# Supplementary Material: Statistical analysis of perceived restorativeness

This supplementary material provides a detailed description of the procedure and the results of the statistical modelling of perceived restorativeness as mentioned in the sections 2.5.2 and 3.5 of the manuscript.

## Variable description

**Geodata**

|  |  |
| --- | --- |
| **Acronym** | **Explanation** |
| LCARTIF | proportion of artificial surfaces within 250 m buffer |
| LCFOREST | proportion of forest within 250 m buffer |
| HETER | land cover heterogeneity in 250 m buffer |
| OVDIST | distance to the nearest public transport stop |
| VIS5K | percentage of visible area within a radius of 5 km |
| RL\_NDVI | mean NDVI in 250 m buffer around RL |
| RL\_NOISE | mean road traffic noise in 250 m buffer around RL |
| DISTKM | Euclidean distance between home and RL |
| JNYTIME | travel time from home to RL (as indicated in the survey) |
| STRIMP123 | length of roads with high traffic intensity |
| STRIMP999 | length of other roads (low traffic intensity) |

**Mediator variables (questionnaire responses)**

|  |  |
| --- | --- |
| **Acronym** | **Explanation** |
| FEELNAT | Feeling of being in nature (1=strongly disagree; 7 = strongly agree) |
| LNOISE | Overall soundscape quality (1 = very bad; 5 = very good) |
| LOC\_SENS | Sensations; e.g. wind in hair (1=strongly disagree; 5 = strongly agree) |
| LOC\_SOUN | Sounds; e.g. birds (1=strongly disagree; 5 = strongly agree) |
| LOC\_SCEN | Scents and odours (1=strongly disagree; 5 = strongly agree) |
| LOC\_VISE | Visual elements (1=strongly disagree; 5 = strongly agree) |
| LOC\_VEGE | Vegetation and its changes (1=strongly disagree; 5 = strongly agree) |
| LOC\_FAUN | Wild animals (1=strongly disagree; 5 = strongly agree) |

**Response variables: Perceived Restorativeness Scale PRS[[1]](#footnote-2)**

|  |  |
| --- | --- |
| **Acronym** | **Explanation** |
| MEAN | Aggregated mean of all PRS dimensions (1=strongly disagree; 7 = strongly agree) |
| FA | Fascination (1=strongly disagree; 7 = strongly agree) |
| BA | Being away (1=strongly disagree; 7 = strongly agree |
| EC | Extent and coherence (1=strongly disagree; 7 = strongly agree) |
| ES | Scope and compatibility (1=strongly disagree; 7 = strongly agree) |

## Data preparation

We first cleaned the database from records with missing information. We further eliminated records where respondents (1) reported wearing headphones during the outdoor activity, (2) reported activity duration of more than 2 hours, or (3) indicated a restorative place farther away than 33 km from home, as beyond everyday recreation. The thresholds for criteria (2) and (3) equalled to 90th percentile of the given values. Finally, we kept 1494 out of 2206 observations; the number of matches per filter criteria are given in Table S1, whereas one record might include several criteria matches (source("R/data\_prep.R")).

*Table S1 Number of matches per filter criteria*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Headphone | PRS\_all\_NA | Distance | Activity\_NA | Duration | HM\_Noise\_NA | JourneyTime |
| 303 | 226 | 221 | 102 | 96 | 96 | 20 |

Next we splitted the data into a training and a test set (50/50) before testing the hypothesis to ensure valid inference after feature selection. Missing values were imputed using MissForest (doi:10.1093/bioinformatics/btr597). This method leverages conditional dependencies between variables to predict missing values through an iterative random forest approach. To avoid introducing spurious correlations between different variable sets, we imputed the following data groups separately: PRS variables on the complete dataset; mediators on training data only; geodata on training data only; mediators for prediction analysis; geodata for prediction analysis; PRS variables for prediction analysis.

Mediators and geodata were intentionally not imputed on the test set to maintain valid inference, as MissForest does not provide a mechanism to propagate imputation uncertainty. Missing values in the test set predictors remained untreated, which is justified under the missing completely at random (MCAR) assumption, where missing values occur independently of all other variables.

For the prediction analysis, fewer statistical assumptions were required, so using the MissForest approach did not violate any assumptions. PRS variables could have been imputed separately for training/test sets and prediction analysis, but we prioritized simplicity as these variables serve only as response variables. Additionally, we compared MissForest with simpler imputation methods (variable-wise and observationwise mean imputation) for the PRS variables. Results confirmed that MissForest consistently outperformed these alternatives.

## Prediction analysis

We applied multiple machine learning models using the mlr3 framework (doi:10.21105/joss.01903)

* Linear models (Lm; baseline)
* XGBoost (gradinent boosting with tree-based models and hyperparameter tuning for learning rate and tree depth (arxiv: 1603.02754)
* Random Forest (RF; with default parameters) (doi: 10.1023/A:1010933404324)

**3.1 Benchmark helper function**

Ein Bild, das Text, Screenshot, Dokument, Schrift enthält.

KI-generierte Inhalte können fehlerhaft sein.

**3.2 Results of prediction analysis**

The following tables present the variance of the predicted parameters explained by the different model approaches.

*Table S2 Predicting PRS with geodata (explained variance)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lm | xgboost | RF |
| MEAN | -0.002 | 0.005 | -0.036 |
| FA | -0.002 | 0.013 | 0.002 |
| BA | -0.006 | -0.009 | -0.021 |
| EC | 0.001 | -0.027 | -0.042 |
| ES | 0.048 | 0.041 | 0.026 |

Code:

Ein Bild, das Text, Schrift, Screenshot, Reihe enthält.

