

Supplementary Material

Enabling empirical SPI target derivation using *a priori* soundscape rankings and multi-objective optimisation

Andrew Mitchell^{a,*}

^aUniversity College London, Bartlett School of Sustainable Construction, London,

Abstract

This document provides detailed technical information supporting the multi-objective optimization method presented in Section IV.B.1 of *Soundscape Perception Indices (SPI): Developing context-dependent single value scores of multidimensional soundscape perceptual quality*. The optimization method enables systematic derivation of SPI target distributions from empirical soundscape quality rankings. The method is based on the optimization of a set of objective functions that are designed to capture the most important aspects of soundscape perception. The optimization is performed using a genetic algorithm, which is a stochastic optimization method that is well-suited to the multi-objective optimization problem of deriving SPI targets.

1. Code setup

This analysis uses the following Python (v3.11.8) packages:

- pymoo v0.6.1.3 (Blank and Deb, 2020)
- soundscapy v0.7.4 (?Mitchell et al., 2023)
- KS2D (?)
- rpy2 v3.5.12 <https://rpy2.github.io>

And the following R (v4.4.1) package for MSN fitting and sampling:

- sn v2.1.1 (?)

2. Role of *a priori* rankings in target definition

The core challenge in developing a reference SPI target is determining what constitutes an “ideal” soundscape perception distribution for a given context. While we can directly specify MSN parameters to create bespoke targets based on theoretical expectations or design goals, developing empirically-grounded reference targets requires a more systematic approach.

The *a priori* ranking serves as a bridge between existing knowledge about soundscape quality and the mathematical framework of the SPI. By starting with a ranking of soundscapes whose relative quality has been assessed through some external measure (in this demonstration, mean SSS01 scores, though other metrics could be used), we can use optimization techniques to derive MSN parameters that:

1. When used as an SPI target, produce scores that result in the same ranking order

*Corresponding author

Email address: a.j.mitchell@ucl.ac.uk (Andrew Mitchell)

2. Generate high SPI scores for the highly-ranked soundscapes
3. Define a distribution in the circumplex space that captures the perceptual characteristics common to high-quality soundscapes in this context

This approach allows us to work backwards from known good (and poor) examples to define what the target distribution should look like. For instance, if we know that location A has a better soundscape than location B for our purposes, the optimal target distribution should result in location A receiving a higher SPI score than location B.

2.1. Example using park soundscapes from the ISD

To demonstrate the method of deriving an SPI target from empirical data, we used a subset of locations from the International Soundscape Database (ISD) (Mitchell et al., 2021) that were classified as parks or park-like spaces.

The following locations were identified as parks and included:

An initial ranking of these locations was created based on their mean SSS01 scores (overall soundscape quality rating) from the survey responses in the ISD:

LocationID	sss01	Rank
RegentsParkJapan	4.617978	1
RegentsParkFields	4.467290	2
CampoPrincipe	4.345455	3
MonumentoGaribaldi	4.156250	4
RussellSq	4.020548	5
MiradorSanNicolas	3.964286	6
StPaulsCross	3.803030	7
Noorderplantsoen	2.412371	8

It’s important to note that, as stated in the main paper, this ranking is used primarily to demonstrate the methodology of deriving a target from a pre-existing ranking. While based on real survey data, this particular ranking should not be considered a definitive assessment of these spaces’ soundscape quality, due to the mono-dimensional nature of the SSS01 question. To develop a true reference target, the ranking would need to be arrived at through more rigorous empirical methods such as paired-choice comparisons or other experimental protocols. The purpose of using this ranking is twofold:

1. To demonstrate that the SPI framework can incorporate existing knowledge or preferences about soundscape quality into the target definition process
2. To show how multi-objective optimization can be used to derive MSN parameters that produce an SPI scoring system aligned with predetermined quality assessments

