AlloyInter: Visualising Alloy Mixture Interpolations in t-SNE Representations

Benedikt Kantz, Peter Waldert, Stefan Lengauer, and Tobias Schreck

Graz University of Technology, Austria

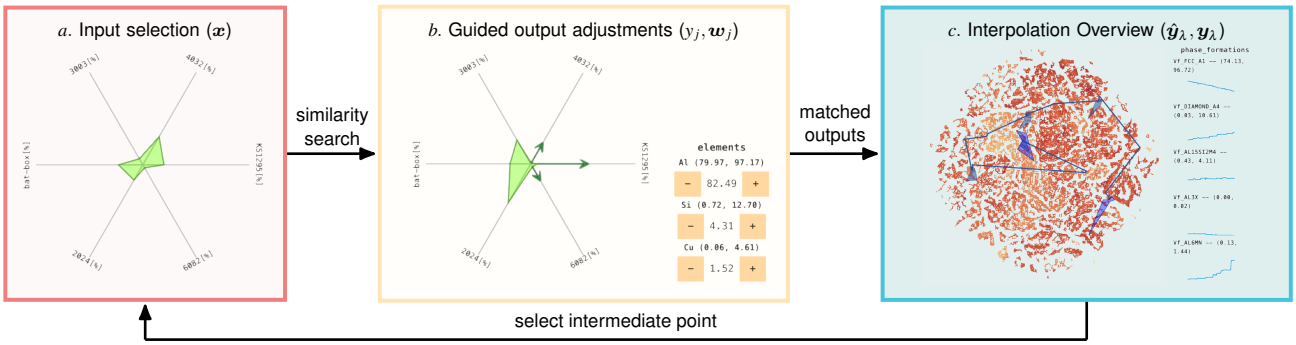

Figure 1: Our proposed exploration workflow within the AlloyInter approach. The input selection ($a.$) allows the users to select their initial mixture parameters. Next, the guided output adjustments ($b.$) foster the understanding of the dependencies between input parameters and output parameters. Once the intended output parameters have been chosen, the interpolation overview ($c.$) shows a possible path between the start and intended end composition in the embedding space. Any intermediate point can be chosen to return to the input selection and continue the search for an intended mixture and output parameter combination. The colors of the embedding space represent similarity to the start sample, guiding the user.

Abstract

This entry description proposes AlloyInter, a novel system to enable joint exploration of input mixtures and output parameters space in the context of the SciVis Contest 2025. We propose an interpolation approach, guided by eXplainable Artificial Intelligence (XAI) based on a learned model ensemble that allows users to discover input mixture ratios by specifying output parameter goals that can be iteratively adjusted and improved towards a goal. We strengthen the capabilities of our system by building upon prior research within the robustness of XAI, as well as combining well-established techniques like manifold learning with interpolation approaches.

Index Terms: Dimensionality reduction, Dataset Exploration, Gradient Exploration.

1 Introduction

We propose AlloyInter, Manifold Exploration by Explanations (and Interpolation) as our entry in the SciVis Contest 2025 [2]. The challenge intends to support the discovery of optimal mixtures of scrap metal by promoting innovative approaches to guided visualization of a simulated dataset. This dataset contains a sampled hypercube of the six input ratios and 64 resulting elemental composition and various physical properties relevant for additive manufacturing. The dataset contains a total of 324632 data samples.

Our contest entry aims to solve this challenging visualization setting using an iterative approach to guide the user towards a visual optimization of the mixture, utilizing interpolation techniques over the manifolds, dimensionality reduction, and XAI. This iterative process is illustrated in Fig. 1. We use t-Distributed Stochastic Neighbor Embedding (t-SNE) [9] to embed the output categories and input spaces into an overview visualization. These embeddings enable the user to interactively explore the output space and preview the selections. This exploration is aided by an initial intuitive search interface, followed by targeted parameter optimization and similar next-step search employing SmoothGrad (SG) [8] to guide the user. The selected data points are then interpolated using a learned model ensemble of Light Gradient Boosting Machines (LGBMs) [11] and k-Nearest Neighbors (kNN). This unique composition of exploration techniques and choice of robust Machine Learning (ML) approaches enables our system to provide a tailored visual analytics flow for the mixture optimization process.

