again. This process was repeated until the number of factors no longer changed, and finally, the characteristic factors were obtained. For the GPR model, this study used automatic relevance determination (ARD) to extract the utilized factors. Specifically, ARD regression was used to calculate the feature importance values, and the factors with parameter values that were less than a threshold were selected as important. Because ARD uses a kernel method, more important factors have smaller parameters. Finally, a factor combination that was suitable for GPR soundscape prediction was obtained.

2.3 Models and evaluation standards

2.3.1 Models

In this study, four models were selected to predict soundscapes: linear regression (LR), a random forest (RF), XGBoost (XGB) and Gaussian process regression (GPR). LR was selected because of its simplicity, transparency, and wide applicability. As a baseline model, LR offers an interpretable benchmark for understanding the linear contribution of each factor to the perceptual outcomes, and it has also been used in other studies as a base model[8,12–15,17,19,20,43]. The RF and XGBoost are ensemble learning methods that are capable of capturing nonlinear relationships and high-order interactions between input variables. Their built-in feature importance mechanisms help identify the key environmental factors that influence soundscape perception, thereby contributing to both performance and interpretability[23,44]. LR, the RF, and XGB models are deterministic models that predict soundscapes from parameter spaces, whereas GPR is a probabilistic model that predicts soundscapes from functional spaces. The GPR model differs from the other models in that it provides a probabilistic modelling framework that does not directly predict the target function but rather predicts the distribution of the target function to describe the associated functional relationship[45]. One can consider a Gaussian process as defining a distribution over functions, where inference takes place directly in the function space, i.e., the function space view. Specifically, GPR first assumes that all functions describe the true function with equal probability (a priori), then optimizes the probabilities of all functions through training data (a posteriori), and finally obtains a function probability distribution that can describe the true function. The GPR output includes not only the mean value as the predicted value but also the associated uncertainty interval, which is particularly useful for analysing the subjective variability exhibited by perceptual responses and has been used in other studies[46,47]. Furthermore, GPR, as a nonparametric probabilistic model, does not require the distribution of each factor to be certain or the same. It only assumes that the noise related to each factor follows a Gaussian distribution[45]. To explore whether probabilistic models perform better than deterministic models do in soundscape prediction tasks conducted at the group level, all machine learning models should demonstrate optimal performance to enable a comparison between the applicability levels of probabilistic and deterministic modelling approaches. This means that each model needs to individually choose the model inputs and hyperparameter settings that will provide the best performance for that model. By comparing the best performances of different models, we can infer whether probabilistic models are more suitable for soundscape prediction than are deterministic models. This method has also been widely used in other studies[48–51].

Because the ISD data were collected from various areas of London, to reduce the impact of regional inequality, this study adopted a stratified 5-fold data splitting strategy based on LocationID, splitting the data into training, validation, and test sets at an 8:1:1 ratio. Stratified K-fold splitting is a cross-validation method and a variation of K-fold splitting that returns stratified folds. The folds were made by preserving the percentage of the samples possessing each LocationID, which means that each set contained approximately the same percentage of samples for each location as that in the ISD. The training set was used for model training purposes, the validation set was used for hyperparameter adjustment, and the test set was used for model evaluation tasks.

2.3.2 Evaluation indicators

Among the traditional evaluation criteria employed for soundscape prediction models, the individual prediction accuracy is often used performance evaluation purposes. This involves metrics such as the coefficient of determination and root mean square error. These indicators assess the accuracy of soundscape predictions from an individual perspective. However, in an urban context, planners, architects, urban designers and others often pay more attention to the soundscape perception evaluation of a group [52]. Therefore, more methods are needed to describe the accuracy of soundscapes prediction model from a group perspective. In an attempt to evaluate the performance of the constructed model at the group level, this study employed a variety of quantitative metrics, thus reflecting the accuracy of soundscape perception prediction at the group level.

Evaluation conducted at the individual level:

To evaluate the individual perspective, commonly used model performance evaluation indicators were selected to examine the performance of each model.

(1). Coefficient of Determination (R^2): This metric reflects the proportion of variance explained by the model and is defined as follows:

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{1}$$

where SSR is the sum of squares of the residuals, which is also called the residual sum of squares; SST is the total sum of squares; y_i denotes the true values; $\hat{y}_i$ represents the predicted values; and $\bar{y}$ is the mean of the true values.

(2). Root Mean Square Error (RMSE): This metric is the square root of the average squared difference between the predicted and true values.

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2} \tag{2}$$

where n is the number of samples; y_i denotes the true values; and $\hat{y}_i$ represents the predicted values.

(3). Mean Absolute Error (MAE): This measure captures the average magnitude of the errors without considering their directions.

$$\mathrm{MAE}(\mathrm{y}, \hat{\mathrm{y}}) = \frac{1}{n}\sum_{\mathrm{i}=0}^{n-1} |\, \mathrm{y}_{\mathrm{i}} - \hat{y}_i \,| \tag{3}$$

where n is the number of samples; y_i denotes the true values; and $\hat{y}_i$ represents the predicted values.

Evaluation conducted at the group level:

To evaluate the performance of the prediction models from a group perspective, this study considered the soundscape perception of a group of people as a whole and used the soundscape perception distribution composed of these soundscape perceptions as the soundscape characteristics of the corresponding scene. Therefore, the degree of matching between the predicted perceived distribution and the true distribution was selected as a basis for judgement, and this study incorporated the Kullback–Leibler (KL) divergence measure (4), the Jensen–Shannon (JS) divergence measure (5), and the distributional mean error (DME) (6) into the evaluation framework.

(1). The KL divergence measure quantifies the difference in information between two probability distributions and is defined as follows:

$$D_{KL}(P \parallel Q) = \sum P(x) log\left(\frac{P(x)}{Q(x)}\right) \tag{4}$$

where $Q(x)$ is the predicted probability distribution; $P(x)$ is the true/target probability distribution; and x represents the soundscape perception index (ISOPleasant or ISOEventful), which ranges from [-1, 1].

The KL divergence indicates the amount of information that is lost when the distribution Q is used to approximate the true distribution P[53]. In this study, the smaller the KL divergence is, the lower the cost of fitting the probability distribution of the model prediction results to the true distribution, and the higher the model accuracy. This metric is particularly useful in model calibration and probabilistic inference tasks, where the distributional shape is of interest rather than just the point estimates.

(2). The JS divergence (JS) is an optimization of KL divergence and is defined as follows:

$$D_{JS}(P \parallel Q) = \frac{1}{2}\sum_{x \in \mathcal{X}}\left[P(x)\, log\frac{2P(x)}{P(x)+Q(x)} + Q(x)\, log\frac{2Q(x)}{P(x)+Q(x)}\right] \tag{5}$$

where $Q(x)$ is the predicted probability distribution; $P(x)$ is the true/target probability distribution; and x represents the soundscape perception index (ISOPleasant or ISOEventful), which ranges from [-1, 1].

Unlike the KL divergence, the JS divergence is symmetric and does not change with changes in the order of the input distribution. Furthermore, owing to its calculation method, the JS divergence has a range of $[0, 1]$, whereas the KL divergence has no upper bound. Furthermore, the KL divergence is sensitive to differences in high-probability events but insensitive to differences in low-probability events. This means that if high-probability events vary significantly, the KL divergence value fluctuates dramatically, while the JS divergence compensates for this effect to some extent. Similar to the KL divergence, the smaller the value of the JS divergence is, the more similar the two compared distributions are and the better the predictive performance of the evaluated model.

