THE UNIVERSITY of TENNESSEE KNOXVILLE

Lecture 24: Deep Kernel Learning

Instructor: Sergei V. Kalinin

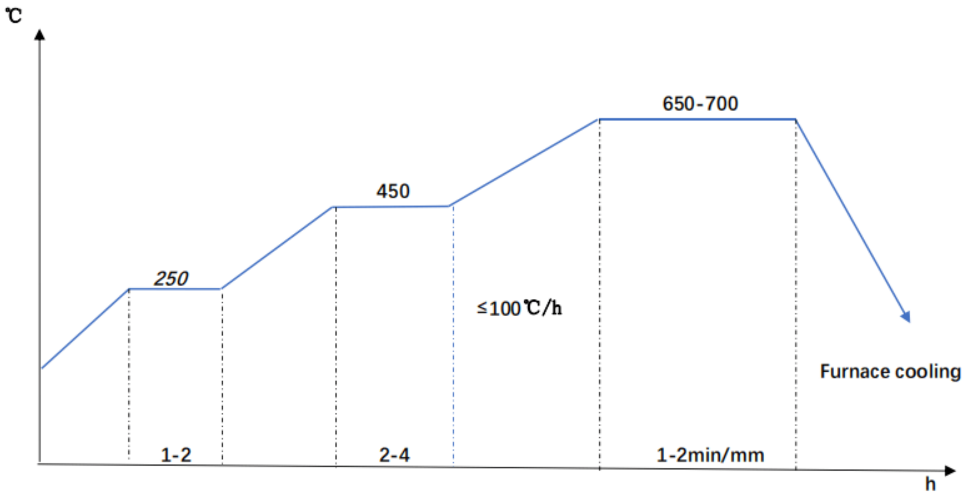
What is the limitation of the GP/BO?

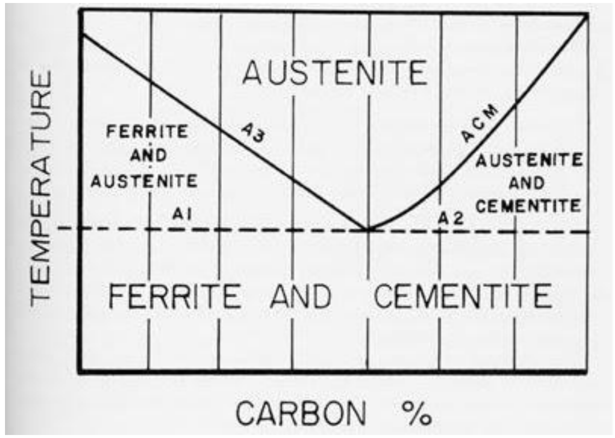
1. Works only in low-dimensional spaces
2. The correlations are defined by the kernel function (very limiting)
3. We do not use any knowledge about physics of the system
4. We do not use cheap information available during the experiment (proxies)

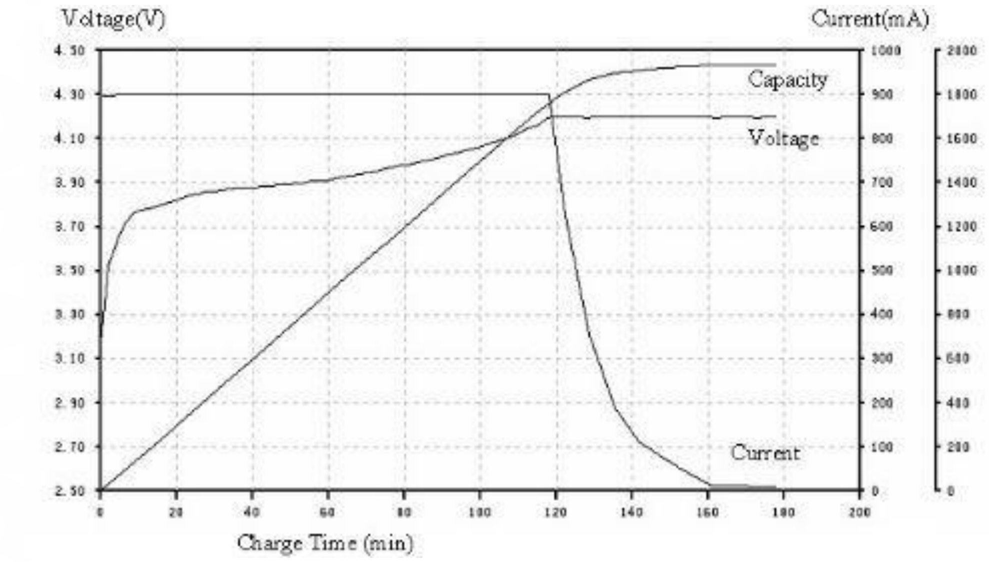
Can we somehow make high dimensional space low-D?

1. Structure-property relationships
2. Molecular discovery and QASR
3. Processing optimization

Making materials: process trajectories



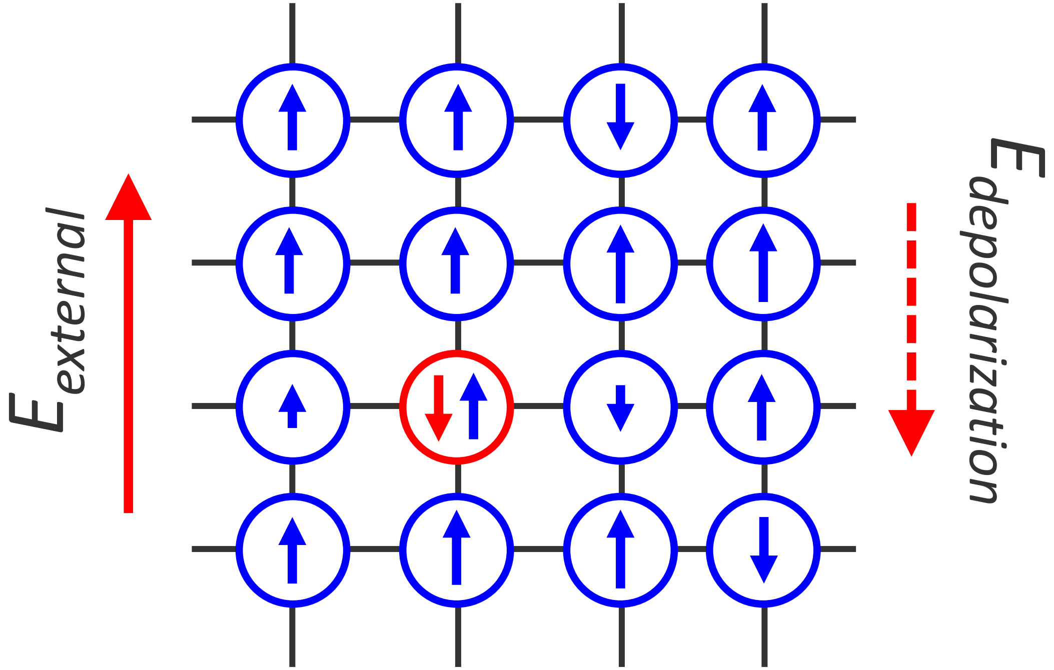


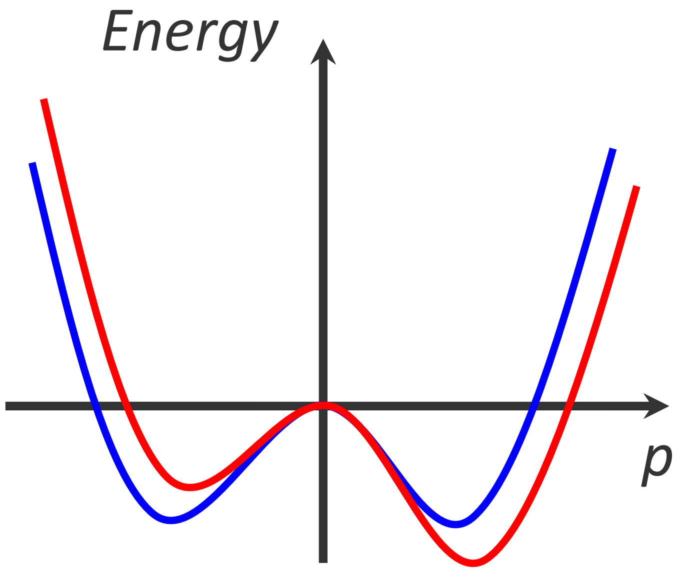


- Making steel: complicated and took a lot of time optimize
- Charging battery: obvious economic impact
- Manufacturing: Annealing hybrid perovskite thin films
- Poling ferroelectric

How do we optimize trajectories if we have (a) only limited or no mechanistic information, (b) our experimental budgets are limited, but (c) we have some access to domain expertise?

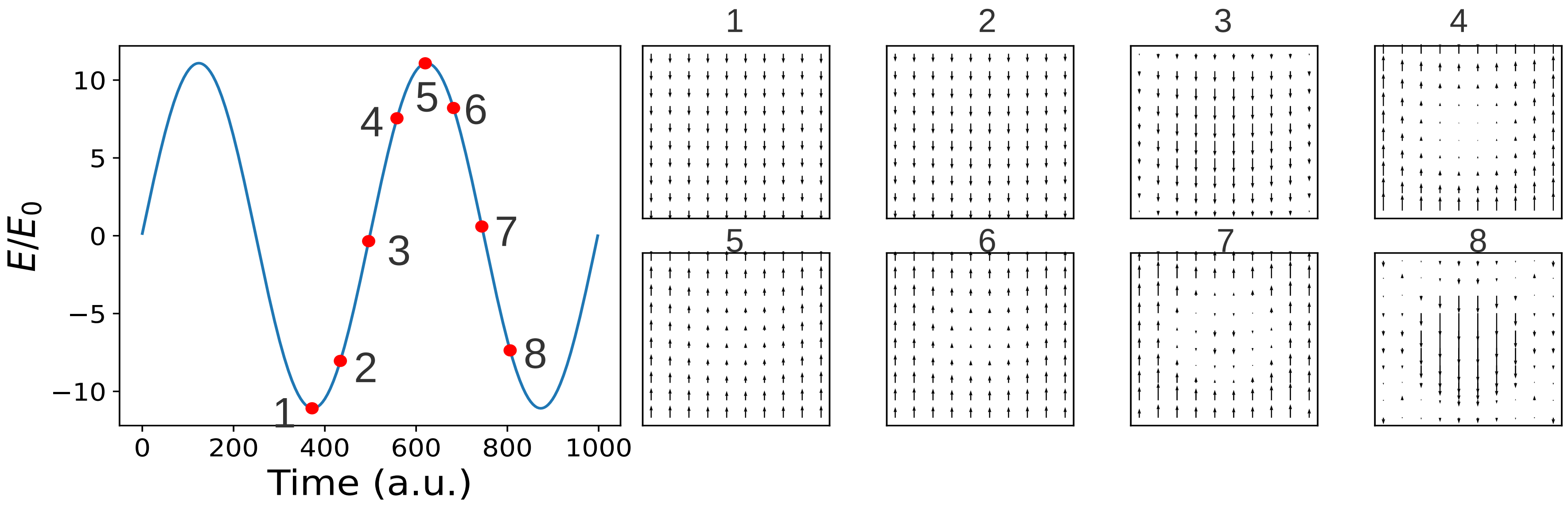
FerroSIM: the simplest interesting ferroelectric