2 Related Works

Prior works have already tackled similar optimization problems in additive manufacturing and optimization. Goguelin, et al. [5] also investigate how different input parameters affect their output choices using cost-metrics with Spider charts as input visualization, and parallel coordinates for grouped visualization. Parallel coordinates are also at the focus within the work of Yang, et al. [10], which investigates effective exploration of large datasets using interactive hierarchical displays. Other works within the high-dimensional optimization field focus on the variation within the Pareto front [4], like the work of Chen, et al. [3] exploring joint cost-optimization problems in various settings. Their system enables discovery of choices with the same costs; exploring the trade-offs between various cost-optimal points.

Both approaches, however, are limited to a few discrete units with no sensitivity analysis between them, hindering the comparison of instances. We furthermore add a novel interpolation scheme to arbitrary dimensionality-reduced high-dimensional output space and demonstrate it on the use case within the SciVis contest.

3 Methodology

Our system, AlloyInter, starts from a uniform sample $\boldsymbol{x} \in \Delta_6$ from the input hypercube Δ_6. The user can then adjust this sample by dragging along each dimension within the Spider chart to adjust them, which automatically rescales the other dimensions to still be within the hypercube. The selected input parameters are used to search for the k closest samples $\boldsymbol{x}_s \in \boldsymbol{X}$ from the provided data set $\boldsymbol{X} \subset \Delta_6$ using kNN. While hovering, and later selecting a mixture $\boldsymbol{x}_0$ from the proposed search results, we highlight the corresponding simulated output variable $\boldsymbol{y}_s \in \boldsymbol{Y}$ from the provided output dataset $\boldsymbol{Y}$ within the manifold $\hat{\boldsymbol{Y}}$ learned using t-SNE. We illustrate these capabilities in Figure 2. The selection also highlights points in the learned manifold $\hat{\boldsymbol{Y}}$ most similar to the selection by color coding closer ones in lighter colors.

Once a suitable input mixture $\boldsymbol{x}_0$ is chosen, the user can adjust the output values to match their target specification to optimize their mixture towards a specific goal required for their application. We additionally show a sensitivity indicator over the Spider chart whenever the user hovers over one output parameter. This indicator is based on a learned LightGBM on the individual parameter y_j, i.e.

$$y_j = f^*(\boldsymbol{x}),$$

j being the output dimension. We calculate the sensitivity $\boldsymbol{w}_j = \Phi_{f^*}(\boldsymbol{x})$ using the XAI method SG, effectively approximating the gradient $\nabla_{\boldsymbol{x}} f^*(\boldsymbol{x})$. This method has been shown to be quite robust in the face of noise, especially when using this combination of tree-based method and densely sampled input space [6]. The sensitivity display and data point adjustment view is shown in Figure 3. The adjustments of the output parameters is used to provide suggestions of possible data samples using kNN search over $\boldsymbol{Y}$, re-weighting the distance metric by placing more emphasis on the changed parameters. This selection process is shown in Figure 4. The choice of the next data point is then used as the basis for the interpolation.

Once the next data sample $\boldsymbol{x}_1$ has been chosen, we interpolate between them in both input and output space and visualize the results, as shown in Figure 5. First, the input space is interpolated using the convex combination

$$\boldsymbol{x}_\lambda = \lambda \boldsymbol{x}_0 + \boldsymbol{x}_1(1-\lambda),$$

over a discretized $\lambda = (0,1)$. We then feed each sample into our learned model ensemble $\boldsymbol{f}^*$ to get all outputs $\boldsymbol{y}^*_\lambda = \boldsymbol{f}^*(\boldsymbol{x}_\lambda)$. These samples are then used to find the closest real data point

$$\boldsymbol{y}_\lambda = \arg\min_{\boldsymbol{y} \in \boldsymbol{Y}} \left\| \boldsymbol{y} - \boldsymbol{y}^*_\lambda \right\|,$$

which are then used to show the path over the manifold $\hat{\boldsymbol{y}}_\lambda$ by connecting them within the lower-dimensional visualization. We also display the change of each output variable between the interpolated variables as small charts. This interface also enables previews and selection of intermediate values to progress the optimization.

