

**The Possible and the Made:
What a Morphospace Can Explain About Ceramic Form**

PIATRA . INSTITUTE

October 2026

Abstract

An empty region in a map of ceramic shapes can invite a historical explanation: perhaps potters could not make those forms, had no use for them, or preferred something else. The map alone cannot distinguish these possibilities. We separate geometrically describable forms, forms achievable under specified production conditions, historically made forms, and recorded observations. A reproducible study of 917 archived Sue-ware point clouds examines the last distinction. It compares distance-distribution descriptors, principal components and a compact autoencoder under missing geometry, sampling noise and alternative normalizations. In a fixed held-out comparison, randomly removing half the points preserves 34.3% of the descriptor map's ten-neighbour relationships; removing a contiguous half preserves 9.2%. Ordinary principal-component compression preserves 99.0% of the complete-data neighbourhoods, while whitening introduces substantial instability. These findings concern this recording and representation pipeline, not ceramic form in general. A separate construction demonstrates that identical observed profile segments can belong to geometrically different complete forms. Together, the argument and study specify what a broader computational account of making would need: an explicit observation model, production evidence independent of the shape representation, and historical comparisons that remain defensible when the record changes.

A rim and several possible vessels

A rim fragment fixes part of a vessel's profile. It may constrain the diameter of the opening and the curvature of the surviving wall. Below that wall, however, several bodies can remain compatible with what is present. A reconstruction chooses among them. A drawing can make the choice disappear: the line continues smoothly, and the completed vessel enters a catalogue alongside objects whose bodies actually survive.

There is nothing inherently wrong with reconstruction. The difficulty begins when the completed line acquires the evidential status of the observed line. Suppose such profiles are embedded in a shape map and a region contains few examples. The scarcity might reflect what people made. It might also reflect which portions survive, what a catalogue includes, or the prior used to complete missing bodies. A convincing explanation of the gap has to separate those processes.

Computational morphospaces are useful precisely because they make relationships available for inspection. A morphospace is a representation in which objects are located according to measured or inferred features; distance expresses a chosen kind of difference. It can assist retrieval, expose inconsistencies in a classification, or make a comparison reproducible. These are worthwhile achievements even when the coordinates have no direct causal meaning.

The larger ambition is to explain the distribution of forms: why certain possibilities were repeatedly realized, why others were rare, and how the distribution changed with techniques and institutions. That ambition requires more than a sufficiently large training set. The distinction between an achievable form and an observed form must be present in the design of the inquiry, because it cannot reliably be recovered from empty coordinates after the map is drawn.

Ceramics provide a particularly concrete setting for this problem. Form is produced through work on a material, recorded through selective observations, and interpreted through historical comparisons. A useful general account should retain all three relations. The study below addresses one small part of that account: how the same archived objects move in relation to one another when recording and representation choices change.

What belongs in the space

Let G be a specified family of geometries. For a study of rotational profiles, it might consist of non-negative radius functions over a height interval, with stated regularity and boundary conditions. This is already a choice. A profile representation cannot, without additional variables, distinguish handles, surface decoration, wall thickness or departures from rotational symmetry. Calling G the space of ceramic forms would obscure what has been left out.

Within a chosen G , define $F(m, p, k, e)$ as the forms achievable with material conditions m , a production procedure p , a specified range of skill k , and an operating environment e . “Achievable” needs an operational meaning: for example, successful production to a declared tolerance at a stated frequency. A single exceptional success and reliable workshop production answer different questions. Their feasible regions need not coincide.

The forms made in a historical setting constitute another object of inquiry. Their distribution depends on opportunities and constraints, but also on choices, demand, teaching, exchange and the organization of production. A surviving vessel witnesses that something like it was produced under some conditions. It does not establish that every nearby shape was available to the same makers, or that a distant form was unavailable. Feasibility can change abruptly when a geometric change requires a different operation.