(3). Distributional Mean Error (DME): To further capture distributional consistency, an accuracy metric based on the MAE between two probability density functions, which is generalized over a continuous interval, is proposed in this work. Given two continuous distributions $Q(x)$ and $P(x)$, the DME is defined as follows:

$$DME = \frac{1}{|b-a|}\int_a^b |Q(x) - P(x)| \tag{6}$$

where $Q(x)$ is the predicted probability distribution; $P(x)$ is the true/target probability distribution; a, b are the integration bounds, which are set to $[-1,1]$ in this study; and x represents the soundscape perception index (ISOPleasant or ISOEventful), which ranges from [-1, 1].

To obtain $Q(x)$ and $P(x)$, kernel density estimation (KDE) is used to smooth and estimate the underlying distributions of both the ground truths and the model predictions. Geometrically, this method first calculates the area between the two distribution curves and then normalizes it over the integration range, yielding an interpretable measure of the average absolute deviation between the two functions. This approach intuitively reflects the average distributional mismatch across the perception spectrum, offering a more holistic evaluation of the predictive mean error in perceptual modelling cases (Figure 3). In terms of values, the closer the KL divergence, JS divergence and DME are to 0, the better the performance of the tested model.

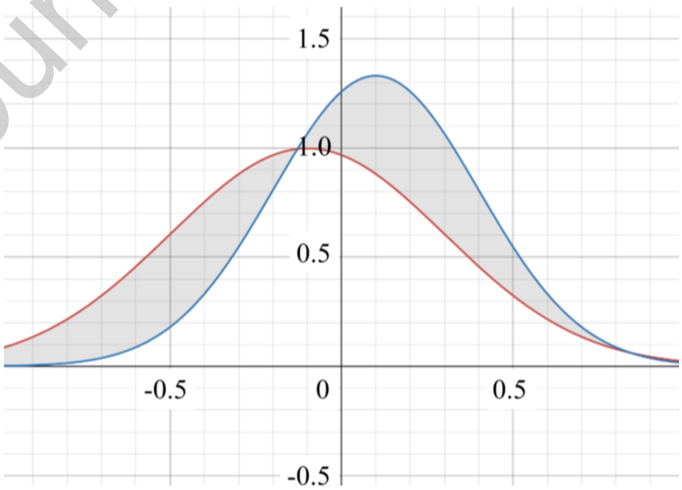


Figure 3 Schematic illustration of the distribution mean error. (Note: The grey area is correlated with the conceptual geometric interpretation of the distribution mean error.)

In summary, in this study, the R^2, $RMSE$ and MAE were used to evaluate the performance of the tested models at the individual level, and the KL, JS divergence and DME were used to evaluate the performance achieved at the group level.

3. Results

3.1 Factor selection

In this study, different feature extraction methods were used for different soundscape prediction models, including correlation analysis, the VIF and feature importance rankings. When the prediction target was ISOPleasant, the selected predictive factors varied across the models (Table 2). For the LR model, the selected factors included $L_{Zeq_{Min}}$, R_{95}, THD_{Max}, I_{50} and RA_{50} as acoustic factors; *vegetation* and *traffic vehicles* as visual factors; *and traffic and public services* as geo-information factors. The RF model incorporated *age* as a demographic factor; $L_{Zeq_{Min,Max,5,10,90}}$, N_{95}, $S_{Max,95}$, $R_{Min,Max,95}$, THD_{Max}, I_{Min} and $SIL4_{Min}$ as acoustic factors; *and vegetation and terrain* as visual factors. For the XGB model, the selected factors included $L_{Zeq_{Mean,90,95}}$, $L_{Ceq_{Mean,Max,5,50,90,95}}$, $L_{Aeq_{Max,5}}$, N_{10}, $R_{Mean,Min,10,90,95}$, THD_{Max} and $SIL4_{Max}$ as acoustic factors and vegetation and terrain as visual factors. Finally, the GPR model included $L_{Zeq_{Min}}$, $L_{Aeq_{50}}$, S_{95}, $R_{90,95}$, $FS_{Mean,50}$, $THD_{Max,50,90}$, $I_{50,95}$, $SIL4_{Mean,50}$ and RA_{50} as acoustic factors; *vegetation, traffic vehicles and human activity* as visual factors; *and traffic and cultural education* as geo-information factors. To summarize the relative importance of the predictive factors across the four models (LR, the RF, XGB, and GPR), the factors were categorized on the basis of their selection frequencies. The acoustic factors R and THD and the visual factor $vegetation$ were selected by all four models, indicating their high importance in soundscape prediction tasks. The factors selected by the three models included the acoustic factors L_{Ceq}, I and $SIL4$, suggesting that these factors were of secondary importance.

Table 2 Table of the important factors selected by the four models (LR, the RF, XGB and GPR) for predicting ISOPleasant; these factors are classified into demographic, acoustic, vision and geographic groups. A check mark indicates that the corresponding factor was selected by the model. For the acoustic factors, the selected factors are presented as percentiles.

Factors		**Models**			
Factor class	**Factor name**	**LR**	**RF**	**XGB**	**GPR**
Demographic	Age		√		
	Gender				
Acoustic	L_{Zeq}	$L_{Zeq_{Min}}$		$L_{Zeq_{Mean,90,95}}$	$L_{Zeq_{Min}}$
	L_{Ceq}		$L_{Ceq_{Min,Max,5,10,90}}$	$L_{Ceq_{Mean,Max,5,50,90,95}}$	
	L_{Aeq}			$L_{Aeq_{Max,5}}$	$L_{Aeq_{50}}$
	N		N_{95}	N_{10}	
	S		$S_{Max,95}$		S_{95}
	R	R_{95}	$R_{Min,Max,95}$	$R_{Mean,Min,10,90,95}$	$R_{90,95}$
	T				
	FS				$FS_{Mean,50}$
	THD	THD_{Max}	THD_{Max}	THD_{Max}	$THD_{Max,50,90}$
	I	I_{50}	I_{Min}		$I_{50,95}$
	SIL4		$SIL4_{Min}$	$SIL4_{Max}$	$SIL4_{Mean,50}$
	RA	RA_{50}			RA_{50}
Visual	Vegetation	√	√	√	√
	Terrain		√	√	
	Sky				
	Traffic Vehicles	√			√
	Traffic Infrastructure				
	Human Activity				√
	Buildings				
Geographic	Traffic	√			√

	Business				
	Cultural/Educational				√
	Public Services	√			
	Medical Health				
	Religious Special				
	Other				

Note: X_n represents X factors with n percentiles, which can be Min, Mean, Max,5, 10, 50, 90, or 95.

When the prediction target was ISOEventful, the employed predictive factors varied across models (Table 3). For the LR model, the selected factors included $L_{Ceq_{10}}$, S_{Mean}, R_{Mean}, FS_{Mean}, THD_{10} and RA_{Min} as acoustic factors *and business and medical health* as geo-information factors. The RF model incorporated *age* as a demographic factor; $L_{Ceq_{Mean}}$, $R_{Min,50,90,95}$, $T_{Max,50,90,95}$, $THD_{Mean,5,10}$, I_{Min} and RA_{Min} as acoustic factors; *vegetation* and *sky* as visual factors; *and business* as a geo-information factors. For the XGB model, the selected factors included $L_{Ceq_{Mean}}$, $L_{Aeq_{Mean}}$, N_{10}, $R_{Min,Mean,5,50,90,95}$, THD_5, $SIL4_5$ and RA_{Min} as acoustic factors. Finally, the GPR model included $L_{Ceq_{10}}$, L_{Aeq_5}, N_{10}, S_{10}, R_{Min}, T_{10}, I_{Min}, $SIL4_5$ and $RA_{Min,5}$ as acoustic factors; *terrain* as a visual factor; and *business* and *medical health* as geo-information factors. To summarize the relative importance levels of the predictive factors across the four models (LR, the RF, XGB, and GPR), the factors were categorized on the basis of their selection frequencies. The acoustic factors L_{Ceq}, R and RA were selected by all four models, indicating their high importance in soundscape prediction tasks. The factors selected by the three models included the acoustic factor THD and the geo-factor $business$, suggesting that these factors were of secondary importance.