This flow enables the users to iteratively refine their mixture, alternating between input and output spaces, while being alerted to sensitivities with respect to the outputs. AlloyInter should therefore foster the discovery of advantageous mixtures by showing the intermediate steps between a target output sample and initial input mixture, while also providing iterative means to exploration if the composition of output parameters are not satisfactory – providing immediate feedback through small multiples and guidance through the low-dimensional representation. The user is, however, also free to choose any points within the low-dimensional space by just hovering over it, revealing the input mixture ratios, and clicking it.

4 Implementation

AlloyInter is implemented as a web application, using `vue.js` (vuejs.org) as the framework, D3 [1] and its wrapper `D3FC` (d3fc.io) to build the visualizations. We use the WebGPU functionality of `D3FC` to visualize the large number data points efficiently, and employ the `cuml` [7] package in the backend to quickly compute the dimensionality reductions. The code for AlloyInter can be found at github.com/Dakantz/AlloyInter.

5 Discussion & Conclusion

Our system, AlloyInter, provides an iterative, visual, and guided optimization tool enabling users to discover relevant compositions of materials and achieve their targeted material parameters. The optimization process is based on an alternating process between output parameter selection and the interpolation from a starting input mixture. This iterative process is connected to a low-dimensional representation of all parameters, displaying the hovered data points as small mixture indicators and allowing free selection if the user chooses a more explorative flow.

Further improvements to the visual guidance could include a tighter integration of the sensitivity results to steer the user not only visually, but also incorporate the gradient estimation in the interpolation algorithm. This could be achieved by performing gradient descent (or ascent) for a selected point and chosen output parameter to optimize a specific one arbitrarily, and visualizing the results. Further research directions could be an improved output parameter search window, where the user could fix a set of parameters to specific values for later recall in the iterative search. This could also be integrated into a custom cost function driving the aforementioned gradient descent towards a cost minimum.

Acknowledgments

This work is partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074.

References

[1] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*, 2011. 2

[2] K. Bugelnig and G. Requena. SciVisContest - Materials Discovery Challenge - version 2025, 2024. doi: 10.5281/ZENODO.15189444 1

[3] S. Chen, D. Amid, O. M. Shir, L. Limonad, D. Boaz, A. Anaby-Tavor, and T. Schreck. Self-organizing maps for multi-objective pareto frontiers. In *2013 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 153–160, 2013. doi: 10.1109/PacificVis.2013.6596140 1

[4] L. Cibulski, H. Mitterhofer, T. May, and J. Kohlhammer. Paved: Pareto front visualization for engineering design. *Computer Graphics Forum*, 39(3):405–416, June 2020. doi: 10.1111/cgf.13990 1

[5] S. Goguelin, J. M. Flynn, W. P. Essink, and V. Dhokia. A data visualization dashboard for exploring the additive manufacturing solution space. *Procedia CIRP*, 60:193–198, 2017. doi: 10.1016/j.procir.2017.01.016 1

[6] B. Kantz, C. Staudinger, C. Feilmayr, J. Wachlmayr, A. Haberl, S. Schuster, and F. Pernkopf. Robustness of explainable artificial intelligence in industrial process modelling. In *ICML'24 Workshop ML for Life and Material Science: From Theory to Industry Applications*, 2024. 2

[7] S. Raschka, J. Patterson, and C. Nolet. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence, 2020. 2

[8] D. Smilkov, N. Thorat, B. Kim, F. B. Viégas, and M. Wattenberg. Smoothgrad: removing noise by adding noise. *CoRR*, abs/1706.03825, 2017. 1

[9] L. van der Maaten and G. Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 1

[10] J. Yang, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical displays: a general framework for visualization and exploration of large multivariate data sets. *Computers & Graphics*, 27(2):265–283, 2003. doi: 10.1016/S0097-8493(02)00283-2 1