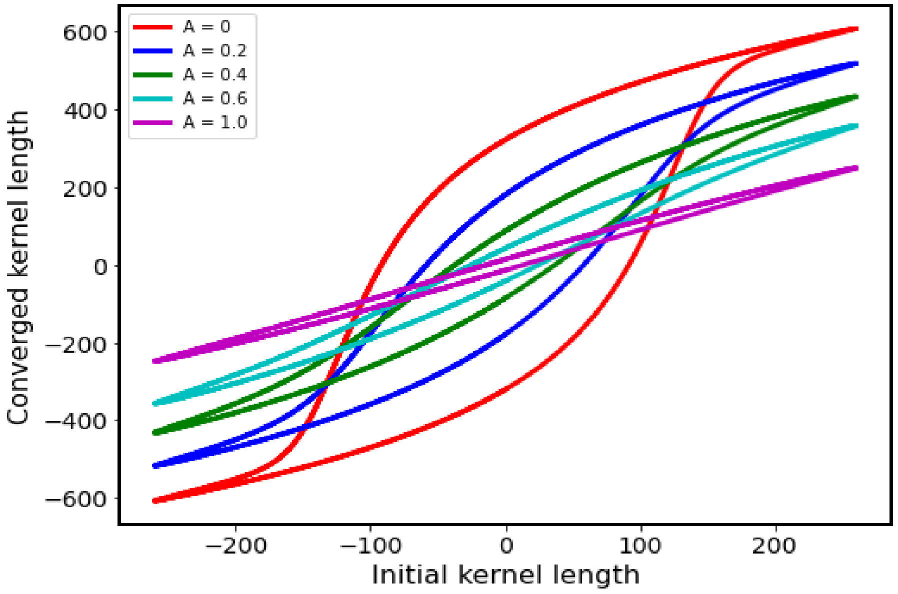


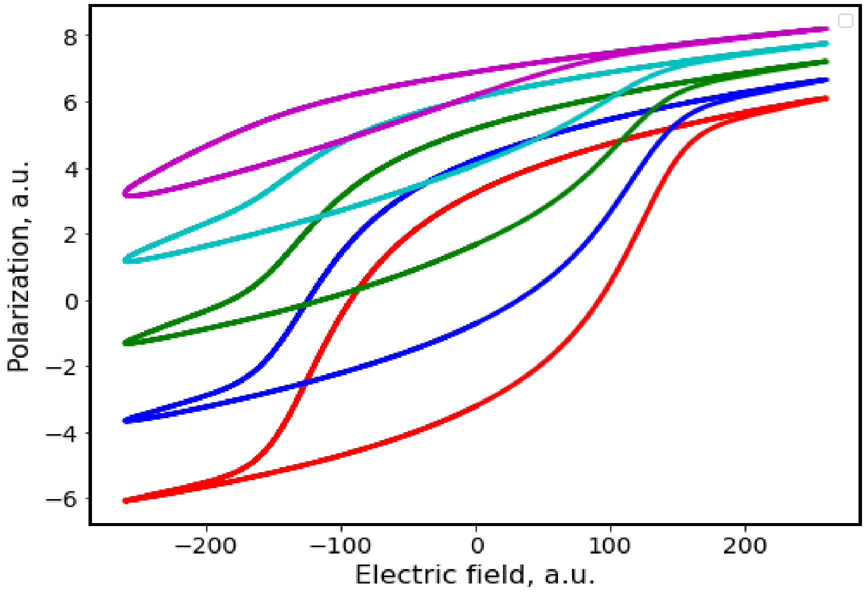


- A discrete square lattice where a continuous polarization vector resides at each lattice site
- The local free energy at each site takes the GLD form:
  - $F_{ij} = \alpha_1 \left(p_{x_{ij}}^2 + p_{y_{ij}}^2\right) + \alpha_2 \left(p_{x_{ij}}^4 + p_{y_{ij}}^4\right) + \alpha_3 p_{x_{ij}}^2 p_{y_{ij}}^2 - E_{loc_{x_{ij}}} p_{x_{ij}} - E_{loc_{y_{ij}}} p_{y_{ij}}$
  - Where, $E_{loc} = E_{ext} + E_{dep} + E_d(i,j)$ and $E_d = -\alpha_{dep} < p >$
- The total free energy is the sum of local free energies and coupling terms:
  - $F = \sum_{i,j}^{N} F_{ij} + K \sum_{k,l} (p_{x_{ij}} - p_{x_{i+k,j+l}})^2 + K \sum_{k,l} (p_{y_{ij}} - p_{y_{i+k,j+l}})^2$
- Polarization at each lattice site is updated to decrease the free energy using $\frac{\gamma d p_{i,j}}{dt} = -\frac{\partial F}{\partial p_{i,j}}$

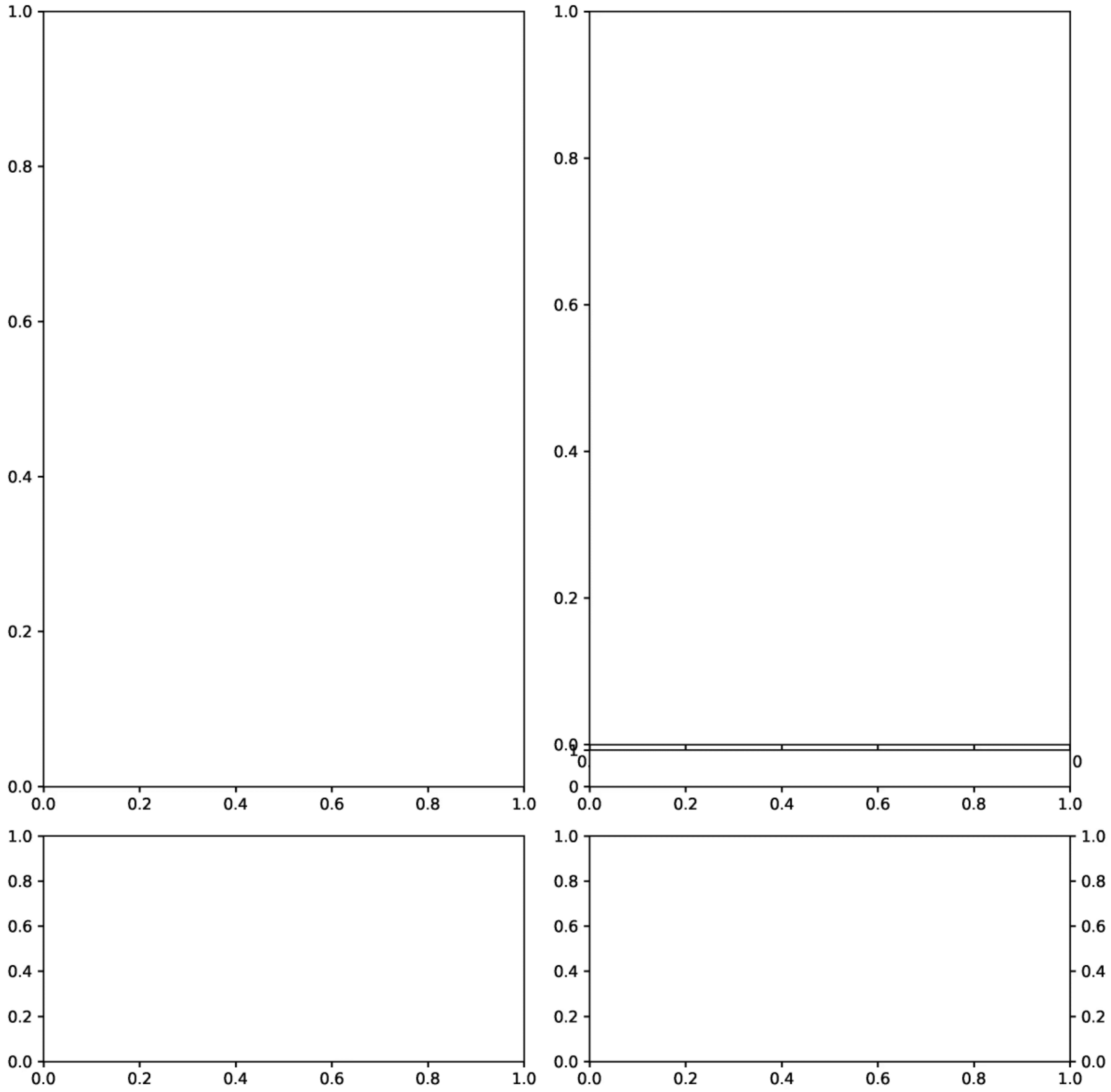
Microstates and Macroscopic Observables







But what about trajectories?



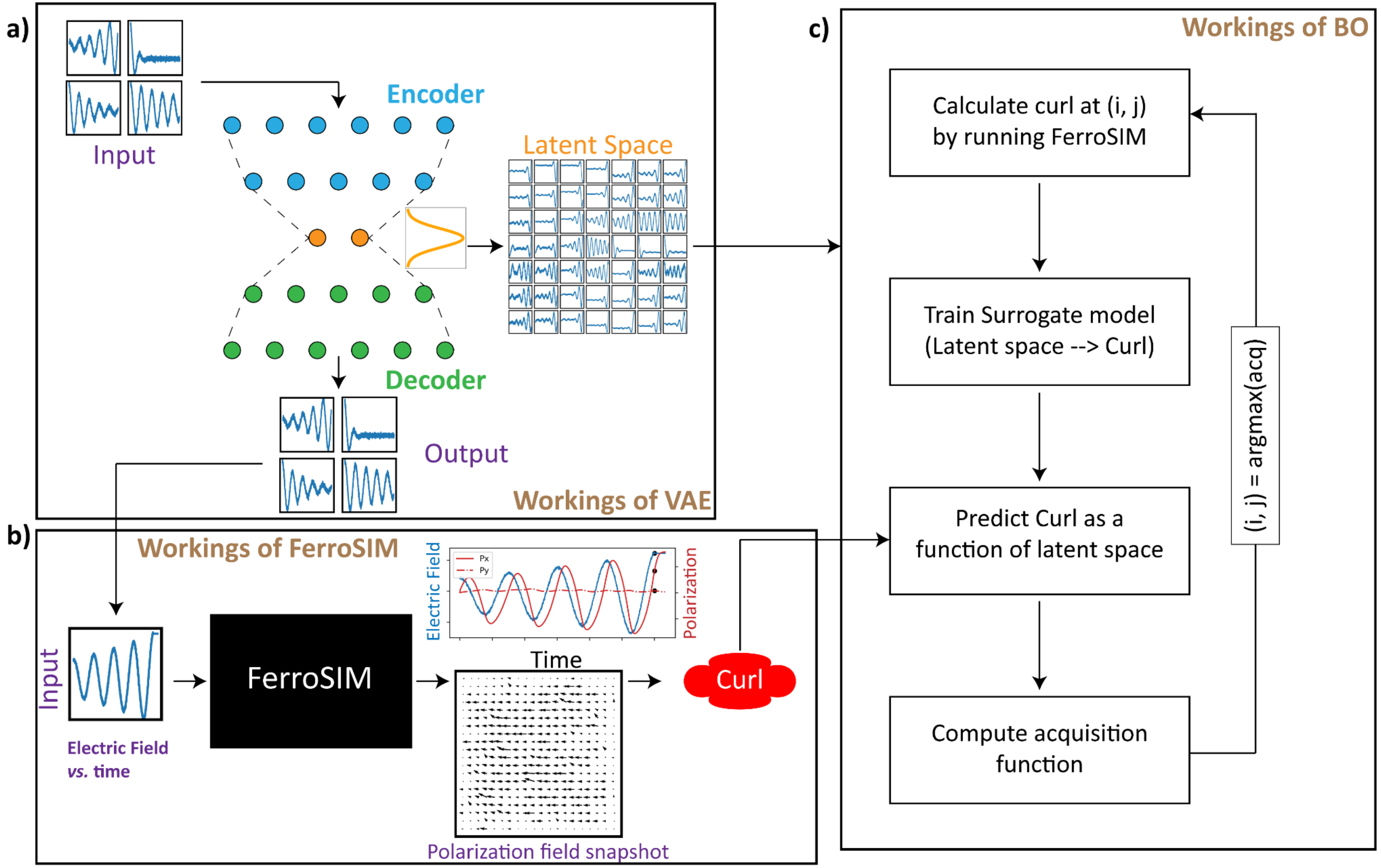
- The model has large number of microstates
- The global state depends on history, i.e. dependence of field vs. time
- Can we somehow optimize the chosen global state in the space of possible histories?

- This space is obviously intractable…

- … however, we are not interested in ALL possible histories. We are interested in relatively simple histories

- **Thought:** what if we start with the histories that make sense from domain perspective, and look for way to simplify them?

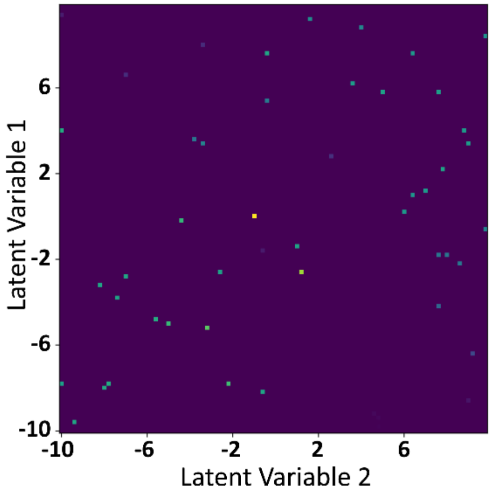
Putting everything together



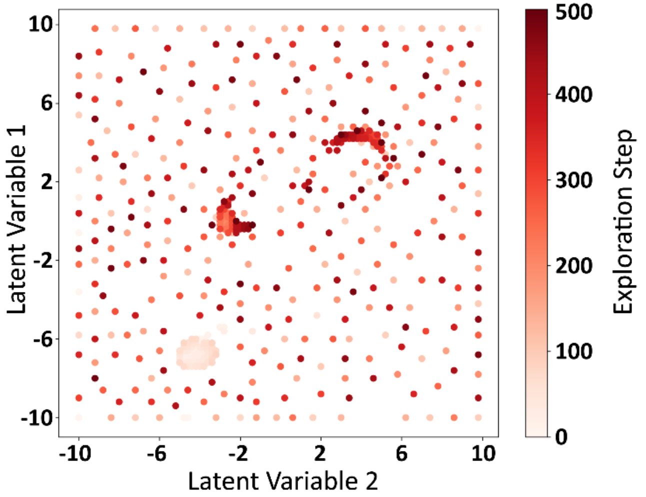
Same approaches are used for molecular discovery, polymers, and biomolecules