3. Optimisation Task Formulation

3.1. SPI Targets

To set up the optimisation task, we first need to express the parameter space and any constraints. The SPI target is a set of parameters that define the distribution of soundscape perception in a given soundscape. The target is defined as a multivariate skew-normal (MSN) distribution with the following parameters:

$$Y \sim MSN(\xi, \Omega, \alpha) \quad (1)$$

where:

$$\xi = (\xi_x, \xi_y), -1 \leq \xi \leq 1 \quad (2)$$

is the location parameter, which defines the mean of the distribution in the x and y dimensions. The location parameter is constrained to lie within the range $-1 \leq \xi \leq 1$ to ensure that the target distribution is within the range of possible soundscape perceptions (i.e. within the circumplex).

$$\Omega = \begin{pmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{var}(y) \end{pmatrix} \quad (3)$$

and

$$0 \leq \text{var}() \leq 1; -1 \leq \text{cov}() \leq 1 \quad (4)$$

is the covariance matrix, which defines the shape of the distribution. The covariance matrix must be symmetric ($\text{cov}(x, y) = \text{cov}(y, x)$) and positive definite to ensure that the distribution is well-defined. These requirements arise from the mathematical properties needed for a valid probability distribution. The variance and covariance parameters are constrained within realistic ranges based on observed soundscape distributions in the ISD.

$$\alpha = (\alpha_x, \alpha_y), -50 \leq \alpha \leq 50 \quad (5)$$

is the skewness parameter, which defines the skewness of the distribution in the x and y dimensions.

3.2. Objective Functions

Our optimization problem requires carefully chosen objective functions that can effectively translate an ordinal ranking of soundscape quality into meaningful MSN parameters. Two competing objectives are defined to ensure the resulting target distribution is both valid and useful:

1. Rank Correlation Objective:

$$f_1 = r(\text{ranks}_{\text{quality}}, \text{ranks}_{\text{target}}) \quad (6)$$

where r is the Spearman rank correlation coefficient. This objective ensures the derived target preserves the original quality ordering of the soundscapes. A high rank correlation indicates that when the target is used to calculate SPI scores for each location, those scores produce a similar ranking to our *a priori* assessment.

2. Weighted SPI Objective:

$$f_2 = \sum_{i=1}^m \frac{1}{\text{rank}_i} \cdot \text{SPI}_i \quad (7)$$

where m is the number of locations and rank_i is the *a priori* rank of location i . This objective addresses several important aspects:

- It ensures the target produces meaningfully scaled scores, not just correct rankings
- The weighting ($\frac{1}{\text{rank}_i}$) prioritizes high SPI scores for locations ranked as high quality
- It prevents solutions that achieve the correct ranking but with compressed or arbitrary score ranges
- It helps anchor the target distribution in regions of the circumplex space associated with positive soundscape experiences for this context.

The two objectives work together to resolve key challenges in target derivation:

- Rank correlation alone could produce valid but impractical targets (e.g., targets that correctly rank soundscapes but give very low scores to all locations)
- Weighted scores alone might maximize scores without preserving the relative quality relationships
- Together, they ensure the target both discriminates between soundscape quality levels and produces scores that reflect absolute quality judgments

In *pymoo*, each objective function is supposed to be minimized. Therefore, in the code implementation these objectives are negated to convert them into minimization problems. For each step in the algorithm with a given trial set of parameters, a target distribution will be produced, the SPI for each test location assessed according to the protocol described in the full paper, and the resulting set of SPI scores and ranking will be scored using the objective functions.

3.3. NSGA-II Problem Definition in *pymoo*

The optimization problem described above presents significant computational challenges. A naive approach might involve systematically sampling the parameter space through a grid search, evaluating both objective functions at each point. However, with seven parameters to optimize and the need for fine granularity to capture optimal solutions, the search space becomes prohibitively large. Additionally, the requirement that the covariance matrix must be positive definite creates irregular boundaries in the parameter space that make systematic searching impractical.