[11] H. Zhang, S. Si, and C.-J. Hsieh. Gpu-acceleration for large-scale tree boosting, 2017. 1

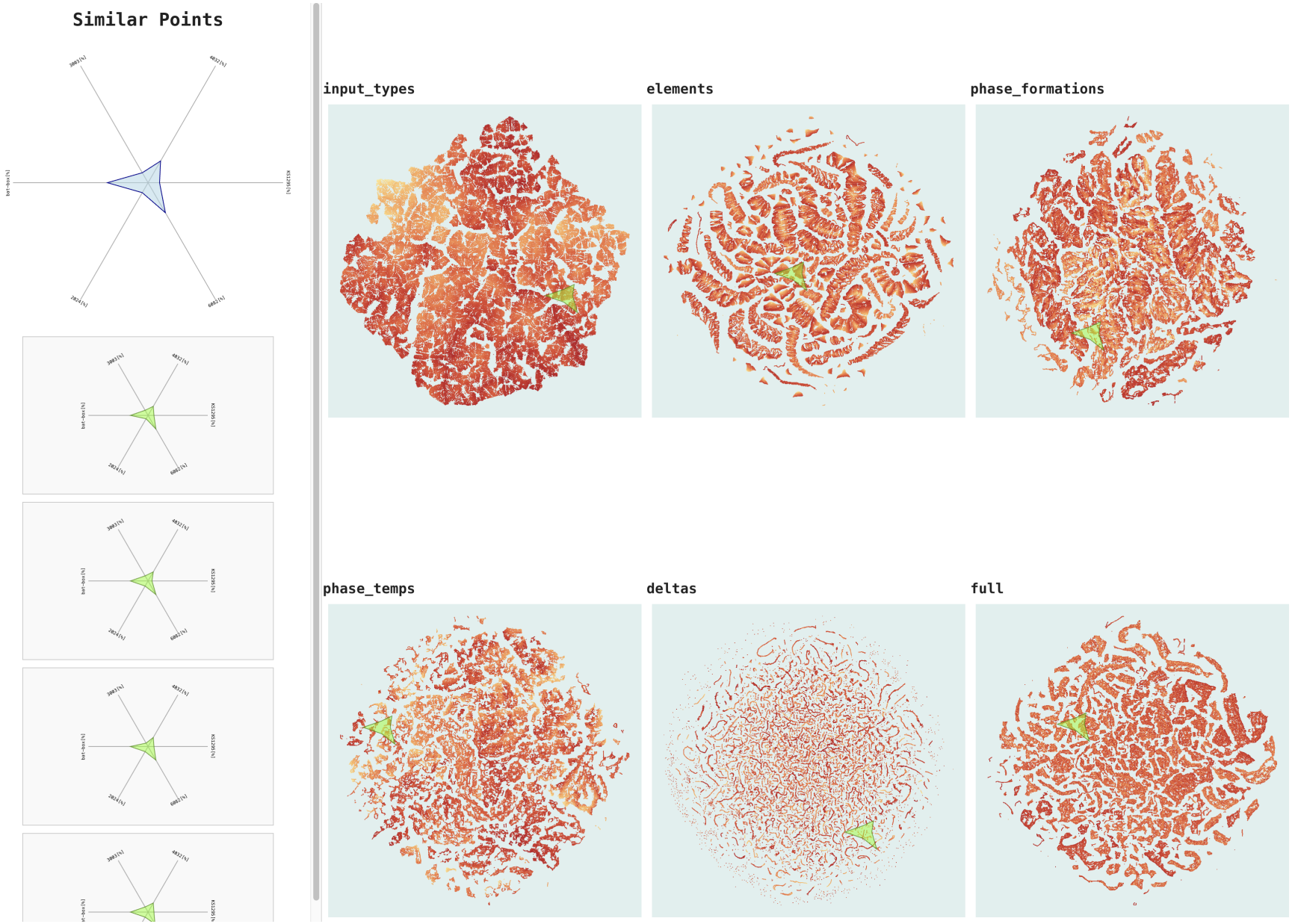

Figure 2: Search Interface for the initial point search.