Bayesian Optimization in the Latent Space

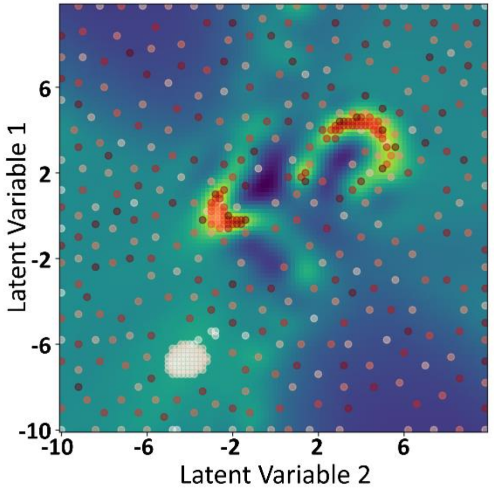
100 initial points



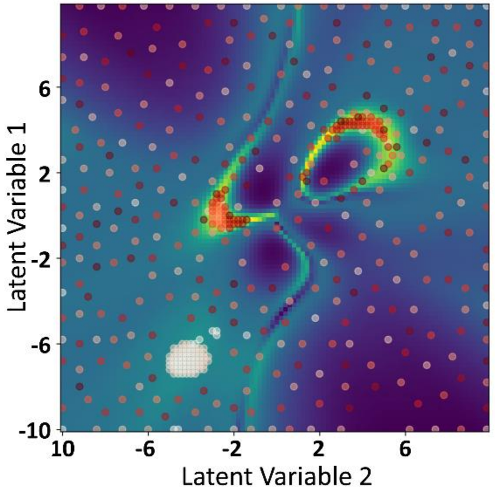
Explored points



Reconstructed curl surface

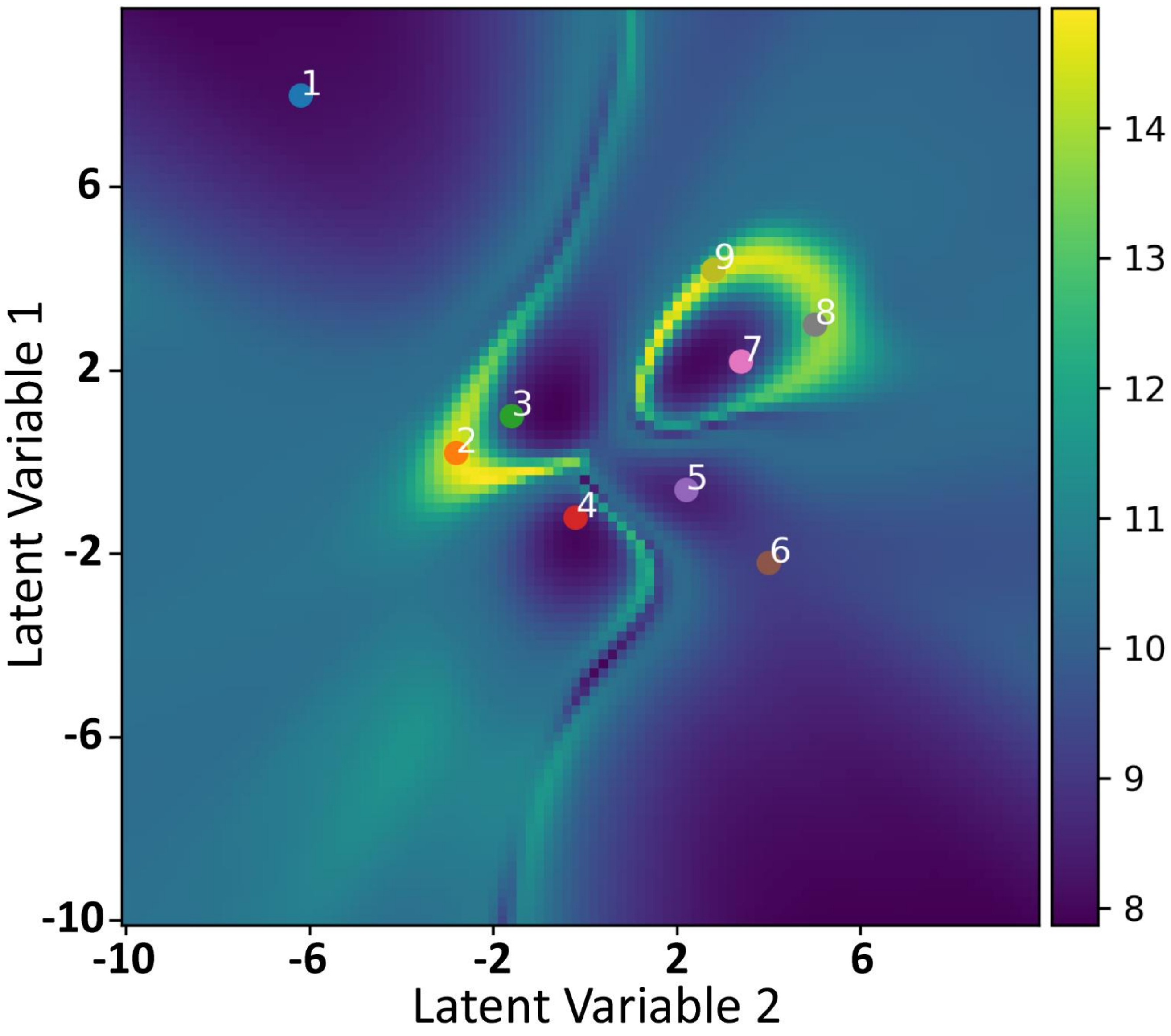


Original curl surface



- 100 initialization points and the BO explored the latent space for the next 500 points
- Acq function: $\mu + 10\sigma$
- So, at the end BO only explored a total of 600 points out of 10,000 points the latent space is divided into
- Caveat: we had to tune the Acq with the ground truth data known

What determines success?

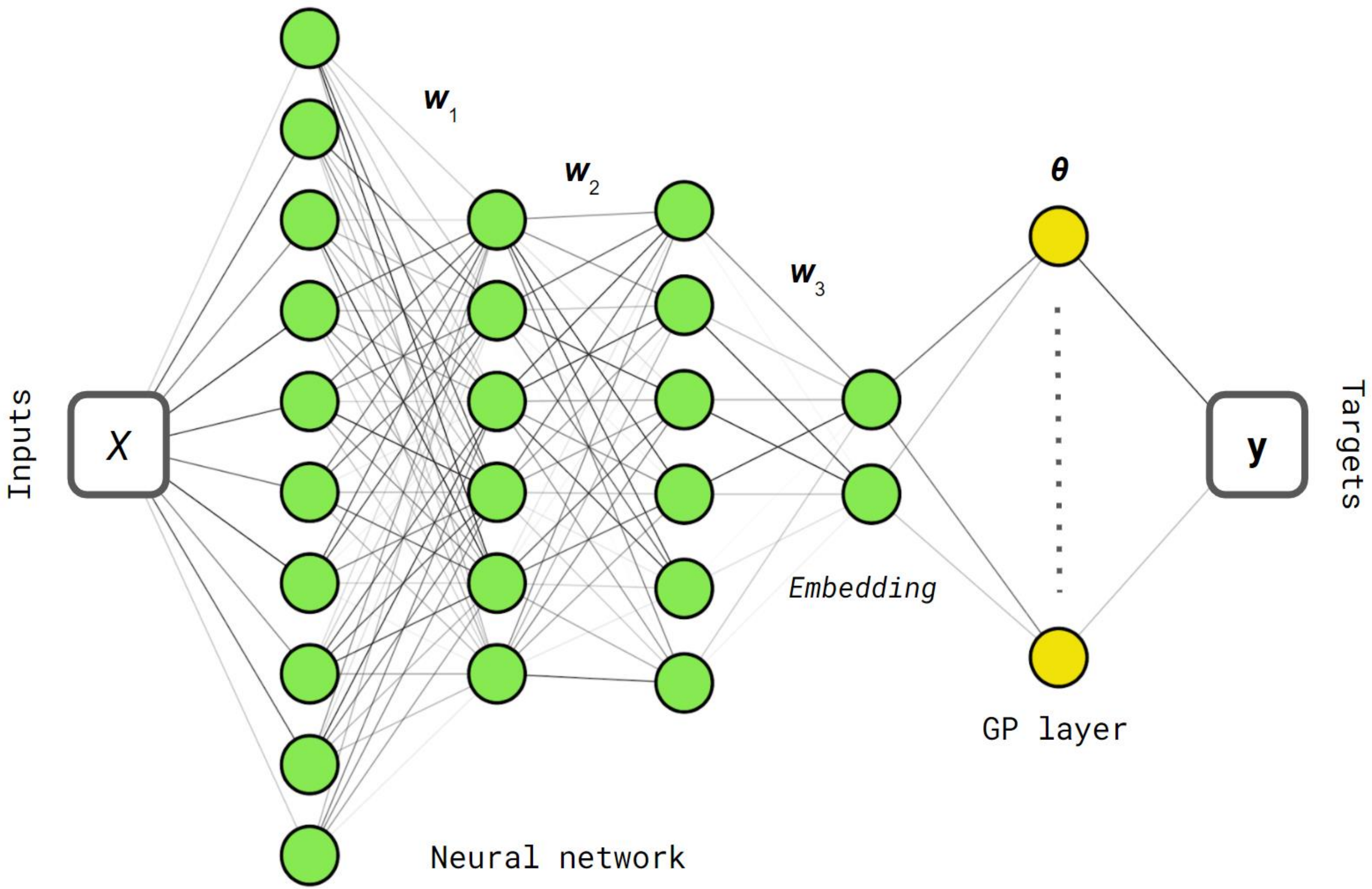


The success of the BO in the latent space clearly depends on the shape on the manifold that points of interest form.

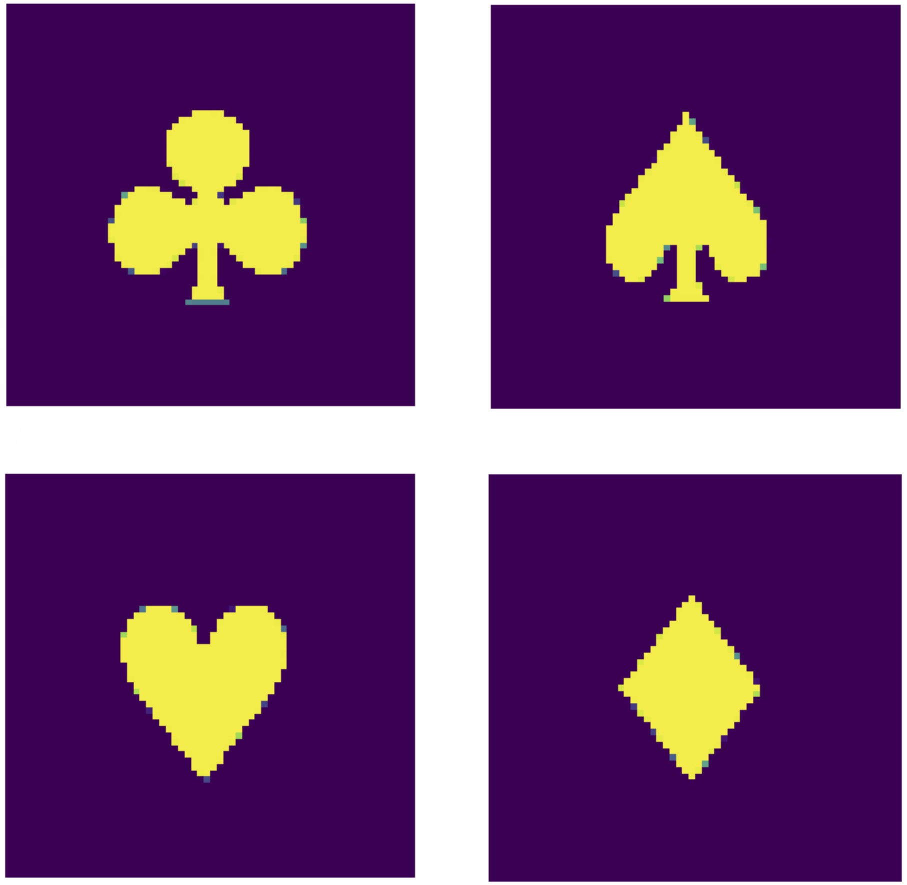
For VAE, the shape of the manifold is determined by the properties of the data only, including

(a) how strong correlations in data reflect in correlation in properties and
(b) weight of the “good” trajectories

