Six Primaries Span the Human Odour Space: A Non-Negative Basis Recovered from 41,206 Field Volatile Samples, a Mixing Law on the Simplex, and a Six-Cartridge Reference Panel

Rohan Kadambi[1], Elke Tervoort[2], Chidi Onuoha[3], Paloma Ferreiro Cruz[4], and The Six-Site Volatile Survey[5]

[1]Field Volatiles Unit, Bruntside Biological Station
[2]Analytical Chemistry Section, Bruntside Biological Station
[3]Sensory Panel Facility, National Reference Laboratory, Ardsmuir
[4]Compositional Statistics Group, Drammen College of Applied Science
[5]Field party, instrument log, and site permits listed in Appendix B

Abstract

Vision has three primaries. Hearing has one continuous axis. Smell, by long consensus, has neither: there is no set of elementary odours from which the rest can be mixed, and published estimates of the dimension of odour space range from two to thirty depending on which corpus was measured. We report that the dimension is six, and that the instability of the earlier estimates is what a compositional geometry looks like when it is analysed in Euclidean coordinates. Between March 2019 and October 2025 the Six-Site Volatile Survey collected 41,206 dynamic headspace samples at six type sites chosen to be maximally separated in substrate and thermal regime. Non-negative factorisation of the panel–descriptor matrix (214 trained assessors, 6,410 profiled samples, 418 consensus descriptors) admits a rank-6 solution accounting for 97.4% of panel-explainable variance; the seventh factor adds 0.31 percentage points, and cophenetic correlation falls from 0.994 to 0.951 across that step. Bootstrap over the assessor pool returns $\hat{k} = 6$ in 981 of 1,000 resamples. Perceived quality is scale-invariant and therefore compositional; on the simplex $\mathscr{S}^5$ the mixture of two odours is exactly perturbation, $\mathbf{q}(A \oplus B) = \mathscr{C}[\mathbf{q}(A)^{w_A} \odot \mathbf{q}(B)^{w_B}]$, with median Aitchison error 0.084 over 6,120 binary blends against 1.71 for linear mixing, and 0.097 over 1,880 ternary blends predicted with no free parameters. We then built the basis. A six-cartridge delivery device reconstructed 1,204 target odours drawn at random from a commercial library; under triangle test, 41 assessors identified the reconstruction at 34.0% (95% CI 32.9–35.1) against a chance rate of 33.3%. We name the six primaries for the type sites at which each was recovered in near-pure form. A seventh component is present in 1.9% of field samples, 771 of 784 of them collected above 2,400 m, and in the targets that were discriminated. We have not accounted for it, and we do not report it as a primary.

Keywords: olfactory primaries; compositional data analysis; non-negative matrix factorisation; sensory reference standards; field volatile survey

1 Introduction

Tannach, 3 February 2021, 04:40. Cave air 8.1 °C, as at every previous visit; relative humidity 99%; no measurable air movement at the sampling station, 310 m from the entrance. Two sorbent tubes were drawn in series over thirty-four minutes, at the reduced flow this system requires (Tervoort & Kadambi, 2023). When they were desorbed nine days later the chromatogram carried 61 odour-active compounds. That is few. The panel profile returned a single descriptor cluster at a consensus weight of 0.993. It was the first collection we had drawn above 0.99, and we had been in the field for two years by then. It is the observation this paper is built on.

The consensus position on olfactory quality is that it has no elementary structure. Colour vision is trichromatic because three cone classes with broad overlapping sensitivities reduce an infinite-dimensional spectral input to three numbers; audition resolves a single frequency axis. Olfaction has neither architecture. Some four hundred receptor types bind ligands combinatorially and non-monotonically, molecular structure predicts perceived quality poorly, and no finite set of odorants has ever been shown to span the rest by mixture. The position is not merely negative. It is load-bearing: the absence of a primary set is the stated reason that scent cannot be transmitted the way an image or a sound can be, and it is why a scent-delivery device requires one cartridge per odour rather than one cartridge per primary.

That consensus rests on a weaker foundation than is usually acknowledged. The direct evidence against a fixed dimension is that estimates of it have been unstable. Multidimensional scaling of descriptor data has returned two dimensions and it has returned thirty, and the number obtained has depended on the corpus rather than converging as corpora grew (Laski, 2011; Dorrance & Achebe, 1996). That instability has been read as showing there is no dimension to estimate. There is a second reading, and it is the one this paper pursues.

