Embeddings in the 2-Wasserstein space

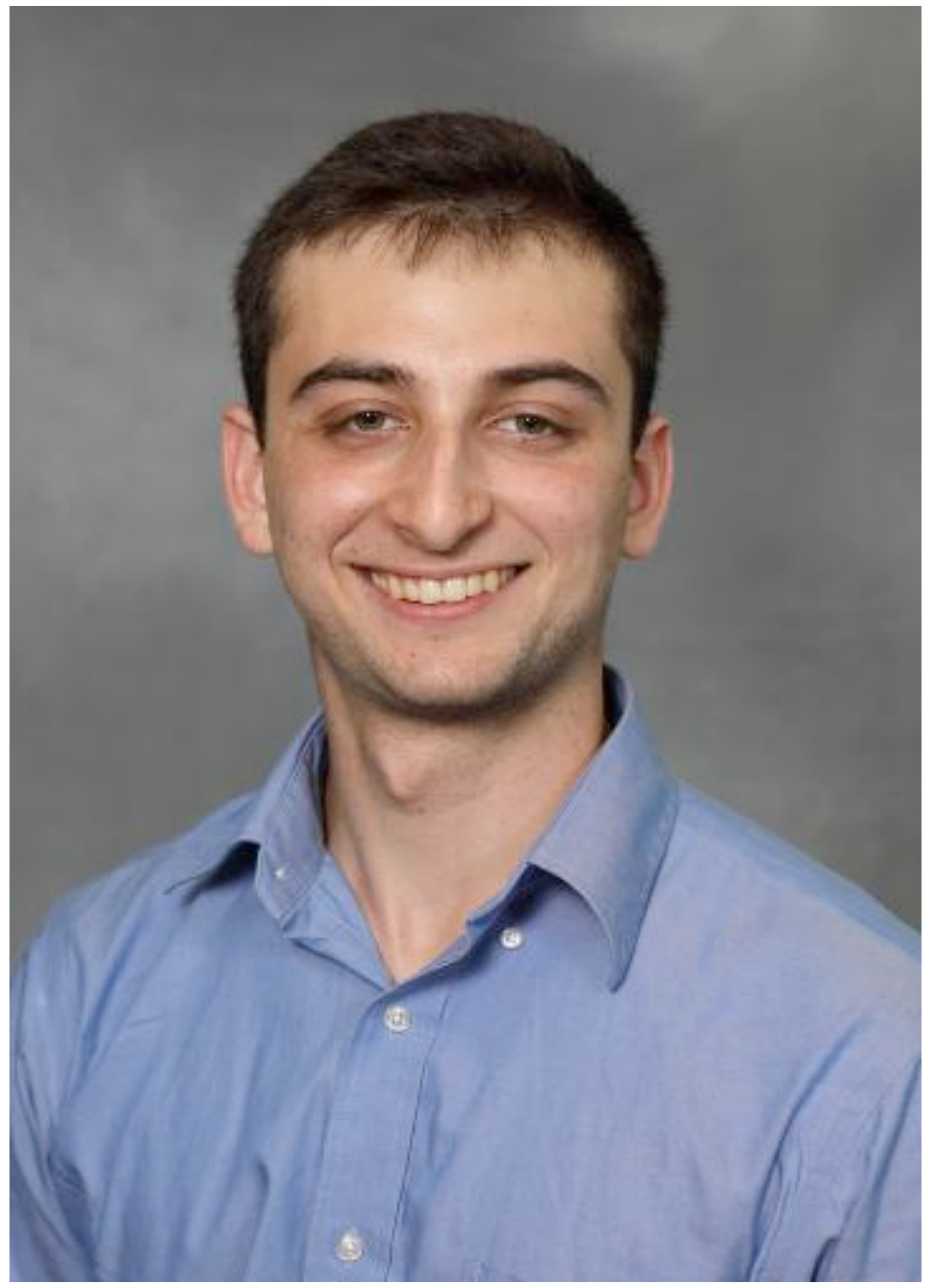
Maria Oprea, MOPTA August, 2024

Joint work with:

Yunan Yang

Goenka Family Assistant Professor in Mathematics[1]

Jonah Botvinick-Greenhouse



PhD candidate at the Center for Applied Mathematics[1]

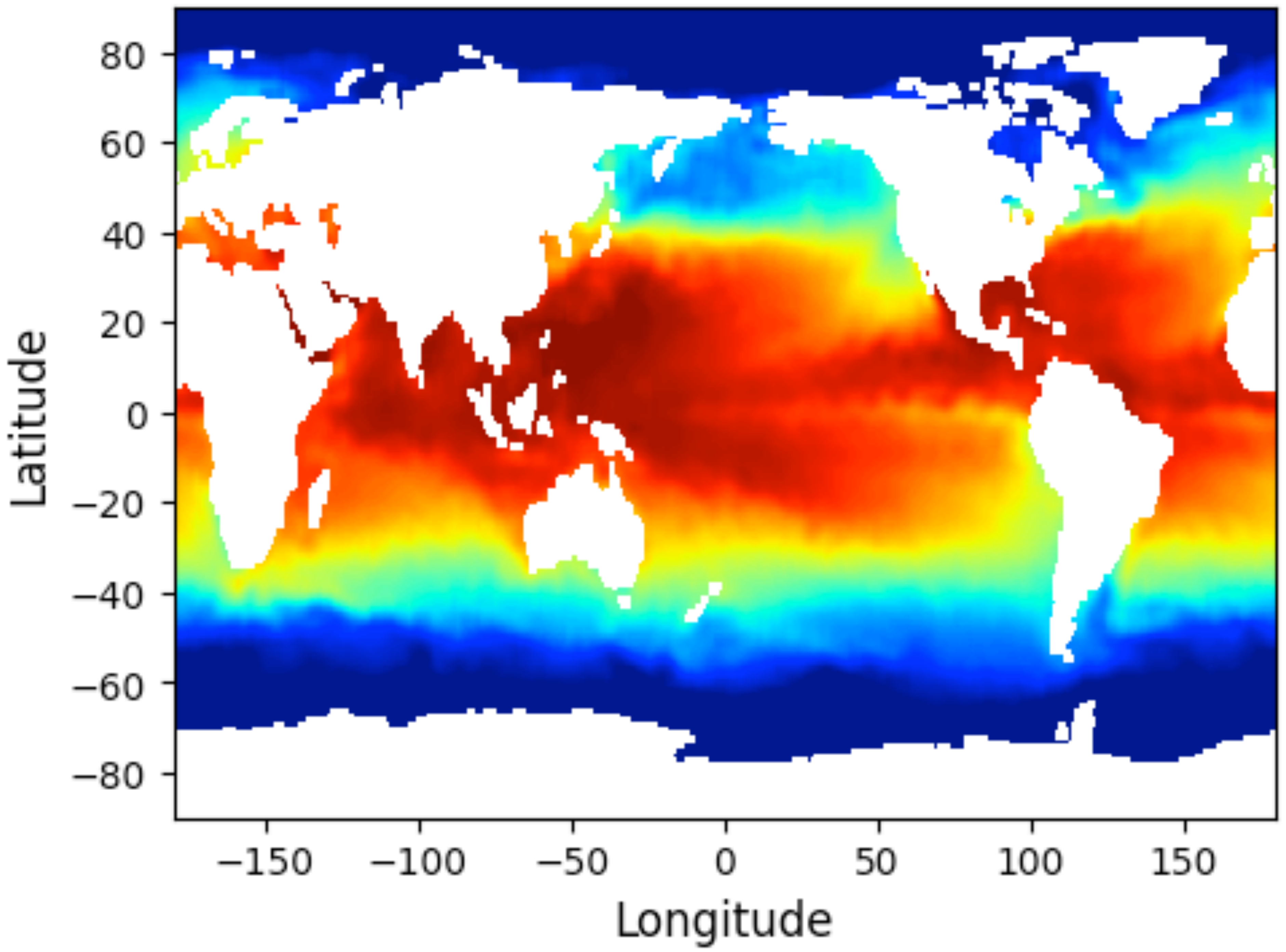
Romit Maulik

Assistant Professor at Penn State College of Information Sciences and Technologies

[1]Cornell University

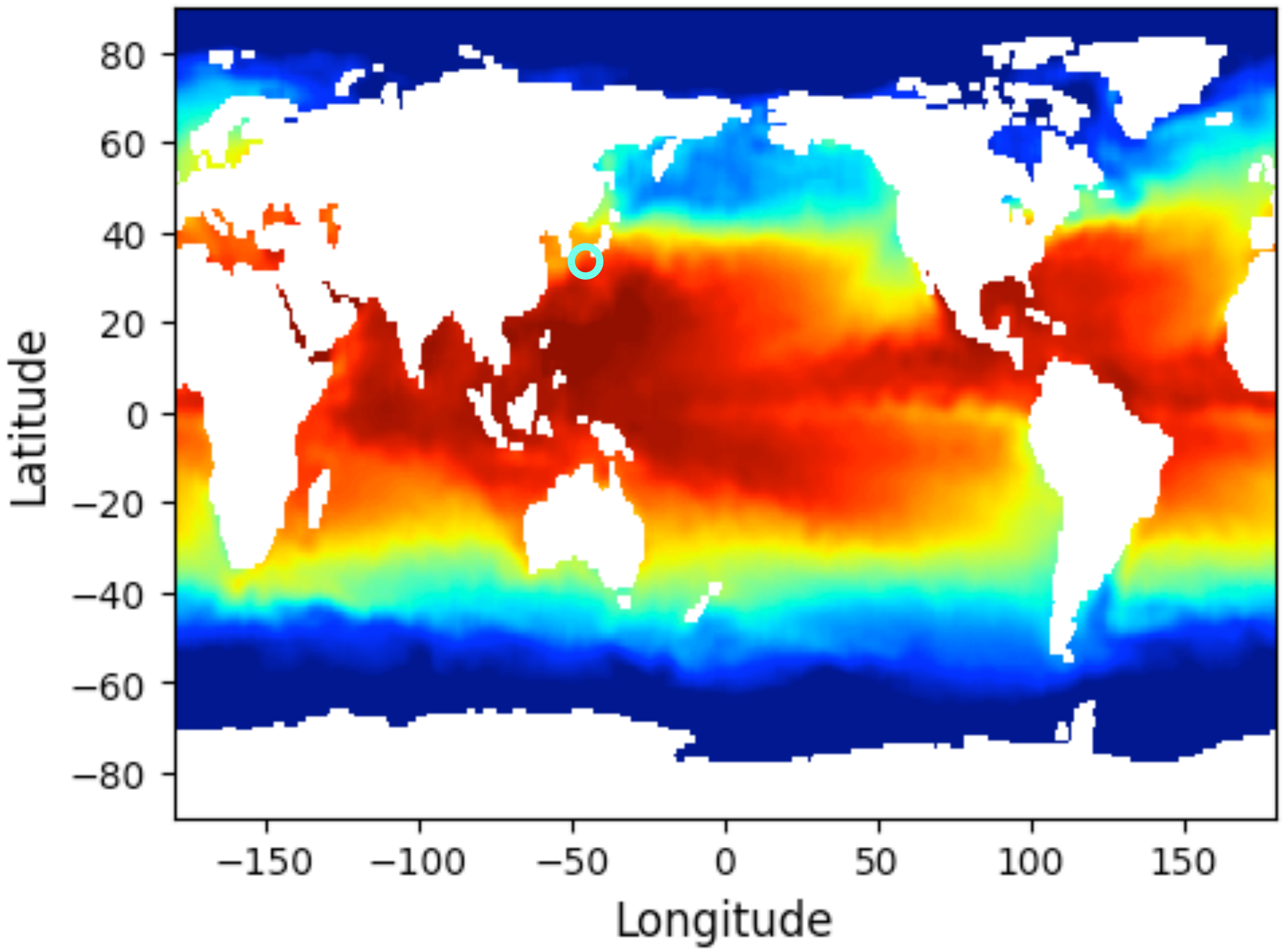
Motivation

Sea surface temperature (NOAA)



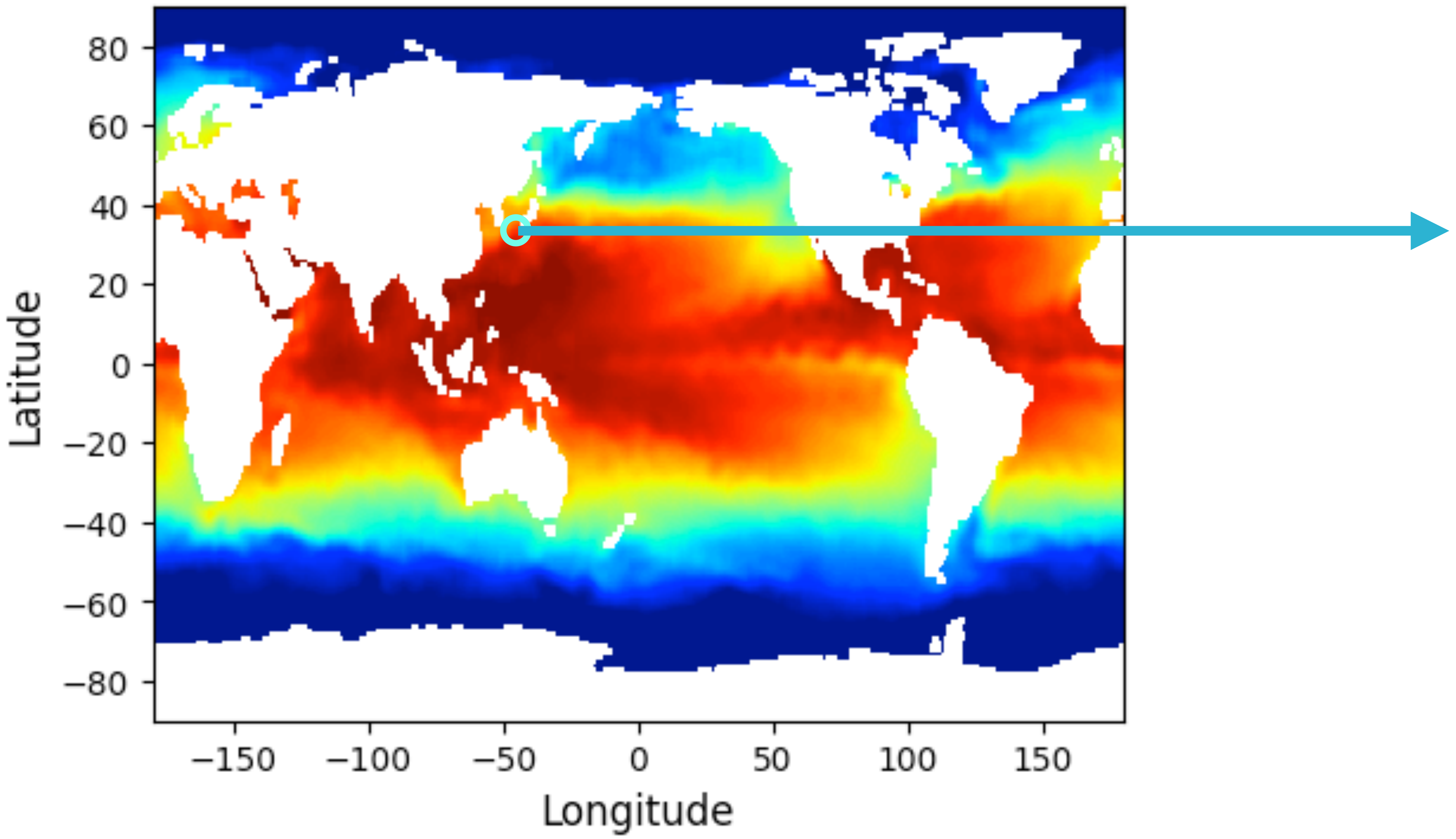
Motivation

Sea surface temperature (NOAA)



