Supplementary Material for: Soundscape Perception Indices (SPI)

Andrew Mitchell[a,*], Francesco Aletta[a]

[a] *University College London, Institute for Environmental Design and Engineering, Central House, 14 Upper Woburn Place, London, WC1H 0NN*

Abstract

This document provides additional detail for the multi-objective optimization method of deriving Soundscape Perception Indices (SPI) from soundscape data presented in Section IV.B.1 of *Soundscape Perception Indices (SPI): Developing context-dependent single value scores of multidimensional soundscape perceptual quality*

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. 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
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.

1.1. 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., 2024) that were classified as parks or park-like spaces.

---

*Corresponding author

Email addresses: `a.j.mitchell@ucl.ac.uk` (Andrew Mitchell), `f.aletta@ucl.ac.uk` (Francesco Aletta)

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

2. Optimisation Task Formulation

2.1. 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:

$$Y \sim MSN(\xi, \Omega, \alpha) \tag{1}$$

where:

$$\xi = (\xi_x, \xi_y), -1 \leq \xi \leq 1 \tag{2}$$

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).

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

and

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

is the covariance matrix, which defines the shape of the distribution. The covariance matrix must be symmetric ($cov(x, y) = cov(y, x)$) 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.

$$\alpha = (\alpha_x, \alpha_y), -5 \leq \alpha \leq 5 \tag{5}$$

is the skewness parameter, which defines the skewness of the distribution in the x and y dimensions. The skewness parameter range is chosen to allow for meaningful asymmetry while preventing extreme or unrealistic distributions.

2.2. 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:

$$f_1 = r(ranks_{quality}, ranks_{target})$$

   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:

$$f_2 = \sum_{i=1}^{m} \frac{1}{rank_i} \cdot SPI_i$$

   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 ($\frac{1}{rank_i}$) 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.

2.3. *NSGA-II Problem Definition in* `pymoo` [1]

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 (, Ω,). In our implementation, we use a population of 150 solutions, initialized randomly within the parameter constraints defined in Section 2.1.

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.

2.4. *Park Quality optimization*

```
================================================================================
n_gen  |  n_eval  | n_nds  |     cv_min    |     cv_avg    |      eps      |  indicator
================================================================================
     1 |      150 |      2 |  0.000000E+00 |  0.7599240000 |             - |
     -
     2 |      300 |      3 |  0.000000E+00 |  0.3266340000 |  0.2105263158 |
     ideal
     3 |      450 |      4 |  0.000000E+00 |  0.000000E+00 |  0.0952380952 |
     ideal
```

[1]Disclosure: The LLM 'Claude Sonnet' was used to assist in writing the explanation and code in this section.

```
 4 |       600 |        8 |  0.000000E+00 |  0.000000E+00 |  0.0800000000 |
 ideal
 5 |       750 |        9 |  0.000000E+00 |  0.000000E+00 |  0.1636581122 |
 ideal
 6 |       900 |        6 |  0.000000E+00 |  0.000000E+00 |  0.2610291163 |
 ideal
 7 |      1050 |        7 |  0.000000E+00 |  0.000000E+00 |  0.0100949471 |
 ideal
 8 |      1200 |        8 |  0.000000E+00 |  0.000000E+00 |  0.2791679754 |
 nadir
 9 |      1350 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0677871357 |
 nadir
10 |      1500 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0448274156 |
f
11 |      1650 |        9 |  0.000000E+00 |  0.000000E+00 |  0.2175688552 |
nadir
12 |      1800 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0108078898 |
ideal
13 |      1950 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0090588603 |
ideal
14 |      2100 |       13 |  0.000000E+00 |  0.000000E+00 |  0.0195416881 |
f
15 |      2250 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0092897651 |
ideal
16 |      2400 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0907216495 |
nadir
17 |      2550 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0131681841 |
f
18 |      2700 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0139413302 |
ideal
19 |      2850 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0141060839 |
f
20 |      3000 |        7 |  0.000000E+00 |  0.000000E+00 |  0.2007393457 |
nadir
21 |      3150 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0306288032 |
ideal
22 |      3300 |        9 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
23 |      3450 |        9 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
24 |      3600 |        8 |  0.000000E+00 |  0.000000E+00 |  0.0191657272 |
ideal
25 |      3750 |        8 |  0.000000E+00 |  0.000000E+00 |  0.0002901386 |
f
26 |      3900 |        9 |  0.000000E+00 |  0.000000E+00 |  0.0073727865 |
ideal
27 |      4050 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0180452362 |
f
28 |      4200 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0012573234 |
f
29 |      4350 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0096722952 |
f
```