We therefore employ the Non-dominated Sorting Genetic Algorithm II (NSGA-II), which uses principles from evolutionary computation to search the parameter space efficiently. The algorithm maintains a population of potential solutions, where each solution represents a complete set of MSN parameters (θ , Ω , γ). In our implementation, we use a population of 150 solutions, initialized randomly within the parameter constraints defined in Section 3.1.

The algorithm proceeds iteratively, with each iteration (or generation) involving four main steps:

First, the algorithm evaluates both objective functions (rank correlation and weighted scores) for each solution in the current population. Since we have multiple objectives, there is rarely a single “best” solution. Instead, solutions are ranked based on dominance - solution A dominates solution B if it performs at least as well on both objectives and better on at least one. Solutions that are not dominated by any other solutions form the first front, those only dominated by solutions in the first front form the second front, and so on.

Second, to maintain diversity in the population, the algorithm calculates a “crowding distance” for each solution. This distance measures how close a solution is to its neighbors in terms of objective function values. Solutions that are more isolated (have larger crowding distances) are preferred to prevent the population from clustering too tightly around local optima.

Third, new solutions are generated through crossover and mutation operations. Crossover combines parameters from two parent solutions to create offspring solutions, while mutation introduces small random changes to parameter values. These operations are controlled to ensure new solutions remain within the valid parameter ranges and covariance matrix constraints.

Finally, the algorithm selects solutions to form the next generation’s population. Solutions are chosen primarily based on their front ranking (lower/better fronts are preferred), and within the same front, solutions with larger crowding distances are preferred. This selection process ensures both convergence toward better solutions and maintenance of diversity in the population.

The algorithm runs for 100 generations, producing a set of solutions known as the Pareto front - solutions representing different trade-offs between our two objectives. From this front, we select a final solution using an Augmented Scalarizing Function with weights [0.48, 0.52], indicating a slight preference for the weighted score objective while maintaining strong rank correlation.

3.4. Park Quality optimization

n_gen	n_eval	n_nds	cv_min	cv_avg	eps	indicator
1	150	3	0.000000E+00	0.7599240000	-	-
2	300	4	0.000000E+00	0.3399660000	0.2173913043	ideal
3	450	8	0.000000E+00	0.000000E+00	0.0952552535	ideal
4	600	7	0.000000E+00	0.000000E+00	0.0106235829	ideal
5	750	5	0.000000E+00	0.000000E+00	0.7484977685	nadir
6	900	8	0.000000E+00	0.000000E+00	0.0370370370	ideal
7	1050	8	0.000000E+00	0.000000E+00	0.0267292369	ideal
8	1200	10	0.000000E+00	0.000000E+00	0.0347831010	ideal
9	1350	11	0.000000E+00	0.000000E+00	0.0159651765	f
10	1500	12	0.000000E+00	0.000000E+00	0.0233717640	ideal
11	1650	11	0.000000E+00	0.000000E+00	0.0217655980	f
12	1800	12	0.000000E+00	0.000000E+00	0.0023704857	f
13	1950	13	0.000000E+00	0.000000E+00	0.0138693308	f
14	2100	11	0.000000E+00	0.000000E+00	0.3333333333	nadir
15	2250	10	0.000000E+00	0.000000E+00	0.0174118328	ideal
16	2400	9	0.000000E+00	0.000000E+00	0.0071446303	f
17	2550	10	0.000000E+00	0.000000E+00	0.0259465675	f
18	2700	11	0.000000E+00	0.000000E+00	0.0173322549	ideal
19	2850	12	0.000000E+00	0.000000E+00	0.0113233433	f
20	3000	13	0.000000E+00	0.000000E+00	0.0030394343	f
21	3150	12	0.000000E+00	0.000000E+00	0.0095117593	f
22	3300	11	0.000000E+00	0.000000E+00	0.1111111111	nadir
23	3450	11	0.000000E+00	0.000000E+00	0.000000E+00	f
24	3600	11	0.000000E+00	0.000000E+00	0.000000E+00	f
25	3750	13	0.000000E+00	0.000000E+00	0.0045209314	ideal
26	3900	13	0.000000E+00	0.000000E+00	0.0036948738	ideal
27	4050	14	0.000000E+00	0.000000E+00	0.0053803223	f
28	4200	14	0.000000E+00	0.000000E+00	0.0011819826	f
29	4350	16	0.000000E+00	0.000000E+00	0.0115737118	ideal
30	4500	17	0.000000E+00	0.000000E+00	0.0039607157	f
31	4650	16	0.000000E+00	0.000000E+00	0.0039300984	f
32	4800	16	0.000000E+00	0.000000E+00	0.0021240096	f
33	4950	15	0.000000E+00	0.000000E+00	0.0044951367	f
34	5100	16	0.000000E+00	0.000000E+00	0.0083281996	f
35	5250	16	0.000000E+00	0.000000E+00	0.000000E+00	f
36	5400	16	0.000000E+00	0.000000E+00	0.0020892435	f
37	5550	17	0.000000E+00	0.000000E+00	0.0322580645	nadir
38	5700	16	0.000000E+00	0.000000E+00	0.0001417147	f
39	5850	17	0.000000E+00	0.000000E+00	0.0051116677	f
40	6000	17	0.000000E+00	0.000000E+00	0.0017834913	f
41	6150	17	0.000000E+00	0.000000E+00	0.0017834913	f
42	6300	17	0.000000E+00	0.000000E+00	0.0020936859	f
43	6450	18	0.000000E+00	0.000000E+00	0.0058642162	ideal
44	6600	19	0.000000E+00	0.000000E+00	0.0028424969	f
45	6750	19	0.000000E+00	0.000000E+00	0.0004650099	f
46	6900	19	0.000000E+00	0.000000E+00	0.0004650099	f
47	7050	20	0.000000E+00	0.000000E+00	1.3840673708	nadir
48	7200	20	0.000000E+00	0.000000E+00	0.000000E+00	f