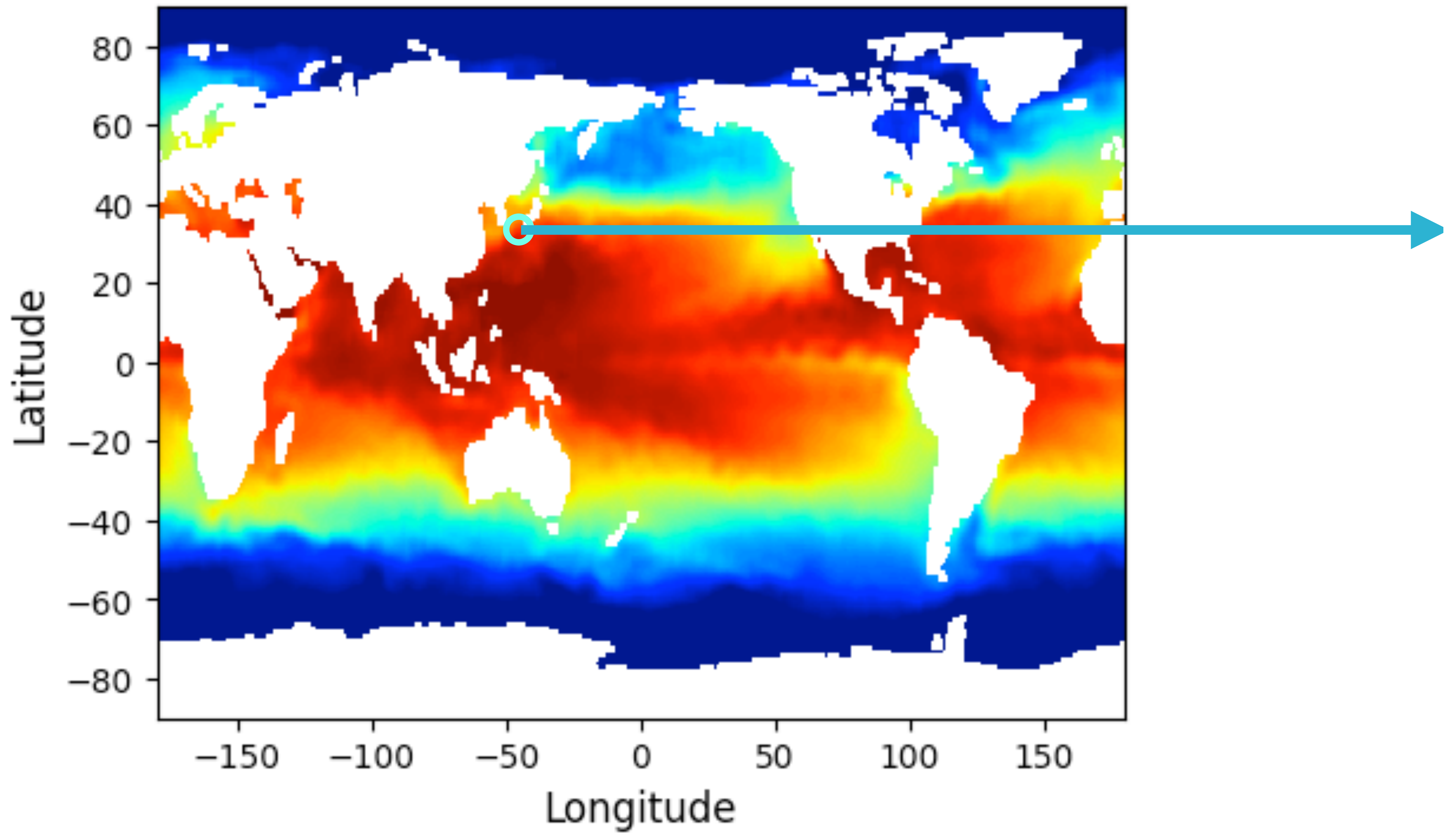
Motivation

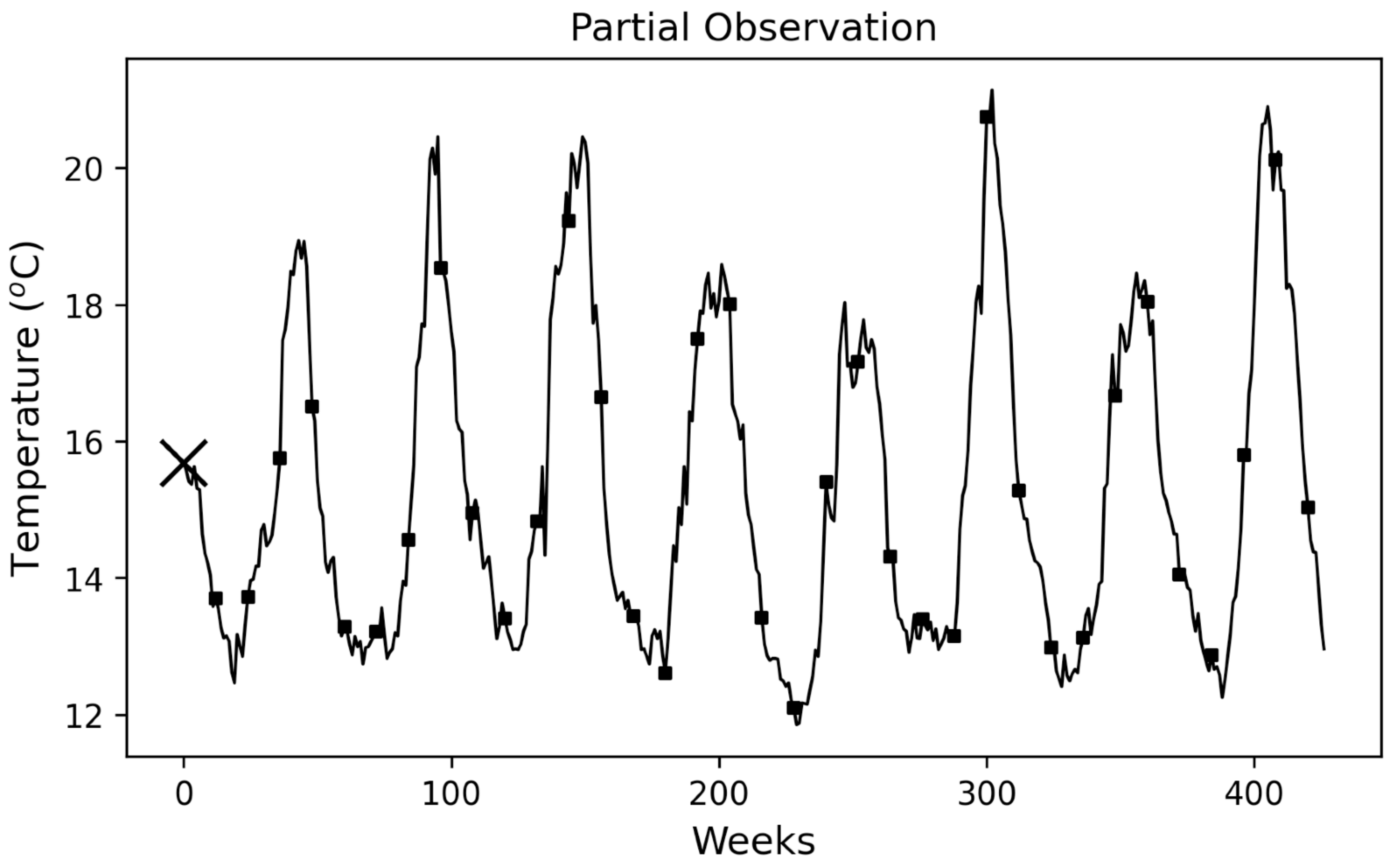
Sea surface temperature (NOAA)



Motivation

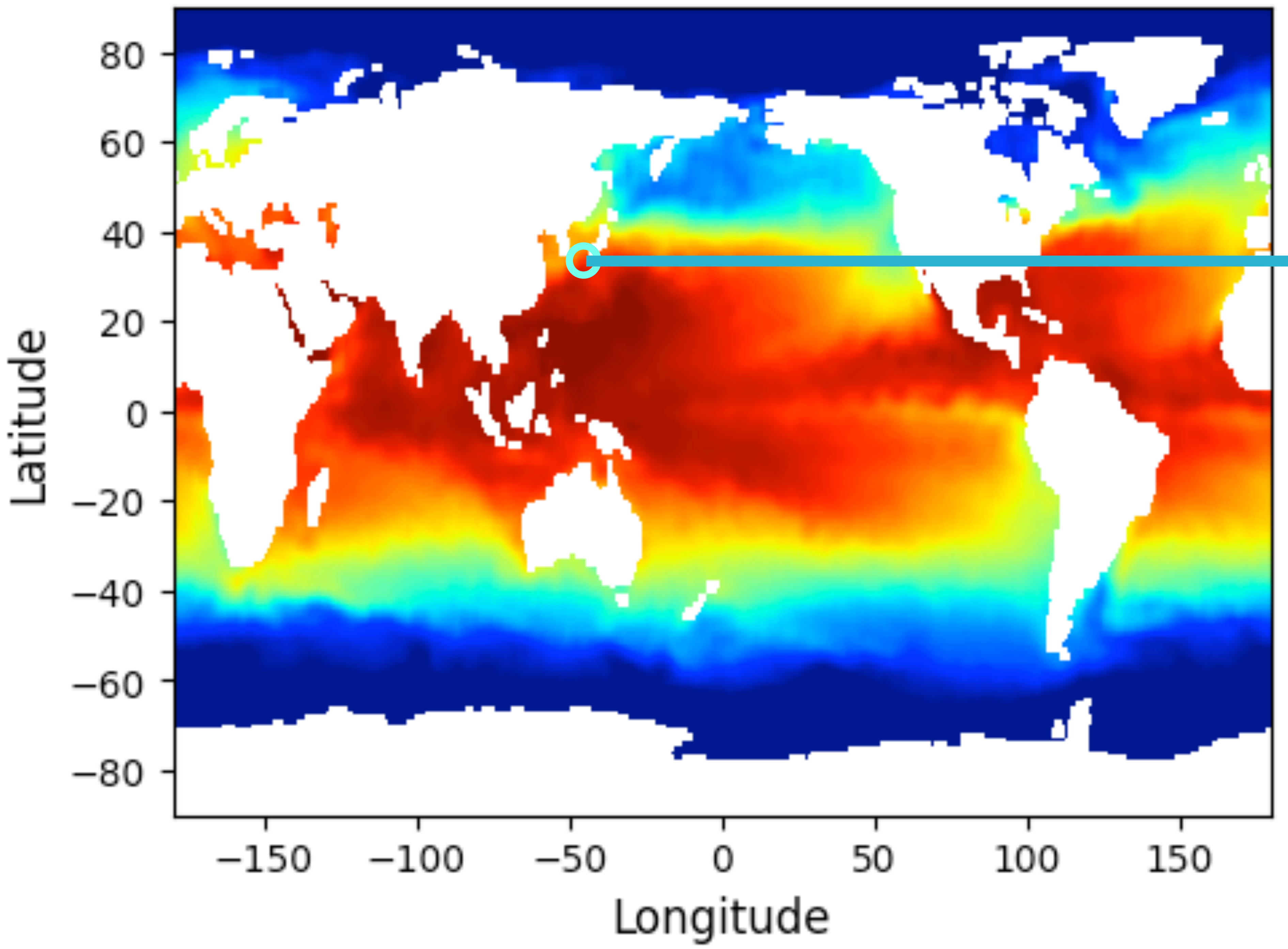
Sea surface temperature (NOAA)

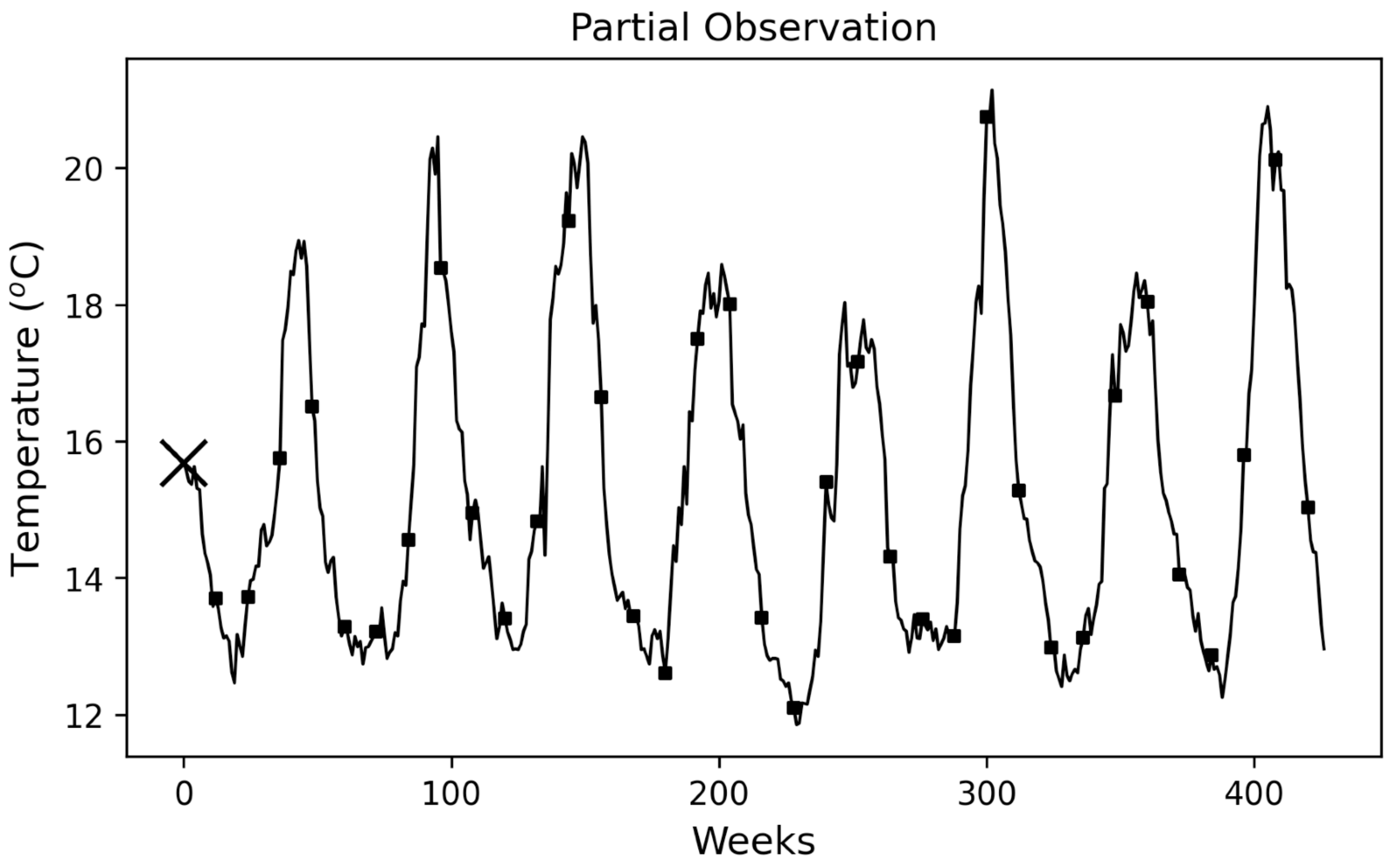




Motivation

Sea surface temperature (NOAA)

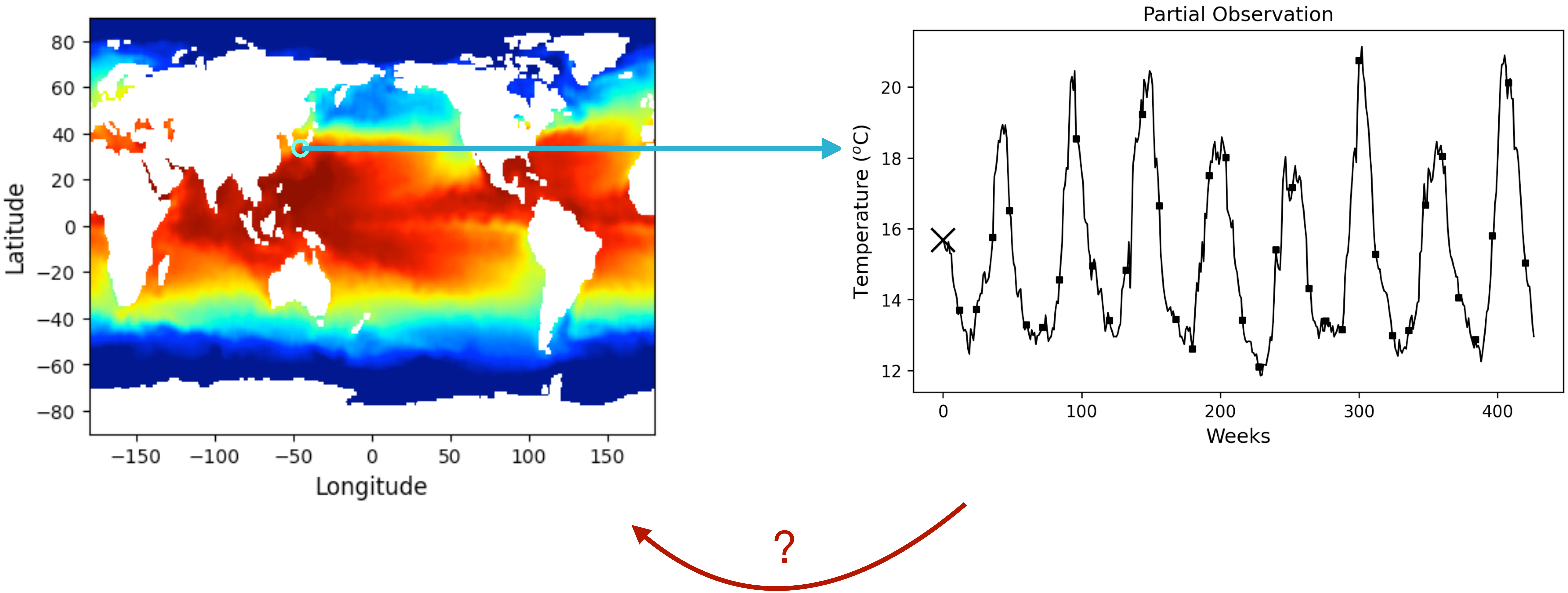




?

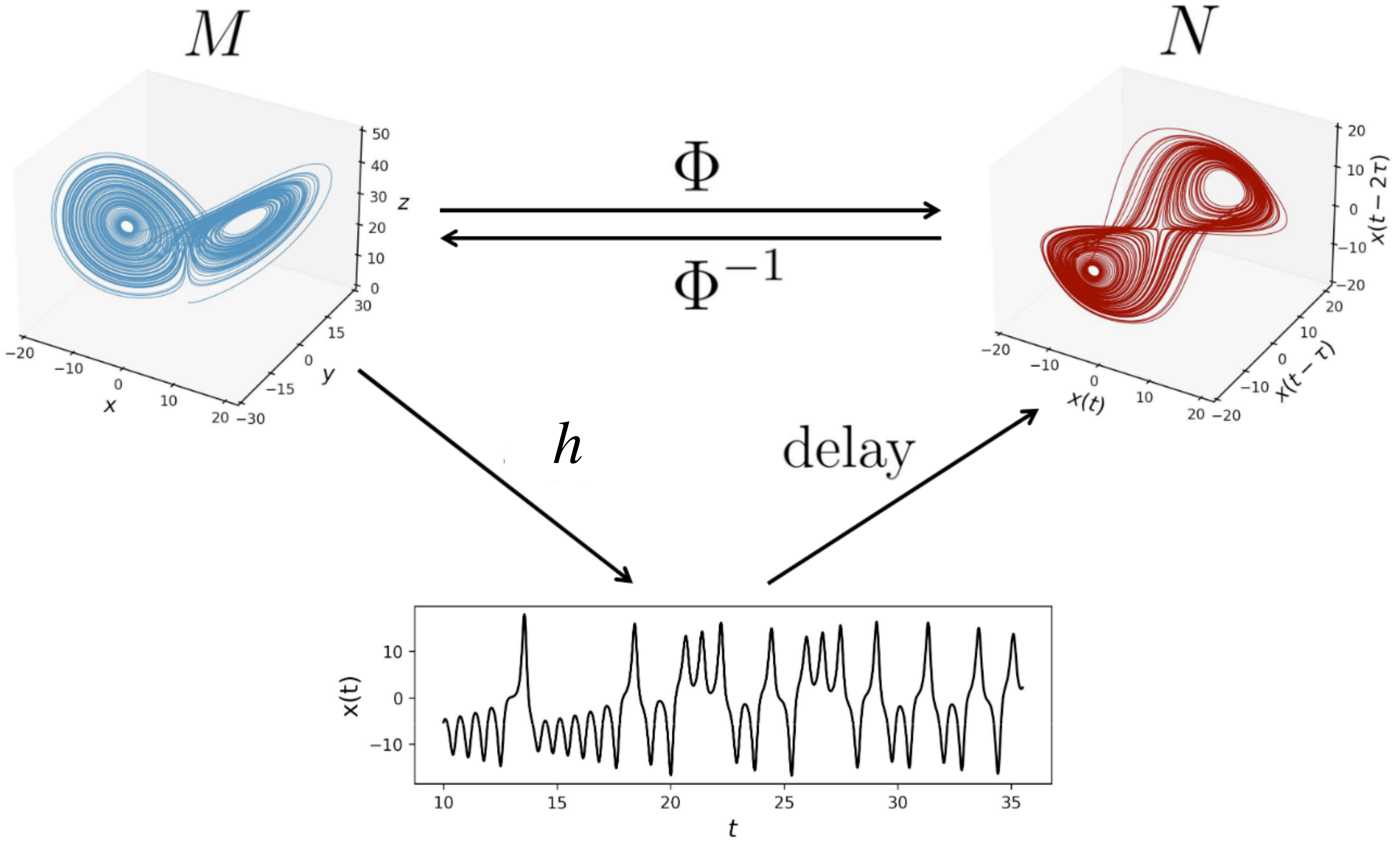
Motivation

Sea surface temperature (NOAA)



- Takens embedding theorem: there is an embedding between the true attractor and the delay reconstruction

$$\Phi_{h,\phi_t}(x) := (h(x), h(\phi_\tau(x)), \ldots h(\phi_{(d-1)\tau}(x))) \in \mathbb{R}^d$$



Embeddings

- Notion of „sameness”

Embeddings

- Notion of „sameness”
- Diffeomorphism that preserves the differential structure