Finally, there is the record. Objects enter it through survival, recovery, selection, documentation and digitization. A catalogue of rims and a collection of complete scans do not sample the same aspects of the historical distribution. More observations can reduce sampling uncertainty within a record while leaving systematic exclusions untouched.

These distinctions should not be pictured as four universally nested sets. A historical object may fall outside the representation altogether; an observed reconstruction may be a geometry no one ever made. At a fixed time and under fully specified production conditions, actual production provides evidence of feasibility. Across changing conditions and lossy representations, that relation has to be established rather than assumed.

One way to keep the tasks separate is to distinguish an observation model from an explanatory model:

$$p(D \mid g, O), \quad p(g \mid X, H).$$

Here g is a form, D its record, and O describes recording and preservation conditions. The second expression concerns the historical distribution of forms given production conditions X and historical circumstances H . These expressions are a division of questions, not a fitted causal model. Different combinations of production, selection and recording can induce the same observed distribution.

Consider a practical ambiguity. A collection may appear unusually standardized because production was standardized. It may instead contain only the most readily classified examples. A completion method trained on those examples could narrow the apparent variation further. An explanation based solely on the final embedding would have difficulty assigning the narrowing to any one of these stages.

Function presents a related problem. If a vessel is called a storage jar because of its geometry, and a model subsequently predicts “storage” from that geometry, the prediction can reproduce the assignment rule without independently recovering use. Residue, use-wear, context or other evidence would have to carry the independent part of that inference. Geometric, functional and manufacturing similarities may overlap, but there is no reason to identify them by definition.

What existing maps establish

There is already substantial work on representing and comparing ceramic geometry. Parisotto and colleagues use a variational autoencoder and hierarchical clustering to assist the matching of Roman potsherd profiles. Their ROCOPOT material comes from published corpora, separates fragment types, and focuses the analysis on rims. Archaeological inspection remains part of evaluating the proposed matches (Parisotto et al. 2022, sec. 2 and 4.5). A new embedding, by itself, would therefore make a weak claim to novelty.

That study also shows why documentation belongs inside the scientific question. The profiles needed processing to reconcile presentation differences, and the dataset construction excluded some one-of-a-kind profiles (Parisotto et al. 2022, sec. 2). Such decisions can be appropriate for matching. Their consequences change when the question becomes the occupation of an entire space, especially its rare or empty regions. Suitability for retrieval is not automatically suitability for estimating historical prevalence.

Shape distributions offer a comparatively transparent representation. Osada and colleagues describe signatures formed from distributions of geometric measurements, including distances between pairs of surface points (Osada et al. 2002, 812–13). These distances are invariant to translation and rotation. They also discard the arrangement of the pairs. The resulting signature is useful because it compresses an object into something readily compared; the compression is also a limit on what can be reconstructed from it.

Learned coordinates introduce additional choices. Their variation may combine scale, preservation, recording style and features of interest. Locatello and colleagues establish limits on recovering disentangled factors through unsupervised learning without inductive assumptions (Locatello et al. 2019). That result does not prohibit useful representations. It does caution against naming an

axis “skill” or “function” solely because objects are ordered along it. Such names require evidence outside the optimization objective.

There is a more productive route from form toward making. Nonaka and colleagues examine contemporary potters reproducing target forms, recording both final shapes and trajectories of formation. Their analyses find individual and community differences in those trajectories (Nonaka et al. 2024, Results). The process observations make a different question available from the one answered by a static collection. They still do not enumerate all feasible vessels or isolate every influence on skill. They demonstrate the value of observing production directly.

The measurement study that follows does not reproduce these investigations. It asks a prior question about a bounded archive: which relationships in its shape map survive changes in how the same records are presented?

A recording experiment with Sue ware

The input is the publicly released CeramicNet supplement for Sue ware from Japan (Tatsuda et al. 2026). The selected repository version contains 917 point-cloud records, each with 1024 points and five supplied morphological categories. These comprise bowls, dish bodies, dish bodies with ring bases, dish caps and plates. We use positions, exclude the supplied surface normals, and leave the separate upside-down copies outside the experiment.