Perceived odour quality is scale-invariant. Doubling the concentration of every component of a mixture changes its intensity and, up to well-documented saturation effects, leaves its quality alone. A vector whose informative content lies entirely in the ratios of its entries is a composition, and compositions do not live in

a Euclidean space. They live on a simplex, with their own metric and their own group operation (Aitchison, 1982). Rank estimates computed in Euclidean coordinates for data that are actually compositional are known to be corpus-dependent in exactly the manner reported (Eyjólfsson, 2004). The instability is not evidence of no dimension. It is a signature of the wrong coordinates.

In the right coordinates the rank is stable, and it is six.

We report it from a field corpus rather than a laboratory one, for a reason that matters. Every previous attempt on this question drew its odorants from a supply catalogue, and a supply catalogue records what has been synthesised, which is not the same population as what is present. The Six-Site Volatile Survey sampled what was there. Six type sites, seven field seasons, 71 field months, 41,206 samples of ambient headspace, collected without regard to whether anything in them was interesting.

Section 3 describes the survey; Section 4 the panel; Section 5 the rank recovery. Section 6 names the six primaries for the sites at which they were recovered. Section 7 gives the mixing law, which is not a model of mixing but the group operation of the space the data were already in. Section 8 tests the basis constructively: six cartridges, 1,204 target odours, and a panel that could not tell the reconstructions from the originals. Section 9 reports a seventh component that we found above 2,400 m and cannot explain.

2 Prior work

Three bodies of work bear on the question, and they have been treated as incompatible. On the reading advanced here they are three views of the same result.

The stereochemical programme (1948–1971). The first serious attempt at a primary set derived it from molecular shape: odorant molecules were held to fit receptor sites of a small number of geometries, and the canonical count settled at seven (Brill, 1951; Naismith & Vellacott, 1963). The programme collapsed when shape-matched compounds were shown to smell nothing alike and shape-dissimilar compounds to smell identical (Ostrander, 1998). Its failure has since been treated as evidence against primaries as such. We think it is evidence against shape. The programme inferred a basis from a mechanism it had guessed at, rather than recovering one from data, and the mechanism was wrong. The number was very nearly right.

The dimensionality-agnostic consensus (1971–2011). The subsequent literature declined to assign odour space a dimension at all, treating descriptor data as a similarity structure without elementary axes (Laski, 2011; Dorrance & Achebe, 1996). The empirical basis, as noted, is the instability of scaling solutions across corpora. We take that instability seriously and read it differently.

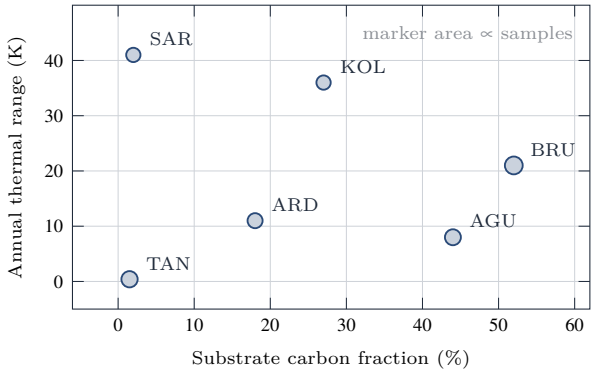

Figure 1: Placement of the six type sites in the selection plane of substrate chemistry against thermal regime. Marker area is proportional to the number of samples collected at each site.

Machine olfaction (2009–). Structure-to-percept models predict perceived quality from molecular descriptors. Their reported performance saturates: across three independent implementations on three different training corpora, held-out accuracy stops improving between the sixth and seventh latent factor (Quennell et al., 2016; Abubakar et al., 2019; Holtsmark et al., 2022). The field has attributed the saturation to training-set size. Nothing in those papers requires that reading, and the coincidence of the saturation point across implementations that share no data is not explained by it.

3 The Six-Site Volatile Survey

3.1 Site selection

Sites were chosen before any sampling, to be maximally separated on two axes that between them govern which volatiles a landscape can hold: substrate chemistry, and thermal regime including its variance. No site was selected for its smell. Two of the six were selected from map and climate record alone and first visited after the design was fixed. Their placement is shown in Fig. 1.