Embeddings

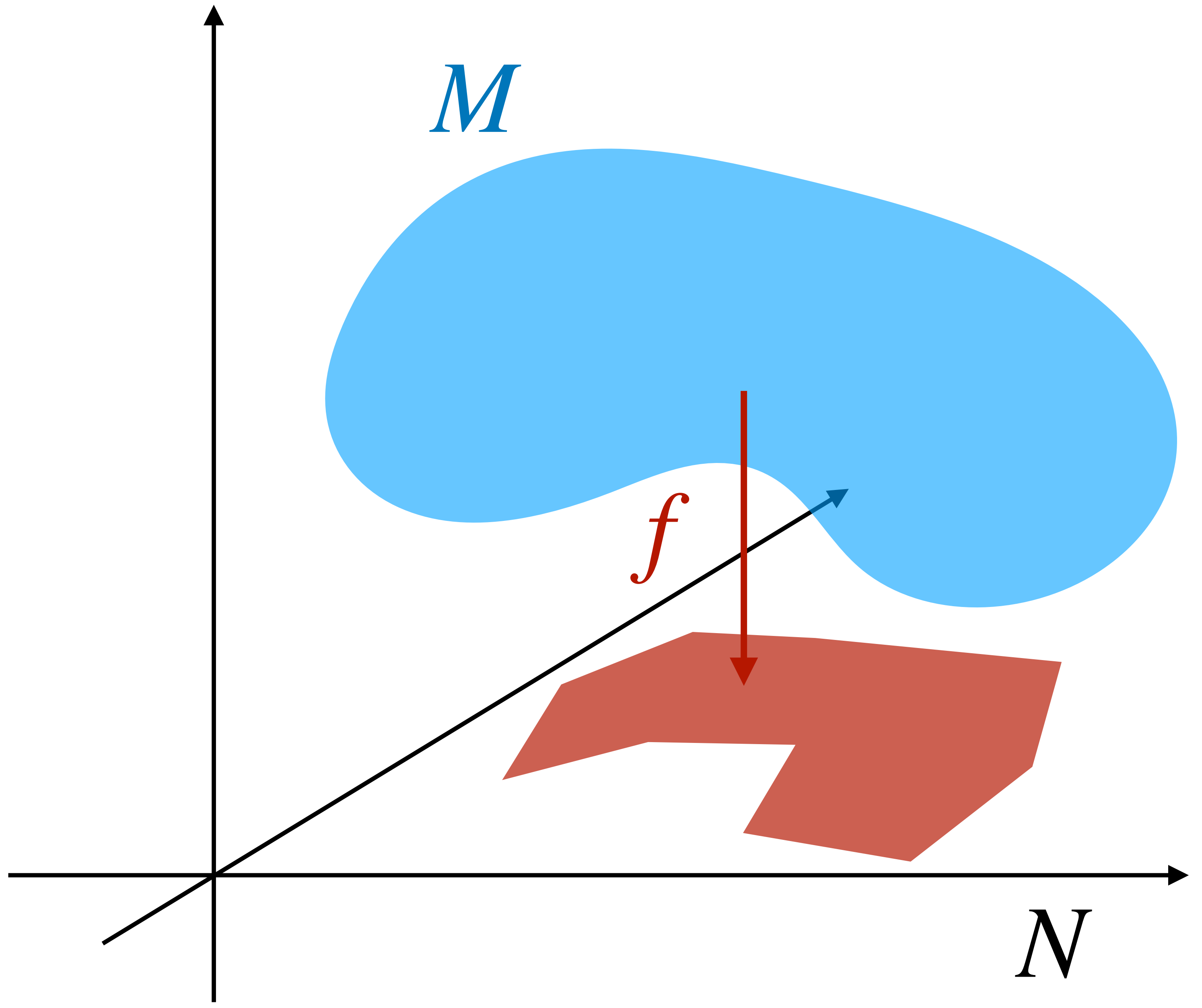
- Notion of „sameness”
- Diffeomorphism that preserves the differential structure

Definition: $f : M \to N$ is an embedding if:

- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$

Embeddings

- Notion of „sameness”
- Diffeomorphism that preserves the differential structure



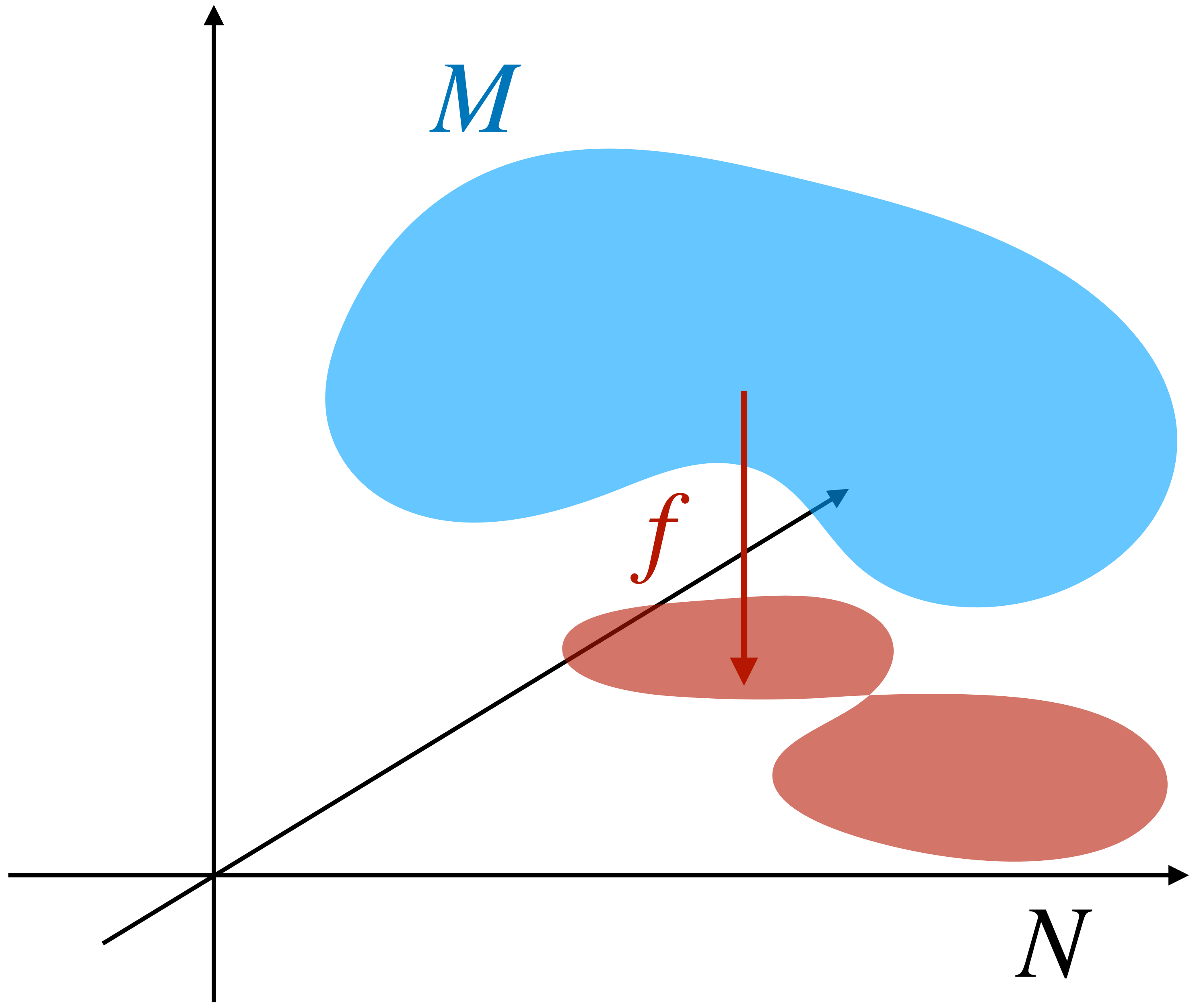
Definition: $f: M \to N$ is an embedding if:

- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$

Embeddings

- Notion of „sameness”
- Diffeomorphism that preserves the differential structure

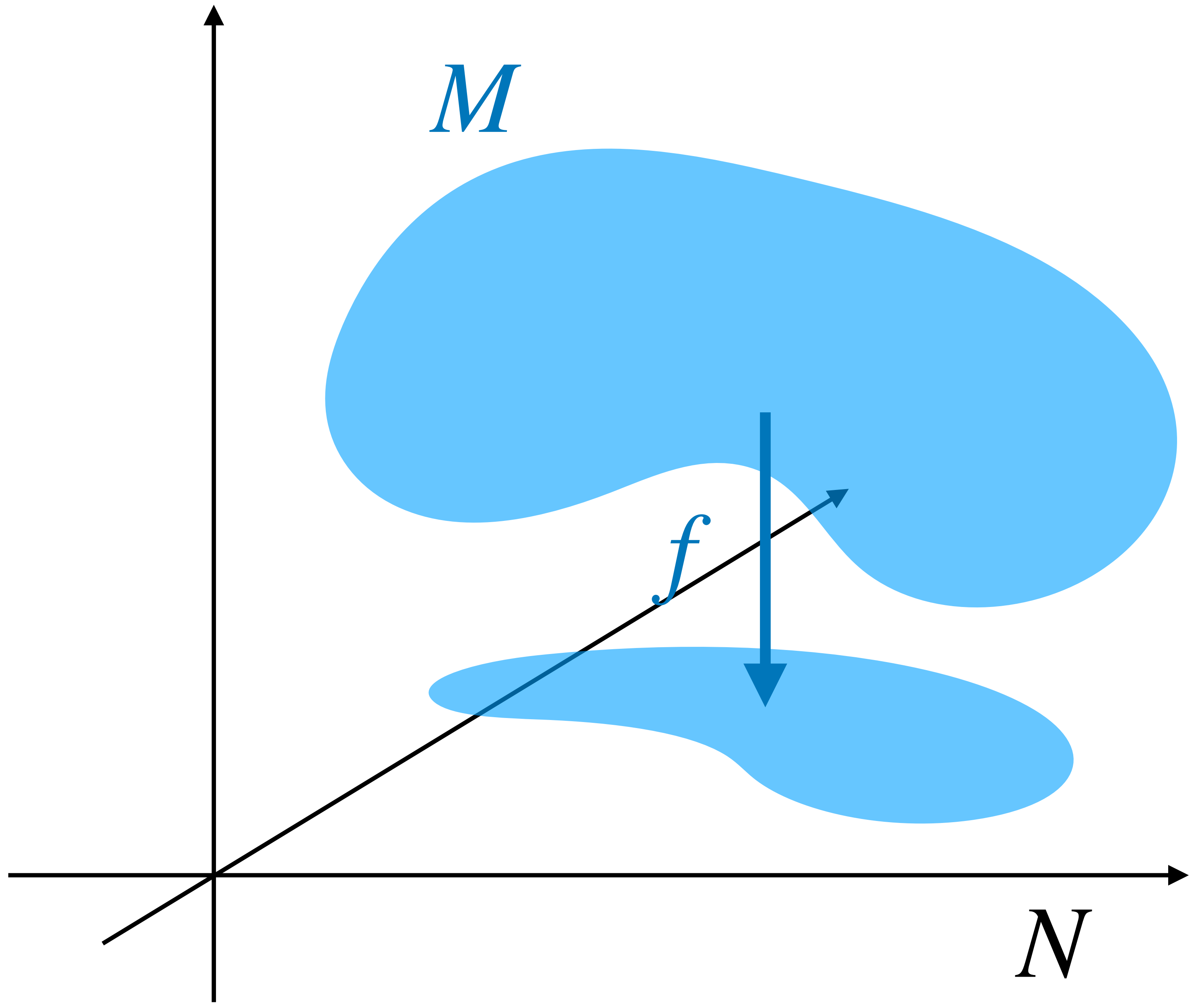
[Definition]: $f : M \to N$ is an embedding if:
- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$



Embeddings

- Notion of „sameness”

- Diffeomorphism that preserves the differential structure



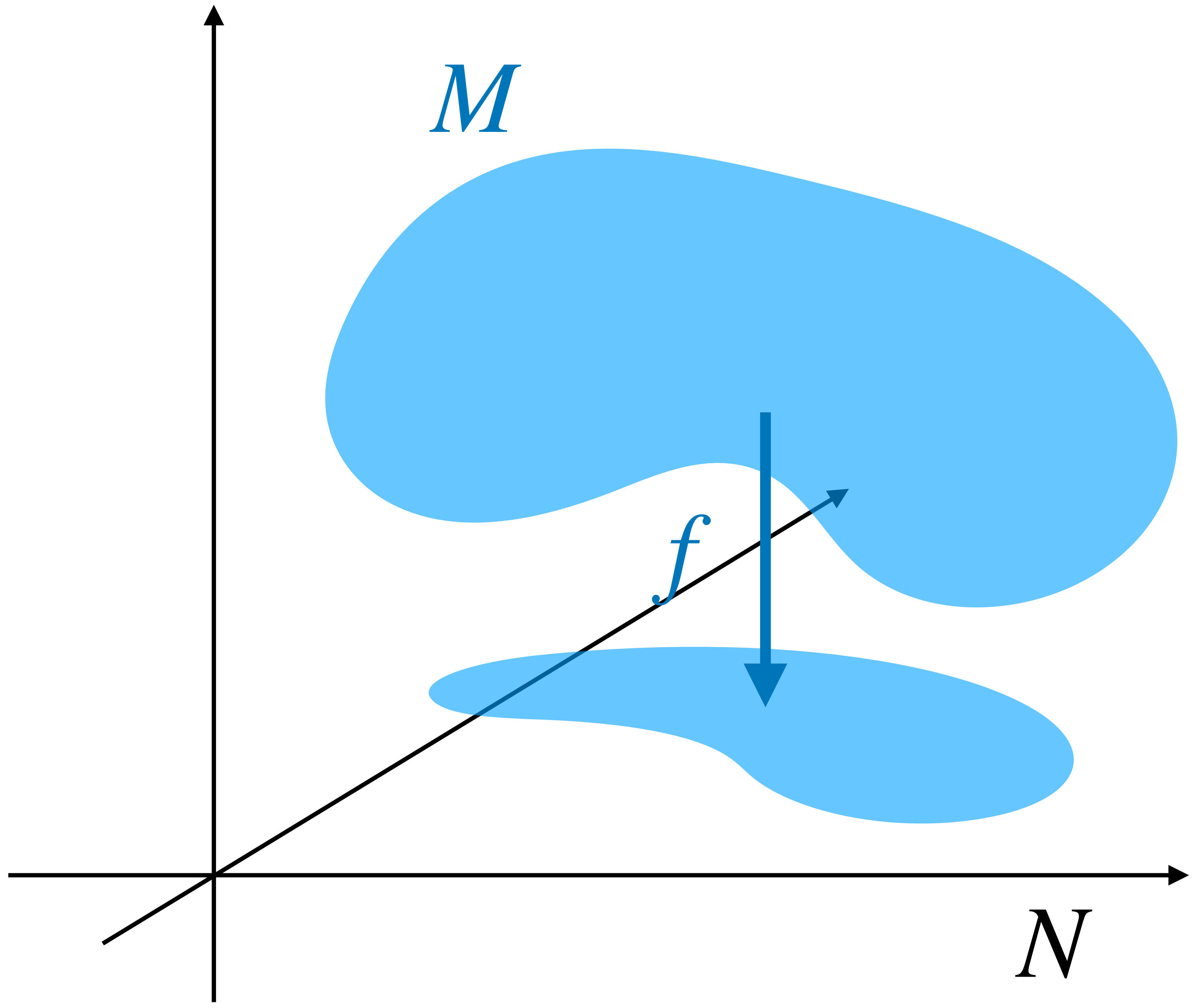
Definition: $f : M \to N$ is an embedding if:
- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$

Embeddings

- Notion of „sameness”
- Diffeomorphism that preserves the differential structure

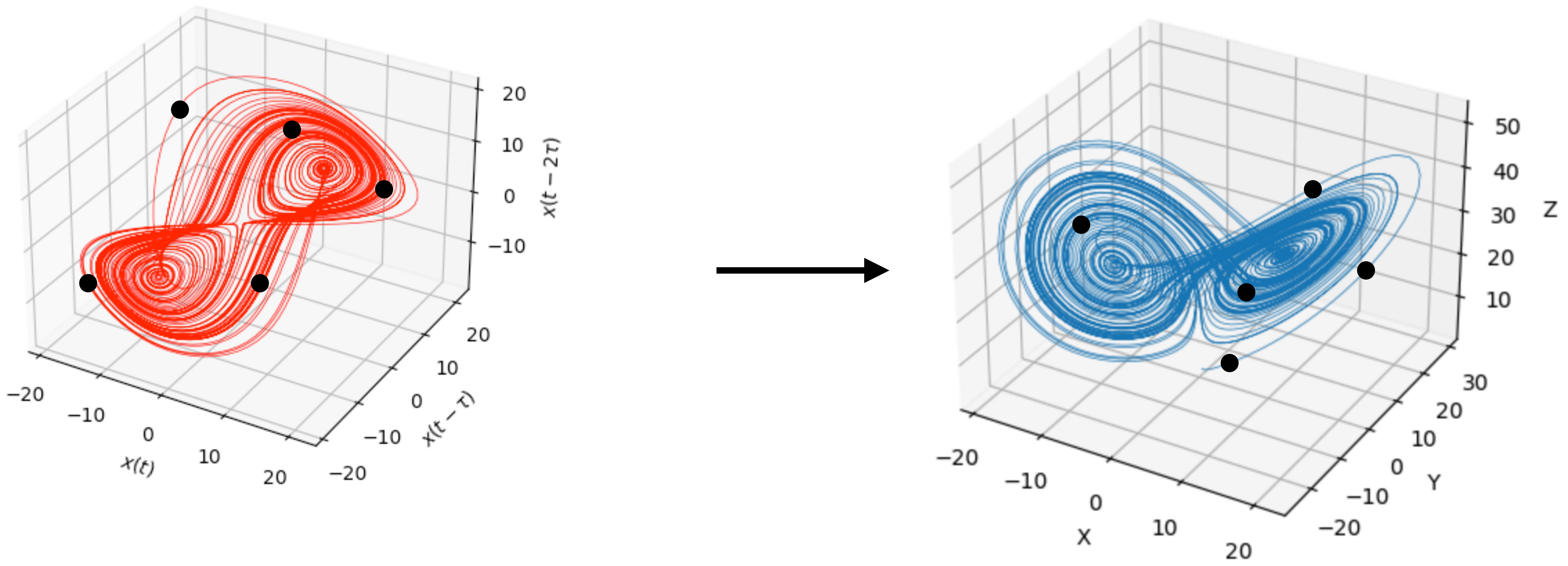
Definition: $f : M \to N$ is an embedding if:
- f is bijective onto $f(M)$
- f is differentiable
- Df_x is injective $\forall x \in M$