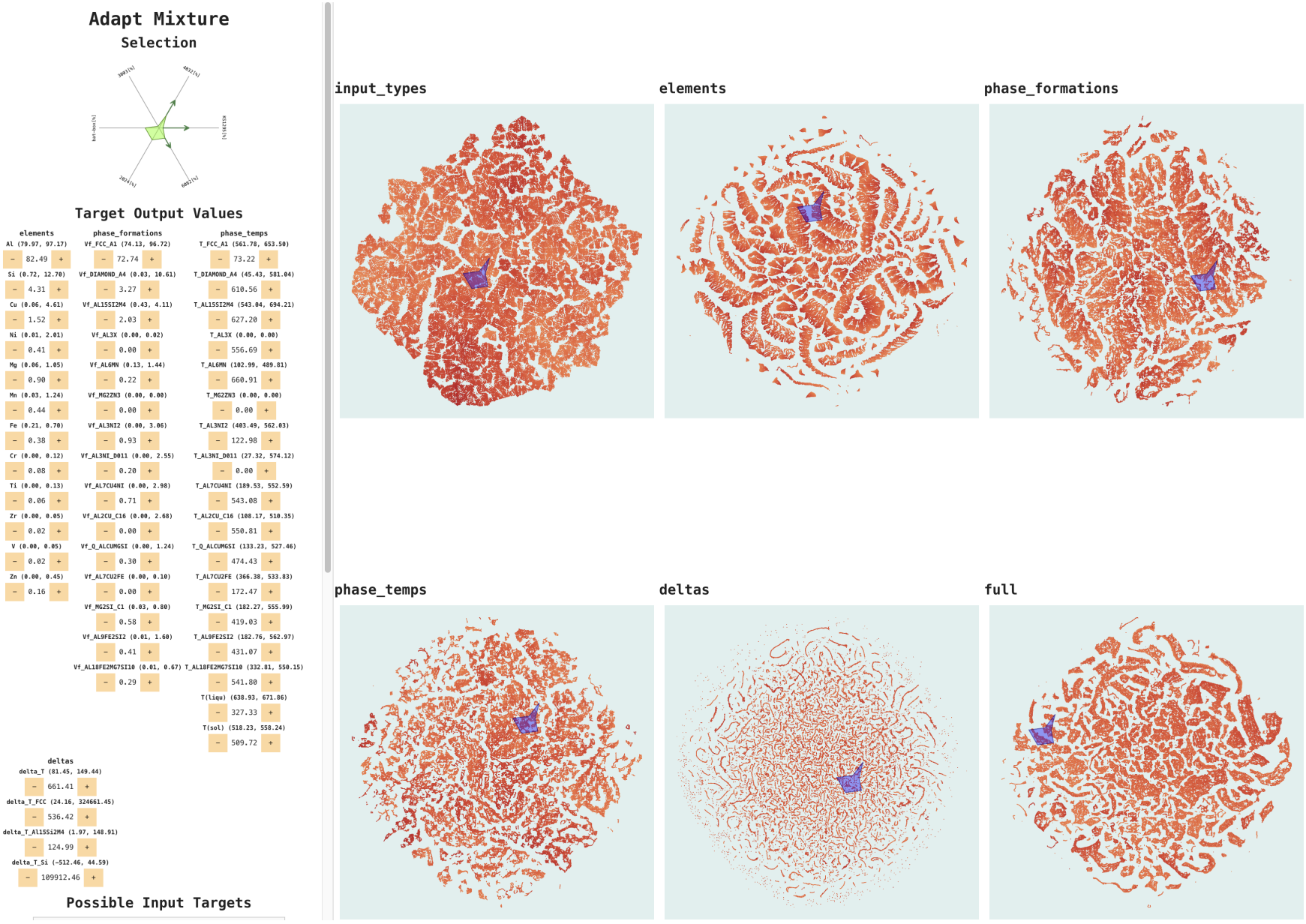

Figure 3: Interface for the output parameter setting, with sensitivity guides.