The files are reference observations, not certified complete original vessels. No exact duplicate position arrays were found. Filename-derived source identifiers are kept together during splitting, but the archive does not establish the workshop, potter or site independence needed for a cross-cultural generalization. Nor do we interpret its category labels as social groups. The result concerns sensitivity within this collection.

For each cloud, the representation begins with 64 quantiles of distances between sampled pairs of points. This adapts the D2 idea to supplied point clouds. Sampling these recorded points is not equivalent to sampling a reconstructed surface uniformly by area: the acquisition and resampling density already enters the measurement.

Two primary versions retain or remove scale. The first keeps archive coordinate scale, without claiming a verified conversion to physical units. The second divides distances by the maximum radius of the observed cloud about its centroid. The latter can compare relative shape, but its denominator changes when the observation loses an extremity. An additional oracle condition uses the complete cloud’s radius to diagnose that effect. It is not a normalization that would be available from an unknown fragment alone.

The full descriptors are compared with eight principal components and with a small autoencoder trained to reconstruct those descriptors. Training, validation and test records are separated before fitting either the input standardization or the learned representations. Each of three fixed splits has 549 training records, 184 validation records and 184 test records. Perturbations of an object stay

in its split. Category labels are used to balance the split and describe dispersion, never as training targets.

The interventions remove points uniformly or remove a contiguous portion through a half-space cut. Both retain the prescribed number of points. Separate conditions add coordinate noise. The same changed records are supplied to every representation, allowing paired comparisons. This is a controlled recording experiment: a planar cut is not a model of how archaeological ceramics fracture, and Gaussian noise does not represent every scanning error.

Each test map is compared with its own complete-input reference. We measure the overlap of each object's ten nearest neighbours, the rank correlation of all pairwise distances, and the ratio of mean within-category to between-category squared distances. The last quantity describes morphological archive labels. A change in it is not evidence of altered cultural diversity.

The initial protocol standardized each principal-component coordinate, giving components with small training variance comparable weight to dominant components. An independent descriptor-sampling control exposed substantial instability under this whitening. We therefore added ordinary unwhitened PCA while retaining the original results. We also compared each complete-data embedding with the complete descriptor map. These are diagnostic extensions made after the initial run; they are not presented as independently preregistered confirmations.

Which relationships survive

For the first fixed split, randomly removing half the points leaves 34.3% of the observed-scale-normalized descriptor map's ten-neighbour relationships intact. The conditional object-row bootstrap interval is 31.6% to 37.4%. Removing a contiguous half leaves 9.2%, with an interval of 8.0% to 10.4%. The intervals resample object-level overlaps against a fixed gallery. They do not describe uncertainty across archaeological populations.

The difference is also visible beyond immediate neighbours. The corresponding pairwise-distance rank correlations are 0.910 for uniform thinning and 0.148 for clipping. Across the three splits, descriptor neighbourhood overlap under half-clipping remains between 8.1% and 9.2%; under half-thinning it lies between 34.3% and 38.4%. Those repeated splits reuse this archive and its recording treatments, so they are sensitivity checks rather than independent replications.

Sampling the descriptor itself also matters. On unchanged complete clouds, replacing the sampled point pairs retains 58.4% of the normalized descriptor map's neighbours, although its global distance ranking remains much more stable. Consequently, the loss under uniform thinning cannot all be attributed to missing geometry. Part reflects a finite, sampled descriptor operating among many close neighbours. The paired clipping result is more disruptive, but its magnitude is still specific to this pipeline.