Deep Kernel Learning



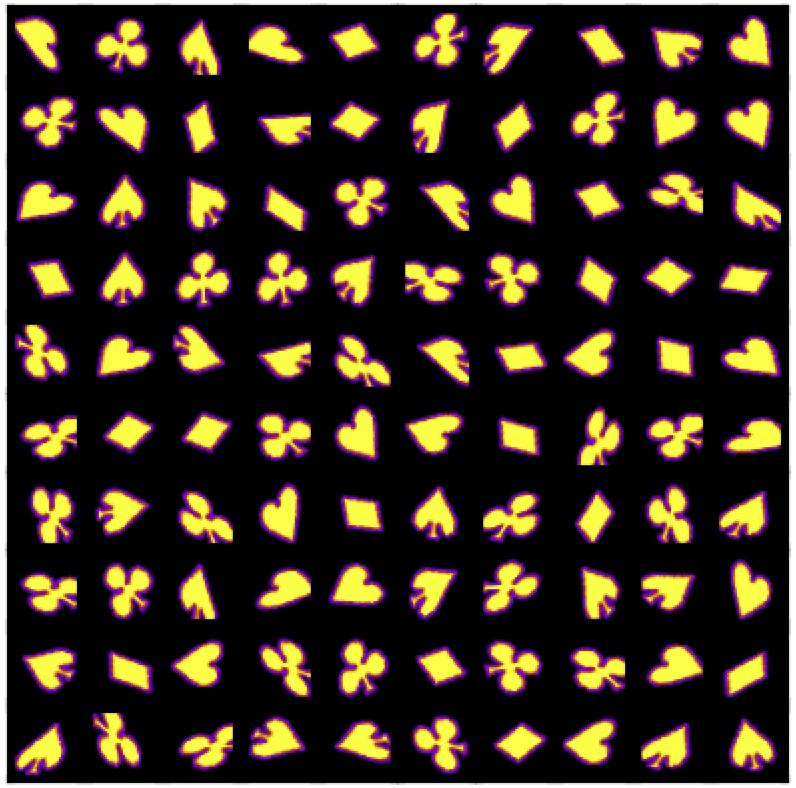
Card data set



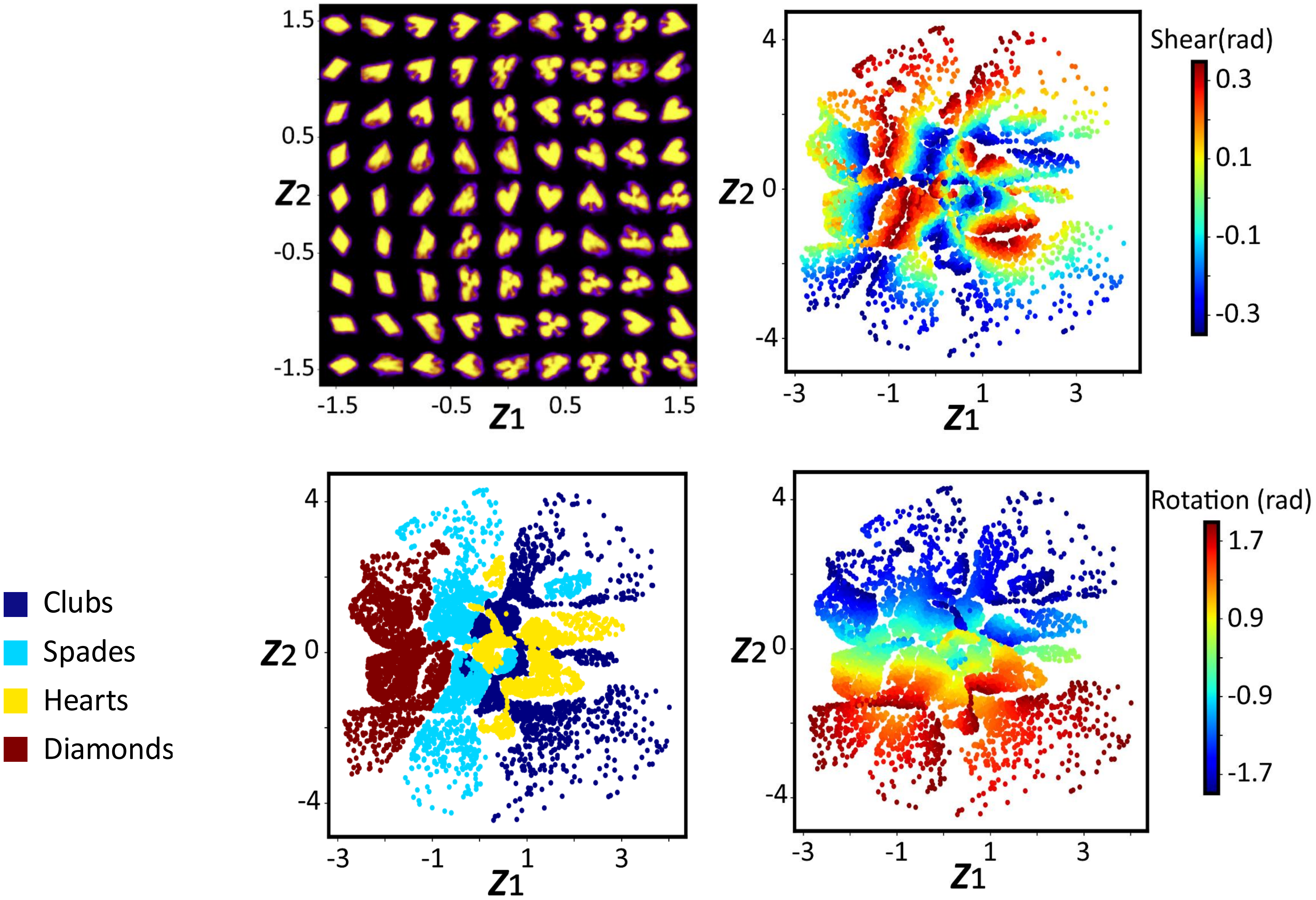
Rotations:
[-120°, 120°]

Shear:
[-20°, 20°]

Randomly picked samples

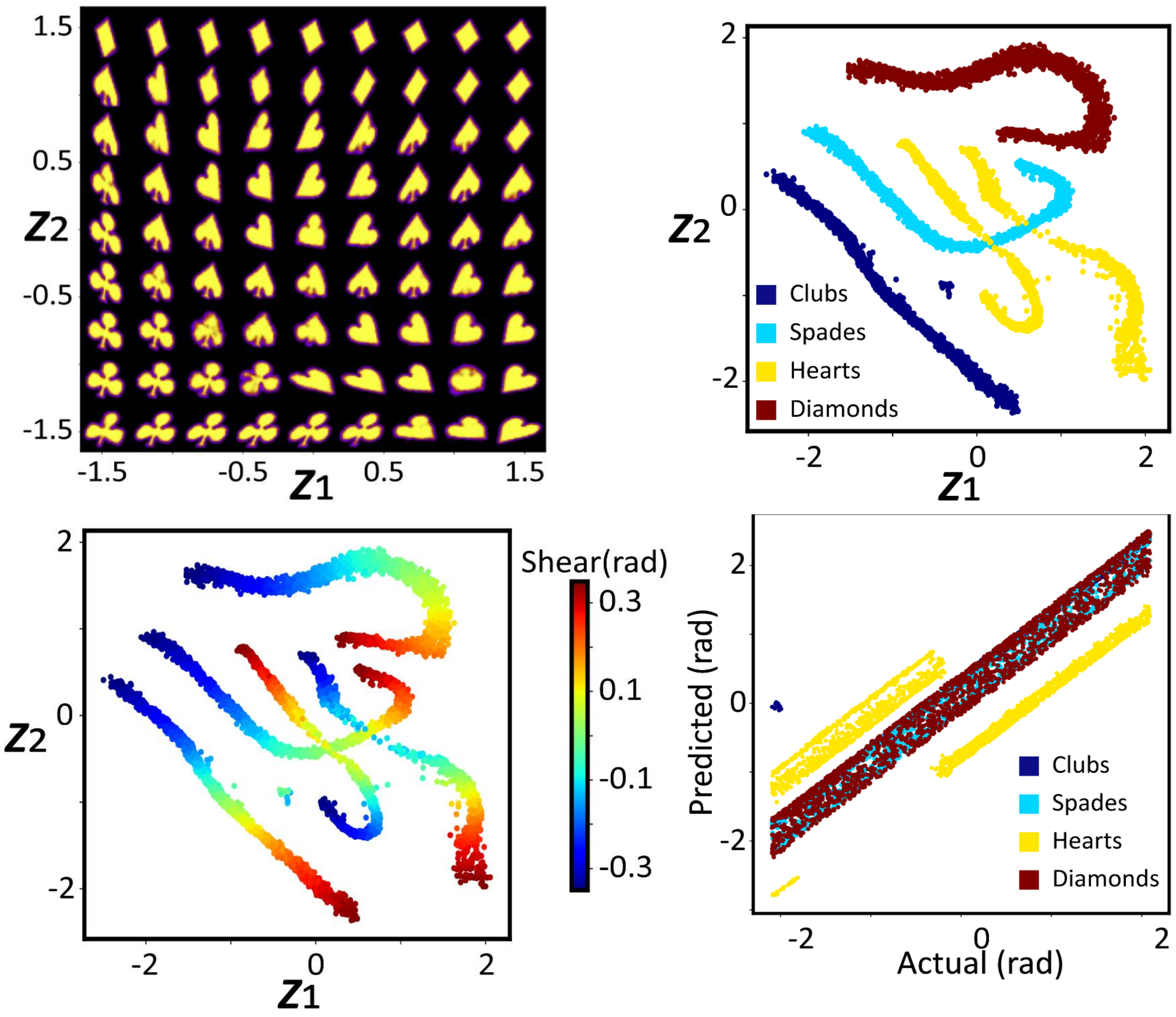


VAE on Cards

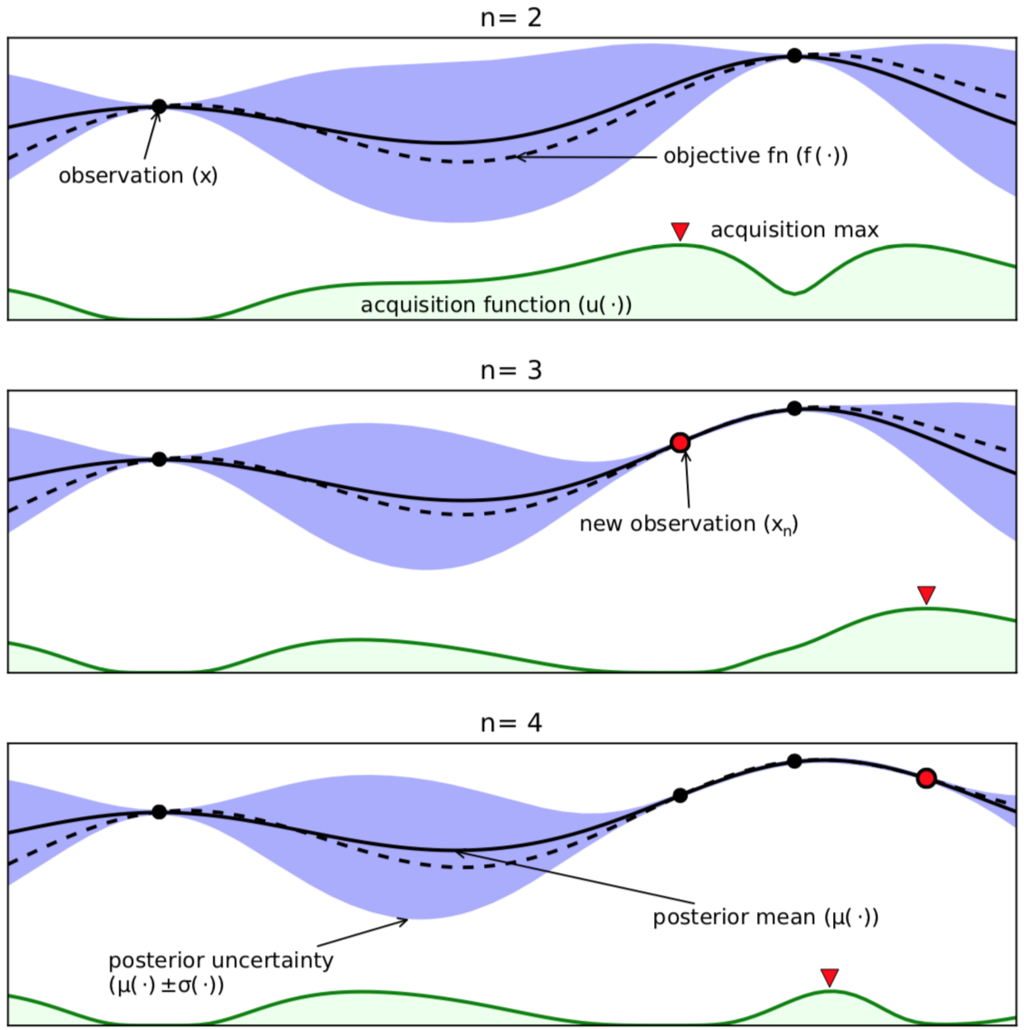


1.5
0.5
-0.5
-1.5
Z2
-1.5
-0.5
0.5
1.5
Z1
Shear(rad)
0.3
0.1
-0.1
-0.3
4
0
-4
-3
-1
1
3
Clubs
Spades
Hearts
Diamonds
Rotation (rad)
1.7
0.9
-0.9
-1.7