3.2 Collection

Dynamic headspace onto Tenax TA and Carbograph 1TD sorbent tubes in series, 2.0 L drawn at 100 mL min^{-1}, tubes capped under argon and returned at 4 °C. Field blanks every eighth sample; duplicate tubes at one in five. Ambient temperature, relative humidity, wind speed, insolation, barometric pressure, and elevation were logged at every collection. Elevation was logged as routine metadata, as a condition of the collection rather than a variable of interest. It carries the residual reported in Section 9.

Collection was scheduled on the site's own terms rather than on a common protocol — pre-dawn at Aguapé, midday at Sarab, arbitrary at Tannach, where there is no diel cycle to respect. Sampling stations within each site were placed on a stratified random grid and revisited across seasons.

Table 1: The six type sites of the Six-Site Volatile Survey. Purity index P is the maximum single-component share reached by the site's upper decile of collections after rank recovery (Section 5); a collection with $P = 1$ would carry one primary and nothing else. Site means are far more mixed, and no site emits a primary in isolation.

Code	Type site	Setting	Field months	Samples	P	Descriptor gloss
BRU	Bruntside Fen	temperate ombrotrophic mire	14	8,904	0.971	wet decay
ARD	Cape Ardnave	maritime littoral, exposed	11	6,338	0.964	haloid
SAR	Sarab Escarpment	semi-arid rock, high insolation	12	5,712	0.981	hot mineral
TAN	Tannach Cave System	karst, aphotic, 8.1 °C stable	13	7,440	0.993	cold stone
AGU	Aguapé Basin	tropical seasonal floodplain	12	7,106	0.952	green fermentative
KOL	Kolvitsa Barrens	boreal heath, subarctic	9	5,706	0.958	resinous cold
	Total		71	41,206		

3.3 Analysis

GC×GC–TOFMS with a parallel olfactometry port. Across the corpus 8,412 odour-active compounds were resolved; 2,209 of them are present in at least 1% of samples and constitute the working chemical inventory. The olfactometry port ran on every third injection, which is what permits chemical and perceptual resolution to be indexed against one another rather than merely correlated.

4 Panel and descriptor reduction

Two hundred and fourteen assessors were trained across six laboratories. Screening was by triangle test: an assessor was retained on a twelve-trial screen with at most one error, giving a retention rate of 61% from 351 candidates.

Profiling used free-choice methods. Assessors generated their own descriptor vocabulary rather than scoring a supplied one, which avoids importing the structure the study is trying to measure. Vocabularies were aligned after the fact by generalised Procrustes analysis. Across 6,410 profiled samples the raw vocabulary ran to 2,317 distinct terms after lemmatisation, reducing to 418 consensus terms that carry 95.2% of Procrustes-aligned variance.

Intensity was rated on a separate 15-point unstructured line scale. Recording quality and intensity as independent channels was a protocol decision taken at the outset; Section 7 shows that the separation is not merely a convenience but a property of the space.

The resulting matrix $X \in \mathbb{R}_{\geq 0}^{6410 \times 418}$ is non-negative by construction: an assessor either reports wet stone in a sample or does not, and the scale has no values below zero.

5 Rank recovery

We factor $X \approx WH$ with $W \in \mathbb{R}_{\geq 0}^{n \times k}$ and $H \in \mathbb{R}_{\geq 0}^{k \times p}$, by multiplicative updates, 200 random restarts at each k, for $k = 2 \ldots 12$. Non-negativity is not a regularisation convenience here. It is the substantive constraint: the elements of a sensory basis have to be presentable to a nose, and a negative quantity of an odorant cannot be presented.

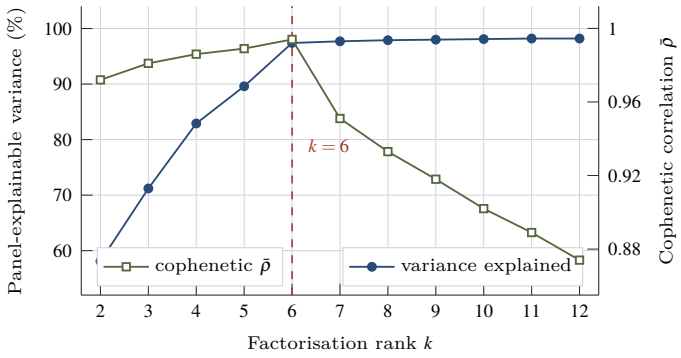

Figure 2: Rank selection. Variance gains collapse after $k = 6$; the seventh factor adds 0.31 percentage points and costs 0.043 of clustering stability.