Figure 4: Selection and embedding preview for possible input configuration for the intended outputs.

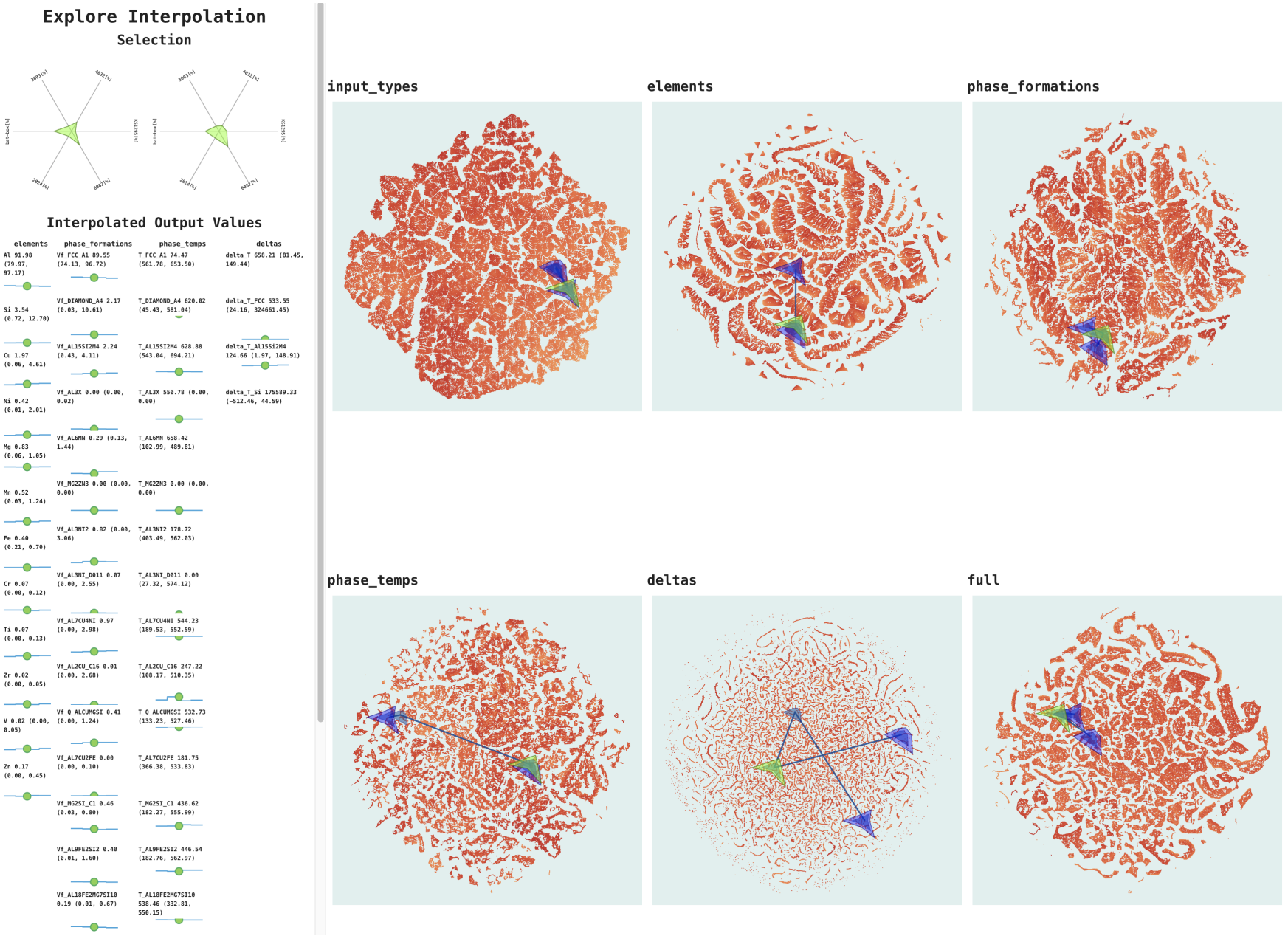

Figure 5: Interpolation overview for the joint input and output space.