Table 3 Table of the important factors selected by the four models (LR, the RF, XGB and GPR) for predicting ISOEventful; these factors are classified into demographic, acoustic, vision and geographic groups. A check mark indicates that the corresponding factor was selected by the model. For the acoustic factors, the selected factors are presented as percentiles.

Factors		**Models**			
Factor class	**Factor name**	**LR**	**RF**	**XGB**	**GPR**
Demographic	Age		√		
	Gender				

Acoustic	L_{Zeq}				
	L_{Ceq}	$L_{Ceq_{10}}$	$L_{Ceq_{Mean}}$	$L_{Ceq_{Mean}}$	$L_{Ceq_{10}}$
	L_{Aeq}			$L_{Aeq_{Mean}}$	L_{Aeq_5}
	N			N_{10}	N_{10}
	S	S_{Mean}			S_{10}
	R	R_{Mean}	$R_{Min,50,90,95}$	$R_{Min,Mean,5,50,90,95}$	R_{Min}
	T		$T_{Max,50,90,95}$		T_{10}
	FS	FS_{Mean}			
	THD	THD_{10}	$THD_{Mean,5,10}$	THD_5	
	I		I_{Min}		I_{Min}
	SIL4			$SIL4_5$	$SIL4_5$
	RA	RA_{Min}	RA_{Min}	RA_{Min}	$RA_{Min,5}$
Visual	Vegetation		√		
	Terrain				√
	Sky		√		
	Traffic Vehicles				
	Traffic Infrastructure				
	Human Activity				
	Buildings				
Geographic	Traffic				
	Business	√	√		√
	Cultural/Educational				
	Public Services				
	Medical Health	√			√
	Religious Special				
	Other				

Note: X_n represents X factors with n percentiles, which can be Min, Mean, Max,5, 10, 50, 90, or 95.

3.2 Soundscape prediction model comparison

In this study, the LR, RF, XGB, and GPR models were used to predict soundscape perceptions based on the ISD dataset. To ensure the stability of the performance and prediction results of the models, K-fold cross-validation was used to evaluate the models on the basis of the results produced by all the models. Furthermore, to provide practical implications for soundscape design work, this study also attempted to evaluate the model results at the group level.

On the test set, the performance of each model was different from that attained of the training set. When the prediction target was ISOPleasant, the GPR model outperformed the other three models at the individual level, achieving an R^2 of 0.3649, an MAE of 0.2604, and an $RMSE$ of 0.3296. At the group level, GPR had the best results in terms of two metrics, with a KL divergence of 0.8096 and a DME of 0.2587, whereas the RF model achieved the best performance in terms of the JS divergence ($JS = 0.1095$). Similarly, when predicting ISOEventful, GPR again exhibited better performance at the individual level, with an R^2 of 0.1759, an MAE of 0.2303, and an $RMSE$ of 0.2872. At the group level, the performances of the models differed across various evaluation metrics: XGB achieved the lowest KL divergence ($KL = 1.7055$), the RF performed best in terms of the JS divergence ($JS = 0.1513$), and GPR resulted in the smallest DME ($DME = 0.3812$). Overall, on the test set, the performance of the GPR model was superior to that of the other three models.

Table 4 Tabular performance of the four models (LR, RF, XGB and GPR) at the individual and group levels when predicting ISOPleasant and ISOEventful.

Targets	Models	Model performance based on different evaluation indicators					
		R2	**MAE**	**RMSE**	**KL**	**JS**	**DME**
ISOPleasant	LR	0.3525	0.2638	0.3329	0.8567	0.1118	0.2821
	RF	0.3566	0.2644	0.3318	0.8648	**0.1095**	0.3107
	XGB	0.3281	0.2718	0.3391	0.8526	0.1383	0.3652
	GPR	**0.3649**	**0.2604**	**0.3296**	**0.8096**	0.1137	**0.2587**
ISOEventful	LR	0.1684	0.2308	0.2885	2.2243	0.1590	0.3985
	RF	0.1547	0.2330	0.2909	2.0849	**0.1513**	0.4017
	XGB	0.1216	0.2371	0.2965	**1.7055**	0.1656	0.4926
	GPR	**0.1759**	**0.2303**	**0.2872**	2.0144	0.1612	**0.3812**

Note: best results are highlighted.

In addition, in this study, the kernel density estimation method was used to visualize the soundscape perception distribution of the test set and the prediction results. For the single-dimensional ISOPleasant and ISOEventful indicators, their corresponding probability density functions were visualized as soundscape perception distributions at the corresponding scale to reflect the soundscape perceptions at the group level (Figure 4). The red curve in the figure represents the probability distribution of the real data, and the blue curve represents the probability distribution of the predicted data. To better describe the soundscape prediction results, this study aimed to evaluate the results from a group perspective. The closer the probability distribution of the predicted values was to the probability distribution of the real data, the closer the soundscape perceptions predicted by the model were to the actual group-level soundscape perceptions, indicating that the prediction performance of this model was better. When the probability density functions derived from the kernel density estimation results were analysed, two key characteristics were of primary interest: the peak value on the Y-axis and the spread along the X-axis. Since the total area under a probability density function is constrained to one, an inverse relationship exists between these two properties. A higher peak value corresponds to a narrower spread, indicating that a majority of a group of people have the same perception of the examined soundscape. Conversely, a lower peak with a wider spread means people have more dispersed or varied perceptual responses. In Figure 4, regarding the prediction of ISOPleasant, the peaks of the distributions predicted by the RF and XGBoost models deviated more from the true values than those of the LR and GPR models did. Furthermore, compared with the GPR model, the LR model had a narrower spread. For the prediction of ISOEventful, the peak and spread of the GPR-predicted distribution aligned much more closely with the true values than those of the other three models did. These results imply that the GPR model possesses a better ability to approximate the true distributions of soundscape perceptions for both the ISOPleasant and ISOEventful attributes.