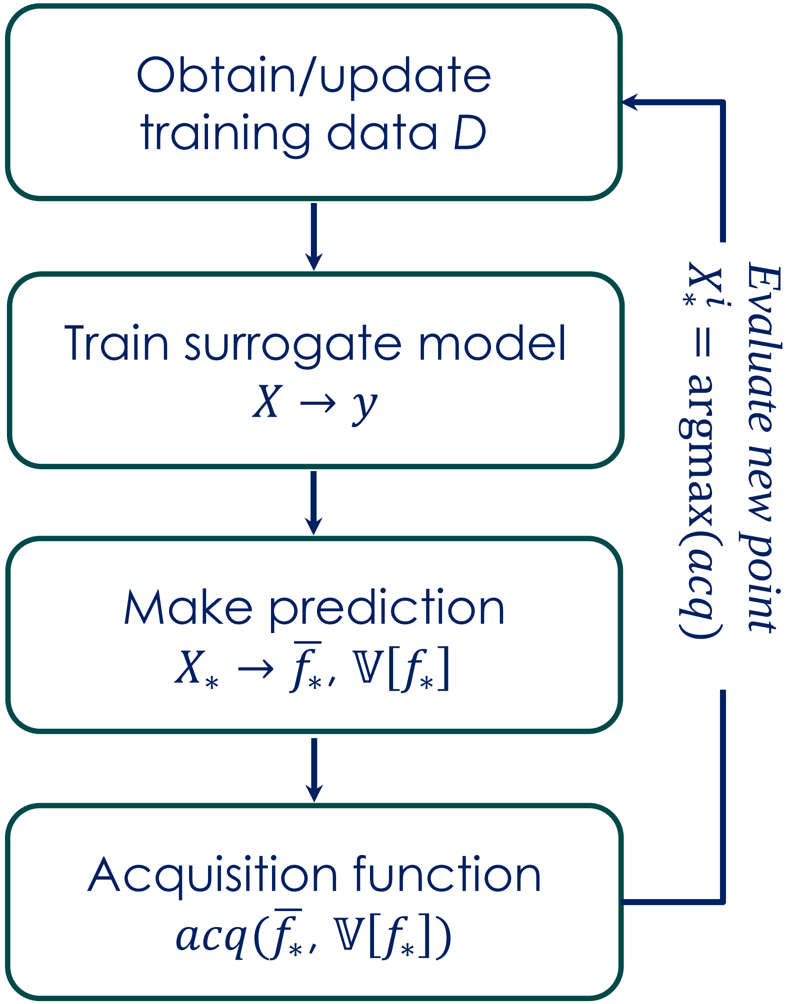
rVAE on Cards



Bayesian Optimization



X, y: (sparse) Training data
X_*: New (not yet evaluated) points

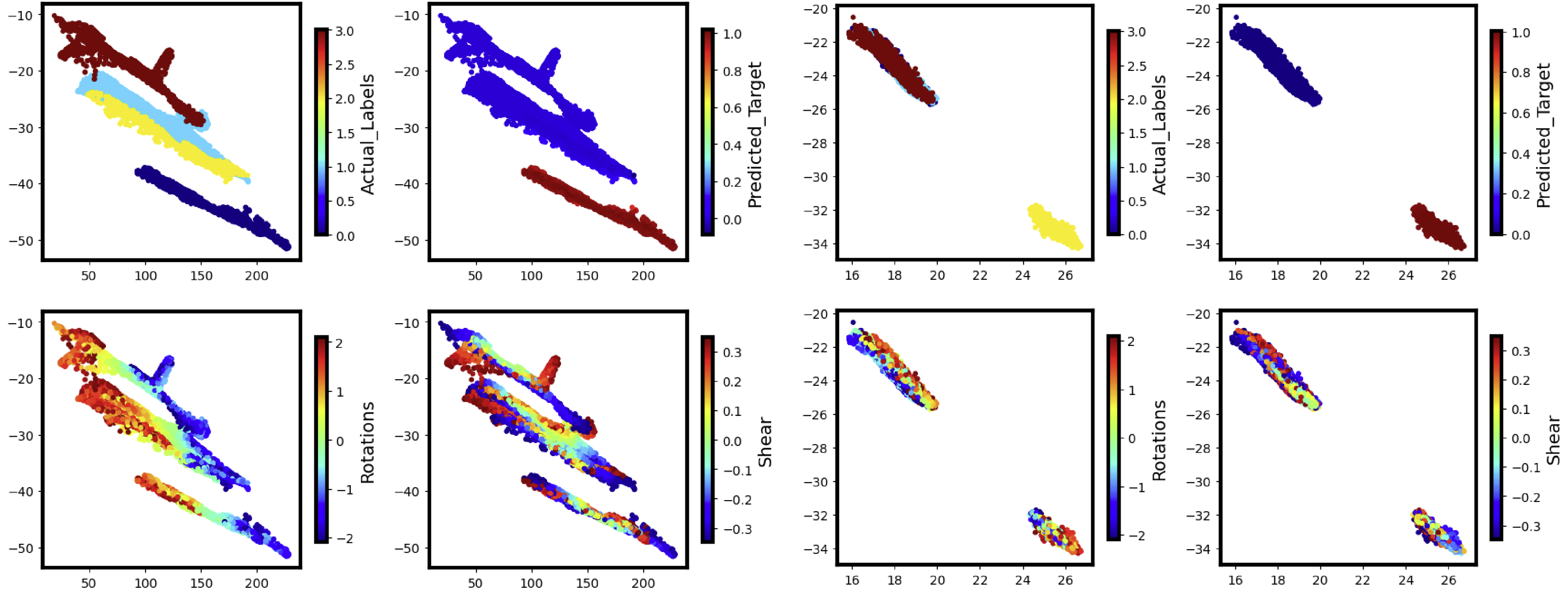


N. de Freitas et al., Taking the Human Out of the Loop: A Review of Bayesian Optimization , *Proceedings of the IEEE* **104**, 148 (2015)

DKL to predict labels

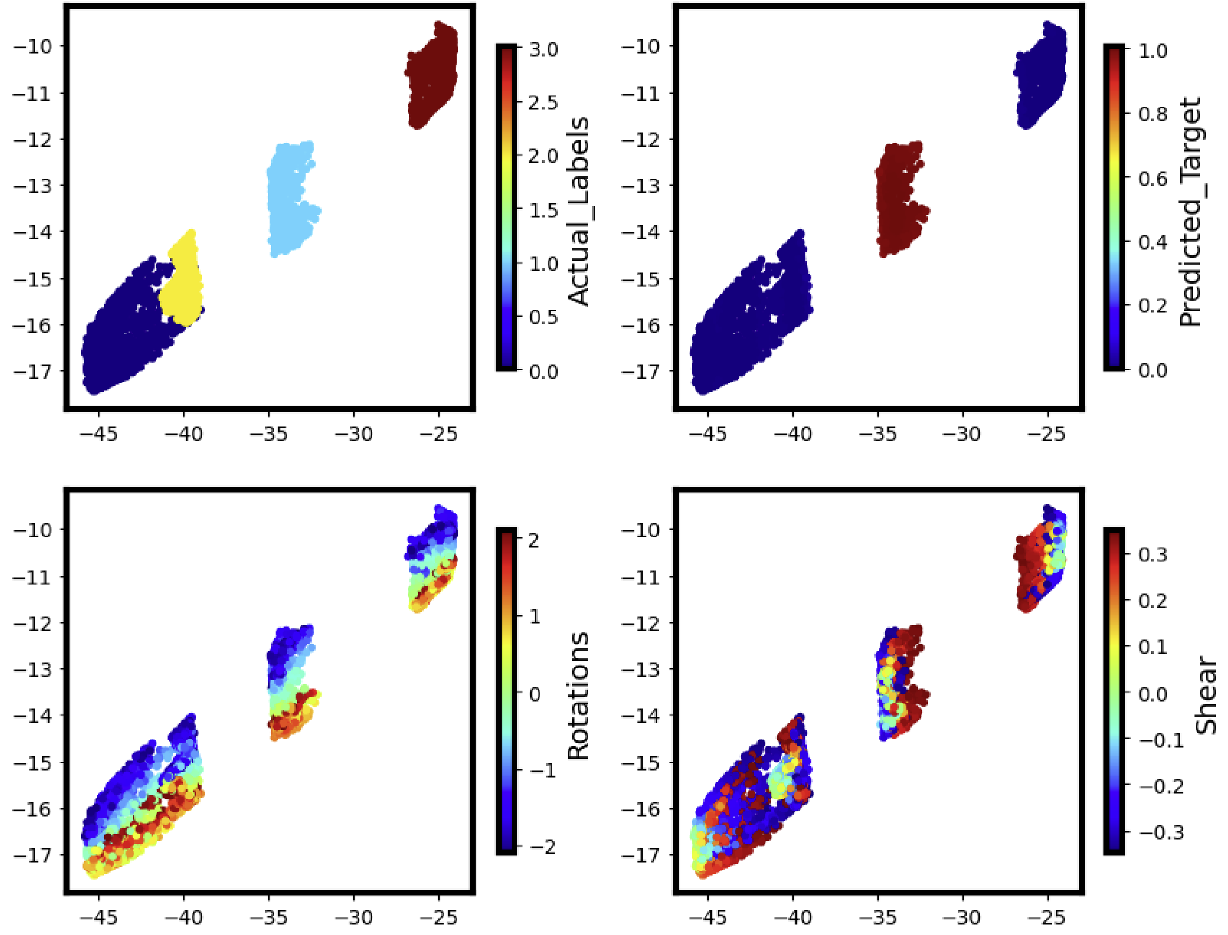
Clubs

Hearts

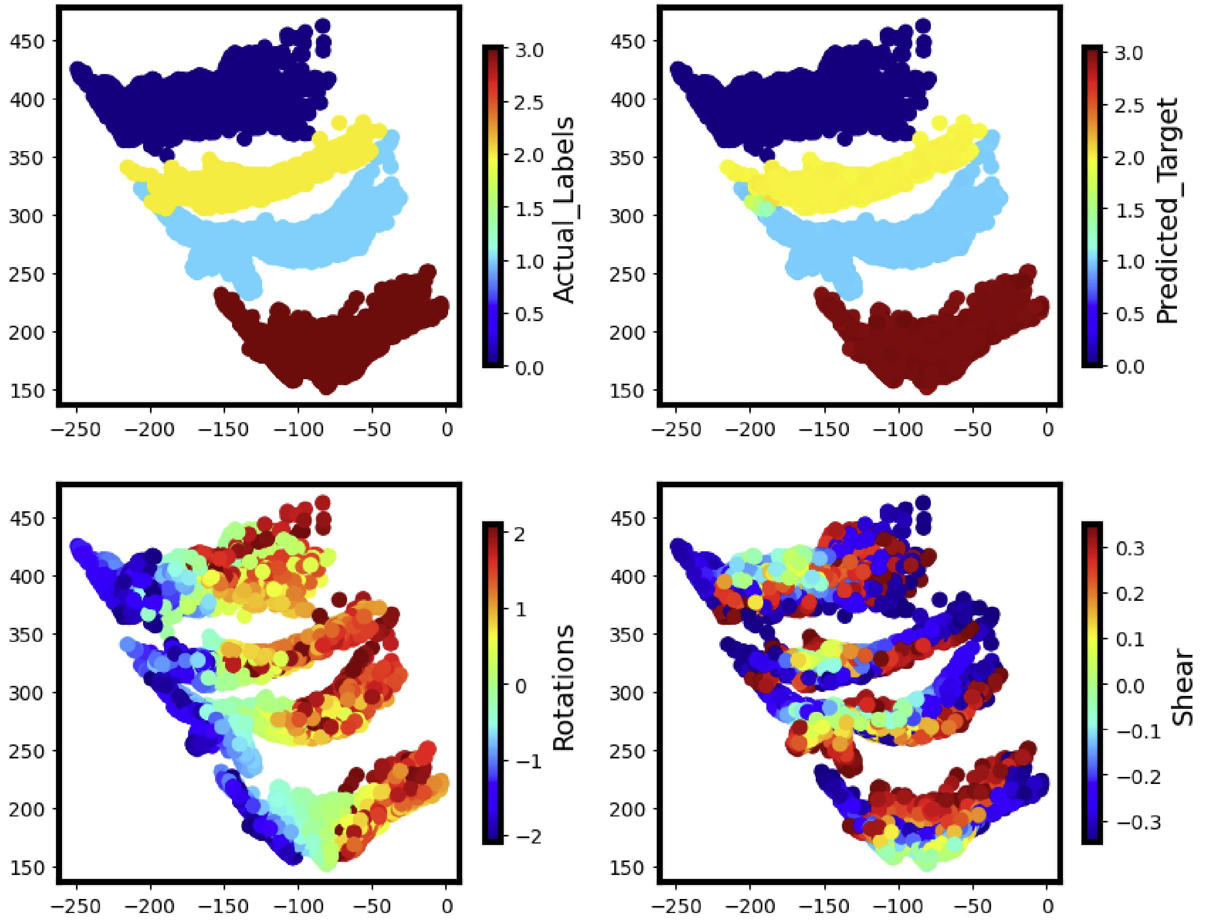


DKL to predict labels

Spades



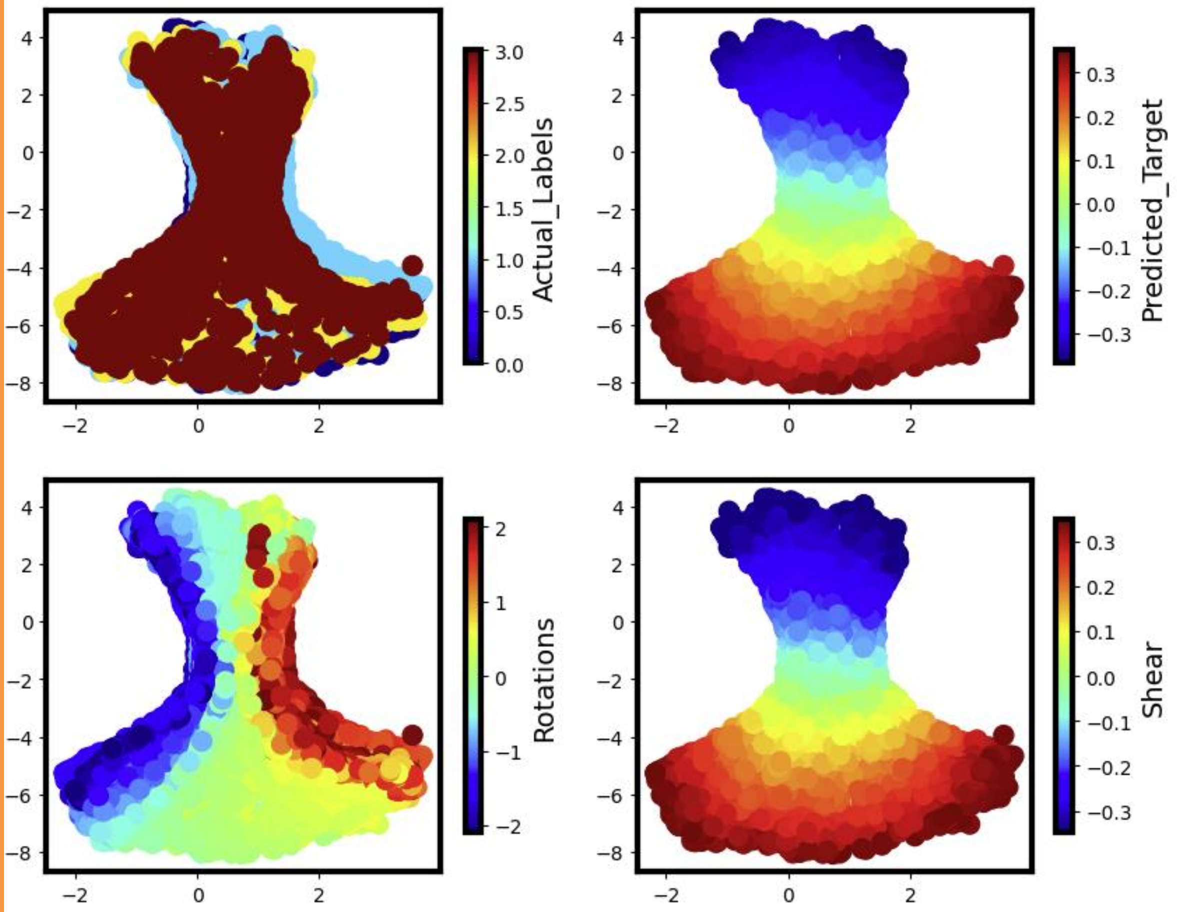
All suits



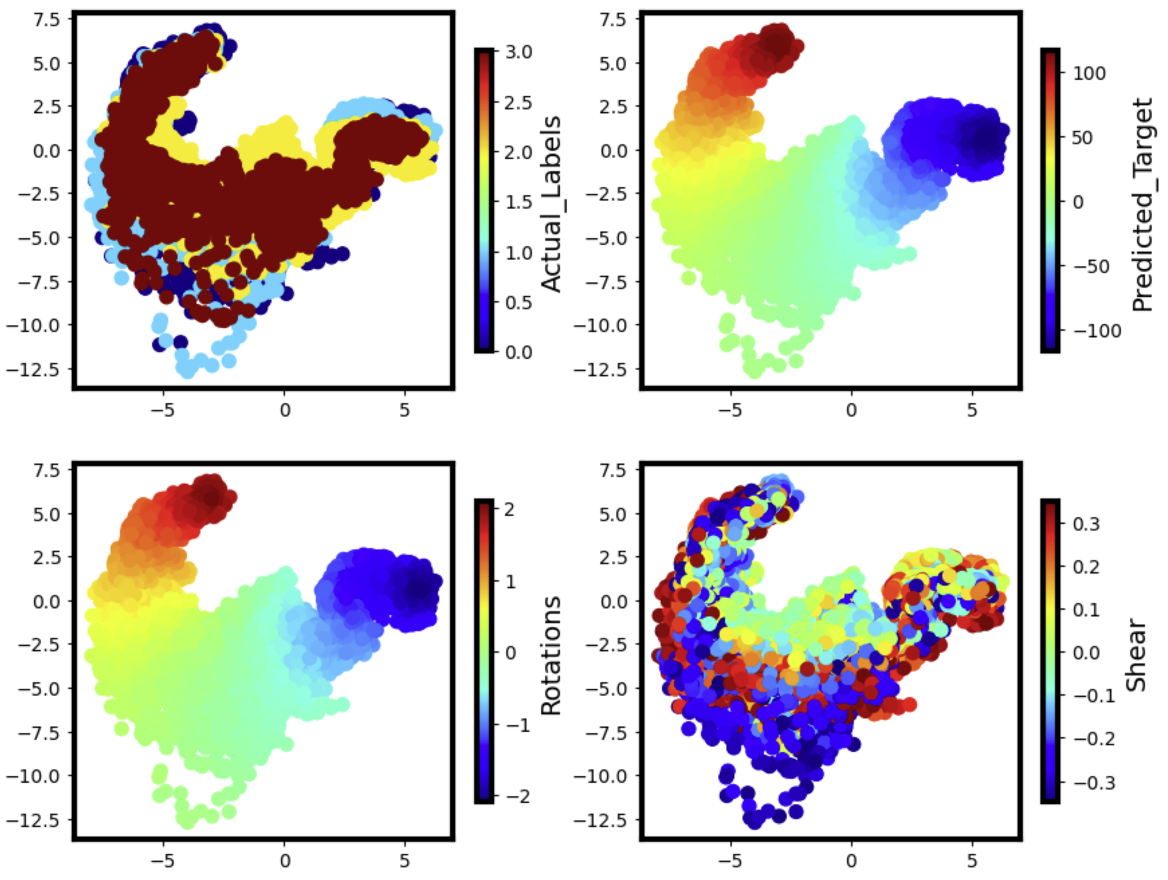
The DKL clearly forms the manifold based on the label!

DKL to predict continuous target function

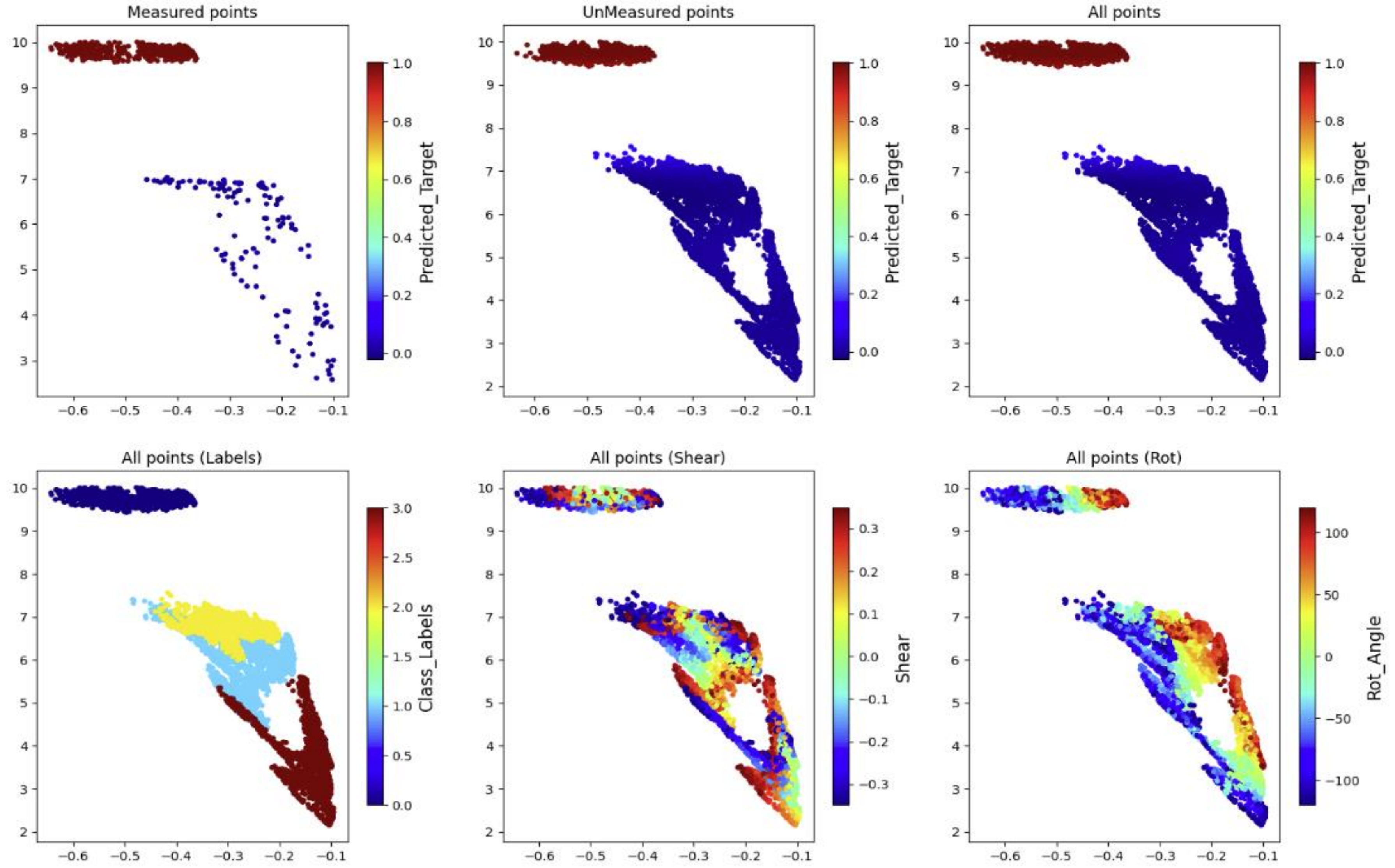
Shear



Rotations

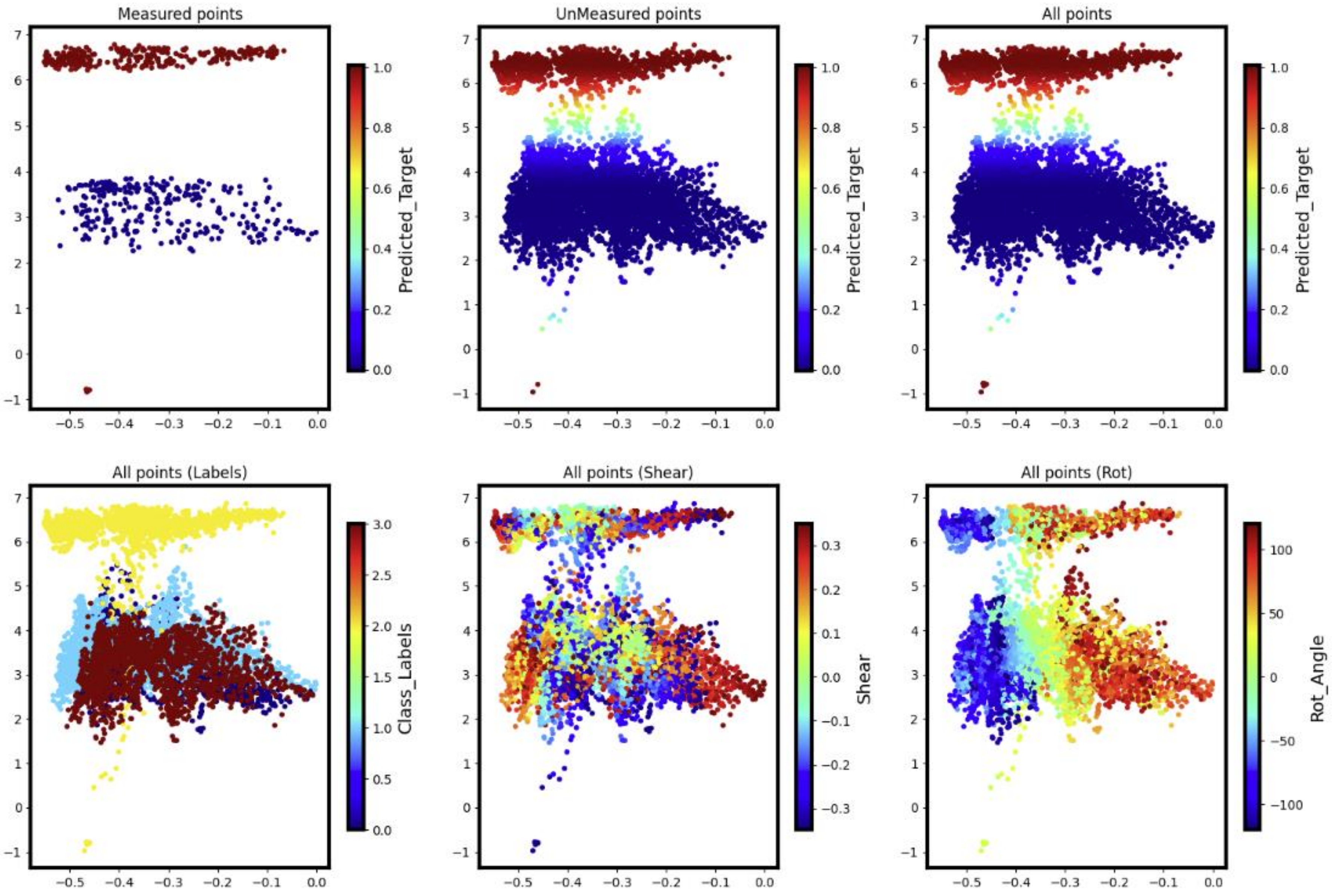


DKL BO: Active Learning

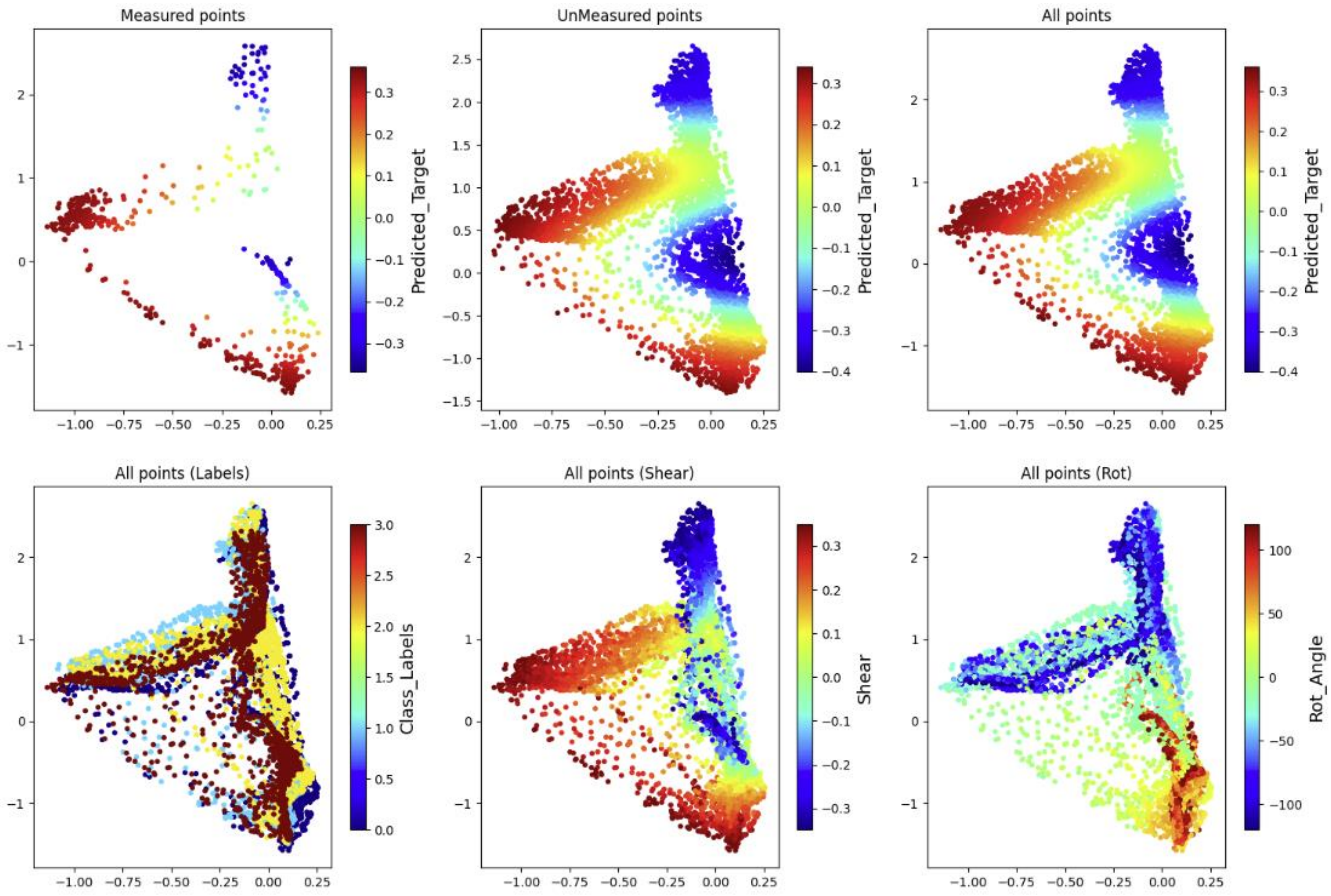


- 100 initialization points and then BO explored 500 points subsequently
- Acquisition function: $\mu + 10\sigma$

DKL BO: Hearts

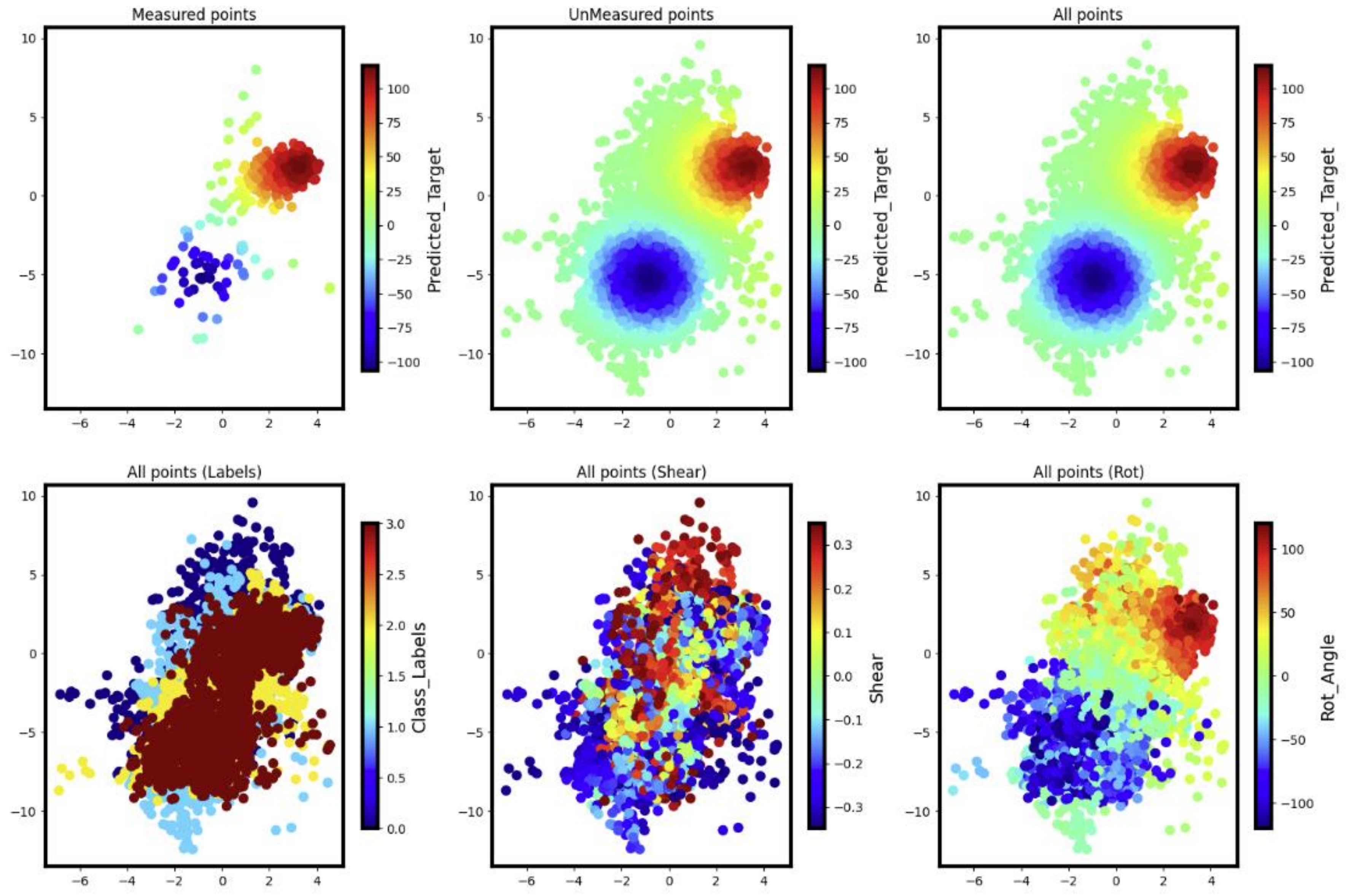


DKL BO: Shear



Measured points
UnMeasured points
All points
All points (Labels)
All points (Shear)
All points (Rot)
Predicted_Target
Class_Labels
Shear
Rot_Angle

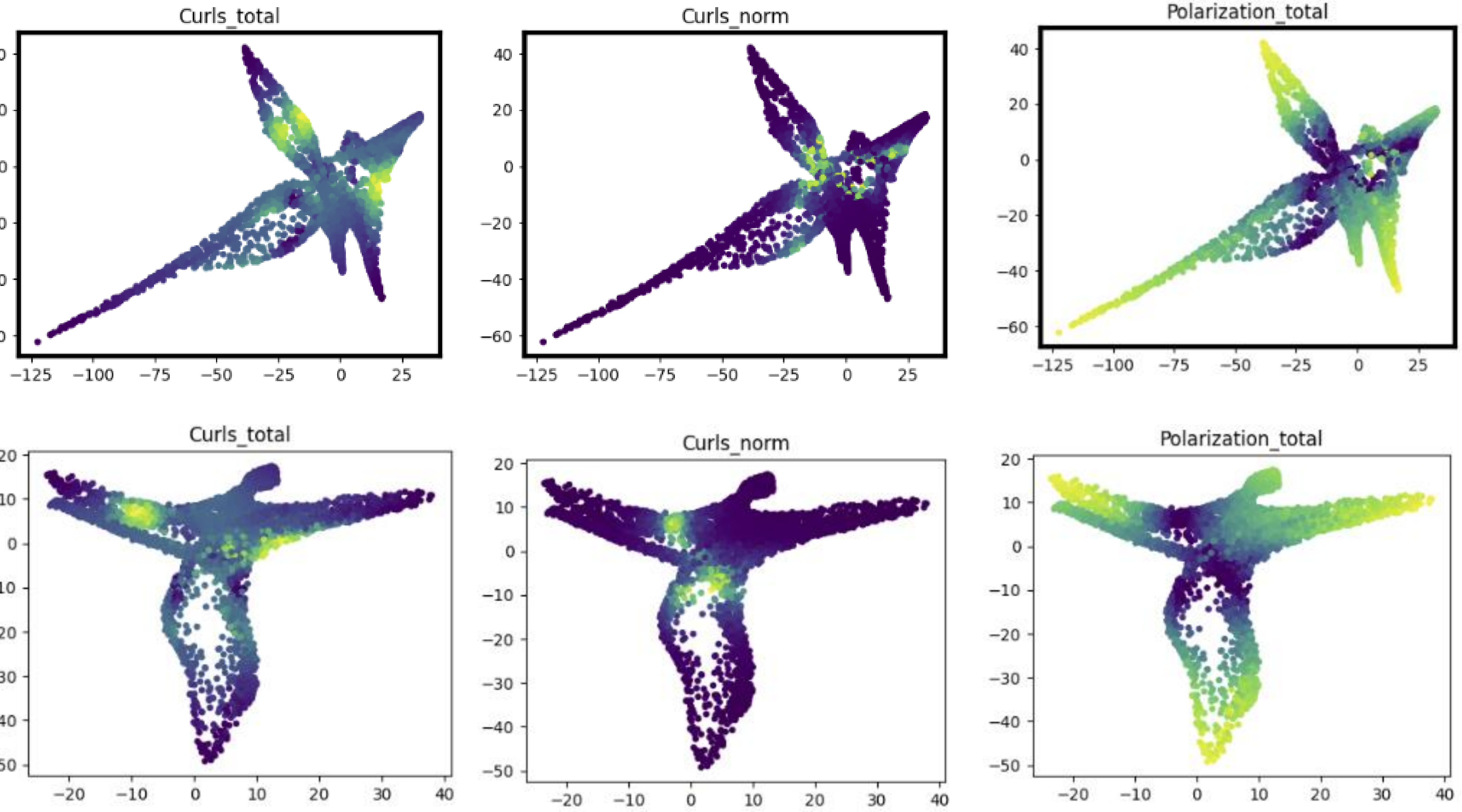
DKL BO: Rotations



DKL on FerroSIM: Static

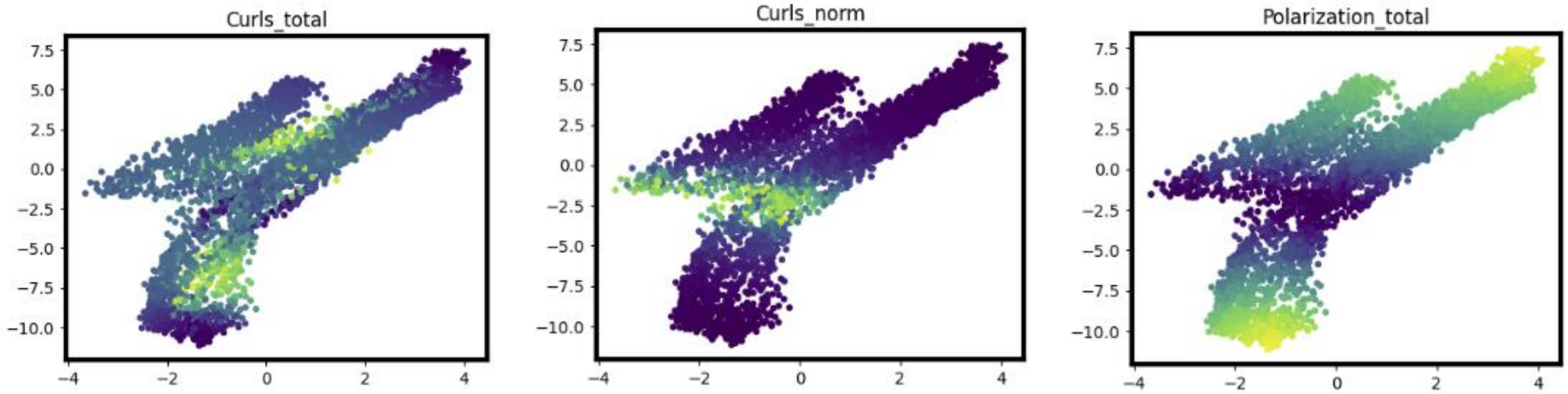
Target function: Curl

Target function: Normalized curl

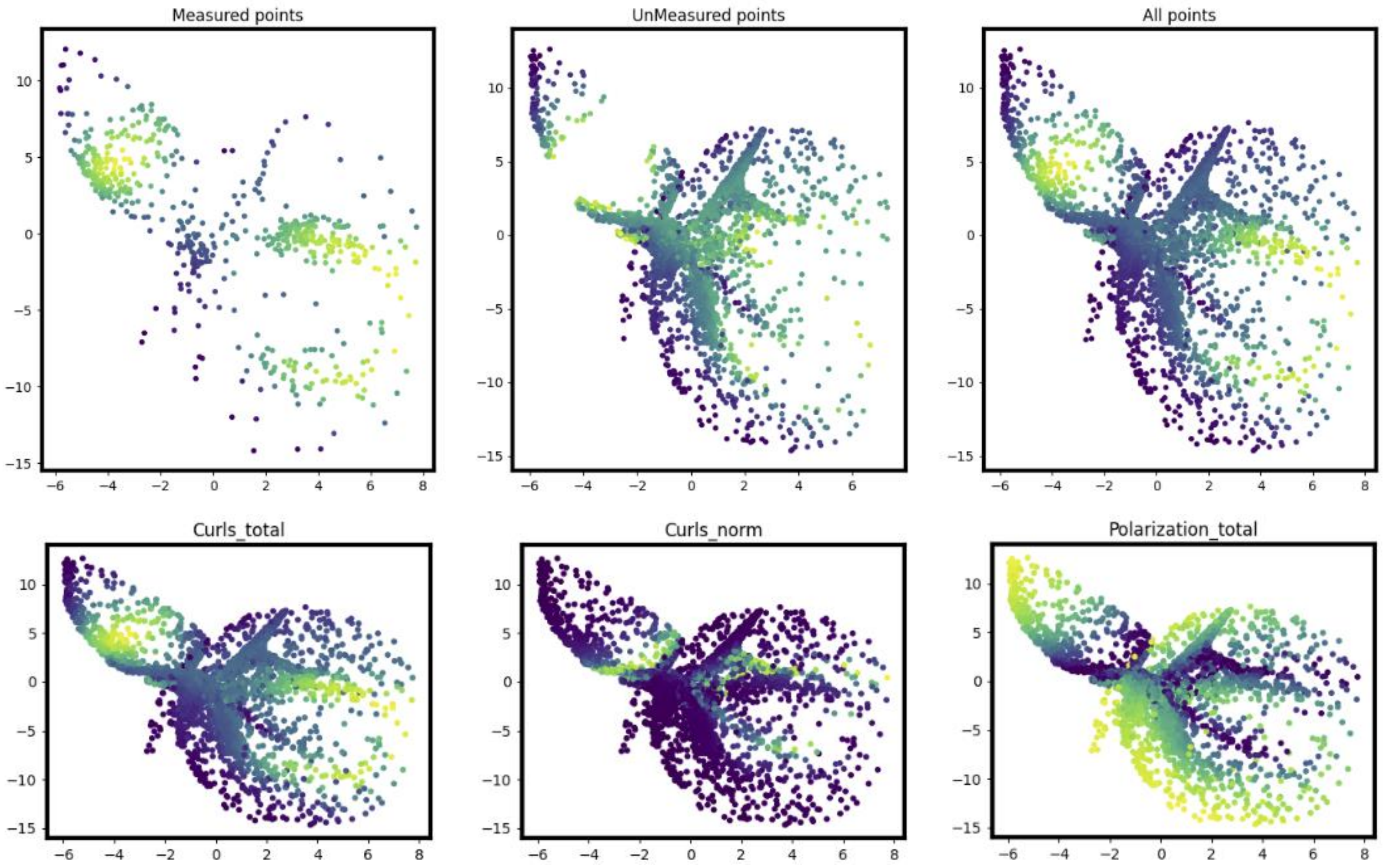


DKL on FerroSIM: Static

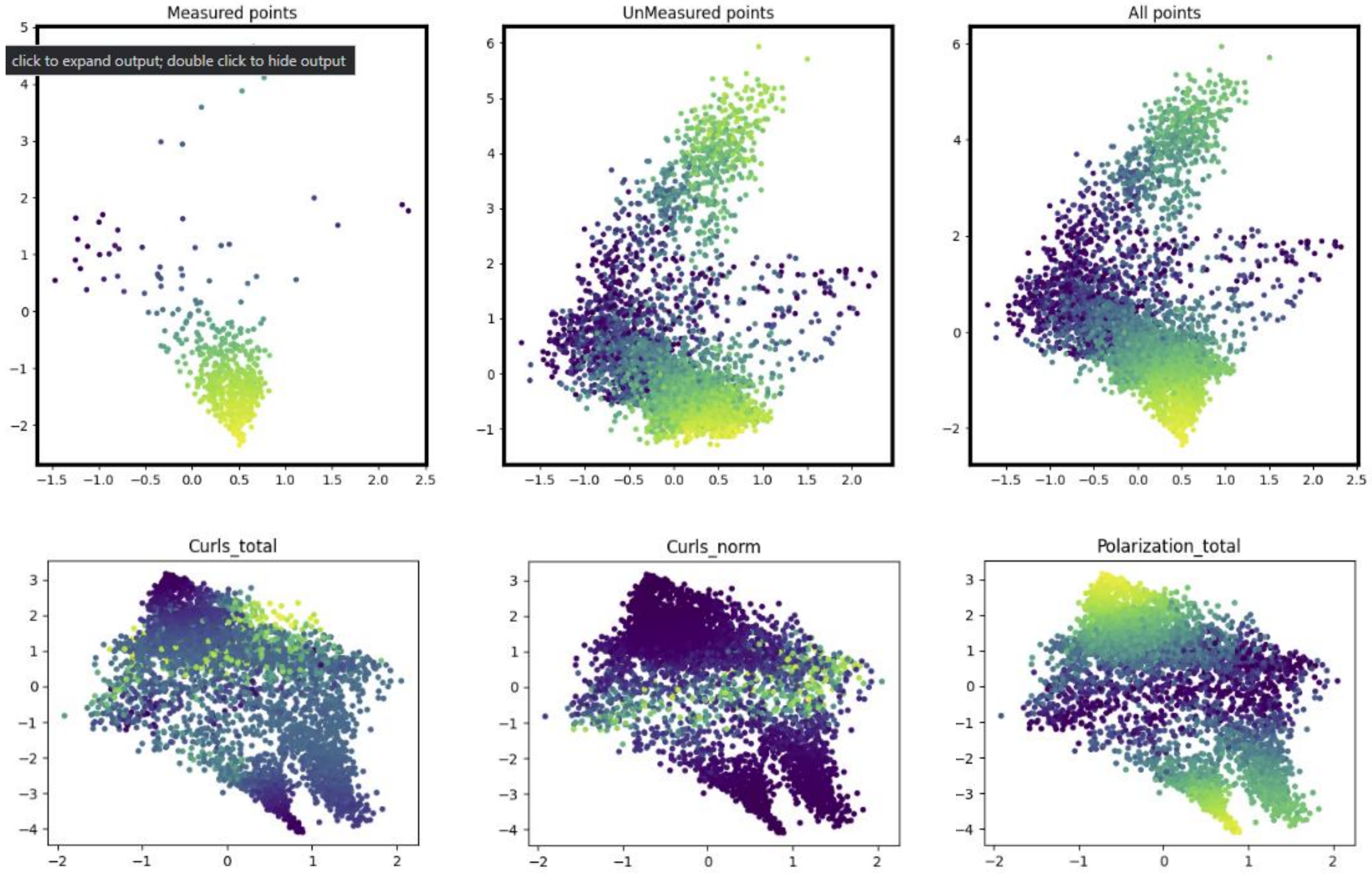
Target function:
Polarization



DKL BO: Active Learning of Curl

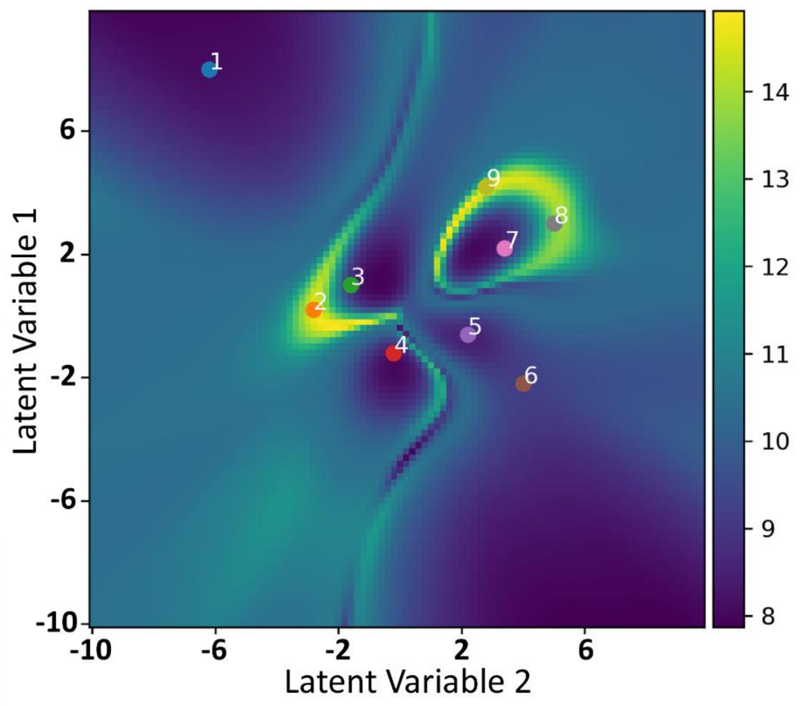


DKL BO: Active Learning of Polarization

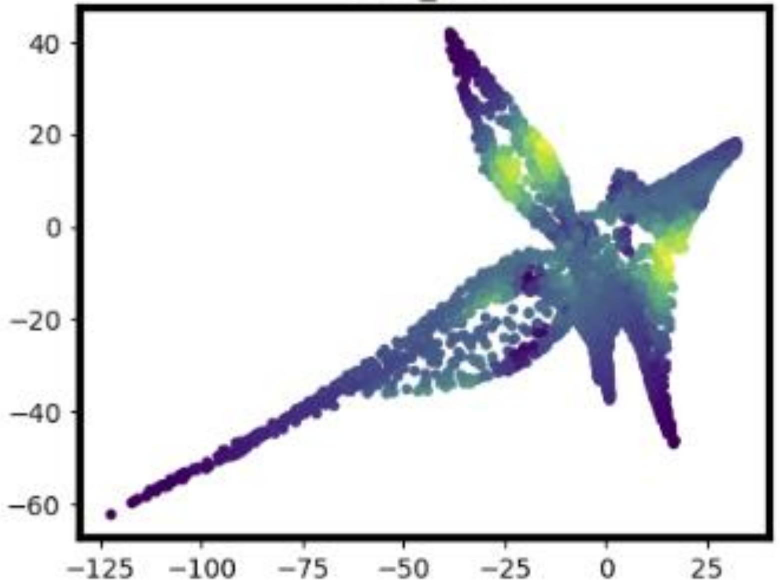


Comparing VAE BO, Static DKL, and Active DKL

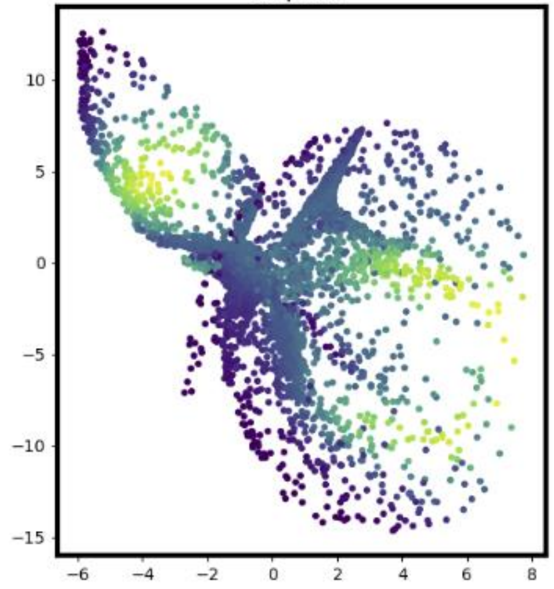
VAE - BO



Static DKL



Active DKL



Summary:

- Manifold structure determines how fast can the unsupervised or active learning work
- For VAEs, the latent structure is determined by the data only. Sometimes property are forming convenient manifolds, most of the time not.
- Static DKL forms much better organized manifolds
- … Active learning produces best manifolds!