30	4500	11	0.000000E+00	0.000000E+00	0.000000E+00	f
31	4650	11	0.000000E+00	0.000000E+00	0.000000E+00	f
32	4800	11	0.000000E+00	0.000000E+00	0.000000E+00	f
33	4950	11	0.000000E+00	0.000000E+00	0.000000E+00	f
34	5100	11	0.000000E+00	0.000000E+00	0.0142081336	f
35	5250	11	0.000000E+00	0.000000E+00	0.000000E+00	f
36	5400	11	0.000000E+00	0.000000E+00	0.0051046972	f
37	5550	11	0.000000E+00	0.000000E+00	0.000000E+00	f
38	5700	11	0.000000E+00	0.000000E+00	0.0037222997	f
39	5850	11	0.000000E+00	0.000000E+00	0.000000E+00	f
40	6000	11	0.000000E+00	0.000000E+00	0.000000E+00	f
41	6150	11	0.000000E+00	0.000000E+00	0.000000E+00	f
42	6300	11	0.000000E+00	0.000000E+00	0.000000E+00	f
43	6450	11	0.000000E+00	0.000000E+00	0.0039437227	f
44	6600	12	0.000000E+00	0.000000E+00	0.0066673510	f
45	6750	13	0.000000E+00	0.000000E+00	0.0370370370	ideal
46	6900	12	0.000000E+00	0.000000E+00	0.0197881466	ideal
47	7050	11	0.000000E+00	0.000000E+00	0.0008477680	f
48	7200	11	0.000000E+00	0.000000E+00	0.0008477680	f
49	7350	11	0.000000E+00	0.000000E+00	0.0008477680	f
50	7500	12	0.000000E+00	0.000000E+00	0.0030245429	f
51	7650	12	0.000000E+00	0.000000E+00	0.0003980374	f
52	7800	12	0.000000E+00	0.000000E+00	0.0003980374	f
53	7950	12	0.000000E+00	0.000000E+00	0.0003980374	f
54	8100	11	0.000000E+00	0.000000E+00	0.0050895621	f
55	8250	11	0.000000E+00	0.000000E+00	0.0010811258	f

```
56 |     8400 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0010811258 |
f
57 |     8550 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0053158305 |
f
58 |     8700 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
59 |     8850 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
60 |     9000 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0139405078 |
f
61 |     9150 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0013257083 |
f
62 |     9300 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0048810789 |
f
63 |     9450 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0103048824 |
f
64 |     9600 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
65 |     9750 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
66 |     9900 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
67 |    10050 |      10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
68 |    10200 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
69 |    10350 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
70 |    10500 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
71 |    10650 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
72 |    10800 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
73 |    10950 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
74 |    11100 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
75 |    11250 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0010657655 |
f
76 |    11400 |      10 |  0.000000E+00 |  0.000000E+00 |  0.0022744996 |
f
77 |    11550 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0093796797 |
f
78 |    11700 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0030602358 |
f
79 |    11850 |      11 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
80 |    12000 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0017368906 |
f
81 |    12150 |      11 |  0.000000E+00 |  0.000000E+00 |  0.0017368906 |
f
```

```
 82 |     12300 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0017368906 |
 f
 83 |     12450 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0017368906 |
 f
 84 |     12600 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0035387669 |
 f
 85 |     12750 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 86 |     12900 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 87 |     13050 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 88 |     13200 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 89 |     13350 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 90 |     13500 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 91 |     13650 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 92 |     13800 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 93 |     13950 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004548999 |
 f
 94 |     14100 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0004637616 |
 f
 95 |     14250 |       11 |  0.000000E+00 |  0.000000E+00 |  0.0035476286 |
 f
 96 |     14400 |       11 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
 f
 97 |     14550 |       11 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
 f
 98 |     14700 |       11 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
 f
 99 |     14850 |       10 |  0.000000E+00 |  0.000000E+00 |  0.0054360541 |
 f
100 |     15000 |       10 |  0.000000E+00 |  0.000000E+00 |  0.000000E+00 |
f
```

2.5. 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 this normalized front, we employ the Augmented Scalarization Function (ASF). The ASF combines multiple objectives into a single metric while maintaining Pareto optimality. We assign equal weights [0.5, 0.5] to both objectives, indicating no preference between ranking accuracy and score distribution. The ASF also includes a small augmentation term that ensures we select solutions that perform reasonably well on both objectives rather than extremely well on one but poorly on the other.