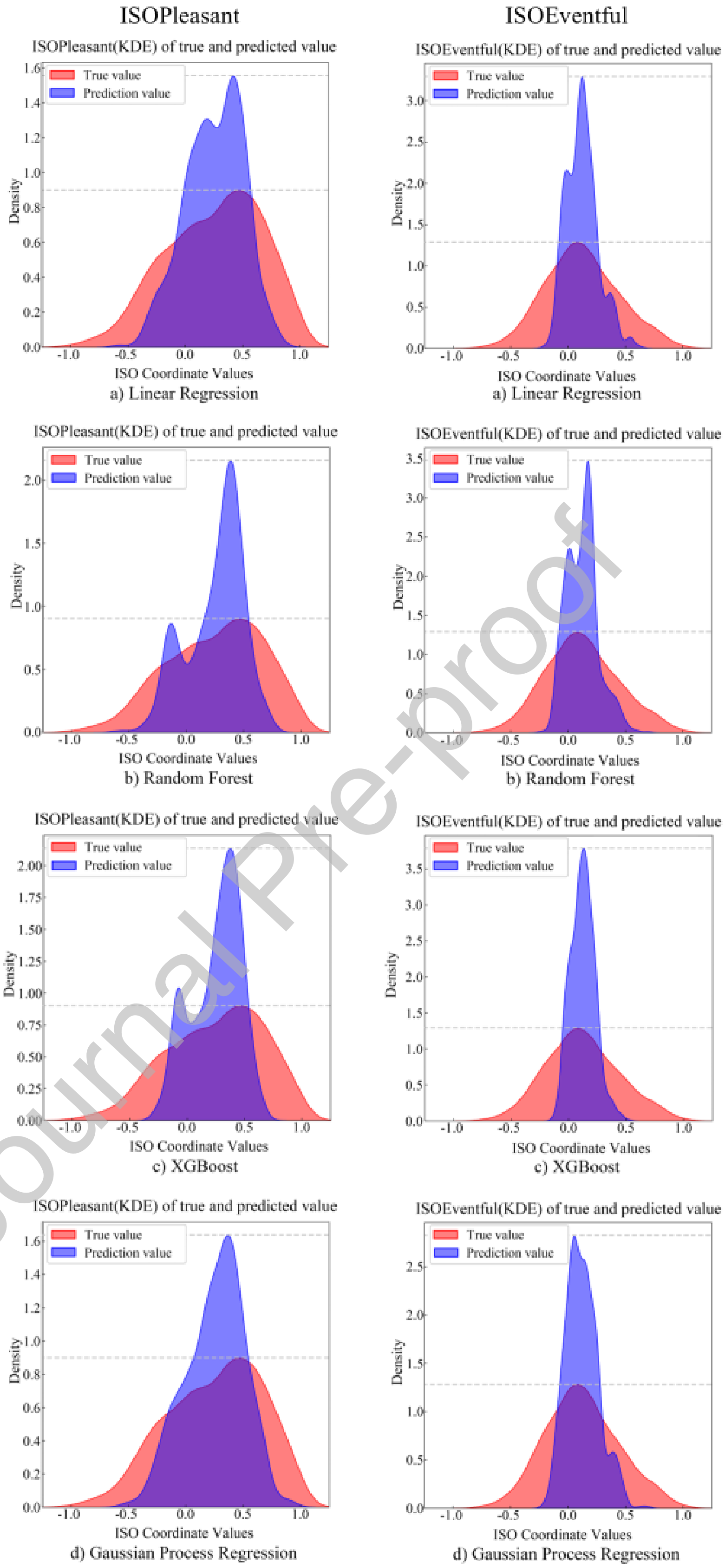


Figure 4 Group-level visualization of the prediction performance achieved by the four models regarding the distributions of ISOPleasant and ISOEventful.

Furthermore, this study also visualized the joint probability distributions of ISOPleasant and ISOEventful (Figure 5). Orange points represent the predicted values, blue points represent the true values, and the enclosed orange area represents the simplified kernel density estimate produced for the predicted points, whereas the enclosed blue area represents the simplified kernel density estimate obtained for the true points. The closer the shapes of the two areas are and the greater their degree of overlap, the closer the predictions of the examined model are to the actual perception of the soundscape by the group, indicating better prediction performance.

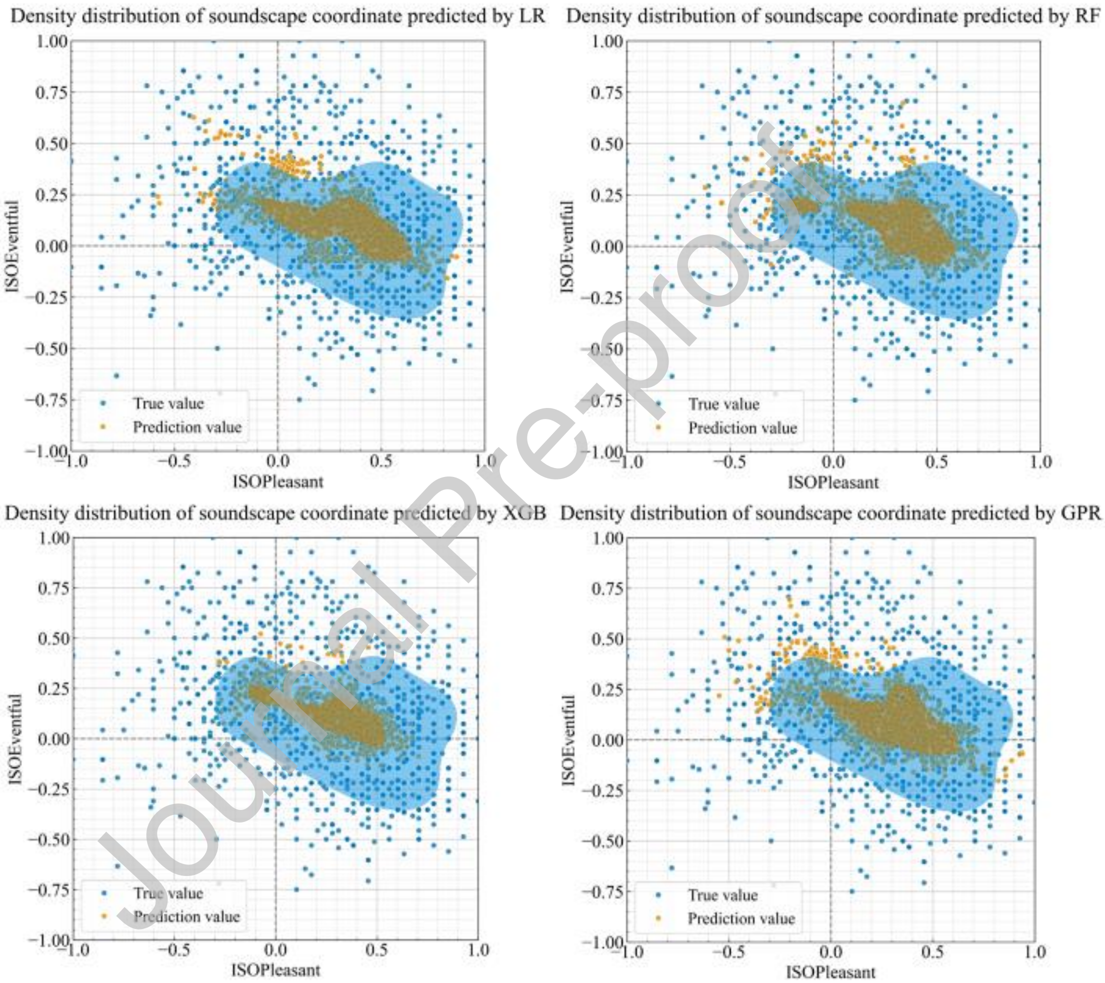


Figure 5 Comparison of true and prediction value based on four models at group-level visualization task.

4. Discussion

4.1 Factors related to soundscape prediction

This study identified the factors which can be used for soundscape prediction (ISOPleasant and ISOEventful). The results indicate that the acoustic factors of the sound pressure level, roughness, total harmonic distortion, and relative approach, as well as the visual factor of vegetation, play important roles in soundscape prediction scenarios.

Specifically, L_{Ceq} was negatively correlated with ISOPleasant ($r \approx -0.36$) and positively correlated with ISOEventful ($r \approx 0.31$). L_{Ceq} is the C-weighted equivalent continuous sound pressure level, which reflects the overall sound energy level of the scene (the intensity of the ambient noise) filtered with the "C-type filter" which preserves the low frequency information but rolls off the high frequencies (unlike the A-type filter which rolls of low both low and high frequencies, or the Z-type filter which preserves the full frequency spectrum) and is highly correlated with people's intuitive perceptions of "quiet/noisy". In 2009, Yu and Kang reported the important influence of the equivalent sound pressure level on soundscape perceptions when an ANN was used to model soundscapes[7]. In addition, Ricciardi and Pheasant et al., used A-weighted equivalent sound pressure levels to predict soundscapes [13,14]. This means that higher sound pressure levels make the sound environment noisier and reduce people's pleasure, whereas high sound pressure levels often mean an increase in the number of sound sources, making a setting feel more eventful.