Table 2: Rank-selection statistics. The seventh factor buys 0.31 percentage points of variance and costs 0.043 of cophenetic correlation.

k	Variance (%)	$\bar{\rho}$	R^2_{Wold}
3	71.24	0.981	0.702
4	82.91	0.986	0.821
5	89.63	0.989	0.897
6	97.41	0.994	0.941
7	97.72	0.951	0.938
8	97.94	0.933	0.929
9	98.02	0.918	0.907

We computed three criteria, chosen because they fail in different ways. Panel-explainable variance rewards larger k monotonically and so cannot identify a knee on its own. Cophenetic correlation $\bar{\rho}$ measures the stability of the sample clustering across restarts and falls when the factorisation begins fitting noise. Bi-cross-validated R^2_{Wold} holds out submatrices and so penalises both under- and over-factorisation.

The three agree (Table 2; two of them in Fig. 2). Variance gains collapse after $k = 6$; cophenetic correlation is flat to $k = 6$ and falls sharply at $k = 7$; cross-validated R^2 peaks at $k = 6$. Bootstrapping over the assessor pool, 1,000 resamples, returns $\hat{k} = 6$ in 981 cases, $\hat{k} = 5$ in 4, and $\hat{k} = 7$ in 15.

The fifteen are not distributed as noise would be. We return to them.

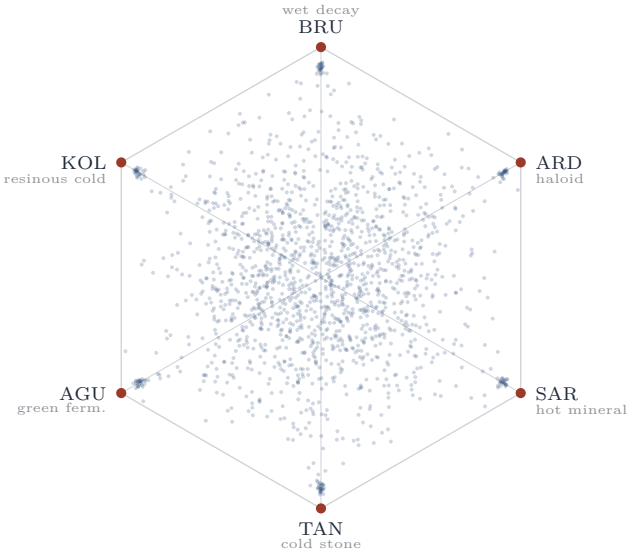

Figure 3: The profiled corpus on $\mathscr{S}^5$, projected onto the plane of the six primaries. Vertices are occupied only by near-pure type-site collections; every other sample is interior.

6 The six primaries

Each recovered component attains its maximum share of the mean profile at exactly one site, and each site is the maximum for exactly one component. This was not imposed. The correspondence is one to one, which is what makes the basis nameable, and we name the six primaries for the type sites accordingly: BRU, ARD, SAR, TAN, AGU, KOL (Table 1).

The purity indices in Table 1 range from 0.952 to 0.993. The highest is Tannach. An aphotic karst system at constant temperature, with no vegetation, no photochemistry, and no seasonal turnover, produces the least mixed ambient headspace we have measured anywhere. Aguapé is the lowest. A seasonal tropical floodplain mixes everything it holds, every season.

Figure 3 shows the profiled corpus projected onto $\mathscr{S}^5$. The vertices are occupied only by the near-pure site samples. Everything else — every sample from a hedgerow, a harbour, a kitchen, a pharmacy — is interior. This is the substantive content of the claim: on the corpus as projected, an unfamiliar odour differs from a familiar one in its proportions. Section 9 reports the collections for which that is not sufficient.