KI-generierte Inhalte können fehlerhaft sein.

*Table S3 Predicting PRS with geodata and mediators (explained variance)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lm | xgboost | RF |
| MEAN | 0.225 | 0.234 | 0.223 |
| FA | 0.236 | 0.260 | 0.248 |
| BA | 0.120 | 0.130 | 0.131 |
| EC | 0.042 | 0.012 | 0.021 |
| ES | 0.150 | 0.168 | 0.158 |

Code:

Ein Bild, das Text, Schrift, Screenshot enthält.

KI-generierte Inhalte können fehlerhaft sein.

*Table S4 Predicting PRS with mediators (explained variance)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lm | xgboost | RF |
| MEAN | 0.224 | 0.235 | 0.191 |
| FA | 0.227 | 0.254 | 0.225 |
| BA | 0.132 | 0.145 | 0.112 |
| EC | 0.036 | 0.025 | -0.020 |
| ES | 0.130 | 0.137 | 0.085 |

Code:

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

KI-generierte Inhalte können fehlerhaft sein.

*Table S5 Predicting mediators with GIS variables (explained variance)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lm | xgboost | RF |
| FEELNAT | 0.126 | 0.136 | 0.111 |
| LNOISE | 0.096 | 0.070 | 0.080 |
| LOC\_SENS | 0.020 | -0.008 | -0.025 |
| LOC\_SOUN | 0.041 | 0.010 | -0.001 |
| LOC\_SCEN | 0.039 | 0.052 | 0.021 |
| LOC\_VISE | 0.008 | -0.029 | -0.047 |
| LOC\_VEGE | 0.052 | 0.025 | 0.032 |
| LOC\_FAUN | 0.057 | 0.062 | 0.047 |

Code:

Ein Bild, das Text, Schrift, Screenshot enthält.

KI-generierte Inhalte können fehlerhaft sein.

## Hypothesis testing with linear modelling

**4.1 Imputation with MissForest on training data**

Number of NA in mediators: sapply(D[Mediator\_vars], \(x) sum(is.na(x)))

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| FEELNAT | LNOISE | LOC\_SENS | LOC\_SOUN | LOC\_SCEN | LOC\_VISE | LOC\_VEGE | LOC\_FAUN |
| 16 | 291 | 28 | 30 | 36 | 62 | 69 | 88 |

Number of NA in GIS variables: sapply(D[GIS\_vars], \(x) sum(is.na(x)))

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| LCARTIF\_sqrt | LCFOREST\_sqrt | HETER | OVDIST\_sqrt | VIS5K\_sqrt | RL\_NDVI | RL\_NOISE | DISTKM\_sqrt | JNYTIME\_sqrt | STRIMP123\_sqrt | STRIMP999\_sqrt |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 86 | 0 | 0 |

Impute missing values using MissForest

Ein Bild, das Text, Screenshot, Schrift, Zahl enthält.

KI-generierte Inhalte können fehlerhaft sein.

**4.2 Scaling test data**

Scaling variables and show original scale:

Ein Bild, das Text, Schrift, Screenshot enthält.

KI-generierte Inhalte können fehlerhaft sein.

*Table S6 Original scale of variables*

|  |  |  |
| --- | --- | --- |
|  | **Mean** | **sd** |
| FEELNAT | 6.142 | 1.055 |
| LNOISE | 4.210 | 0.747 |
| LOC\_SENS | 4.098 | 1.016 |
| LOC\_SOUN | 4.296 | 0.947 |
| LOC\_SCEN | 3.967 | 1.056 |
| LOC\_VISE | 4.080 | 1.027 |
| LOC\_VEGE | 4.343 | 0.859 |
| LOC\_FAUN | 3.298 | 1.365 |
| LCARTIF\_sqrt | 0.271 | 0.269 |
| LCFOREST\_sqrt | 0.454 | 0.311 |
| HETER | 1.305 | 0.402 |
| OVDIST\_sqrt | 21.797 | 10.144 |
| VIS5K\_sqrt | 3.323 | 1.620 |
| RL\_NDVI | 0.635 | 0.202 |
| RL\_NOISE | 41.615 | 9.261 |
| DISTKM\_sqrt | 1.473 | 1.156 |
| JNYTIME\_sqrt | 3.830 | 2.247 |
| STRIMP123\_sqrt | 6.555 | 10.739 |
| STRIMP999\_sqrt | 47.557 | 13.162 |
| PRS | 4.987 | 0.879 |
| FA | 5.266 | 1.111 |
| BA | 5.140 | 1.156 |
| EC | 4.540 | 1.287 |
| ES | 5.006 | 1.426 |

**4.3 Testing variance inflation factor (VIF)**

VIF: PRS (Mediators + GIS vars)^2 (without interaction)

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

KI-generierte Inhalte können fehlerhaft sein.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 1.21 | 1.46 | 1.94 | 2.03 | 2.24 | 4.99 |

VIF: PRS (Mediators + GIS vars)^2 (with interaction)

Ein Bild, das Text, Screenshot, Schrift, Reihe enthält.

KI-generierte Inhalte können fehlerhaft sein.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| 12 | 81 | 147 | 205 | 269 | 1398 |







































































































1. Pasini, M., Berto, R., Brondino, M., Hall, R., Ortner, C. 2014. How to measure the restorative quality of environments: the PRS-11. *Procedia Social and Behavioral Sciences* 159, 293–297. DOI: 10.1016/j.sbspro.2014.12.375 [↑](#footnote-ref-2)