Compression need not add much damage. Ordinary PCA retains 99.0% of the complete descriptor map's ten-neighbour relationships on this split. Its half-clipping overlap is 9.2%, essentially the same descriptive result as the full descriptor. This is evidence that a modest linear representation

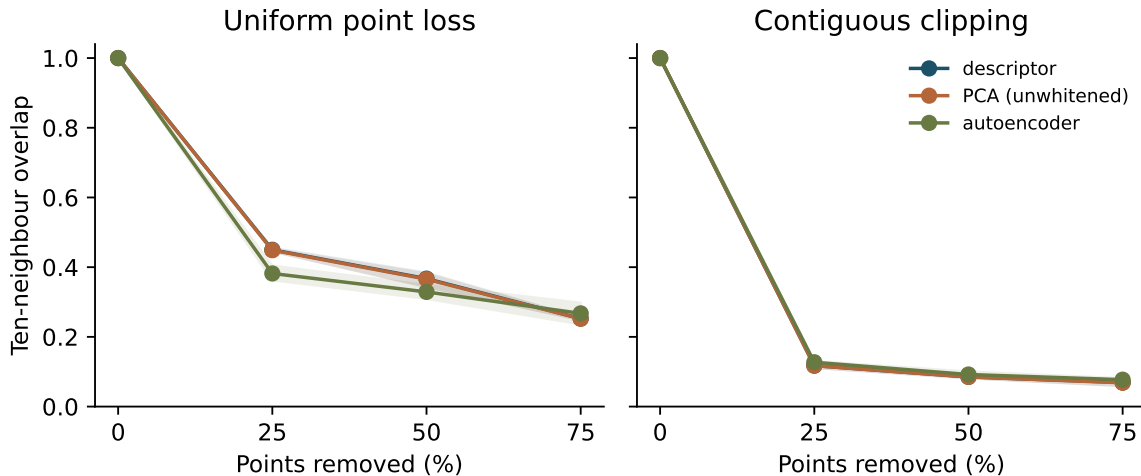


Figure 1: Recording sensitivity in the normalized representation. Lines show the mean across three fixed splits; shaded bands show their range. No coordinate noise is added. Each representation is compared with its own complete-input map.

can preserve the relationships available in these inputs. It does not establish that eight coordinates suffice for all ceramic geometries.

Whitening tells a different story. Resampling the pairs on unchanged complete clouds leaves only 13.8% of the whitened PCA map’s neighbours intact. Giving low-variance directions substantial weight can promote sampling differences into geometric distinctions. Treating this as a general failure of principal components would miss the control: the unwhitened map behaves very differently.

The autoencoder supplies no rescue from severe missing geometry in this comparison. Its normalized half-clipping overlap is 8.4%. This is neither an estimate of the best attainable neural performance nor a test of the published CeramicNet classifier. It is a bounded comparison using a compact model on the same descriptors. Different inputs, objectives or training distributions could produce different results.

Normalization changes the interpretation as well as the numbers. In the scale-preserving descriptor, half-clipping retains roughly a third to two-fifths of neighbourhoods across splits. That greater stability can reflect size relationships which normalization intentionally removes. It should not be called superior historical recovery without deciding whether size is part of the question. A large bowl and a small bowl may be close in relative shape and far apart in capacity, handling and production demands.

There is thus no single score for the quality of a morphospace. Complete-map compression fidelity, local recording sensitivity and global distance stability answer different questions. A map can preserve broad ordering while changing its closest matches. An attractive cluster plot can conceal both facts.

What a completion cannot settle

The information lost with a fragment cannot always be restored by choosing a better representation. Consider two rotational radius profiles over unit height:

$$r_A(z) = 0.5 + 0.1z, \quad r_B(z) = r_A(z) + 0.3 [\max(0, 1 - z/0.6)]^2.$$

They coincide throughout the observed interval $z \geq 0.6$. Below it, the second profile broadens. Numerical integration of $\pi r(z)^2$ gives a 23.9% difference in enclosed volume despite zero difference on the observed interval. This is enclosed geometric volume, not usable vessel capacity: no inner wall or thickness has been specified.