- **Goal: learn the embedding from data**



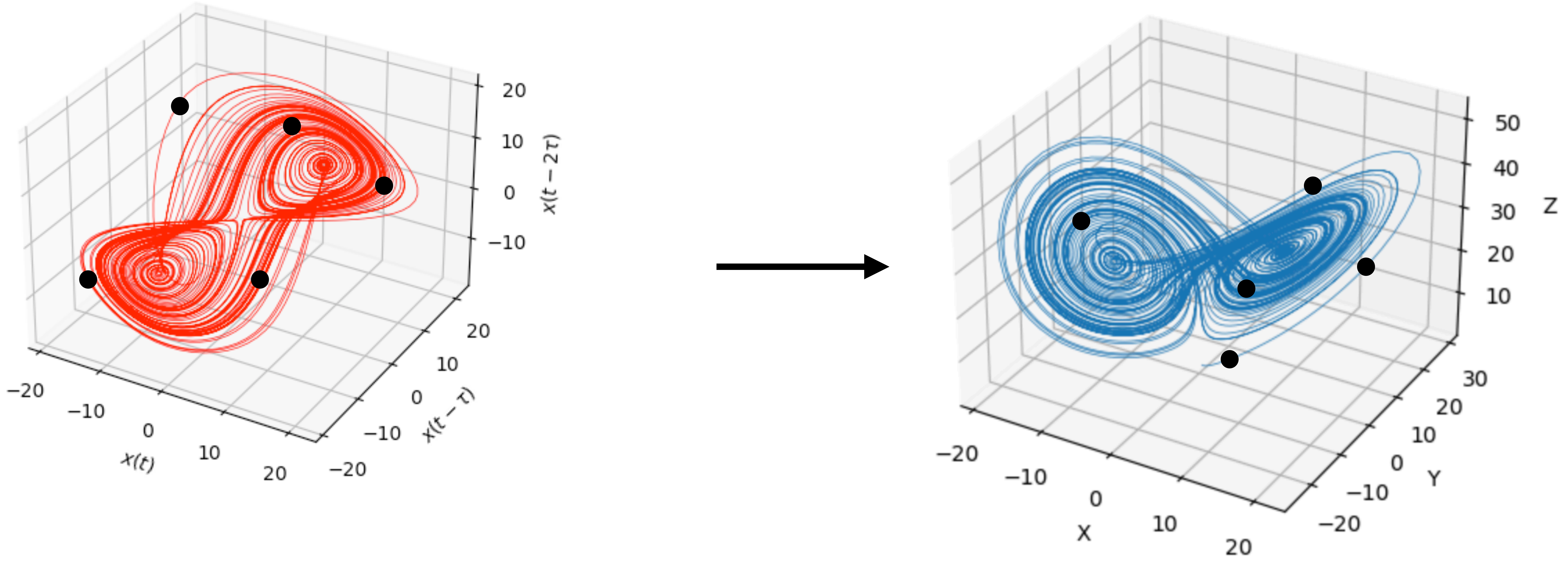
Learning paradigm

- Given samples $\{x_i, y_i = f(x_i)\}_i$ learn f



Learning paradigm

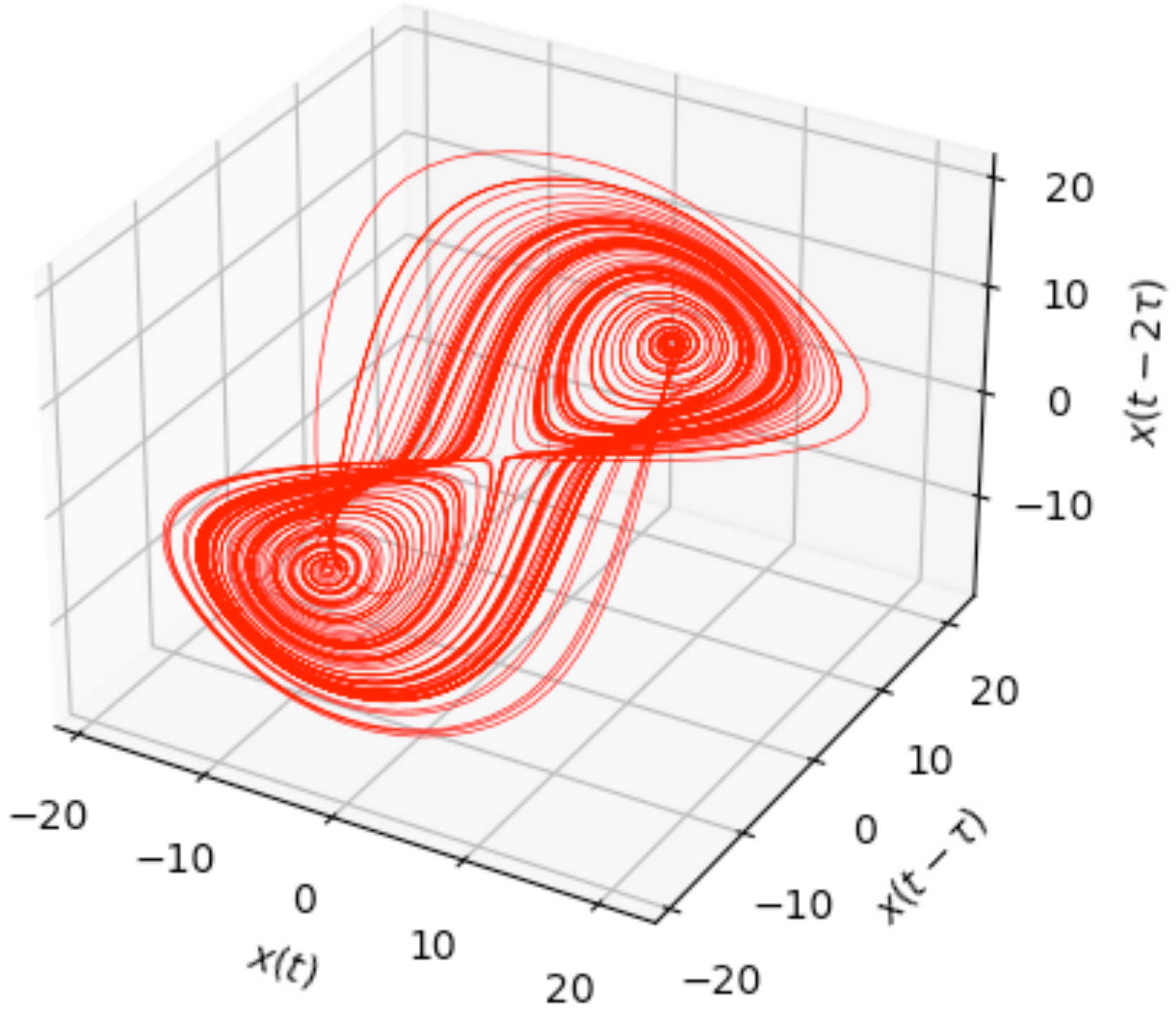
- Given samples $\{x_i, y_i = f(x_i)\}_i$ learn f



- For Takens: long trajectory $\{x_{t_i}, \Phi(x_{t_i})\}$

Limitations

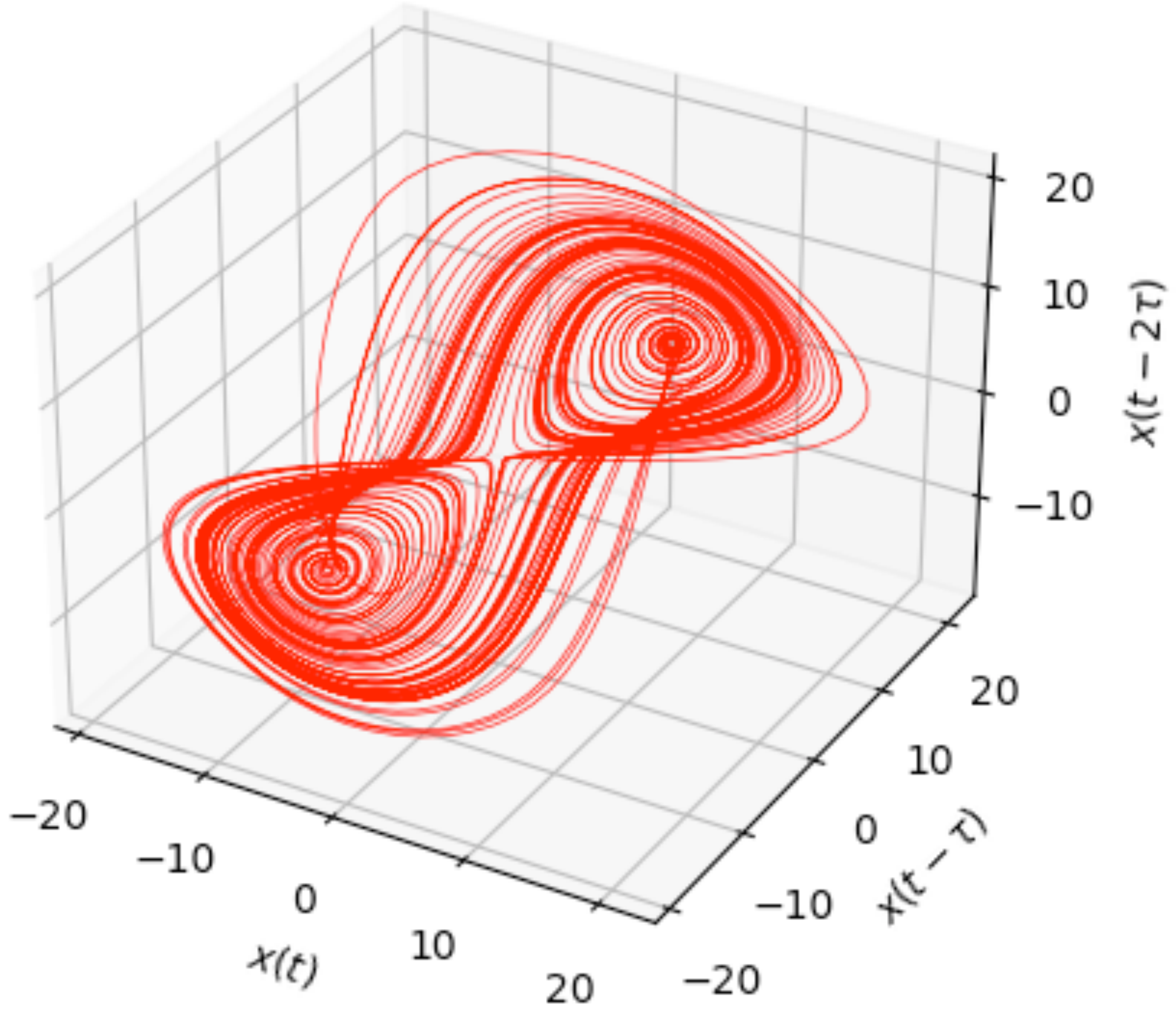
- Takens requires clean data, deterministic setup.



Clean

Limitations

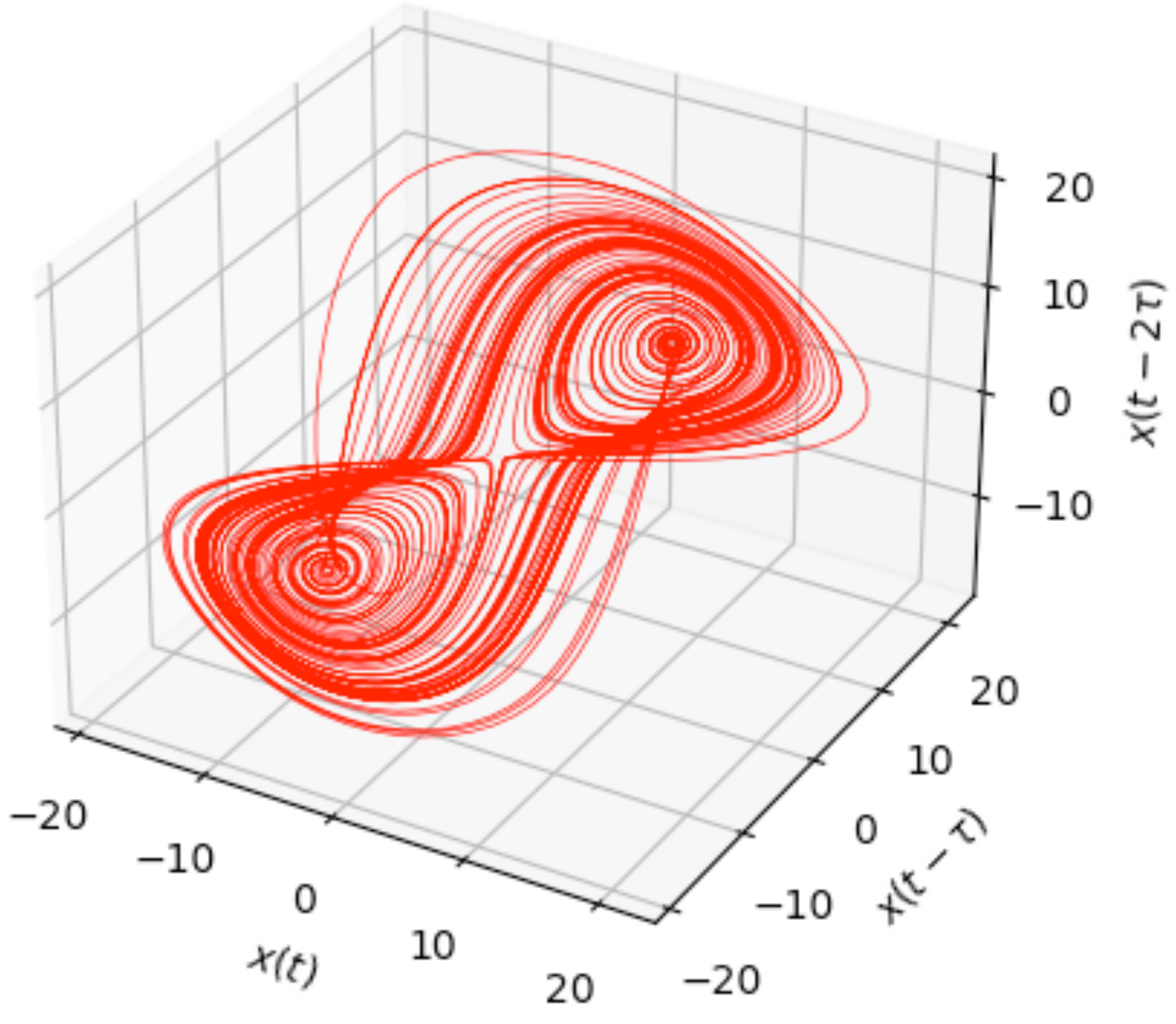
- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$