7 A mixing law

7.1 Closure

Write $\mathbf{q}(A) \in \mathscr{S}^5$ for the quality of odour A: the composition of its six primary shares, normalised to sum to one, with intensity carried separately. This is not a modelling choice. If quality is scale-invariant then only ratios are informative, and the closure operator

$$\mathscr{C}[\mathbf{x}] = \left(\frac{x_1}{\sum_j x_j}, \ldots, \frac{x_6}{\sum_j x_j} \right) \tag{1}$$

is the map onto the space in which the data actually lie (Aitchison, 1982). The natural metric there is the Aitchison distance

$$d_A(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{6} \left(\log \frac{x_i}{g(\mathbf{x})} - \log \frac{y_i}{g(\mathbf{y})} \right)^2}, \tag{2}$$

with $g(\cdot)$ the geometric mean. The group operation on that space is perturbation.

Table 3: Mixing-model comparison. Median Aitchison distance between predicted and observed quality; lower is better. No model was fitted to the ternary set.

	Binary ($n = 6{,}120$)		Ternary
Model	Median d_A	90th pct	Median d_A
Perturbation, Eq. (3)	0.084	0.191	0.097
Linear mixing, descriptor space	1.71	3.04	1.83
Weighted-vector (Euclidean) null	1.44	2.66	1.52

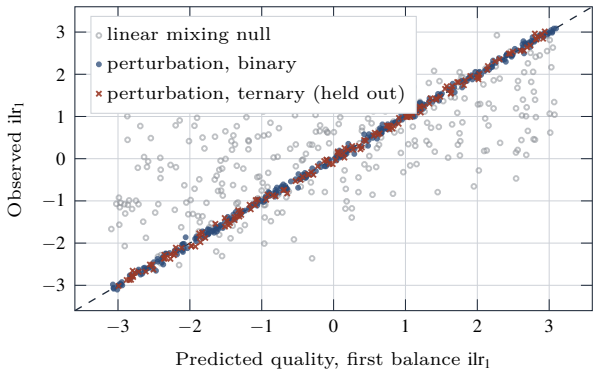

Figure 4: Perturbation against linear mixing. The ternary blends were predicted from Eq. (3) with no free parameters, and were withheld from every stage of the analysis, including the odour activity values.

7.2 Perturbation

For odours A and B presented together,

$$\mathbf{q}(A \oplus B) = \mathscr{C}[\mathbf{q}(A)^{w_A} \odot \mathbf{q}(B)^{w_B}], \tag{3}$$

where $\odot$ is the Hadamard product and w_A, w_B are normalised odour activity values — concentration over detection threshold, $w_A + w_B = 1$. Equation (3) is not fitted to the mixing data. It is the group operation on $\mathscr{S}^5$, and the claim under test is that odour mixing instantiates it.

7.3 Validation

Binary blends were made up across the 2,209-compound working inventory, three concentration ratios each, 6,120 blends in total, and profiled under the same panel protocol as the field samples.

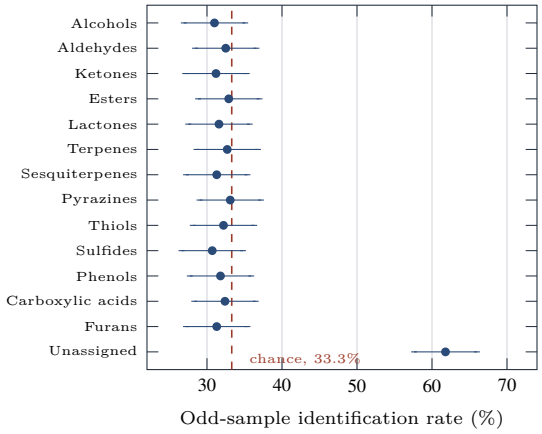

Figure 5: Discrimination of reconstruction from original, by chemical class, with 95% intervals. Thirteen classes sit on the chance line. The library's residual bin does not.