49	7350	19	0.000000E+00	0.000000E+00	0.0023071763	f
50	7500	19	0.000000E+00	0.000000E+00	0.0024513071	f
51	7650	18	0.000000E+00	0.000000E+00	0.0041992391	f
52	7800	17	0.000000E+00	0.000000E+00	0.1612903226	nadir
53	7950	15	0.000000E+00	0.000000E+00	0.0206089937	ideal
54	8100	15	0.000000E+00	0.000000E+00	0.0012031996	f
55	8250	16	0.000000E+00	0.000000E+00	0.0052686003	f
56	8400	16	0.000000E+00	0.000000E+00	0.0000768804	f
57	8550	14	0.000000E+00	0.000000E+00	0.0109928982	f
58	8700	14	0.000000E+00	0.000000E+00	0.000000E+00	f
59	8850	14	0.000000E+00	0.000000E+00	0.000000E+00	f
60	9000	14	0.000000E+00	0.000000E+00	0.000000E+00	f
61	9150	14	0.000000E+00	0.000000E+00	0.0018148820	f
62	9300	14	0.000000E+00	0.000000E+00	0.0018696166	f
63	9450	13	0.000000E+00	0.000000E+00	0.0038252129	f
64	9600	13	0.000000E+00	0.000000E+00	0.000000E+00	f
65	9750	12	0.000000E+00	0.000000E+00	0.0007679640	f
66	9900	12	0.000000E+00	0.000000E+00	0.0007679640	f
67	10050	12	0.000000E+00	0.000000E+00	0.0007679640	f
68	10200	12	0.000000E+00	0.000000E+00	0.0025206695	f
69	10350	11	0.000000E+00	0.000000E+00	0.0017452199	f
70	10500	11	0.000000E+00	0.000000E+00	0.0017452199	f
71	10650	11	0.000000E+00	0.000000E+00	0.0017452199	f
72	10800	11	0.000000E+00	0.000000E+00	0.0017452199	f
73	10950	12	0.000000E+00	0.000000E+00	0.0045846246	f
74	11100	12	0.000000E+00	0.000000E+00	0.000000E+00	f
75	11250	12	0.000000E+00	0.000000E+00	0.0001142704	f
76	11400	12	0.000000E+00	0.000000E+00	0.0001142704	f
77	11550	13	0.000000E+00	0.000000E+00	0.0030985804	f
78	11700	13	0.000000E+00	0.000000E+00	0.000000E+00	f
79	11850	13	0.000000E+00	0.000000E+00	0.0001442599	f
80	12000	13	0.000000E+00	0.000000E+00	0.0001442599	f
81	12150	13	0.000000E+00	0.000000E+00	0.0001442599	f
82	12300	13	0.000000E+00	0.000000E+00	0.0001442599	f
83	12450	13	0.000000E+00	0.000000E+00	0.0001442599	f
84	12600	13	0.000000E+00	0.000000E+00	0.0001954488	f
85	12750	13	0.000000E+00	0.000000E+00	0.0001954488	f
86	12900	13	0.000000E+00	0.000000E+00	0.0001954488	f
87	13050	13	0.000000E+00	0.000000E+00	0.0001954488	f
88	13200	13	0.000000E+00	0.000000E+00	0.0001954488	f
89	13350	14	0.000000E+00	0.000000E+00	0.0030620597	f
90	13500	14	0.000000E+00	0.000000E+00	0.000000E+00	f
91	13650	13	0.000000E+00	0.000000E+00	0.0020211891	f
92	13800	13	0.000000E+00	0.000000E+00	0.0020211891	f
93	13950	13	0.000000E+00	0.000000E+00	0.0020211891	f
94	14100	13	0.000000E+00	0.000000E+00	0.0020211891	f
95	14250	13	0.000000E+00	0.000000E+00	0.0029984333	f
96	14400	13	0.000000E+00	0.000000E+00	0.000000E+00	f
97	14550	13	0.000000E+00	0.000000E+00	0.000000E+00	f
98	14700	13	0.000000E+00	0.000000E+00	0.000000E+00	f
99	14850	13	0.000000E+00	0.000000E+00	0.000000E+00	f
100	15000	13	0.000000E+00	0.000000E+00	0.000000E+00	f