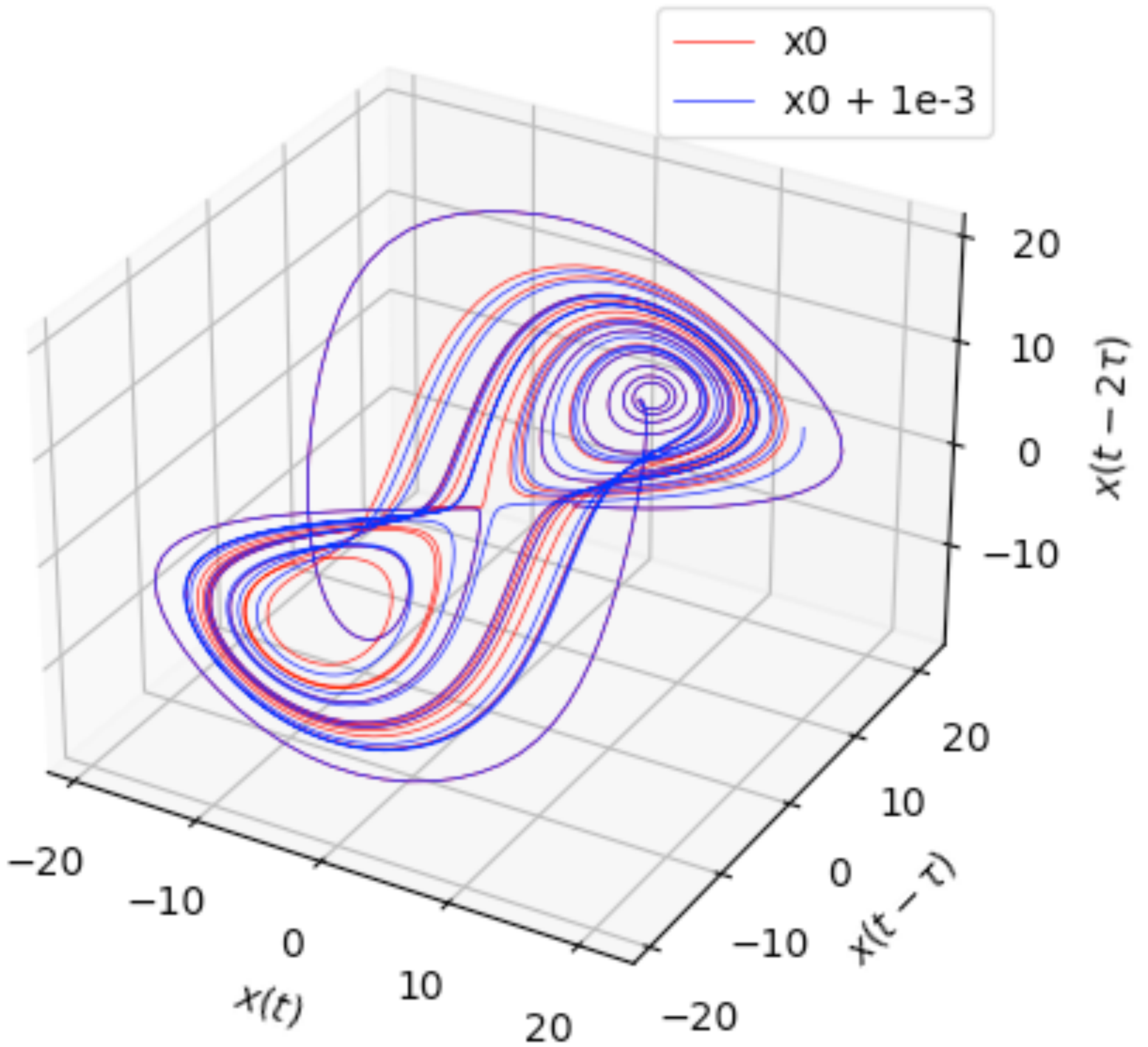
Clean

Limitations

- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$

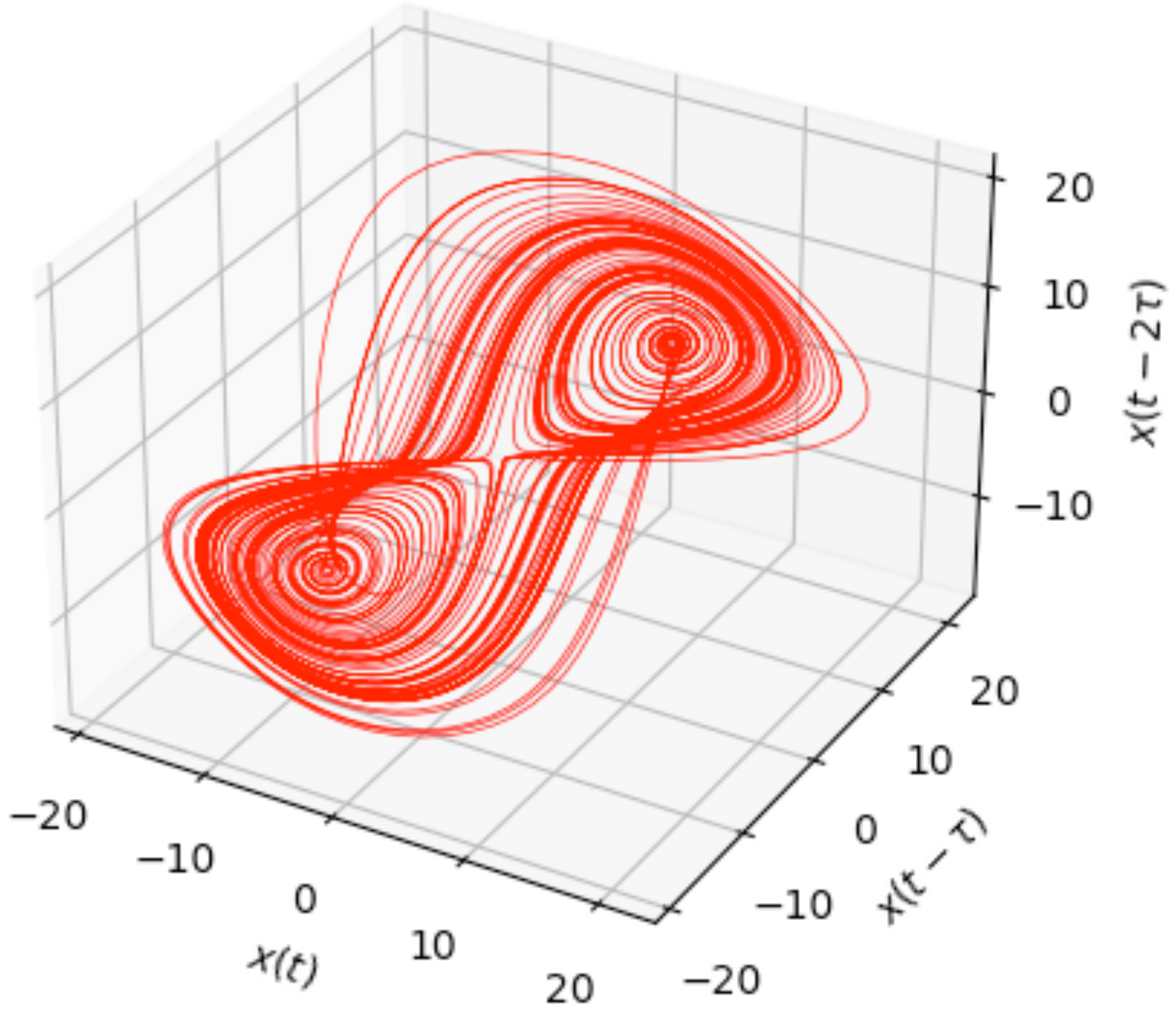


Clean



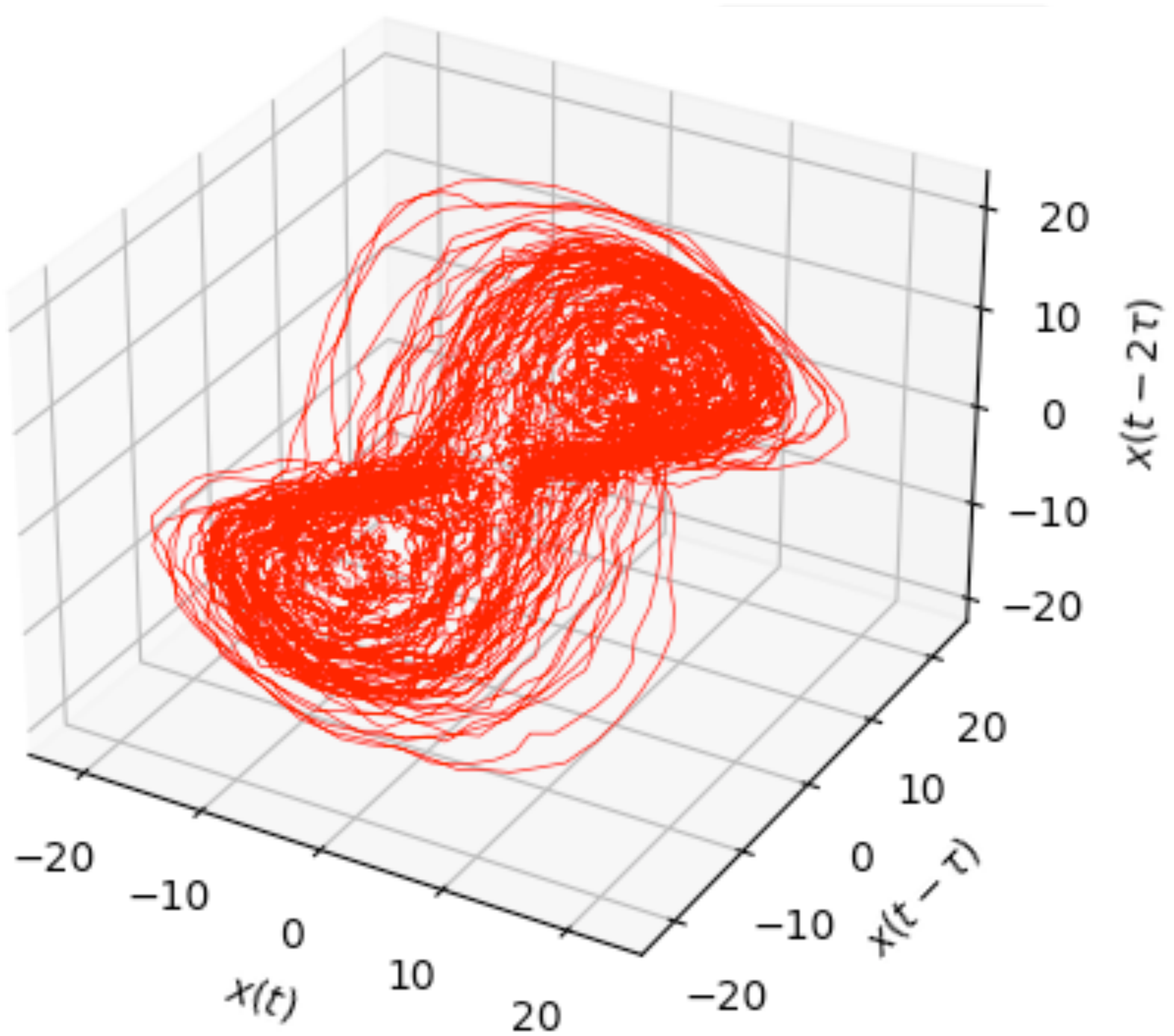
Limitations

- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$
- Stochastic dynamics $\rho_t = law(X_t)$



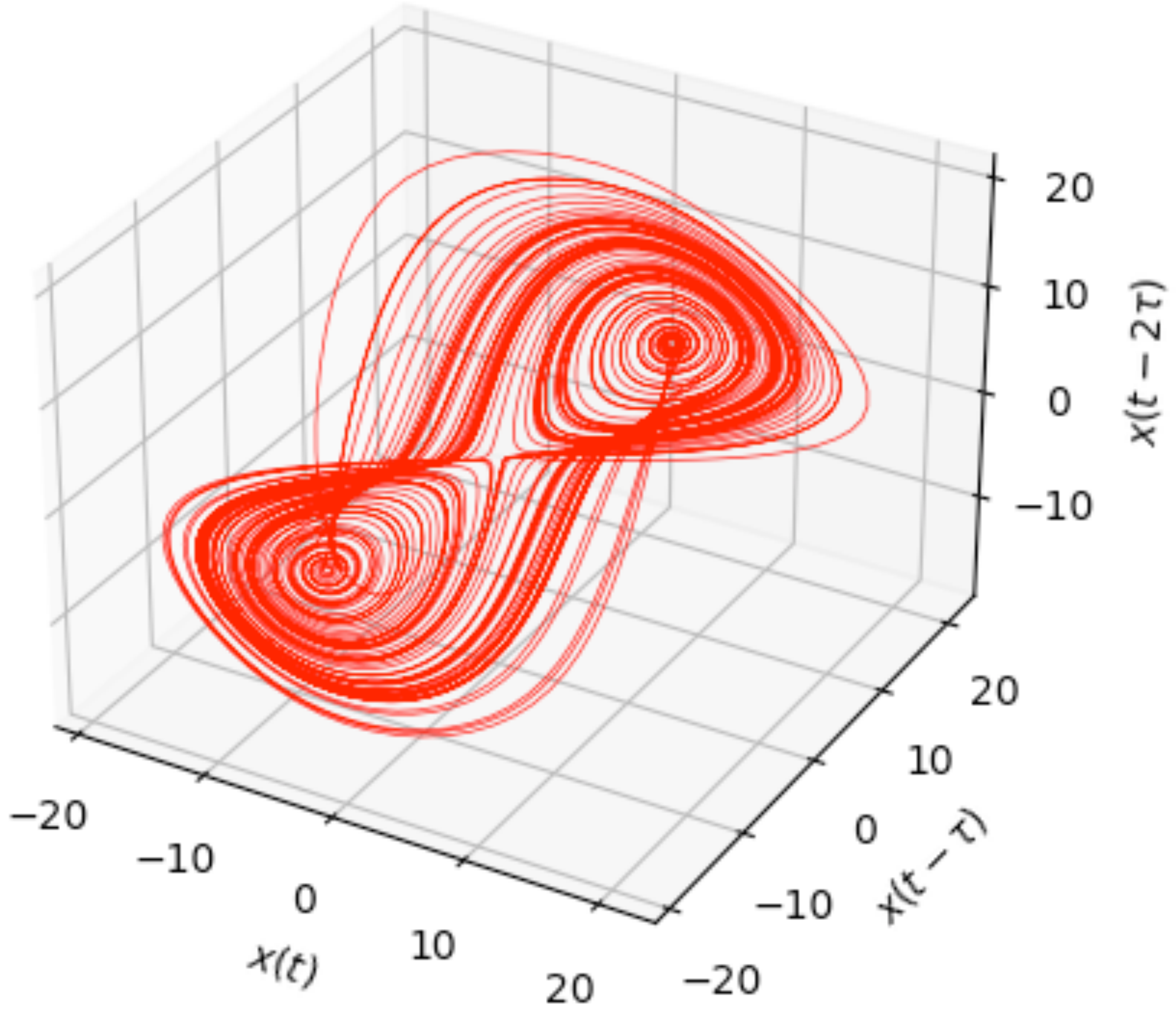
Clean

Noisy

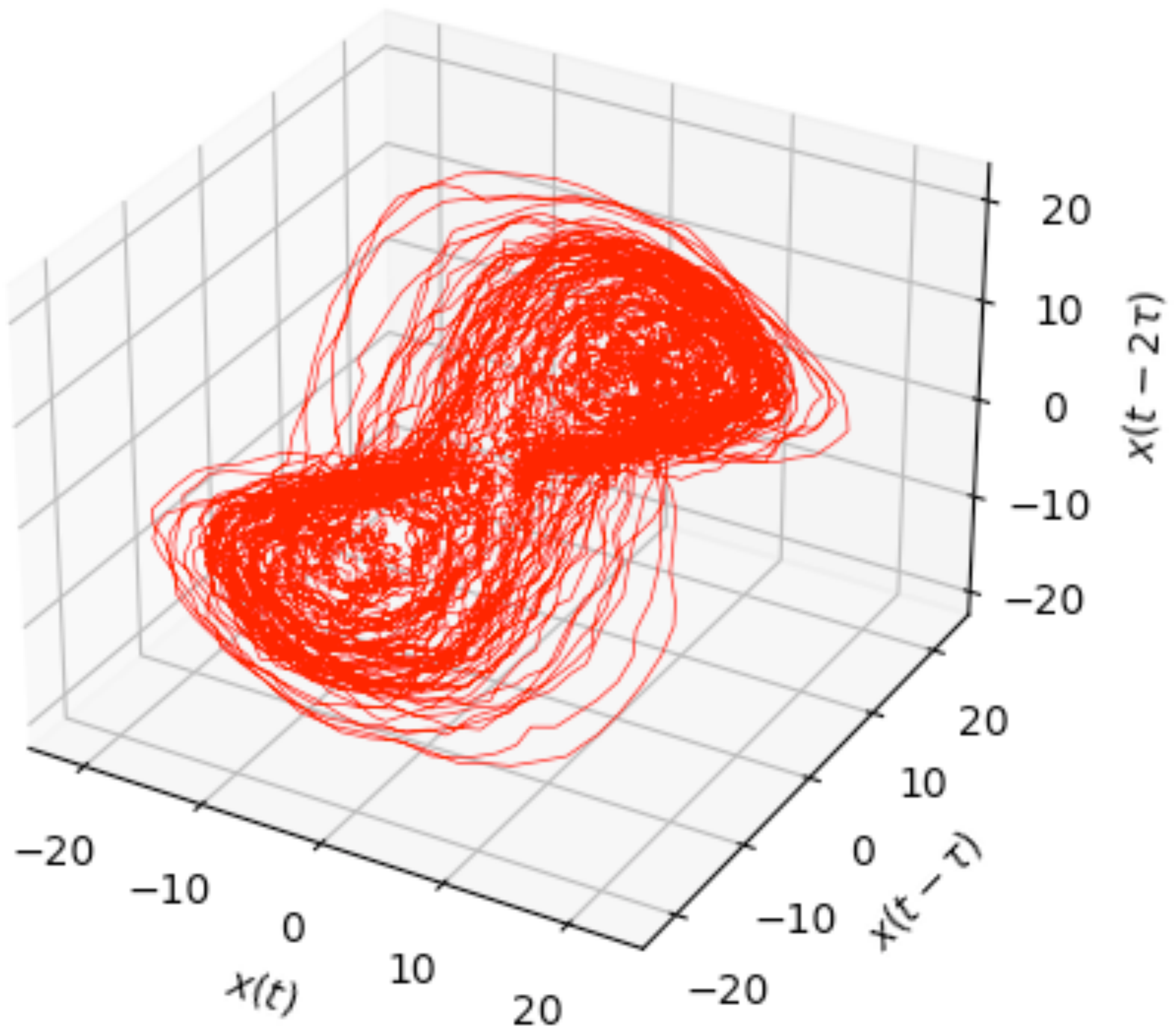


Limitations

- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$
- Stochastic dynamics $\rho_t = law(X_t)$
- Noisy data $\tilde{x}_t = x_t + \mathcal{N}(0,1)$, $\rho_t = \delta_{x_t} * \mathcal{N}(0,1)$