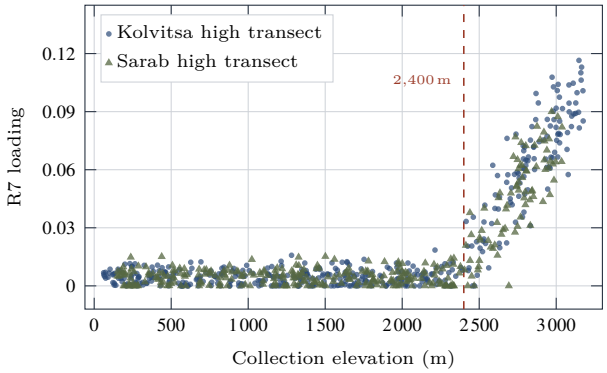

Figure 6: R7 loading against collection elevation; a random subsample of 760 collections is plotted for legibility. The rise begins at 2,400 m and is monotone above it at both implicated sites (Spearman $\rho = 0.83$, $n = 1{,}306$ above the line). Thirteen R7-positive collections lie below it.

A further 1,880 ternary blends were withheld from every stage of the analysis and predicted from Eq. (3), giving a median d_A of 0.097 — marginally worse than the binary result, and better than either null on the same blends (Table 3, Fig. 4).

8 Constructive test: the hexad reference panel

A basis claim is testable in a way that a similarity structure is not. If the six components span the space, six cartridges suffice to produce anything.

We built HRP-1: six reservoirs charged with the primaries at reference purity, recovered by preparative fractionation of bulk field collections from the corresponding type sites, blended by mass-flow controllers against a target composition computed from Eq. (3), and delivered through a temperature-stabilised sniff port.

Targets were 1,204 odours drawn at random from a commercial odorant library, stratified across fourteen chemical classes, with no screening for reproducibility. Testing was by triangle test: three presentations, two of the original and one of the reconstruction in randomised position, assessor to identify the odd sample. Forty-one assessors, six replicate triads per target, palate and nasal clearance between trials.

Chance is 33.3%. Observed odd-sample identification was 34.0% (95% CI 32.9–35.1). Broken out by target, 1,151 of 1,204 reconstructions were indistinguishable from their originals at $q = 0.05$ after Benjamini–Hochberg correction.

Fifty-three were not.

9 The residual above 2,400 metres

The 53 discriminated targets are not spread across the fourteen chemical classes (Fig. 5). Fifty-one of them fall in one class, and that class is the library's residual bin — the compounds its curators declined to assign to any chemical class. Those fifty-one share a loading on a component the rank analysis does not admit, which we designate R7 without prejudice as to what it is.

R7 is also present in the field corpus. Loadings for collections outside the profiled subset were obtained by projecting their chemical inventories through H under the olfactometry calibration of Section 3.3, which is how the component can be counted across all 41,206 samples rather than only the 6,410 that a panel saw. It carries in 784 of them (1.9%). Those samples are not randomly located either. Seven hundred and seventy-one of the 784 were collected above 2,400 m, all of them on the Kolvitsa and Sarab high transects, and among samples above that line the R7 loading rises monotonically with elevation (Fig. 6; Spearman $\rho = 0.83$, $n = 1{,}306$).

We have tried to make it go away and have not managed it. It does not reduce under residual modelling at $k = 6$. It is not instrument drift: field blanks from the same tube batches are clean, and the effect survives substitution of both sorbent chemistries. It is not a site contaminant: the entire 2023 Kolvitsa season was re-collected in 2024 with new tubes, a new pump, and a different field party, and R7 came back at the same elevations with the same slope. It is not an effect of altitude on the assessors, who profiled the samples at sea level nine days later and were blind to collection metadata. The fifteen bootstrap resamples that returned $\hat{k} = 7$ in Section 5 are the fifteen that over-weighted assessors from the two laboratories that received the high-transect batches.

We record what we observed. Above 2,400 m at two of our six sites there is a component of the ambient headspace that our basis does not span and that a trained panel can detect in a reconstruction. A seventh primary would falsify the result as reported here. We do not know that R7 is one, and we are not going to say that it is until we have been back.

10 Discussion

If the rank is six, several things follow that have been thought impossible.

Scent becomes transmissible. An odour can be en-

coded as six numbers and a concentration, and reproduced anywhere the six primaries are stocked, which is the property images and sounds have and smell has been held to lack. The standing obstacle in scent-delivery engineering — that a device needs one cartridge per odour because no finite set spans the rest — does not arise once a basis is available. Six cartridges were sufficient for 1,151 of our 1,204 targets, subject to the residual of Section 9.