Roughness was negatively correlated with ISOPleasant ($r \approx -0.44$) and positively correlated with ISOEventful ($r \approx 0.34$). Roughness reflects the harshness or instability caused by rapid amplitude/frequency fluctuations in sound. Sounds with high roughness, such as brake sounds, chain saw sounds, and harsh metal friction sounds, appear harsh, tense, mechanical, or discordant. Sounds with low roughness, such as wind sounds and light music, are softer and smoother. Roughness reflects the instability of a sound texture and is closely related to the intensity of "irritation/harshness" in a scene. When exploring changes in soundscape perceptions during an epidemic, Mitchell et al., reported that roughness was strongly correlated with soundscape perceptions ($|r| > 0.28$) [24,54]. In his study, Zhang also reported that roughness affects sound perceptions[55]. In natural protected areas, Oberman found roughness to be positively correlated with ISOEventful but found no significant associations with ISOPleasant[56].

Total harmonious distortion (THD) was negatively correlated with ISOPleasant ($r \approx -0.32$) and positively correlated with ISOEventful ($r \approx 0.31$). THD measures the loss of sound waveforms. A high THD level results in severe sound loss, poor fidelity, and blurry sounds, such as distortion caused by low-quality speakers or mechanical vibrations. A low THD level results in high-fidelity sound reproduction, closer to the natural/original sound, such as in high-fidelity audio or quiet scenes. THD reflects the fidelity and distortion of the sound in a scene and is closely related to the "clarity or distortion" characteristics of the scene, which in turn influence people's perceptions of pleasure and eventfulness.

Relative approach (RA) was negatively correlated with ISOPleasant ($r \approx -0.30$) and positively correlated with ISOEventful ($r \approx 0.32$). RA often refers to the perceived proximity of a moving sound source (such as a vehicle, an aircraft, or a drone) relative to the listener and is typically characterized by the Doppler effect, intensity change rate, or directionality of the sound source. It reflects the motion and proximity of the sound source within the scene, embodying the dynamic and eventful nature of the sound [57–59]. A high RA value indicates that dynamic elements are more likely to be present in the scene, making the scene appear noisy and distracting, leading to greater perceptions of eventfulness and less of reservations [60–63].

Vegetation, as a key visual factor, was significantly positively correlated with ISOPleasant ($r \approx 0.48$). This relationship may be attributed to two main mechanisms. First, vertical vegetation coverage enhances sound absorption through the repeated reflection of sound waves within dense plant structures, effectively reducing noise and contributing to a more tranquil environment. Second, the presence of greenery promotes a positive audiovisual integration experience, likely evoking a greater sense of pleasure and strengthening the perceived pleasantness of the soundscape[8,15,17,64,65]. These combined effects suggest that vegetation not only improves acoustic environments but also enhances subjective pleasantness perceptions, thereby elevating the ISO pleasantness rating.

In addition to the above factors, acoustic factors including the speech interference level (SIL) and impulsiveness and the geographic factor business were selected as predictive factors by the three models for their respective prediction tasks. The SIL is a measure of the degree to which noise interferes with the intelligibility of speech. It was negatively correlated with ISOPleasant ($r \approx -0.33$) and positively correlated with ISOEventful ($r \approx 0.31$). A higher SIL often indicates a greater number of other sound sources in the target scene, increasing the eventfulness of the scene at the expense of pleasantness. Impulsiveness describes whether a sound contains short, intense transient energy, such as explosions, knocks, and sudden horns. It is also correlated with the ISOPleasant dimension of soundscape perception. Business, as a geographic factor, indicates the commercial environment of the surrounding environment. If the commercial environment factor is high, the number of commercial POIs increases, the flow of people in the scene increases, and the number of sound sources increases, which in turn affects people's perceptions of the sound environment. Although some studies have reported that demographic factors and urban form factors affect soundscape perceptions[34,58,66], they were not sufficiently correlated in this study except for age, which may have been because when the four factors were considered, the importance levels of demographic factors and urban form factors were lower than those of the acoustic factors (such as the sound pressure level, roughness, and loudness) and visual factors (such as the green view rate). Therefore, the demographic and urban form factors were replaced by other more important factors when a certain number of modelling factors were selected.

4.2 Soundscape prediction performance at the individual and group level

To explore the potential for predicting soundscape perception at the group levels, this study compared the performance of deterministic models (at the individual level) using LR, an RF, and XGB as examples and probabilistic models (at the group level) using GPR as an example for the soundscape prediction task. The results show that the GPR model outperformed the other models at the individual scale, whether it was used to predict ISOPleasant or ISOEventful. Moreover, the GPR model outperformed the RF ($R_{pl}^2 = 0.2500, R_{ev}^2 = 0.0928$) and XGB ($R_{pl}^2 = 0.2561, R_{ev}^2 = 0.1016$) models used in Versumer's 2025 experiments on the ISD dataset[23], with $R_{pl}^2 = 0.3649$ and $R_{ev}^2 = 0.1759$ in this study. At the group level, the GPR model also performed relatively well in terms of predicting pleasantness (its KL and DME values were better than those of the other methods). When predicting eventfulness, although the performance achieved by GPR in terms of KL and JS was limited, its DME performance was still better than that of the other approaches. In addition, the probability distribution results obtained for the predictions of each model revealed that the distribution of the prediction results yielded by the GPR model was closer to the shape of the true distribution than those of the other models were. When predicting pleasantness and eventfulness, the peak value and spread of the probability distribution of the prediction results produced by the GPR model were closer to the probability distribution of the true values (Figure 4). When pleasantness and eventfulness were considered together, the distribution shape of the prediction results produced by the GPR model was closer to the true distribution than those of the other deterministic models were (Figure 5). These results indicate that predicting soundscape perceptions at the group level has certain advantages over predicting soundscape perceptions at the individual level. The advantages of modelling at the group level over modelling at the individual level have also been reflected in other fields[67], such as solar output forecasting[68] and runoff prediction in hydrological research[69].

There may be three reasons why using GPR (a probabilistic model) to predict soundscape perceptions at the group level presented certain advantages. First, GPR is a nonparametric

Bayesian method that can flexibly fit complex nonlinear relationships. The prediction uncertainty of GPR can reflect the inherent volatility in these data, thereby improving the reliability of the model [70,71]. Moreover, GPR can use kernel functions (e.g., RBF kernels and dot product kernels) to model the correlations among different scales and features. This makes GPR suitable for tasks such as soundscape perception, which may exhibit periodicity (e.g., day-to-night variations) and spatial autocorrelations. Second, soundscape data are often highly subjective and noisy (perceptions vary from person to person). In real-world contexts, soundscape perceptions are influenced by environmental variables (e.g., temperature and wind) and personal factors (e.g., mood and experience), which are difficult to replicate experimentally. These factors integrate noise and bias into real-world soundscape data, complicating prediction efforts [23]. The probabilistic modelling method has a higher tolerance for such noise and deviations, which gives this method a certain advantage in terms of predicting the distributions of real soundscapes. Third, soundscape perception data often have limited sample sizes but a wide range of feature dimensions (e.g., acoustic factors, environmental factors, and demographic data). Even with a small sample size, GPR can still reliably fit nonlinear relationships, whereas random forests and XGB may overfit the data or produce unstable predictions when the sample size is insufficient [45,72]. These advantages may explain why probabilistic models are better at predicting soundscape perceptions at the group level, which also means that the idea of modelling and predicting soundscape perceptions at the group level has certain advantages.

From the perspective of urban planning, when addressing the question of how to design environments that optimize soundscape perceptions, the focus should be placed on the perceptions of groups rather than individuals. Cities are built to serve communities, not averaged ideal individuals. Therefore, in soundscape prediction tasks, researchers should shift the modelling paradigm from the traditional individual level to the group level. This approach would not only enable soundscape prediction to be performed with a smaller amount of data but also mitigate the influence of individual variability, thereby yielding more robust and reliable predictive results. Ultimately, soundscape prediction models can be further applied to infer which factors most strongly affect soundscape perceptions and to analyse how these factors interact with urban environments. Such insights can guide urban designers and planners to create more acoustically pleasant and perceptually balanced public spaces.