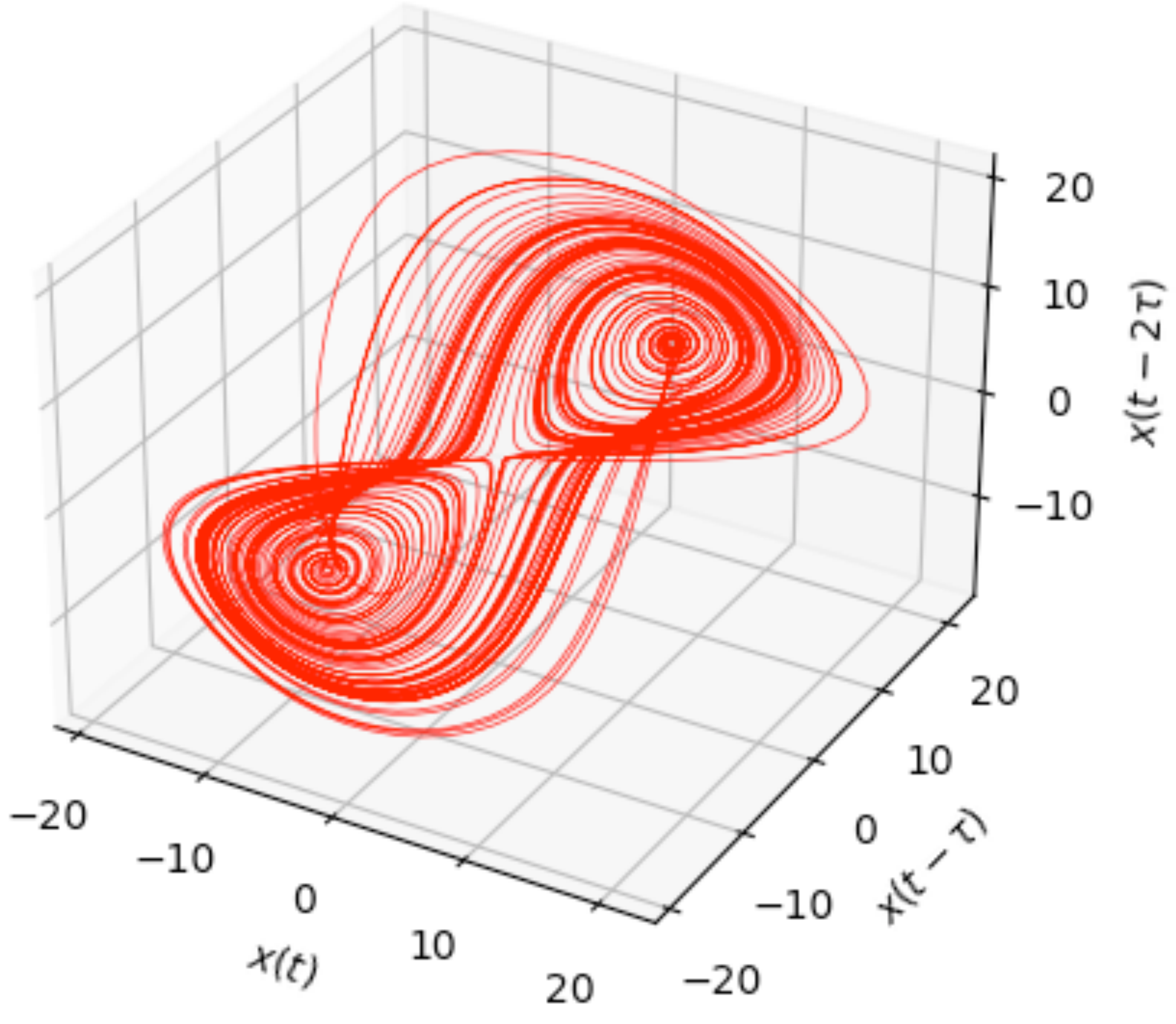
Clean



Noisy

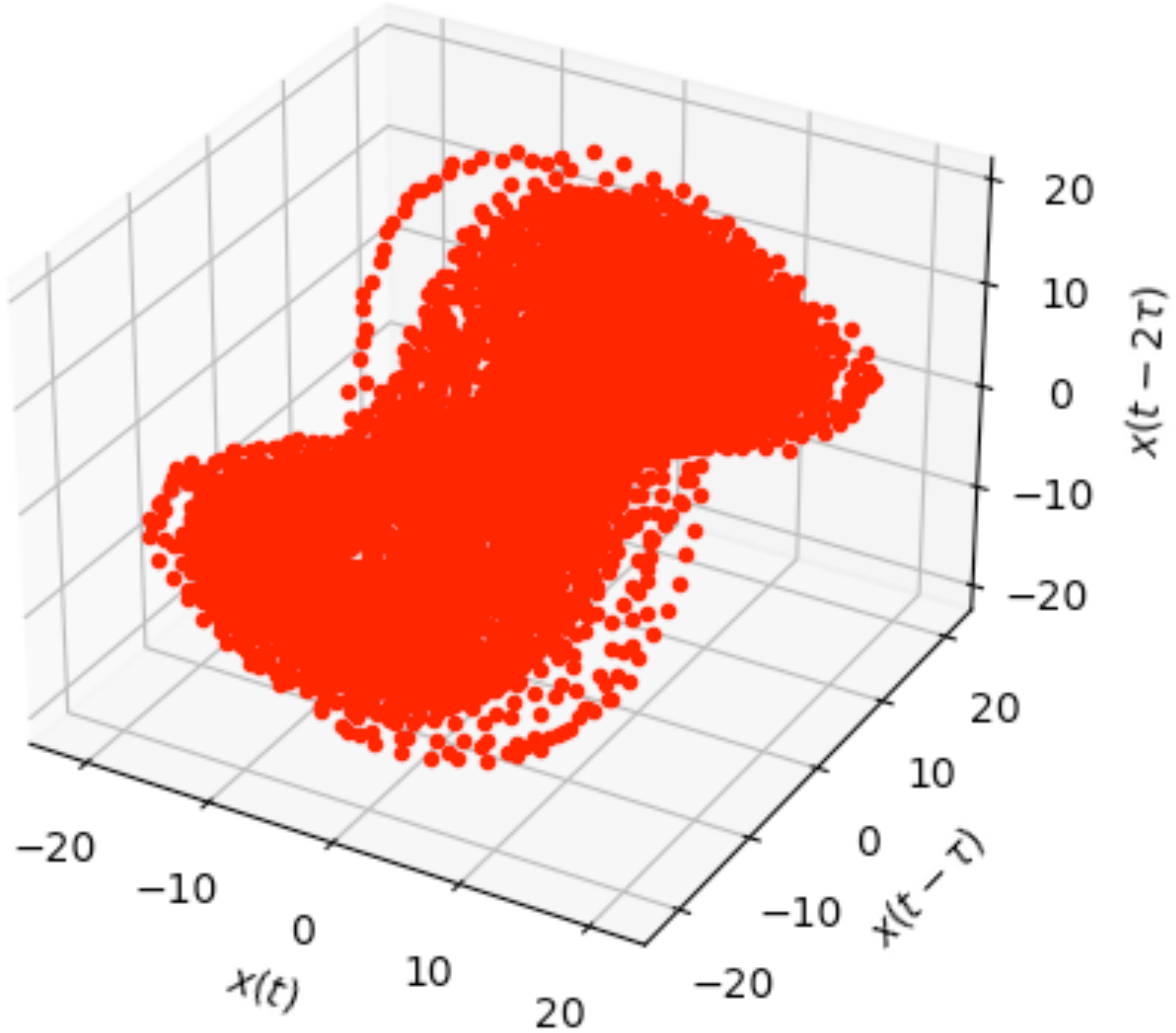
Limitations

- Takens requires clean data, deterministic setup.
- Chaos + Uncertainty in the initial position $\rho_0 \implies \phi_t \# \rho_0$
- Stochastic dynamics $\rho_t = law(X_t)$
- Noisy data $\tilde{x}_t = x_t + \mathcal{N}(0,1)$, $\rho_t = \delta_{x_t} * \mathcal{N}(0,1)$
- Tracking information might not be available



Clean

Noisy



Realistic data assumption

- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - y_i\|_2^2$

Realistic data assumption

- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - y_i\|_2^2$

Good in the noise free deterministic scenario

Realistic data assumption

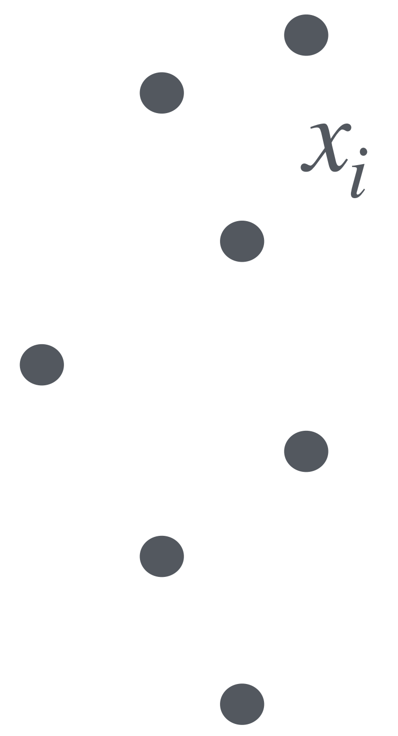
- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - y_i\|_2^2$
- $L_{true}(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - f_{true}(x_i)\|_2^2$

Good in the noise free deterministic scenario

Realistic data assumption

- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - y_i\|_2^2$
- $L_{true}(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - f_{true}(x_i)\|_2^2$

Good in the noise free deterministic scenario



Realistic data assumption

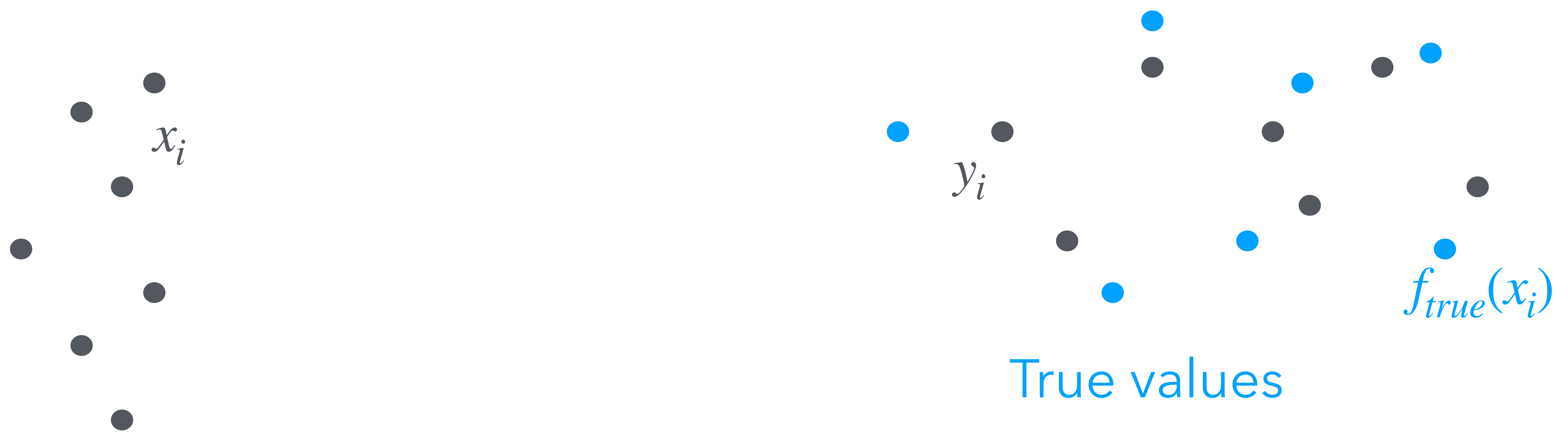
- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - y_i\|_2^2$
- $L_{true}(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - f_{true}(x_i)\|_2^2$

Good in the noise free deterministic scenario

Realistic data assumption

- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - y_i\|_2^2$
- $L_{true}(\theta) = \frac{1}{N}\sum_i \|f_\theta(x_i) - f_{true}(x_i)\|_2^2$

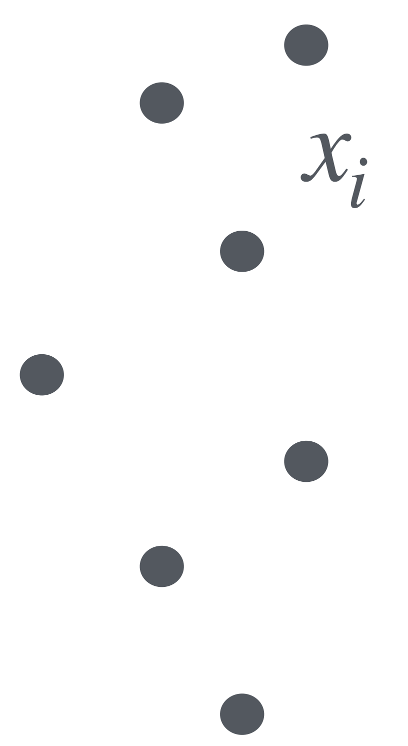
Good in the noise free deterministic scenario

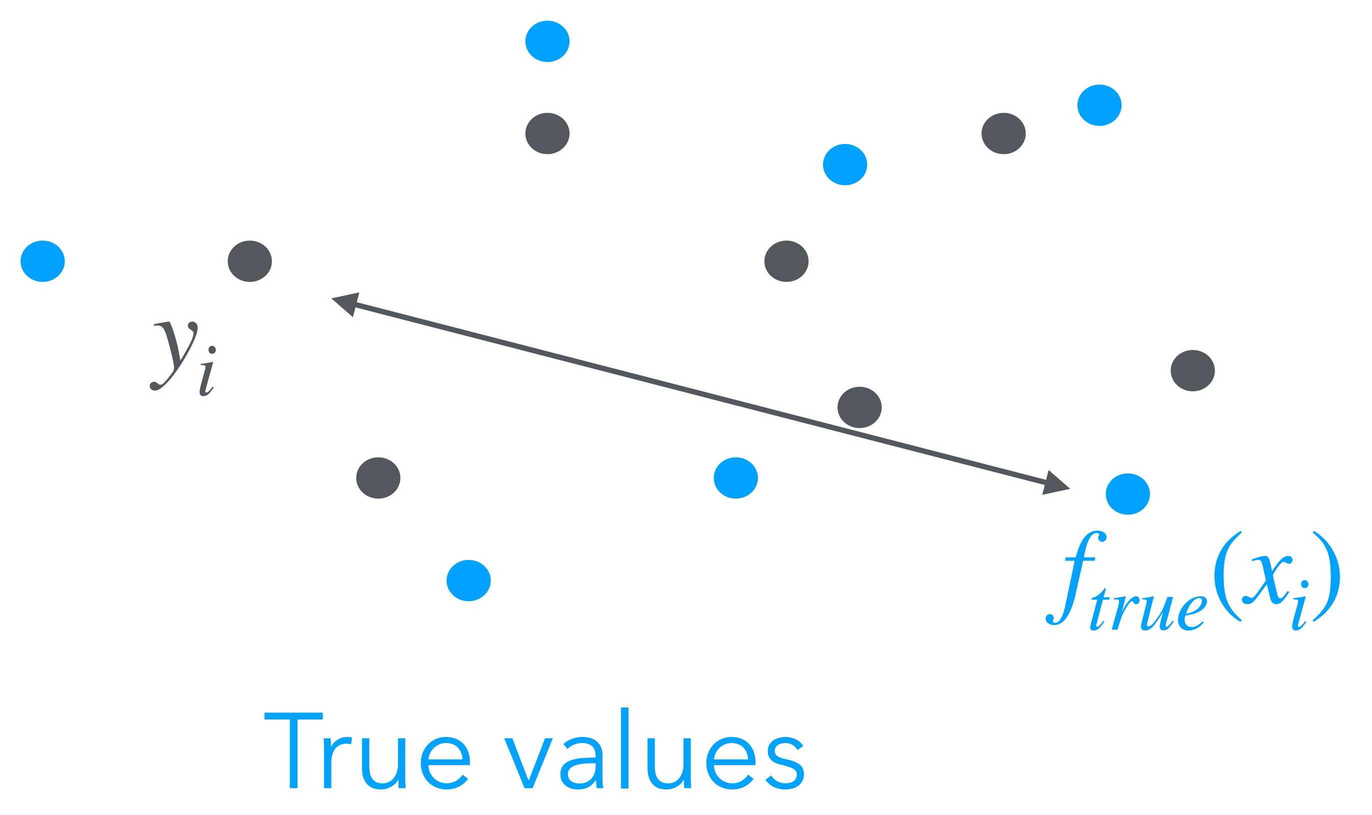


Realistic data assumption

- Pairs $\{x_i, y_i\}$ where $x_i \sim \rho_x$ and $y_i \sim \rho_y$
- Point-wise reconstruction: $L(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - y_i\|_2^2$
- $L_{true}(\theta) = \frac{1}{N} \sum_i \|f_\theta(x_i) - f_{true}(x_i)\|_2^2$

Good in the noise free deterministic scenario





Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$

Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \dfrac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
- If we can show:

Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
- If we can show:

  ➡ F_θ^* is an embedding

Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
- If we can show:
  - F_θ^* is an embedding
  - F_θ^* is directly related to f_{true}

Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
- If we can show:
  - ➜ F_θ^* is an embedding
  - ➜ F_θ^* is directly related to f_{true}
  - ➜ F_θ^* is robuster to noise

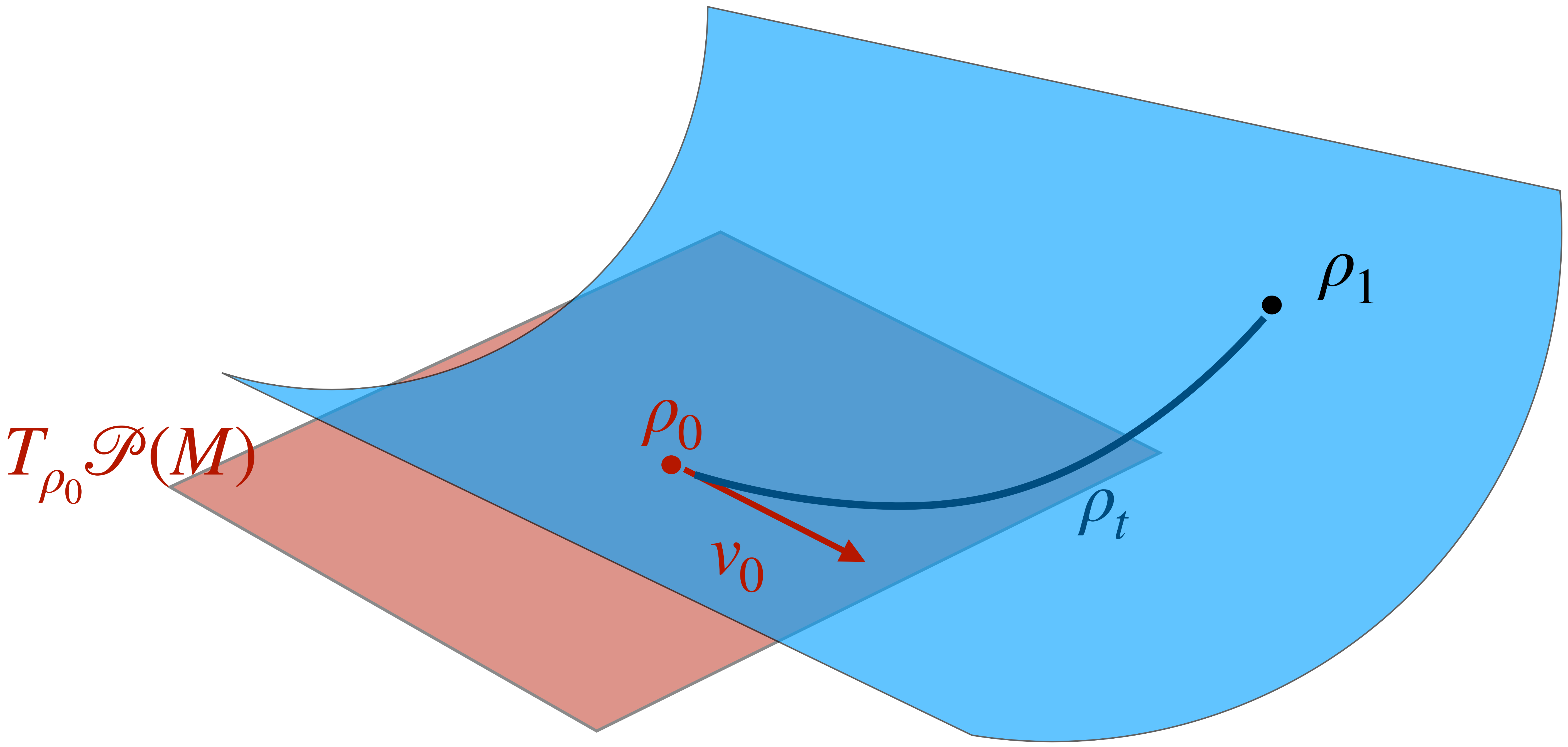
Our approach

- Sparse and noisy data can be represented as $\{\rho_x^i, \rho_y^i\}$
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
- If we can show:
  - F_θ^* is an embedding
  - F_θ^* is directly related to f_{true}
  - F_θ^* is robuster to noise

What is an embedding in $\mathscr{P}_2(M)$?

Eulerian framework - Optimal transport

$$\mathscr{P}_2(M) = \left\{ \rho \,\middle|\, \int |x|^2 \rho < \infty \right\}$$



Geodesic!