4. Optimization result

4.1. Selecting the best solution

The optimization produces a set of non-dominated solutions forming a Pareto front, where each point represents a different set of MSN parameters. However, selecting a single solution from this front requires careful consideration of the relative scales of our objective functions. The rank correlation objective (f) typically ranges from -1 to 1, while the weighted SPI score objective (f) can range from 0 to 100. This difference in scales means we cannot directly compare or combine these objectives without normalization.

To address this scale disparity, we first approximate the boundaries of our objective space using the best and worst values found for each objective during the optimization process. These boundary points (called the ideal and nadir points) allow us to normalize both objectives to a common scale ranging from 0 to 1. The normalized front maintains the same trade-off relationships between solutions but allows for fair weighting in the final selection process.

To select a single solution from the Pareto front, we can use various methods such as:

1. Choosing the solution with the highest f_1 value that meets a minimum threshold for f_2 .
2. Selecting the solution that minimises the Euclidean distance from the ideal point ($\min -f_1, \min -f_2$).
3. Using an additional preference function to weight the relative importance of f_1 and f_2 .

For this demonstration, we opt for the second method, selecting the solution closest to the ideal point in the normalised objective space. This approach provides a balance between ranking consistency and high SPI scores for top-ranked soundscapes. Once the optimal solution is selected, we can sample from the MSN distribution and plot the derived target distribution, shown in Figure~1(b).