4.3 Applications

In the field of soundscape mapping, GPR can construct a dynamically updated probabilistic soundscape map that reflects the spatial distribution characteristics of sound comfort levels in different areas, and this feature makes it particularly suitable for the preassessment of soundscapes in urban renewal projects.

In addition, the results of this study demonstrate that the GPR model can effectively predict group-level soundscape perceptions. This capability offers practical value for guiding sound-oriented environmental planning and design tasks. When a group of people enters a specific environment, their soundscape perception can be anticipated in advance. By doing so, the model helps reduce the reliance on large-scale data collection processes, shortens the required training time, decreases the incurred computational costs, and supports efficient soundscape design decisions with minimal training samples.

4.4 Limitations and future work

Future work could focus on extending the applicability and robustness of the modelling framework. On the one hand, rather than relying solely on group-level performance, efforts should be directed towards scene-specific optimization strategies. By adapting the model to a specific environment, such as an urban park, localized acoustic characteristics and contextual factors can be better captured, allowing the model to provide more accurate and actionable predictions in diverse urban environments. This approach would enhance the capacity of the model to support site-specific soundscape planning and design activities. On the other hand, future research could attempt to directly use the distribution of soundscape perceptions at the group level as the prediction target of the constructed model. That is, more advanced probabilistic models, such as Bayesian models, latent variable analysis or deep neural networks, can be used to directly predict the probability distributions of soundscape perceptions. By treating perceptual outcomes as probability distributions, such models can better account for the individual variability and uncertainty that are inherent in soundscape perceptions. This probabilistic framework also facilitates more targeted model selection and parameter tuning procedures. Together, these potential future research directions could strengthen the ability of soundscape prediction models to provide reliable, context-aware, and uncertainty-informed insights for both research and urban planning practice.

5. Conclusions

This study investigates which factors can be used for soundscape prediction and how the machine learning models performs at the group level, by using LR, RF, XGB, and GPR models constructed based on the ISD.

In this study, key factors of soundscape perceptions were identified by applying four distinct models to screen factors on the basis of correlation, multicollinearity, and importance metrics. Acoustic factors, including the equivalent continuous sound pressure level ($|r| > 0.31$), roughness ($|r| > 0.34$), total harmonic distortion ($|r| > 0.31$), and relative approach ($|r| > 0.30$), alongside the visual factor vegetation ($|r| > 0.48$), were consistently selected across all the models, underscoring their critical role in predicting both ISOPleasant and ISOEventful. Vegetation was positively correlated with ISOPleasant, highlighting the calming effect of natural visual elements. Additional factors such as the speech interference level, impulsiveness, and business-related geographic attributes also contributed to the model, although demographic and urban form factors were largely overshadowed by the acoustic and visual factors. GPR model outperformed the other models at both the individual and group levels. GPR demonstrated a unique advantage in terms of matching the predicted distributions of ISOPleasant and ISOEventful with the actual perceived distributions. By modelling soundscape perceptions in a function space and accounting for uncertainty, the GPR model provided predictions that more closely approximated the "true collective perceptions", i.e. predictions at the group-level.

This distribution-based evaluation approach thus provides a more intuitive and robust strategy for understanding and predicting soundscape perceptions, offering insights for sound environment design and management tasks. This study adopted a probability distribution-based approach to evaluate the performance of the tested models at the group level. This method allowed the predictions to better capture the perceptual diversity among different populations, thereby improving the applicability of the models in real-world acoustic environments. This contribution could help to improve soundscape research and guiding the design and planning processes concerning urban sound environment.

Funding

The authors would like to acknowledge the supports from "National Natural Science Foundation of China (No. 52308060)", "Guangdong Philosophy and Social Sciences Project (No.GD23XSH20)", "Shenzhen Science and Technology Program (No. RCBS20231211090747080)", "European Research Council (Grant No. 740696, project title Soundscape Indices—SSID)".

References:

[1] J. Kang, F. Aletta, T.T. Gjestland, L.A. Brown, D. Botteldooren, B. Schulte-Fortkamp, P. Lercher, I. Van Kamp, K. Genuit, A. Fiebig, J.L. Bento Coelho, L. Maffei, L. Lavia, Ten questions on the soundscapes of the built environment, Building and Environment 108 (2016) 284–294. https://doi.org/10.1016/j.buildenv.2016.08.011.

[2] S.R. Payne, The production of a Perceived Restorativeness Soundscape Scale, Applied Acoustics 74 (2013) 255–263. https://doi.org/10.1016/j.apacoust.2011.11.005.

[3] ISO 12913-1:2014 Acoustics-soundscape-part 1: definition and conceptual framework. Standard, International Organization for Standardization, Geneva, CH; 2014., n.d.

[4] ISO/TS 12913-3:2025 Acoustics — Soundscape Part 3: Data analysis. Standard, International Organization for Standardization., n.d.

[5] A. Mitchell, Predictive Modelling of Complex Urban Soundscapes: Enabling an engineering approach to soundscape design, (2022). https://doi.org/10.13140/RG.2.2.15590.50245.

[6] L. Jiang, A. Bristow, J. Kang, F. Aletta, R. Thomas, H. Notley, A. Thomas, J. Nellthorp, Ten questions concerning soundscape valuation, Building and Environment 219 (2022) 109231. https://doi.org/10.1016/j.buildenv.2022.109231.

[7] L. Yu, J. Kang, Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach, The Journal of the Acoustical Society of America 126 (2009) 1163–1174. https://doi.org/10.1121/1.3183377.

[8] M.S. Tse, C.K. Chau, Y.S. Choy, W.K. Tsui, C.N. Chan, S.K. Tang, Perception of urban park soundscape, The Journal of the Acoustical Society of America 131 (2012) 2762–2771. https://doi.org/10.1121/1.3693644.

[9] X. Fang, T. Gao, M. Hedblom, N. Xu, Y. Xiang, M. Hu, Y. Chen, L. Qiu, Soundscape Perceptions and Preferences for Different Groups of Users in Urban Recreational Forest Parks, Forests 12 (2021) 468. https://doi.org/10.3390/f12040468.

[10] X. Guo, S.-Y. Jiang, J. Liu, Z. Chen, X.-C. Hong, Understanding the Role of Visitor Behavior in Soundscape Restorative Experiences in Urban Parks, Forests 15 (2024) 1751. https://doi.org/10.3390/f15101751.

[11] J.Y. Hong, J.Y. Jeon, Influence of urban contexts on soundscape perceptions: A structural equation modeling approach, Landscape and Urban Planning 141 (2015) 78–87. https://doi.org/10.1016/j.landurbplan.2015.05.004.

[12] G. Brambilla, L. Maffei, M. Di Gabriele, V. Gallo, Merging physical parameters and laboratory subjective ratings for the soundscape assessment of urban squares, The Journal

of the Acoustical Society of America 134 (2013) 782–790. https://doi.org/10.1121/1.4768792.

[13] P. Ricciardi, P. Delaitre, C. Lavandier, F. Torchia, P. Aumond, Sound quality indicators for urban places in Paris cross-validated by Milan data, The Journal of the Acoustical Society of America 138 (2015) 2337–2348. https://doi.org/10.1121/1.4929747.

[14] R. Pheasant, K. Horoshenkov, G. Watts, B. Barrett, The acoustic and visual factors influencing the construction of tranquil space in urban and rural environments tranquil spaces-quiet places?, The Journal of the Acoustical Society of America 123 (2008) 1446–1457. https://doi.org/10.1121/1.2831735.