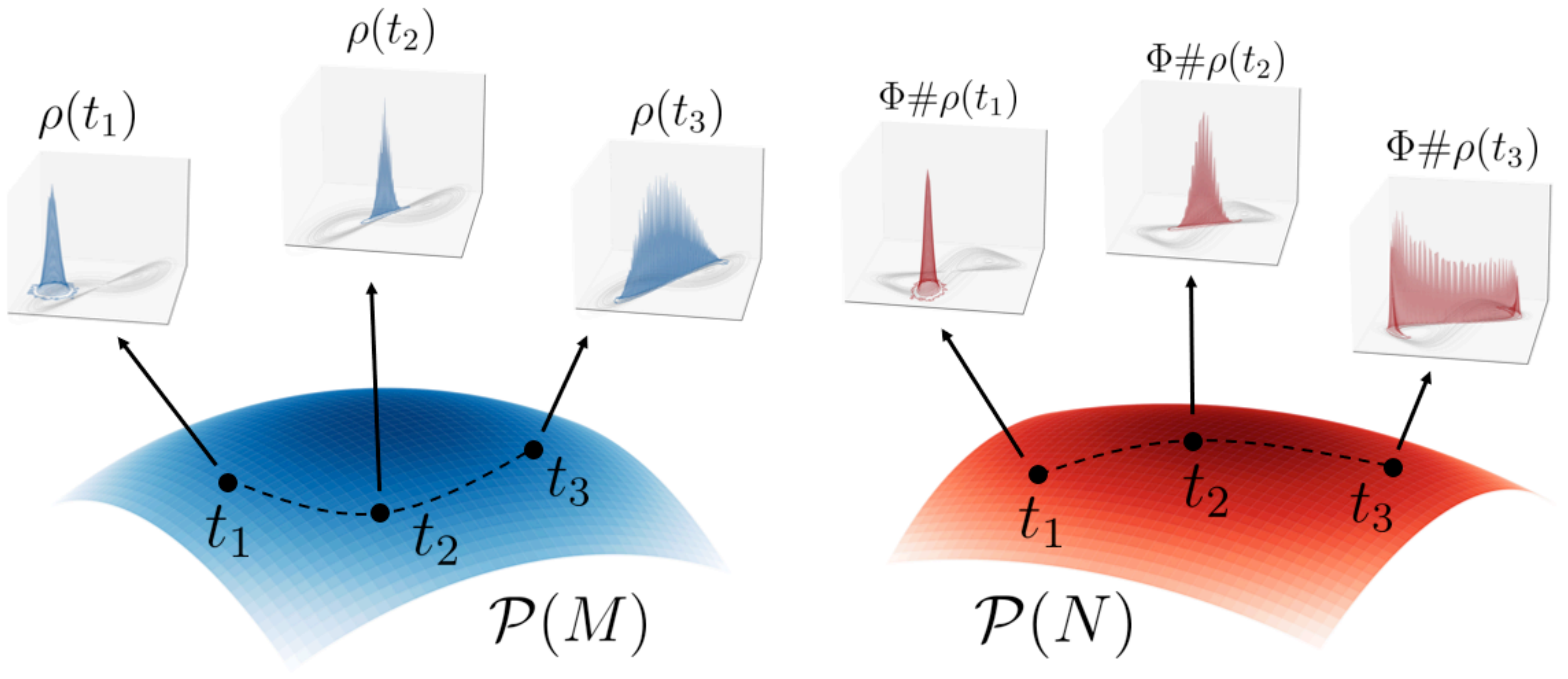
$$W_2^2(\rho_0, \rho_1) = \min_{(\rho_t, v_t)} \int_0^1 \int \|v_t\|^2 d\mu_t dt \,\Big|\, \partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$$

Kinetic energy

ρ_t connects ρ_0 & ρ_1 and has velocity $v_t \iff \dot{x}_t = v(x)$

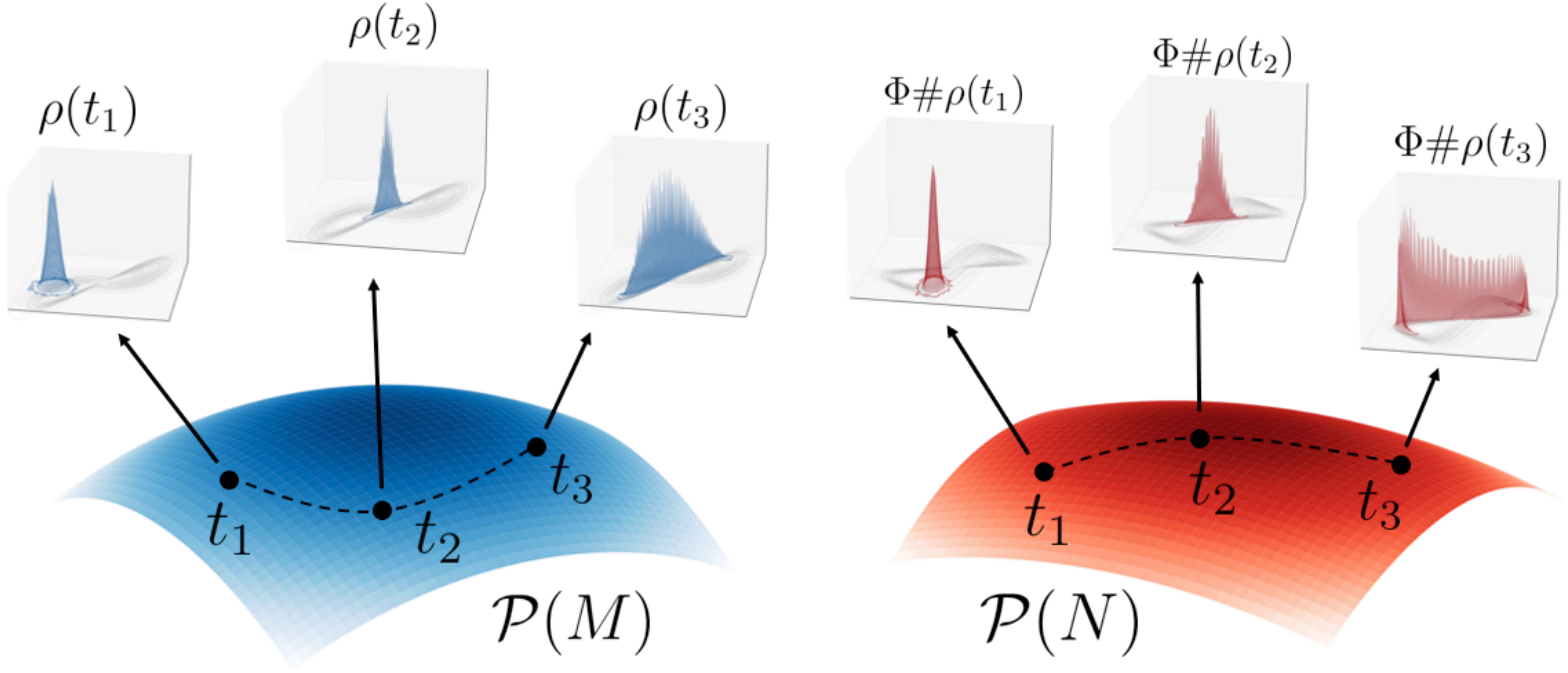
Theorem: Let $f : M \to N$ be an embedding and denote by $F = f\# : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$ its push-forward. Then F is an embedding with respect to the Wasserstein geometry.

Theorem: Let $f : M \to N$ be an embedding and denote by $F = f\# : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$ its push-forward. Then F is an embedding with respect to the Wasserstein geometry.



Theorem: Let $f : M \to N$ be an embedding and denote by $F = f\# : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$ its push-forward. Then F is an embedding with respect to the Wasserstein geometry.

✔ F_θ^* is an embedding



Measure based reconstruction

- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

Measure based reconstruction

- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(\cancel{F_\theta}(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

  $f_{\theta\#}$

Measure based reconstruction

- Optimize $\tilde{L}(\theta) = \frac{1}{K} \sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

  $f_\theta\#$ $\quad f_\theta : M \to N$

Measure based reconstruction

- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

  $f_\theta\#$ $\quad f_\theta : M \to N$

- $\tilde{L}_{true}(\theta) = \frac{1}{K}\sum_{i} \mathscr{D}(f_\theta\#\rho_x^i, f_{true}\#\rho_x^i)$ has $f_\theta^* = f_{true}$ as a minimizer

Measure based reconstruction

- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$

  $f_\theta\#$ $\qquad f_\theta : M \to N$

- $\tilde{L}_{true}(\theta) = \frac{1}{K}\sum_{i} \mathscr{D}(f_\theta\#\rho_x^i, f_{true}\#\rho_x^i)$ has $f_\theta^* = f_{true}$ as a minimizer

✓ $F_\theta^* = f_{true}\#$ in the noise free setting

Measure based reconstruction

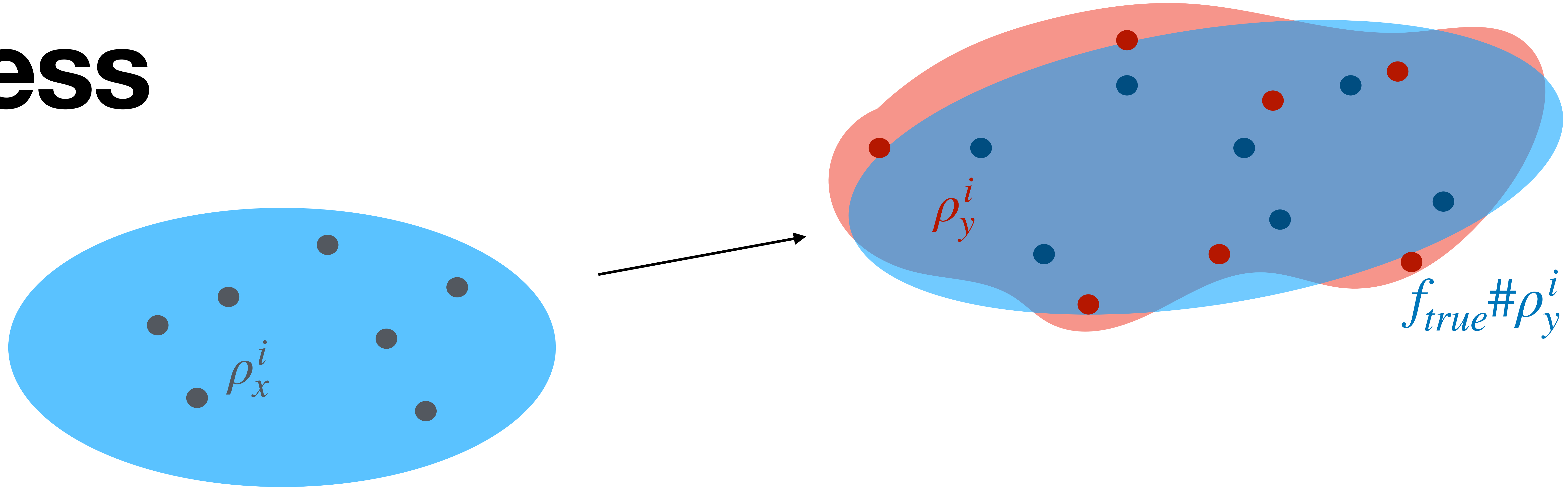
- Optimize $\tilde{L}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathscr{D}(F_\theta(\rho_x^i), \rho_y^i)$ for $F_\theta : \mathscr{P}_2(M) \to \mathscr{P}_2(N)$
  $f_\theta\#$ $\quad f_\theta : M \to N$

- $\tilde{L}_{true}(\theta) = \frac{1}{K}\sum_i \mathscr{D}(f_\theta\#\rho_x^i, f_{true}\#\rho_x^i)$ has $f_\theta^* = f_{true}$ as a minimizer

  $F_\theta^* = f_{true}\#$ in the noise free setting

- When $\rho_x^i = \delta_{x_i}$ then $\tilde{L}(\theta) = L(\theta) \implies$ generalization of the point-wise setting

Robustness



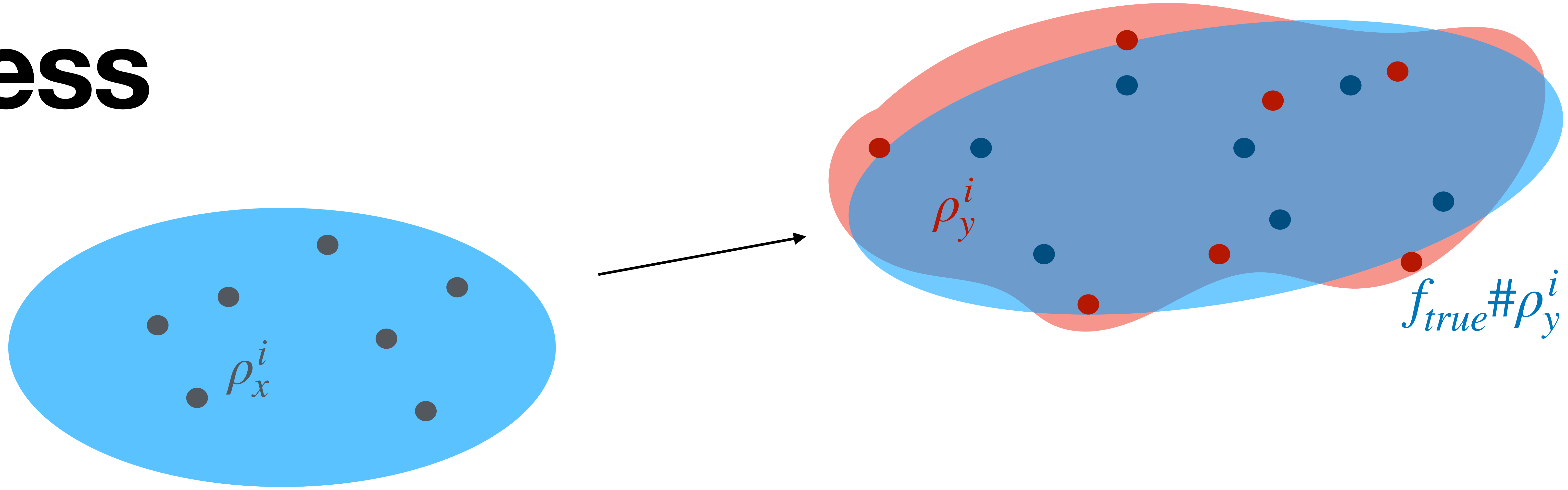
Pointwise

- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$

Measure-based

- ρ_x^i, ρ_y^i

Robustness



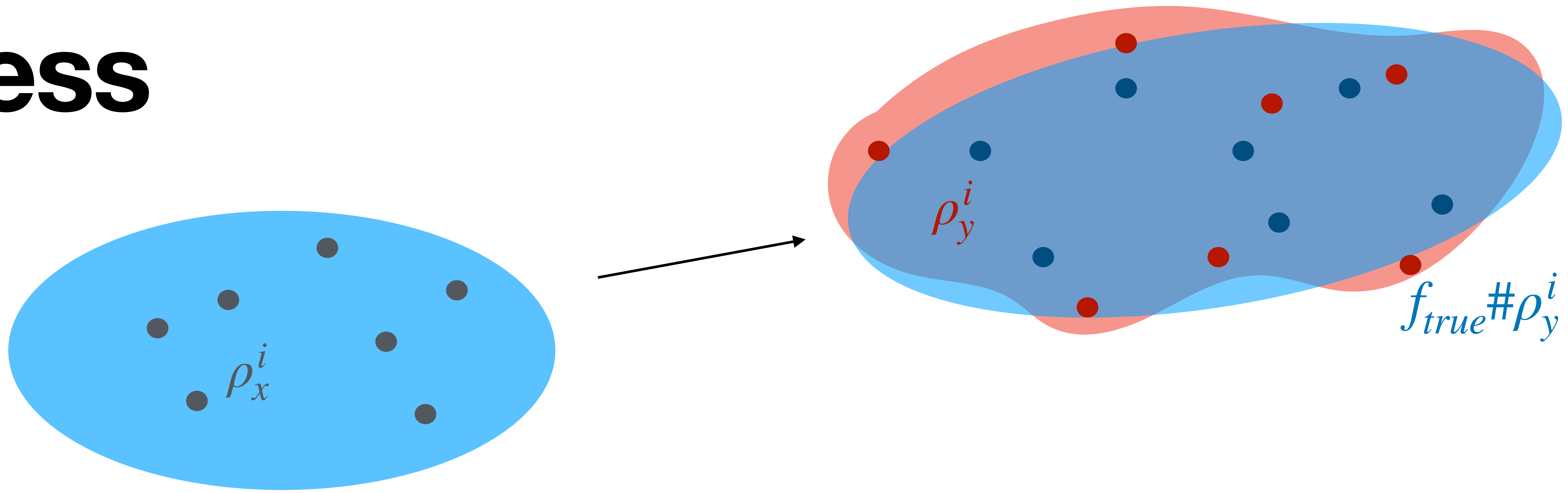
Pointwise

- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average $\mathbb{E}_{\rho_x^i,\rho_y^i} \| f_{true}(x_i) - y_i \|_2^2$ away.

Measure-based

- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.

Robustness



Pointwise

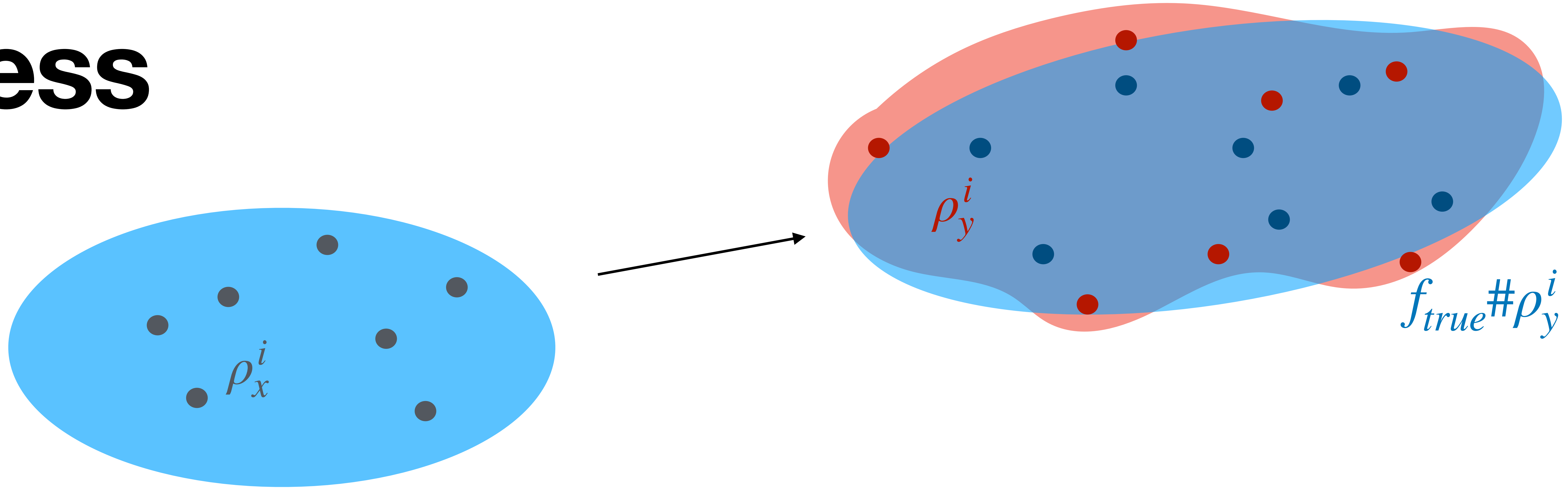
- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average $\mathbb{E}_{\rho_x^i,\rho_y^i}\|f_{true}(x_i) - y_i\|_2^2$ away.

$\geq$

Measure-based

- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.