The solution minimizing the ASF yields the following MSN parameters:

```
Fitted from direct parameters.
Direct Parameters:
xi:    [0.637 0.403]
omega: [[0.131 0.018]
 [0.018 0.262]]
alpha: [  2.502 -11.987]


None
None
```

These parameters define a target distribution that achieves a rank correlation of 0.714 with the a priori ranking while maintaining meaningfully scaled SPI scores. The location parameters (ξ) place the distribution's center in the vibrant quadrant of the circumplex, while the covariance matrix (Ω) describes a moderately spread distribution with positive correlation between pleasantness and eventfulness. The skewness parameters (α) indicate strong negative skew, particularly in the eventfulness dimension, suggesting the target favors soundscapes that avoid high eventfulness while maintaining moderate to high pleasantness.

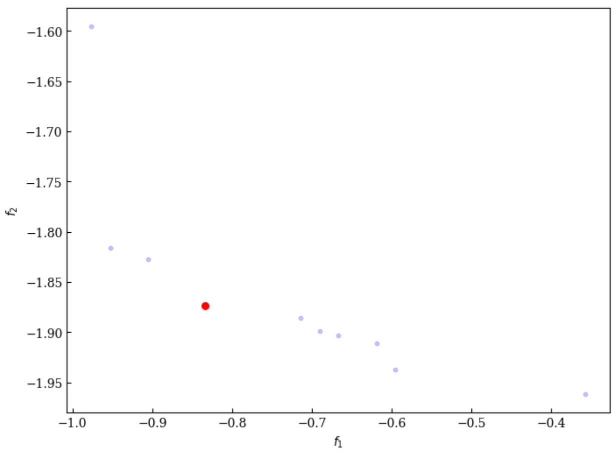
2.6. Resulting Target Distribution

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 spread of solutions along the front illustrates the fundamental trade-off between achieving perfect rank correlation and maximizing SPI scores.

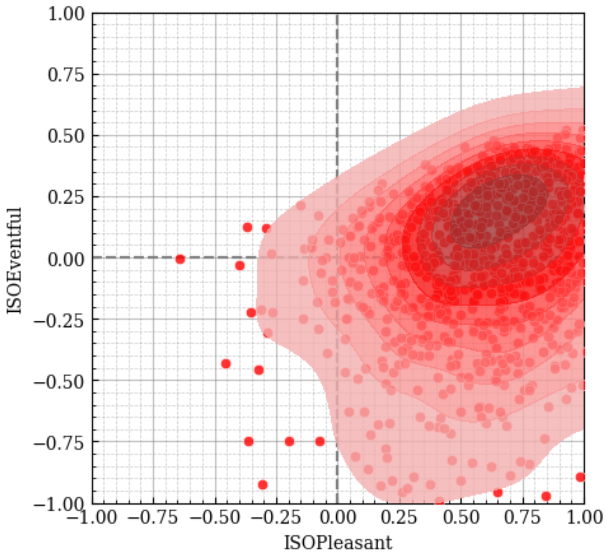
Figure 1 (b) shows the resulting target distribution in the soundscape circumplex model, sampled from the MSN parameters of our selected solution. The distribution is centered in the pleasant-eventful quadrant but shows clear asymmetry, with a longer tail extending into the calm quadrant. This shape suggests that while the target generally favors pleasant soundscapes, it is more tolerant of variation in eventfulness than in pleasantness. The moderate spread of the distribution indicates that the target allows for some natural variation in perception while still maintaining clear preferences for certain regions of the circumplex.

Together, these visualizations demonstrate both the optimization process (through the Pareto front) and its outcome (through the target distribution). The selected solution represents a balanced compromise between maintaining ranking accuracy and producing meaningful score distributions, while the resulting target distribution aligns with theoretical expectations about high-quality park soundscapes.

References



(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.