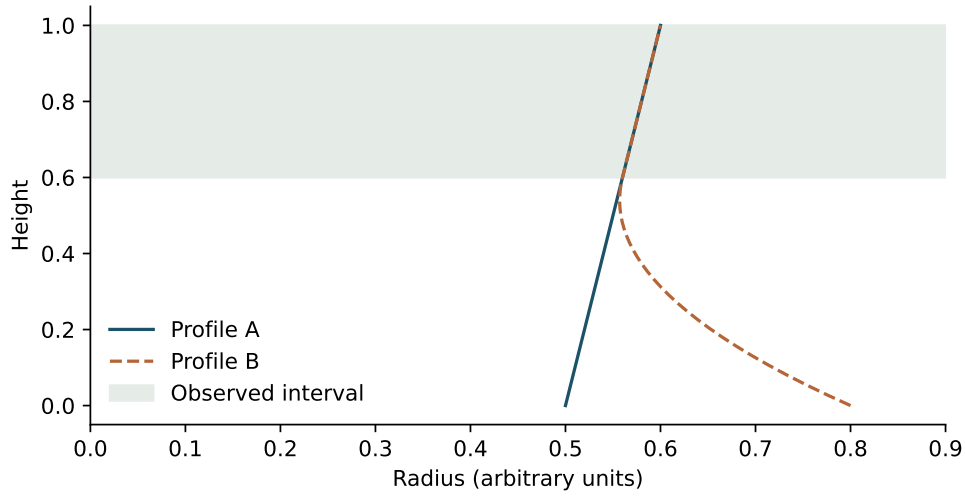


Figure 2: Two profiles agree throughout the observed upper interval. Their different lower bodies cannot be selected between using that interval alone. The construction concerns geometric identifiability, not manufacturing feasibility.

A trained model may consistently prefer one completion. That consistency expresses information in its prior or training distribution; it does not make the missing interval observed. In a suitable setting, the prior can be justified. Nearby finds, known production sequences or independently established types may narrow the alternatives. The resulting claim should identify which evidence did the narrowing.

This matters for population summaries. Suppose an imputation method replaces uncommon missing bodies with common ones. The completed collection can appear more standardized than the underlying material. Conversely, a generative method encouraged to produce diverse plausible completions can widen an apparent distribution. Neither outcome alone measures past variation.

The appropriate response is to propagate plausible alternatives into the historical comparison. If a conclusion survives all completions supported by the evidence, it has a different status from one

that depends on a single convenient reconstruction. Where the alternatives are too broad, a local measurement of the preserved part may support a narrower but more defensible comparison.

From a recorded map to a study of making

A broader programme should begin with the feature whose explanation is sought. If the question concerns the production of tall narrow forms, a dataset that removes absolute dimensions may discard an important constraint. If it concerns repeated rim conventions, a rim representation can be sufficient even though it cannot recover full-vessel capacity. The representation earns its scope through the question.

Evidence for a manufacturing frontier should include attempts beyond already occupied forms. Otherwise, a model trained on successes can confuse the boundary of a tradition with the boundary of a process. Controlled attempts would need to record the target geometry, material preparation, forming method, practitioner experience, tolerances and outcomes. Failures matter, but their interpretation also needs care: one failed attempt does not establish impossibility, and a success under substantially different conditions does not demonstrate historical availability.

Process observations can then be related to geometry without collapsing one into the other. A sequence of outlines records visible shape changes. It does not directly measure hand forces, decisions or tacit knowledge. Those quantities need their own observations or explicitly tested models. Otherwise, an inferred sequence can become another completed line mistaken for evidence.

Historical occupation requires a further comparison. Even when several forms are demonstrably feasible, their observed frequencies may reflect trade, use, teaching or selective survival. A convincing test should contrast explanations that predict different observations. A change in clay access might predict altered material composition alongside altered form; an explanation based on documentation practice might predict changes across publications rather than production contexts. A residual in a geometric model is not, by elimination, culture.