[15] G. Watts, A. Miah, R. Pheasant, Tranquillity and Soundscapes in Urban Green Spaces—Predicted and Actual Assessments from a Questionnaire Survey, Environ Plann B Plann Des 40 (2013) 170–181. https://doi.org/10.1068/b38061.

[16] C. Lavandier, P. Aumond, S. Gomez, C. Dominguès, Urban soundscape maps modelled withgeo-referenced data, Noise Mapping 3 (2016). https://doi.org/10.1515/noise-2016-0020.

[17] V. Puyana Romero, L. Maffei, G. Brambilla, G. Ciaburro, Modelling the soundscape quality of urban waterfronts by artificial neural networks, Applied Acoustics 111 (2016) 121–128. https://doi.org/10.1016/j.apacoust.2016.04.019.

[18] D. Çakır Aydın, S. Yılmaz, Assessment of sound environment pleasantness by sound quality metrics in urban spaces, AZ 13 (2016) 87–99. https://doi.org/10.5505/itujfa.2016.75547.

[19] J. Fan, M. Thorogood, P. Pasquier, Automatic Soundscape Affect Recognition Using A Dimensional Approach, J. Audio Eng. Soc. 64 (2016) 646–653. https://doi.org/10.17743/jaes.2016.0044.

[20] P. Lindborg, A. Friberg, Personality Traits Bias the Perceived Quality of Sonic Environments, Applied Sciences 6 (2016) 405. https://doi.org/10.3390/app6120405.

[21] A. Maristany, M.R. López, C.A. Rivera, Soundscape quality analysis by fuzzy logic: A field study in Cordoba, Argentina, Applied Acoustics 111 (2016) 106–115. https://doi.org/10.1016/j.apacoust.2016.04.013.

[22] F. Aletta, J. Kang, Towards an Urban Vibrancy Model: A Soundscape Approach, IJERPH 15 (2018) 1712. https://doi.org/10.3390/ijerph15081712.

[23] S. Versümer, P. Blättermann, F. Rosenthal, S. Weinzierl, A comparison of methods for modeling soundscape dimensions based on different datasets, The Journal of the Acoustical Society of America 157 (2025) 234–255. https://doi.org/10.1121/10.0034849.

[24] A. Mitchell, T. Oberman, F. Aletta, M. Kachlicka, M. Lionello, M. Erfanian, J. Kang, Investigating urban soundscapes of the COVID-19 lockdown: A predictive soundscape modeling approach, The Journal of the Acoustical Society of America 150 (2021) 4474–4488. https://doi.org/10.1121/10.0008928.

[25] X.-C. Hong, S. Cheng, J. Liu, E. Dang, J.-B. Wang, Y. Cheng, The Physiological Restorative Role of Soundscape in Different Forest Structures, Forests 13 (2022) 1920. https://doi.org/10.3390/f13111920.

[26] M. Erfanian, A. Mitchell, F. Aletta, J. Kang, Psychological well-being and demographic factors can mediate soundscape pleasantness and eventfulness: A large sample study,

Journal of Environmental Psychology 77 (2021) 101660. https://doi.org/10.1016/j.jenvp.2021.101660.
[27] X. Guo, J. Liu, C. Albert, X.-C. Hong, Audio-visual interaction and visitor characteristics affect perceived soundscape restorativeness: Case study in five parks in China, Urban Forestry & Urban Greening 77 (2022) 127738. https://doi.org/10.1016/j.ufug.2022.127738.
[28] G. Watts, A. Khan, R. Pheasant, Influence of soundscape and interior design on anxiety and perceived tranquillity of patients in a healthcare setting, Applied Acoustics 104 (2016) 135–141. https://doi.org/10.1016/j.apacoust.2015.11.007.
[29] T. Zhao, X. Liang, W. Tu, Z. Huang, F. Biljecki, Sensing urban soundscapes from street view imagery, Computers, Environment and Urban Systems 99 (2023) 101915. https://doi.org/10.1016/j.compenvurbsys.2022.101915.
[30] J. Kang, F. Aletta, E. Margaritis, M. Yang, A model for implementing soundscape maps in smart cities, Noise Mapping 5 (2018) 46–59. https://doi.org/10.1515/noise-2018-0004.
[31] J. Fan, M. Thorogood, P. Pasquier, Automatic Soundscape Affect Recognition Using A Dimensional Approach, J. Audio Eng. Soc. 64 (2016) 646–653. https://doi.org/10.17743/jaes.2016.0044.
[32] T. Giannakopoulos, M. Orfanidi, S. Perantonis, Athens Urban Soundscape (ATHUS): A Dataset for Urban Soundscape Quality Recognition, in: I. Kompatsiaris, B. Huet, V. Mezaris, C. Gurrin, W.-H. Cheng, S. Vrochidis (Eds.), MultiMedia Modeling, Springer International Publishing, Cham, 2019: pp. 338–348. https://doi.org/10.1007/978-3-030-05710-7_28.
[33] P. Lundén, Ö. Axelsson, M. Hurtig, On Urban Soundscape Mapping: A Computer Can Predict the Outcome of Soundscape Assessments, (2016).
[34] L. Yu, J. Kang, Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach, The Journal of the Acoustical Society of America 126 (2009) 1163–1174. https://doi.org/10.1121/1.3183377.
[35] M. Lionello, F. Aletta, J. Kang, A systematic review of prediction models for the experience of urban soundscapes, Applied Acoustics 170 (2020) 107479. https://doi.org/10.1016/j.apacoust.2020.107479.
[36] Ö. Axelsson, M.E. Nilsson, B. Berglund, A principal components model of soundscape perception, The Journal of the Acoustical Society of America 128 (2010) 2836–2846. https://doi.org/10.1121/1.3493436.
[37] F. Aletta, J. Kang, Ö. Axelsson, Soundscape descriptors and a conceptual framework for developing predictive soundscape models, Landscape and Urban Planning 149 (2016) 65–74. https://doi.org/10.1016/j.landurbplan.2016.02.001.
[38] A. Mitchell, F. Aletta, J. Kang, How to analyse and represent quantitative soundscape data, JASA Express Letters 2 (2022) 037201. https://doi.org/10.1121/10.0009794.
[39] A. Mitchell, 《The International Soundscape Database: An integrated multimedia database of urban soundscape surveys -- questionnaires with acoustical and contextual information》. Zenodo, Nov 09, 2021. doi: 10.5281/zenodo.5705908., (n.d.).
[40] A. Mitchell, T. Oberman, F. Aletta, M. Erfanian, M. Kachlicka, M. Lionello, J. Kang, The Soundscape Indices (SSID) Protocol: A Method for Urban Soundscape