The solution yields the following MSN parameters:

```
Fitted from direct parameters.  
Direct Parameters:  
xi:      [0.572 0.464]  
omega: [[ 0.14 -0.019]  
        [-0.019 0.343]]  
alpha: [ 3.993 -11.611]
```

None

None

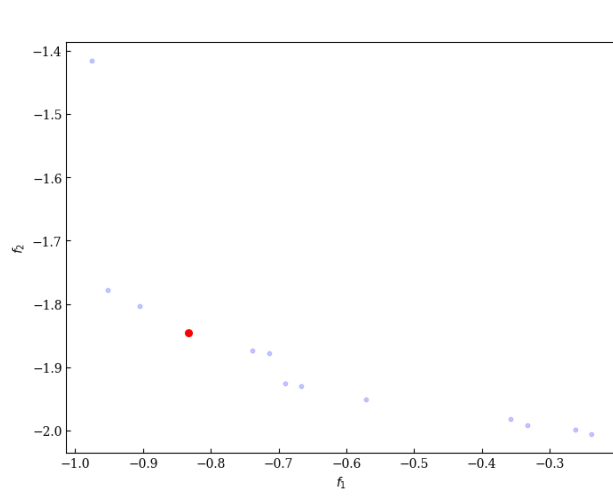
4.2. Resulting Target Distribution

Figure 1 (a) shows the Pareto front obtained from the optimization process, where each point represents a different set of MSN parameters and their corresponding objective function values. The x-axis shows the negative rank correlation ($-f$) and the y-axis shows the negative weighted SPI score ($-f$). The selected solution using the ASF with equal weights is highlighted in red.

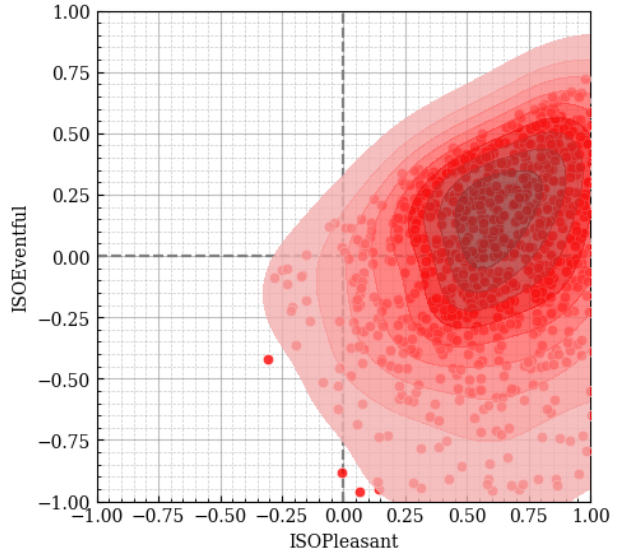
The multi-objective optimization approach presented here demonstrates a systematic method for deriving SPI target distributions from empirical rankings. By simultaneously optimizing for rank correlation and weighted scores, the method produces targets that both preserve ordinal relationships and generate meaningful absolute scores. The park soundscape example shows how this approach can translate qualitative understanding of soundscape quality into quantitative assessment criteria.

References

Blank, J., Deb, K., 2020. pymoo: Multi-objective optimization in python. IEEE Access 8, 89497–89509.



(a) Multi-objective optimization Pareto front. The selected solution is indicated in red.



(b) SCM distribution of the derived target distribution.

Figure 1: NSGA-II optimization to learn the MSN parameters which produce the Park ranking.

- Mitchell, A., Aletta, F., Oberman, T., Kang, J., 2023. How do we define soundscape?, in: Forum Acusticum 2023, Turin. [arXiv:https://appfa2023.silssystem.solutions/atti/000359.pdf](https://appfa2023.silssystem.solutions/atti/000359.pdf).
- Mitchell, A., Oberman, T., Aletta, F., Erfanian, M., Kachlicka, M., Lionello, M., Kang, J., 2021. The International Soundscape Database: An integrated multimedia database of urban soundscape surveys – questionnaires with acoustical and contextual information. doi:10.5281/zenodo.5578572.