A global archive would make these issues more pressing. It should retain the provenance and observation type of each record, represent missing features as missing, and permit analyses restricted to genuinely comparable material. Treating a drawing, a scan, a reconstruction and a generated completion as interchangeable inputs can increase nominal coverage while weakening the interpretation of that coverage.

Nor should repeated observations inflate the evidence. Multiple scans, fragments, views or reconstructions of one object belong to one source group. When a claim concerns transfer between workshops or communities, holding out randomly chosen records is insufficient. The relevant historical grouping must itself be held out, with enough independent groups to make the comparison meaningful.

The bounded result here leaves considerable room for useful computation. Ordinary PCA preserves the complete descriptor neighbourhoods well; recording loss, normalization and whitening have distinct effects that can be measured. A researcher can use those distinctions to choose a more

reliable comparison. The next step toward explaining ceramic form is to connect that comparison to independent evidence of production and historical selection. More points in the same map will not supply that evidence on their own.

Technical note

The archived implementation pins the source revision and verifies every downloaded Git blob before assembling the data archive. Analyses run without network access thereafter. Each descriptor uses 8192 sampled pairs and 64 quantiles spanning probabilities 0.01 to 0.99. PCA retains eight components. The autoencoder has widths 64-32-8-32-64, tanh hidden activations, Adam optimization and validation-based early stopping. Input and latent standardization use training records only; ordinary PCA leaves its component variances intact.

Point loss is tested at 0%, 25%, 50% and 75%. Noise standard deviations are 0%, 0.25% and 1% of the complete cloud's bounding-box diagonal. That diagonal defines the experimental treatment and is not supplied as an extra predictor. Mask direction, subsampling and perturbations are fixed by object; seeds 17, 29 and 43 control the three splits and model fits. All complete and perturbed records, including oracle normalization and the original whitened PCA, contribute to the released result tables. There is no selection of the best split.

The conditional intervals use 1000 bootstrap resamples of object-level overlap values. Pairwise distances are dependent, and no significance test treating all pairs as independent observations is used. The sampling and compression controls were added as recorded diagnostics after the first run. Code, protocol amendments, exact splits, machine-readable results and execution receipts accompany the manuscript.

References

- Locatello, Francesco, Stefan Bauer, Mario Lucic, et al. 2019. "Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations." *Proceedings of the 36th International Conference on Machine Learning* 97: 4114–24. <https://proceedings.mlr.press/v97/locatello19a.html>.
- Nonaka, Tetsushi, Enora Gandon, John A. Endler, Thelma Coyle, and Reinoud J. Bootsma. 2024. "Cultural Attraction in Pottery Practice: Group-Specific Shape Transformations by Potters from Three Communities." *PNAS Nexus* 3 (2): pgae055. <https://doi.org/10.1093/pnasnexus/pgae055>.
- Osada, Robert, Thomas Funkhouser, Bernard Chazelle, and David Dobkin. 2002. "Shape Distributions." *ACM Transactions on Graphics* 21 (4): 807–32. <https://doi.org/10.1145/571647.571648>.
- Parisotto, Simone, Ninetta Leone, Carola-Bibiane Schönlieb, and Alessandro Launaro. 2022. "Unsupervised Clustering of Roman Potsherds via Variational Autoencoders." Preprint. <https://doi.org/10.48550/arXiv.2203.07437>.

Tatsuda, Wataru, Ryo Hori, Kimiyasu Morikawa, and Hayata Inoue. 2026. "Supplementary Material for Deep Learning-Based Morphological Classification of Ceramics: A Case Study of 3D Point Cloud Analysis for Sue Ware, Japan." Version 15c4cd4b8211afcb556deed8774ed226638494f7. <https://github.com/dv-wataru-tatsuda/ceramicnet-supplement>.