Surveys—Questionnaires with Acoustical and Contextual Information, Applied Sciences 10 (2020) 2397. https://doi.org/10.3390/app10072397.
[41] ISO/TS 12913-2:2018 Acoustics — Soundscape Part 2: Data collection and reporting requirements, (n.d.).
[42] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J.M. Alvarez, P. Luo, SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers, (2021). https://doi.org/10.48550/arXiv.2105.15203.
[43] P. Aumond, A. Can, B. De Coensel, D. Botteldooren, C. Ribeiro, C. Lavandier, Modeling Soundscape Pleasantness Using perceptual Assessments and Acoustic Measurements Along Paths in Urban Context, Acta Acustica United with Acustica 103 (2017) 430–443. https://doi.org/10.3813/AAA.919073.
[44] R. Benocci, A. Afify, A. Potenza, H.E. Roman, G. Zambon, Toward the Definition of a Soundscape Ranking Index (SRI) in an Urban Park Using Machine Learning Techniques, Sensors 23 (2023) 4797. https://doi.org/10.3390/s23104797.
[45] C.E. Rasmussen, C.K.I. Williams, Gaussian processes for machine learning, 3. print, MIT Press, Cambridge, Mass., 2008.
[46] Z. Jiao, R. Tao, Geographical Gaussian Process Regression: A Spatial Machine-Learning Model Based on Spatial Similarity, Geographical Analysis (2025) gean.12423. https://doi.org/10.1111/gean.12423.
[47] K. Pedersen, M.K. Transtrum, K.L. Gee, B.A. Butler, M.M. James, A.R. Salton, Machine learning-based ensemble model predictions of outdoor ambient sound levels, in: Bruges, Belgium, 2018: p. 022002. https://doi.org/10.1121/2.0001056.
[48] W. Zhou, Z. Yan, L. Zhang, A comparative study of 11 non-linear regression models highlighting autoencoder, DBN, and SVR, enhanced by SHAP importance analysis in soybean branching prediction, Sci Rep 14 (2024) 5905. https://doi.org/10.1038/s41598-024-55243-x.
[49] R. Tong, Z. Zhu, J. Ling, Comparison of linear and non-linear machine learning models for time-dependent readmission or mortality prediction among hospitalized heart failure patients, Heliyon 9 (2023) e16068. https://doi.org/10.1016/j.heliyon.2023.e16068.
[50] O. Krakovska, G. Christie, A. Sixsmith, M. Ester, S. Moreno, Performance comparison of linear and non-linear feature selection methods for the analysis of large survey datasets, PLoS ONE 14 (2019) e0213584. https://doi.org/10.1371/journal.pone.0213584.
[51] H. Wang, Q. Liang, J.T. Hancock, T.M. Khoshgoftaar, Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods, J Big Data 11 (2024) 44. https://doi.org/10.1186/s40537-024-00905-w.
[52] A.J. Mitchell, Predictive Modelling of Complex Urban Soundscapes, (n.d.).
[53] D.I. Belov, R.D. Armstrong, Distributions of the Kullback–Leibler divergence with applications, Brit J Math & Statis 64 (2011) 291–309. https://doi.org/10.1348/000711010X522227.
[54] M. Lionello, F. Aletta, A. Mitchell, J. Kang, Introducing a Method for Intervals Correction on Multiple Likert Scales: A Case Study on an Urban Soundscape Data Collection Instrument, Front. Psychol. 11 (2021) 602831. https://doi.org/10.3389/fpsyg.2020.602831.

[55] Y. Zhang, J. Kang, B. Yang, Comparison between perceived and felt emotions in the soundscape evaluation of urban open spaces, The Journal of the Acoustical Society of America 155 (2024) 2756–2768. https://doi.org/10.1121/10.0025761.

[56] T. Oberman, A. Latini, F. Aletta, G. Gozzi, J. Kang, S. Torresin, Human sounds and associated tonality disrupting perceived soundscapes in protected natural areas, Sci Rep 15 (2025) 28759. https://doi.org/10.1038/s41598-025-08524-y.

[57] H. Tong, J.L. Warren, J. Kang, M. Li, Using multi-sourced big data to correlate sleep deprivation and road traffic noise: A US county-level ecological study, Environmental Research 220 (2023) 115029. https://doi.org/10.1016/j.envres.2022.115029.

[58] H. Tong, J. Kang, Characteristics of noise complaints and the associations with urban morphology: A comparison across densities, Environmental Research 197 (2021) 111045. https://doi.org/10.1016/j.envres.2021.111045.

[59] H. Tong, J. Kang, Relationships between noise complaints and socio-economic factors in England, Sustainable Cities and Society 65 (2021) 102573. https://doi.org/10.1016/j.scs.2020.102573.

[60] K. De Paiva Vianna, R.M. Rodrigues, M. Alves Cardoso, Noise pollution and annoyance: An urban soundscapes study, Noise Health 17 (2015) 125. https://doi.org/10.4103/1463-1741.155833.

[61] W. Gao, J. Kang, H. Ma, C. Wang, The effects of environmental sensitivity and noise sensitivity on soundscape evaluation, Building and Environment 245 (2023) 110945. https://doi.org/10.1016/j.buildenv.2023.110945.

[62] X. Fu, L. Wang, L. Yuan, H. Hu, T. Li, J. Zhang, Y. Ke, M. Wang, Y. Gao, W. Huo, Y. Chen, W. Zhang, J. Liu, Z. Huang, Y. Zhao, F. Hu, M. Zhang, Y. Liu, X. Sun, D. Hu, Long-Term Exposure to Traffic Noise and Risk of Incident Cardiovascular Diseases: a Systematic Review and Dose-Response Meta-Analysis, J Urban Health 100 (2023) 788–801. https://doi.org/10.1007/s11524-023-00769-0.

[63] F.M.A. Calarco, L. Galbrun, Sound mapping design of water features used over road traffic noise for improving the soundscape, Applied Acoustics 219 (2024) 109947. https://doi.org/10.1016/j.apacoust.2024.109947.

[64] F. Biljecki, T. Zhao, X. Liang, Y. Hou, Sensitivity of measuring the urban form and greenery using street-level imagery: A comparative study of approaches and visual perspectives, International Journal of Applied Earth Observation and Geoinformation 122 (2023) 103385. https://doi.org/10.1016/j.jag.2023.103385.

[65] B. Zhong, H. Xie, Z. Zhang, Y. Wen, Non-linear effects of ecoacoustic indices on urban soundscape assessments based on gradient boosting decision trees in summer Chongqing, China, Building and Environment 278 (2025) 112984. https://doi.org/10.1016/j.buildenv.2025.112984.

[66] C. Lavandier, P. Aumond, S. Gomez, C. Dominguès, Urban soundscape maps modelled withgeo-referenced data, Noise Mapping 3 (2016). https://doi.org/10.1515/noise-2016-0020.

[67] M.S. Caywood, D.M. Roberts, J.B. Colombe, H.S. Greenwald, M.Z. Weiland, Gaussian Process Regression for Predictive But Interpretable Machine Learning Models: An Example of Predicting Mental Workload across Tasks, Front. Hum. Neurosci. 10 (2017).

https://doi.org/10.3389/fnhum.2016.00647.

[68] F. Najibi, D. Apostolopoulou, E. Alonso, Enhanced performance Gaussian process regression for probabilistic short-term solar output forecast, International Journal of Electrical Power & Energy Systems 130 (2021) 106916. https://doi.org/10.1016/j.ijepes.2021.106916.

[69] Q. Shen, L. Mo, G. Liu, Y. Wang, Y. Zhang, Interpretable probabilistic modeling method for runoff prediction: A case study in Yangtze River basin, China, Journal of Hydrology: Regional Studies 52 (2024) 101684. https://doi.org/10.1016/j.ejrh.2024.101684.

[70] T. Yang, P. Shi, Z. Yu, Z. Li, X. Wang, X. Zhou, Probabilistic modeling and uncertainty estimation of urban water consumption under an incompletely informational circumstance, Stoch Environ Res Risk Assess 30 (2016) 725–736. https://doi.org/10.1007/s00477-015-1081-x.

[71] P. Pouya, A.M. Madni, Probabilistic System Modeling for Complex Systems Operating in Uncertain Environments, in: A.M. Madni, B. Boehm, D. Erwin, M. Moghaddam, M. Sievers, M. Wheaton (Eds.), Recent Trends and Advances in Model Based Systems Engineering, Springer International Publishing, Cham, 2022: pp. 113–127. https://doi.org/10.1007/978-3-030-82083-1_11.

[72] E. Camporeale, X. Chu, O.V. Agapitov, J. Bortnik, On the Generation of Probabilistic Forecasts From Deterministic Models, Space Weather 17 (2019) 455–475. https://doi.org/10.1029/2018SW002026.

Declaration of interests

☒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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: