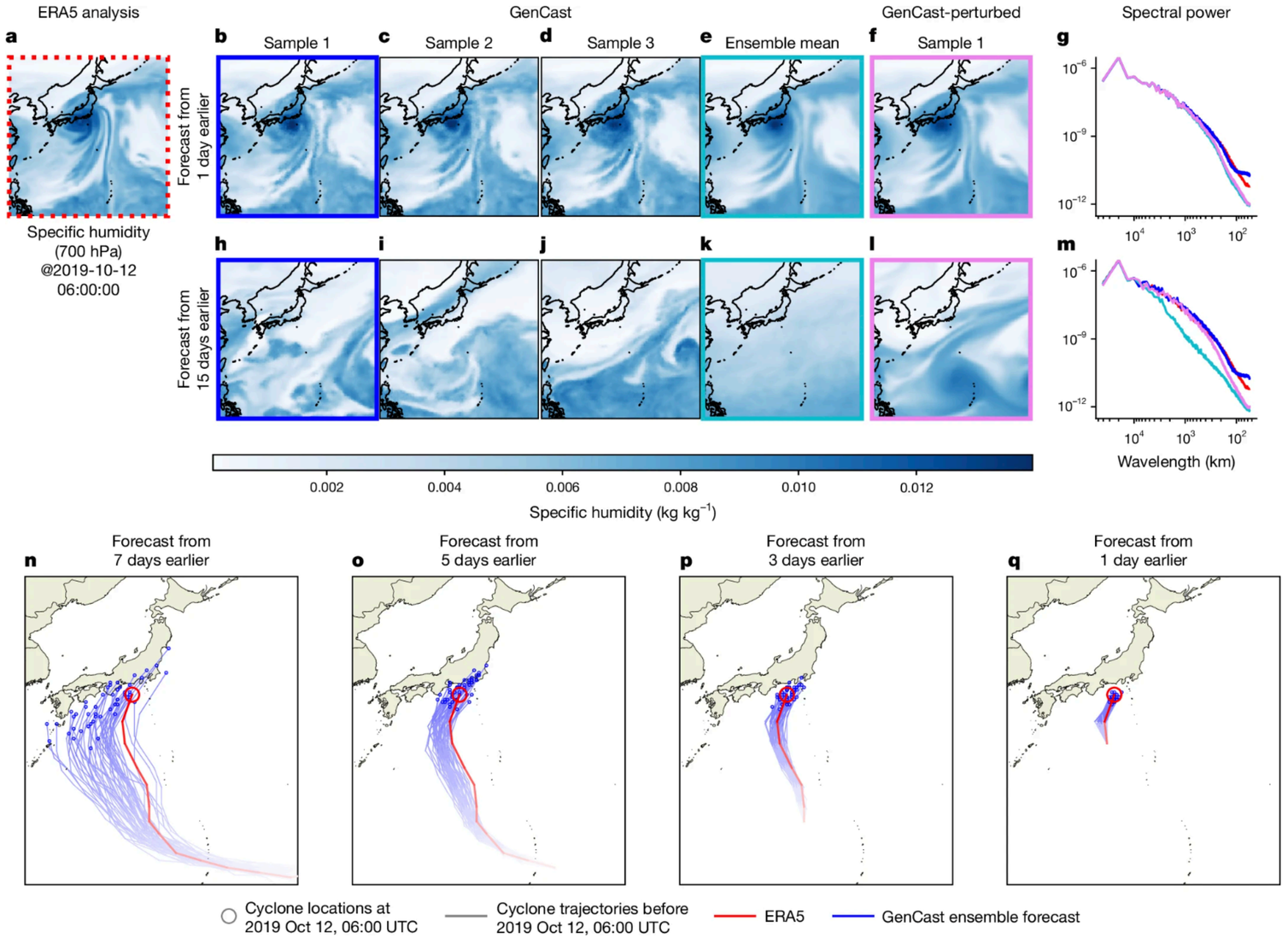
Diffusion models: a sandbox for learning

Nisha Chandramoorthy

The University of Chicago

nishac@uchicago.edu

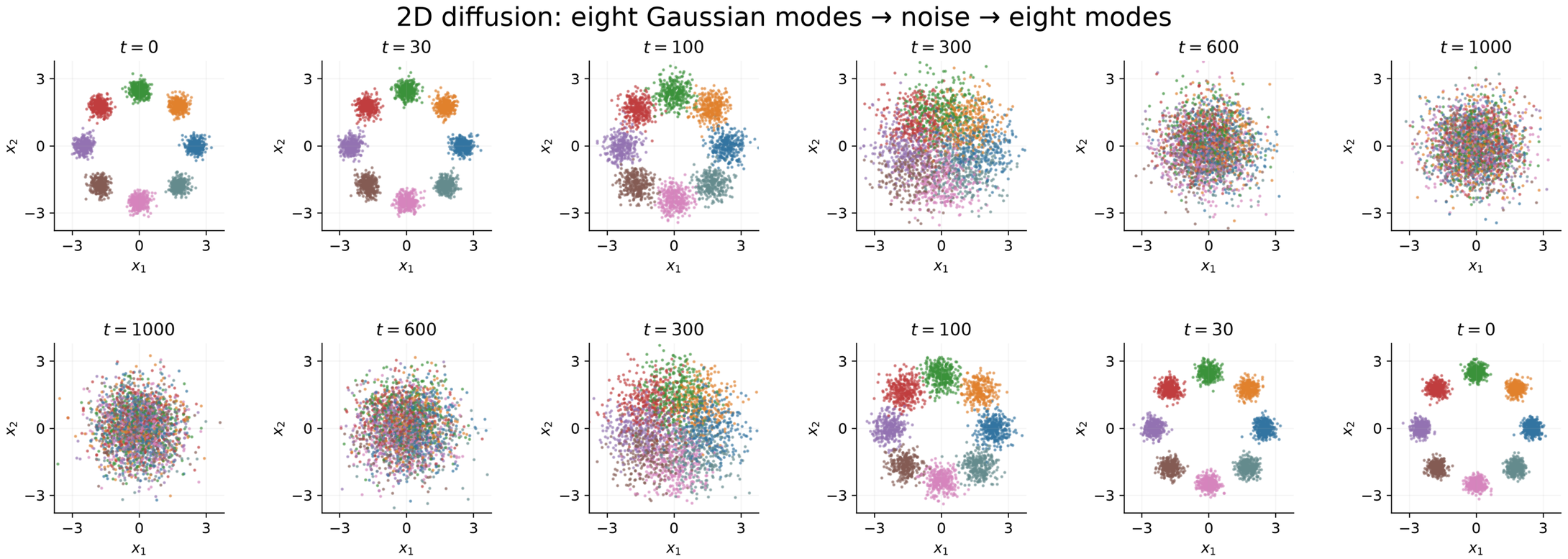
Verifying conditional generative models

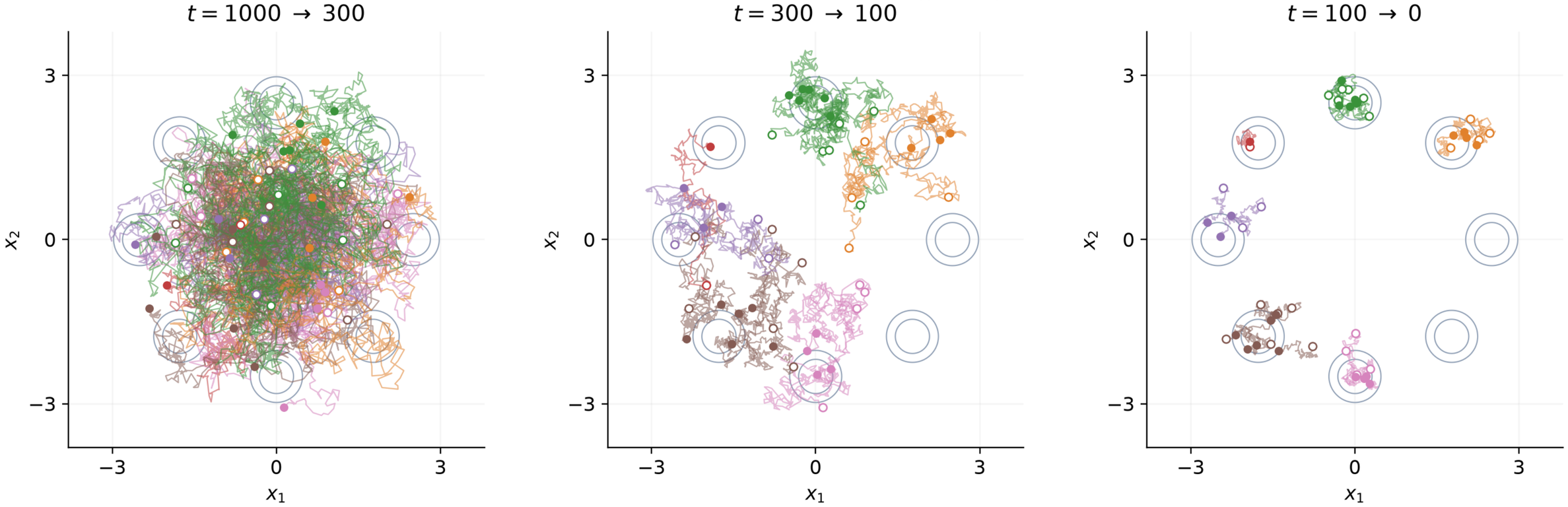


- $X_{\tau}^{(i)} \sim p_{\theta,0}(\,\cdot\,|X_{\tau-1}^{(i)}, X_{\tau-2}^{(i)})$
- 87 million variables
- Not all marginals are accurate, evaluated on complex tasks and statistics

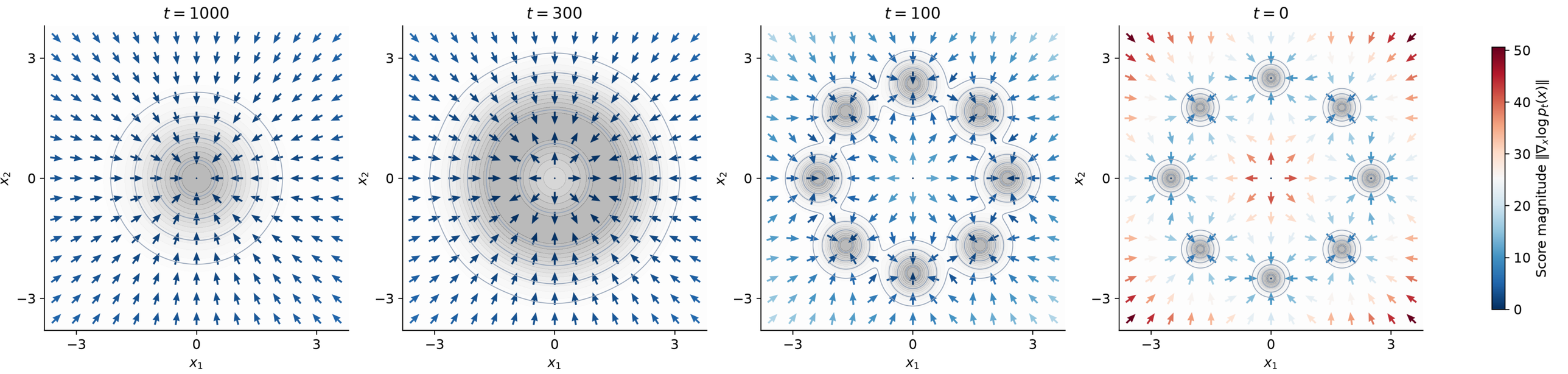
Price et al Nature 2025

Score based generative models [Sohl-Dickstein et al 2015, Song Ermon 2019]





$s_\theta(t, x; X_{\tau-1}, X_{\tau-2}) \approx \nabla_x \log p_{\theta,t}(x; X_{\tau-1}, X_{\tau-2})$ governs the sampling



$s_\theta(t, x; X_{\tau-1}, X_{\tau-2}) \approx \nabla_x \log p_{\theta,t}(x; X_{\tau-1}, X_{\tau-2})$ governs the sampling

At t=0, points toward training samples; at larger noise scales, represents features of the target distribution

Probing the score functions at various noise scales can support learning from emulators

Evaluation in high dimensions: theory

- Convergence results are known in W2/TV distances [Gao, Huang, Jiao and Sheng 2024, Silveri, Conforti, Durmus 2024, Li et al 2024, Bansal et al. (2024), Cheng, Lu, Tan, Xie (2024), Chen et al 2023, 2024, Lee Lu Tan 2023, De Bortoli 2022]
- Integral probability metrics: $\sup_{f \in \mathcal{F}} | E_{x \sim p_0} f(x) - E_{x \sim \hat{p}_{\text{gen}}} f(x) |$
- Minimax error rates to verify: $\mathcal{O}(n^{-1/D})$, not computable!
- In practice, useful to restrict $\mathcal{F}$ to some finite dimensional, known feature space
- Metrics you care about may not be expectations on $\mathcal{F}$

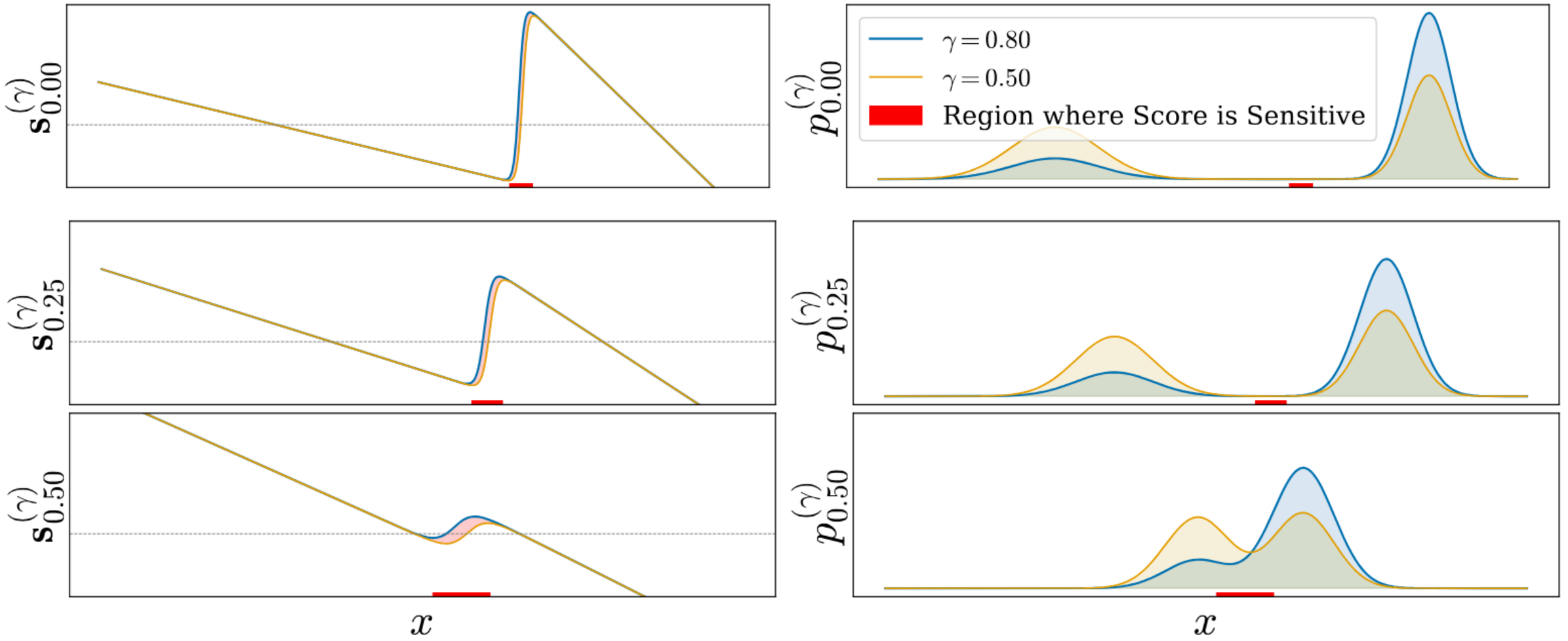
Downstream learning tasks from diffusion models

- When can a GM support accurate solutions of inverse problems? (with Andrew Dennehy, Ramachandran Muthukumar and Rebecca Willett)
- When can a GM accurately model the distribution of an observable?

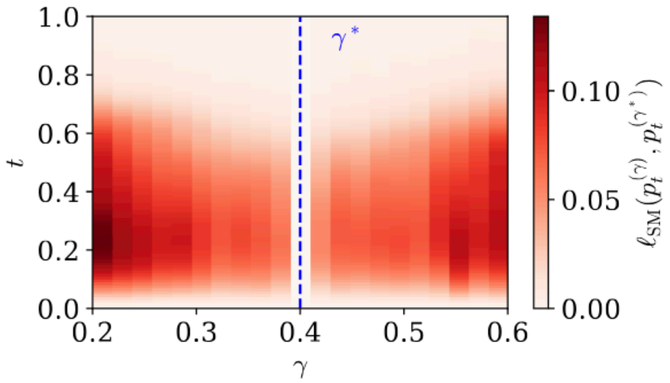
Learning weights of a multimodal distribution

- $p_0 \equiv p^{(\gamma^*)} = \gamma^* q_a + (1 - \gamma^*) q_b$
- Estimate the scalar γ^* from the generated data.
- How well can any parameter be recovered from diffusion-generated samples?
- Applying to weights, when is mode collapse expected? Can rare modes be generated with the correct probabilities?
- Score estimation insufficient to predict weights [Li, van Breugel, and der Schaar 2024; Koehler Hecket Risteski 2023; Chen Chewi Niles-Weed 2021; Chewi Kalavasis Mehrotra Montasser 2025].

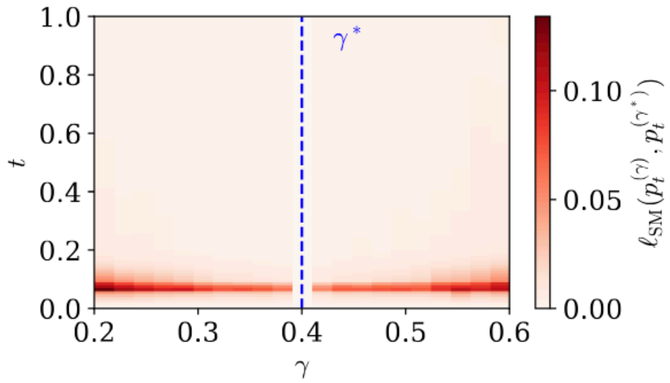
The region of sensitivity is larger for intermediate scores than the target score



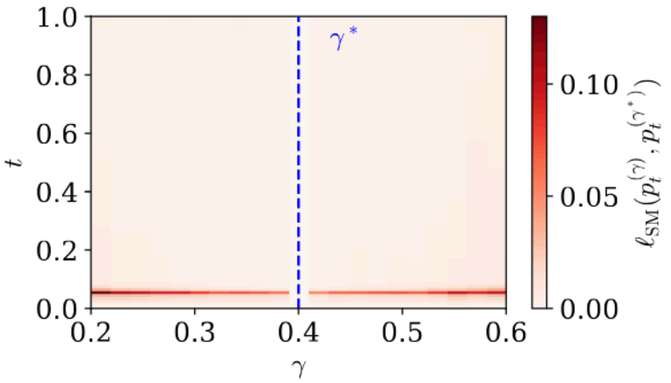
Diffusion score sensitivity determines parameter recovery



$L(\gamma^*) = 1.51$



$L(\gamma^*) = 0.26$



$L(\gamma^*) = 0.15$



$\hat{\gamma} = 0.40$

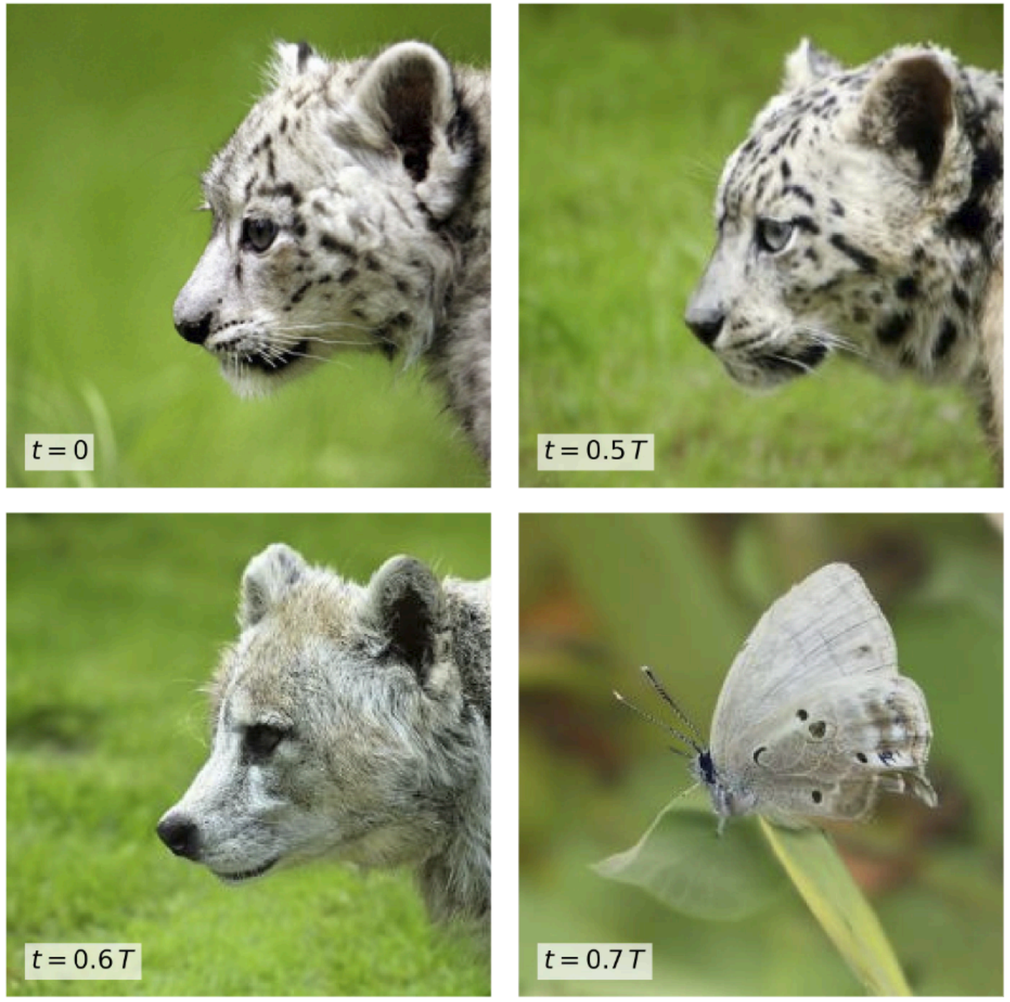


$\hat{\gamma} = 0.37$



$\hat{\gamma} = 0.28$

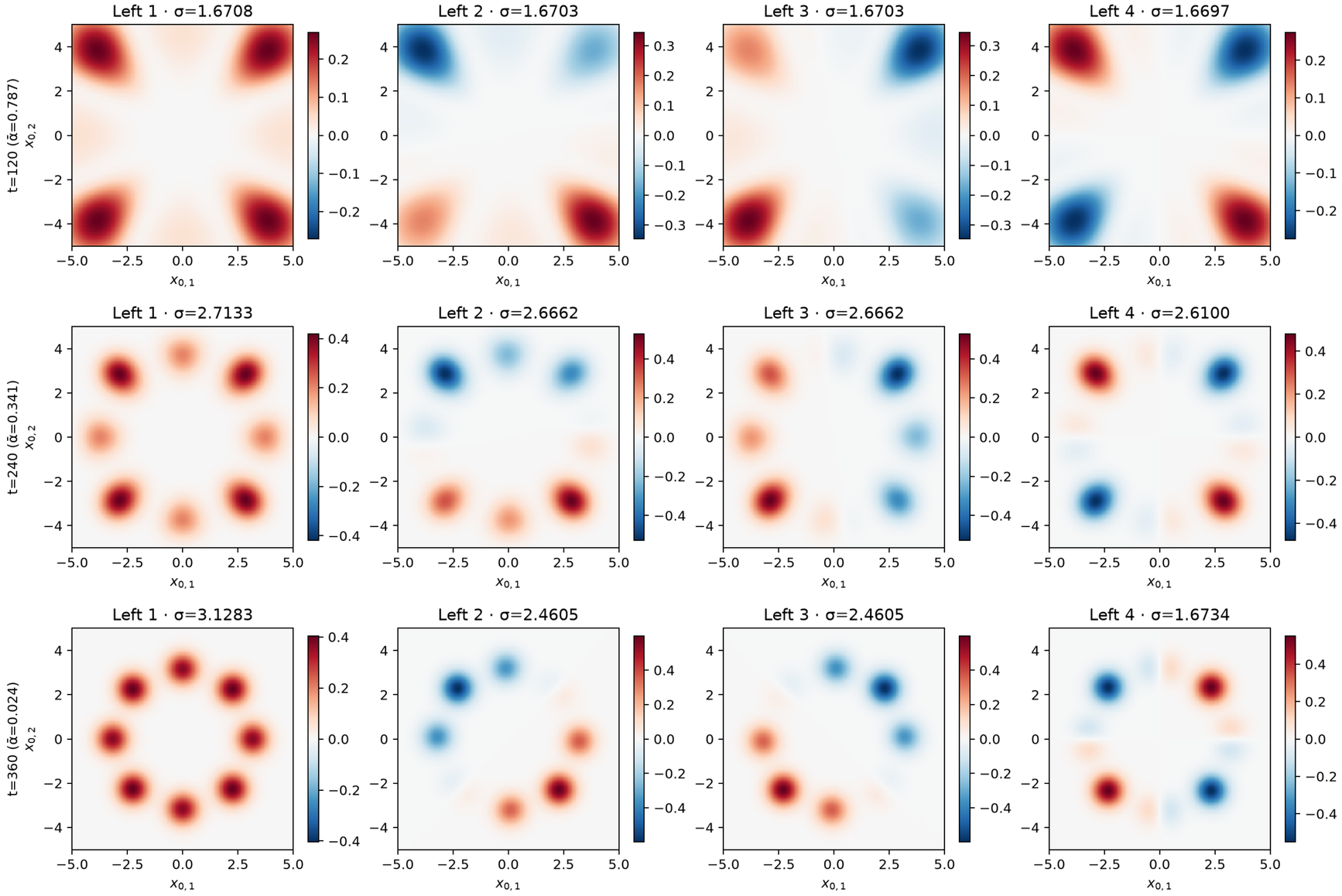
Noising-denoising to reveal the structure of classes



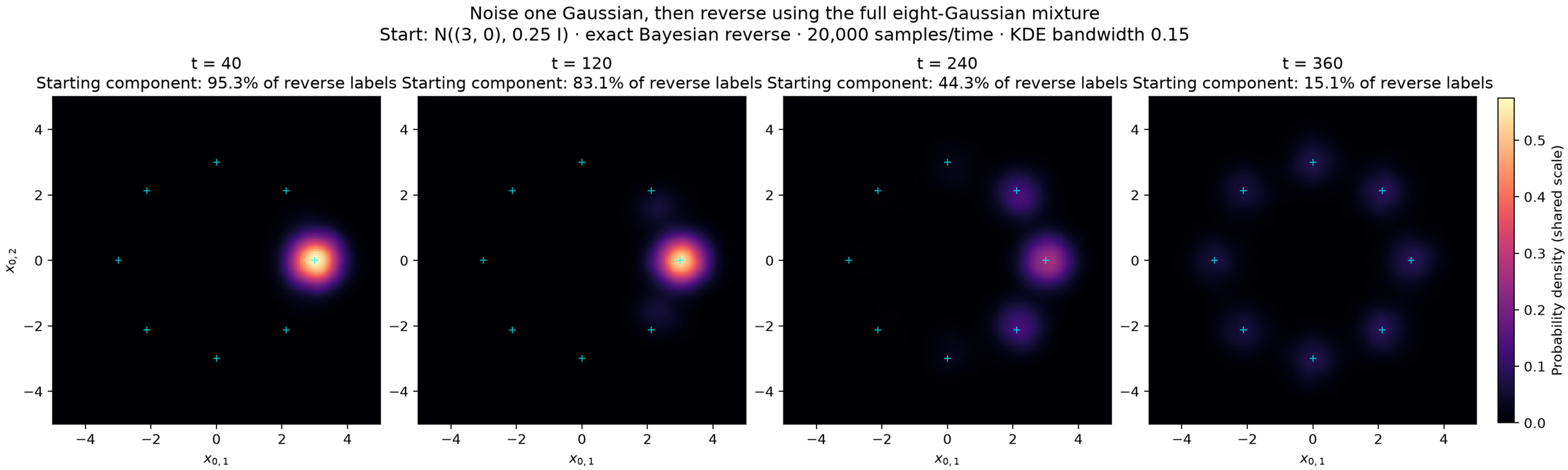
- Denoising operator, a linear operator on probability distributions
- $\mathscr{A}_t \mathscr{N}_t p_0 = \mathscr{A}_t p_t = p_0$
- Two ways of probing structure:
  - Consider perturbations of p_t
  - Obtain some q_t by noising a desired distribution q_0. What is $\mathscr{A}_t q_t$?

Sclocchi, Favero and Wyart 2024

Leading singular vectors of the denoising operator are large-scale perturbations



Denoising after a short time leads to more samples from the same mode



Downstream learning tasks from diffusion models

- When can a GM support accurate solutions of inverse problems? (with Andrew Dennehy, Ramachandran Muthukumar and Rebecca Willett)
- When can a GM accurately model the distribution of an observable?

https://arxiv.org/abs/2607.15485