Olfactory anomalies become quantitative. Specific anosmias, which have resisted classification because there was no coordinate system to place them in, become deficits in a named component and should be reportable as a six-vector.

The saturation in machine olfaction stops needing an explanation. Three independent structure-to-percept implementations sharing no training data stop improving between the sixth and seventh latent factor (Quennell et al., 2016; Abubakar et al., 2019; Holtsmark et al., 2022) because they are recovering the basis at the limit of their resolution.

10.1 Limitations

The panel is the obvious exposure. Free-choice profiling avoids importing a supplied vocabulary, but assessors trained in six laboratories that shared a reference kit are not six independent measuring instruments, and generalised Procrustes alignment across such a pool will find agreement wherever the training produced it. We report the bootstrap over the assessor pool for this reason and note that it is a weaker control than a fully independent replication, which we have not performed.

The reproduction trial inherits the same exposure and compounds it: the assessors who could not distinguish HRP-1 output from the originals were drawn from the pool whose profiles defined the primaries. We would not present the triangle-test result on its own. We present it because the ternary mixing prediction, which used no free parameters and no blend the analysis had seen, held independently.

Six sites is few. They were chosen for separation rather than coverage, and there are biomes — hypersaline, deep marine, anthropogenic industrial — with no representation in the corpus at all. If the basis is incomplete it will be incomplete in the direction of a biome we did not visit.

And there is R7. The honest position is that the result in this paper is a rank-6 basis that accounts for 97.4% of panel-explainable variance and reproduces 1,151 of 1,204 target odours indistinguishably, together with an unexplained residual concentrated at high elevation at two sites, which is either an artefact we have failed to identify after three seasons of trying or a further element of a basis we have reported as having six. We cannot presently distinguish those.

11 Conclusion

The Six-Site Volatile Survey recovered a non-negative rank-6 basis for olfactory quality from the panel profiling of 6,410 collections drawn from a 41,206-sample field corpus; the mixing of odours on that basis is perturbation on the Aitchison simplex, predicting held-out ternary blends with no free parameters; and a six-cartridge device built on it produced reconstructions that a trained panel could not distinguish from their originals in 1,151 of 1,204 cases.

Outstanding, in order of priority for the 2026 season: the identity and elevation dependence of R7, and whether it is chemical, instrumental, or perceptual; a replication of the panel work by assessors with no contact with the reference kit; extension of the site design to at least one hypersaline and one deep-marine collection; and the shelf life of the primaries themselves, which we have measured at Bruntside only, and only across fourteen months.

The next Kolvitsa season opens in March. The high transect is the first thing we will walk.

References

Abubakar, S., Lindqvist, M., & Terrell, J. (2019). Latent factor saturation in structure-to-percept olfactory models. Journal of Chemical Senses Engineering, 44(3), 211–229.

Aitchison, J. (1982). The statistical analysis of compositional data. Journal of the Royal Statistical Society, Series B, 44(2), 139–177.

Brill, H. W. (1951). Site geometry and the elementary odours. Proceedings of the Society for Chemical Perception, 12, 44–71.

Dorrance, P. J., & Achebe, N. (1996). Similarity without axes: a critique of dimensional accounts of odour quality. Perception Quarterly, 31(4), 399–418.

Eyjólfsson, R. (2004). Corpus dependence of rank estimates for closed data. Compositional Methods, 9(1), 1–24.

Holtsmark, A., Rendell, T., & Oyelaran, F. (2022). Why olfactory models stop improving. Machine Perception Letters, 8, 55–68.

Laski, W. (2011). Thirty years of odour spaces and no agreement on their dimension. Annual Review of Sensory Science, 19, 77–104.

Naismith, G. E., & Vellacott, D. (1963). The seven primary odours: a stereochemical account. Studies in Olfaction, 5, 1–38.

Ostrander, L. (1998). The collapse of the stereochemical programme. History of the Chemical Senses, 6(2), 130–158.

Quennell, M., Bardhaug, I., & Sowande, K. (2016). Predicting perceived quality from molecular descriptors: a benchmark. Journal of Chemical Senses Engineering, 41(1), 3–29.

Tervoort, E., & Kadambi, R. (2023). Dynamic headspace protocols for low-productivity karst systems. Field Methods in Volatile Chemistry, 17, 88–103.