Robustness



Pointwise

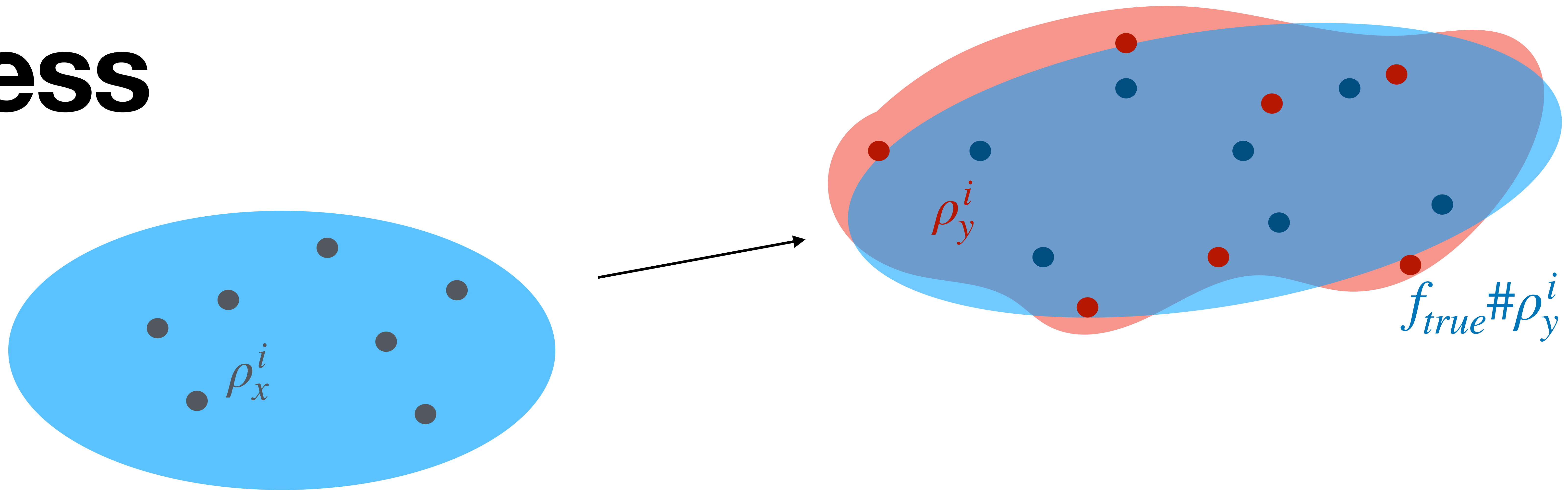
- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average
  $$\mathbb{E}_{\rho_x^i,\rho_y^i} \| f_{true}(x_i) - y_i \|_2^2$$ away.
- Solutions can be highly oscillatory

$\geq$

Measure-based

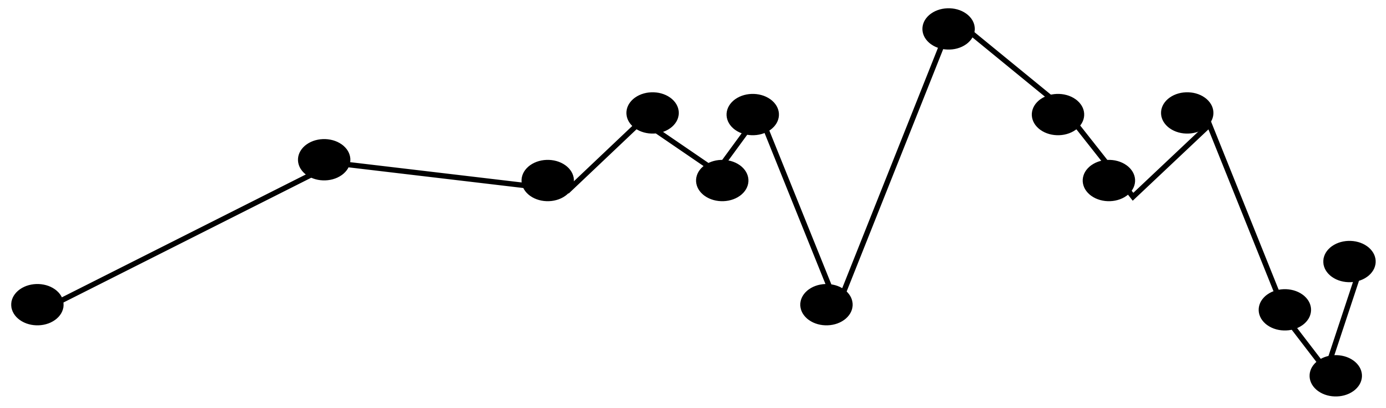
- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.

Robustness



Pointwise

- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average $\mathbb{E}_{\rho_x^i,\rho_y^i}\|f_{true}(x_i) - y_i\|_2^2$ away.
- Solutions can be highly oscillatory

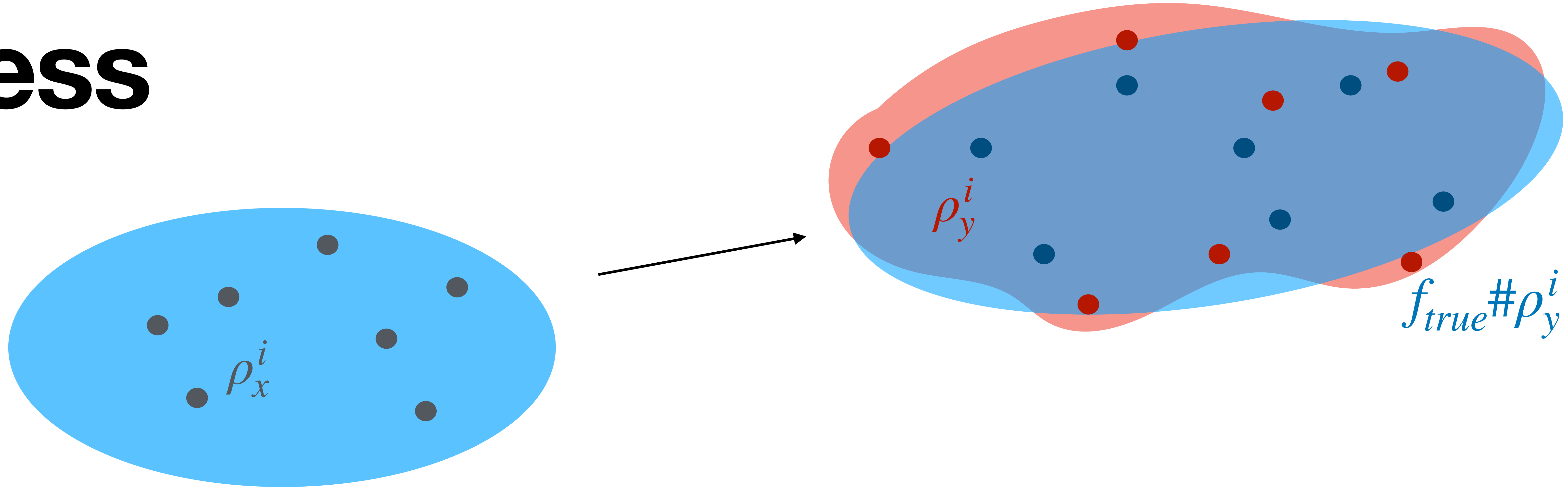


$\geq$

Measure-based

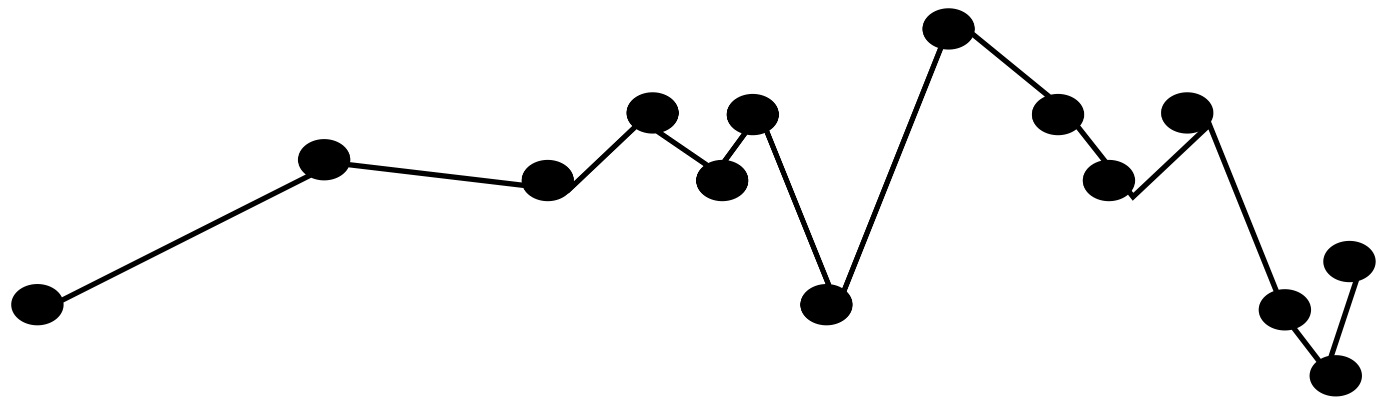
- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.

Robustness



Pointwise

- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average $\mathbb{E}_{\rho_x^i,\rho_y^i} \| f_{true}(x_i) - y_i \|_2^2$ away.
- Solutions can be highly oscillatory

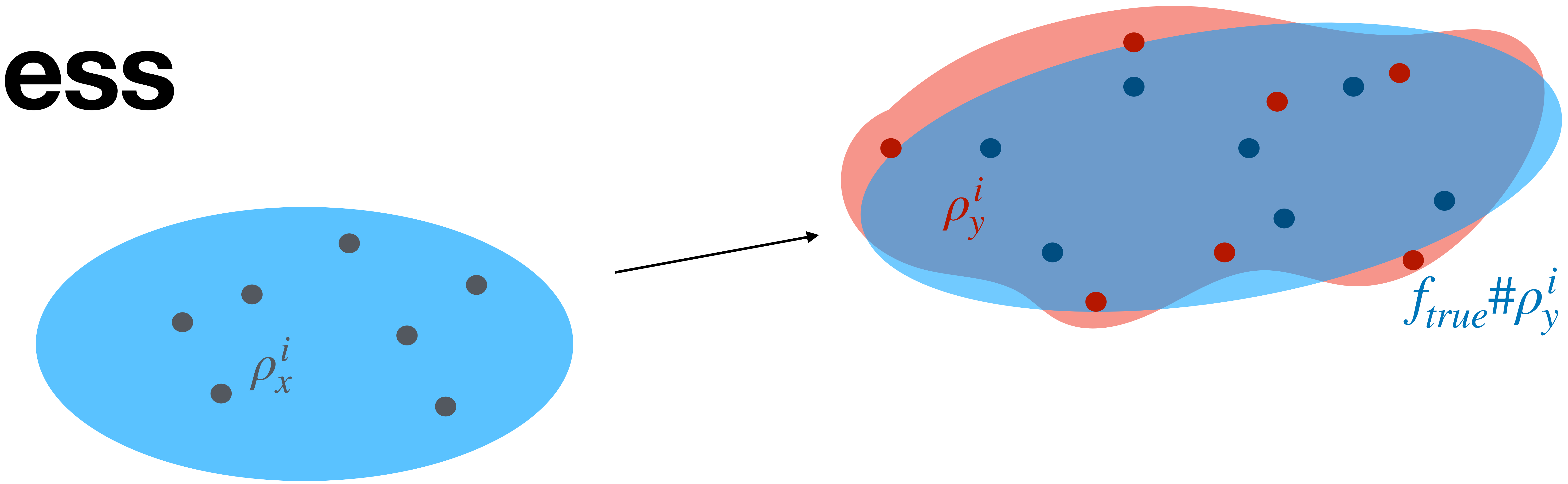


$\geq$

Measure-based

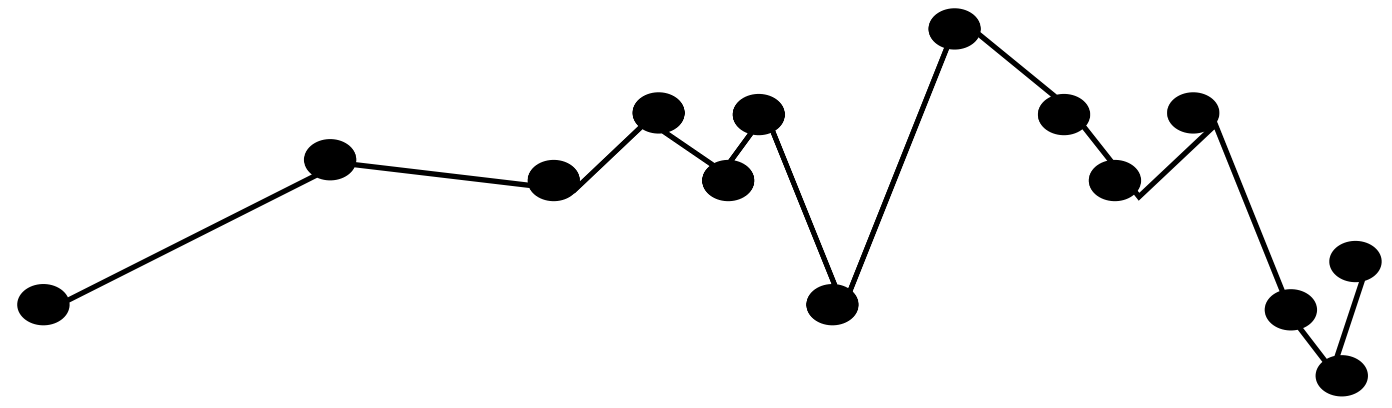
- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.
- Allows wiggle room while keeping the loss low
- NN spectral bias $\Longrightarrow$ implicit regularization

Robustness



Pointwise

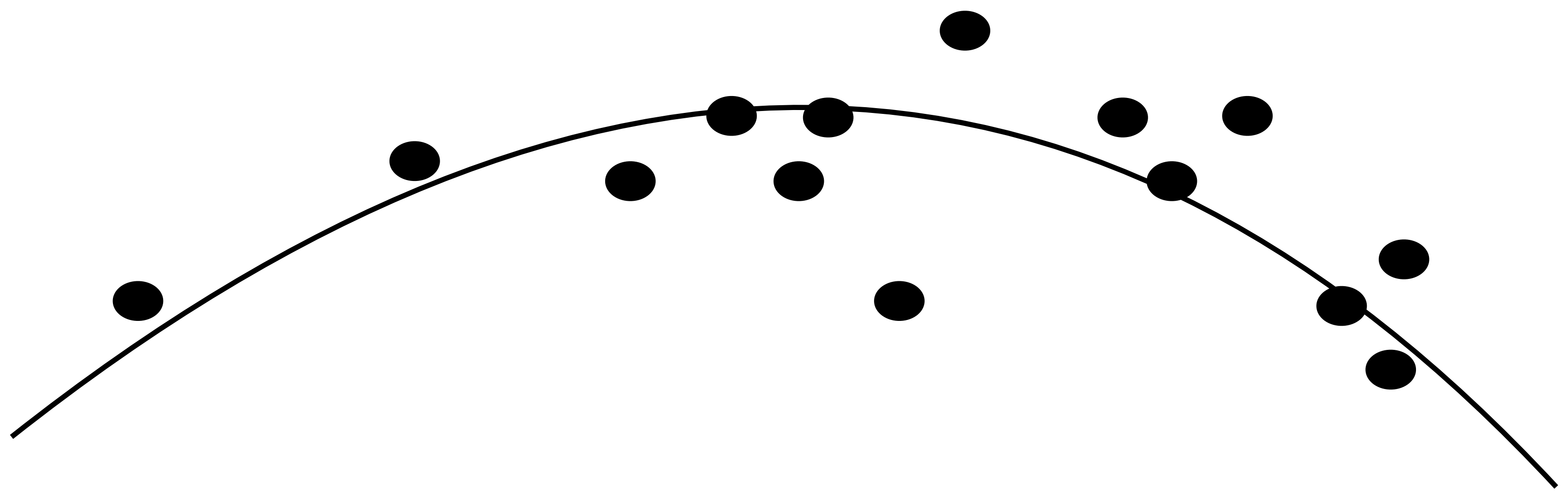
- $x_i \sim \rho_x^i, y_i \sim \rho_y^i$
- Measurements are on average $\mathbb{E}_{\rho_x^i,\rho_y^i} \| f_{true}(x_i) - y_i \|_2^2$ away.
- Solutions can be highly oscillatory



$$\geq$$

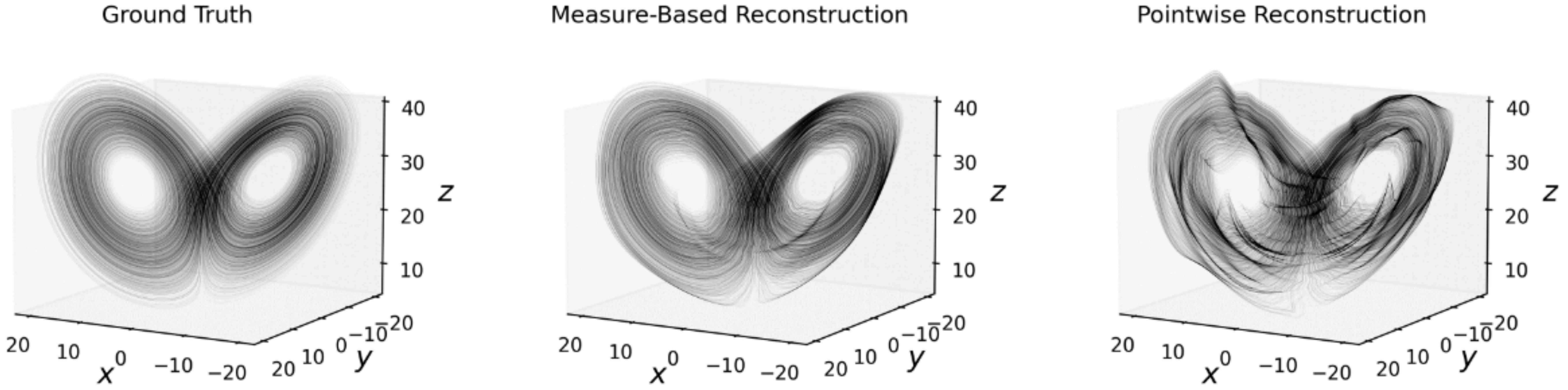
Measure-based

- ρ_x^i, ρ_y^i
- Densities are $W_2^2(f_{true}\#\rho_x^i, \rho_y^i)$ away.
- Allows wiggle room while keeping the loss low
- NN spectral bias $\Longrightarrow$ implicit regularization

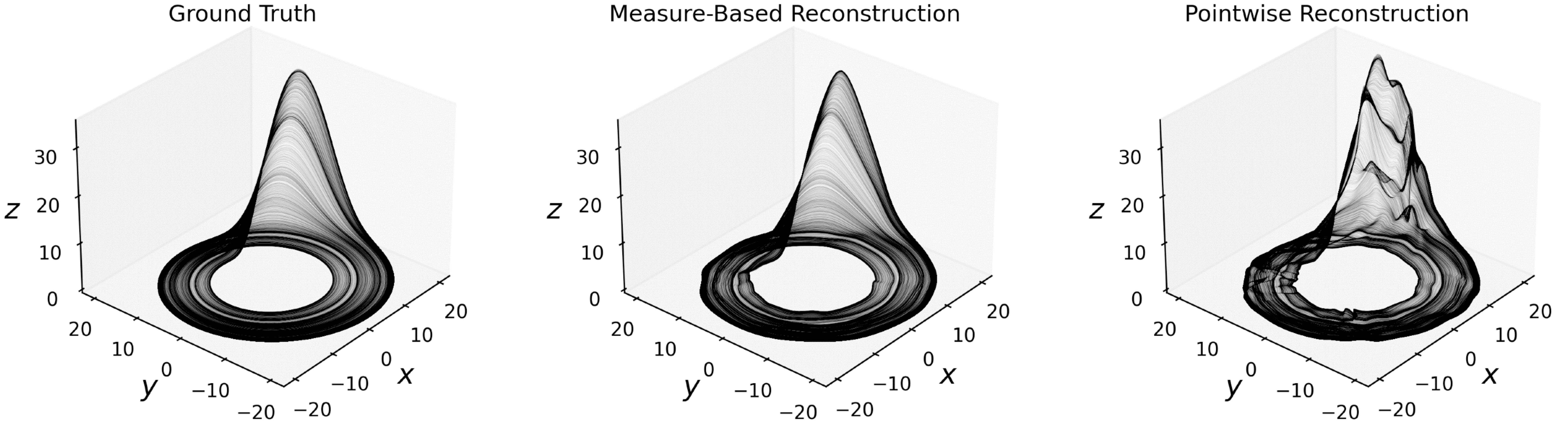


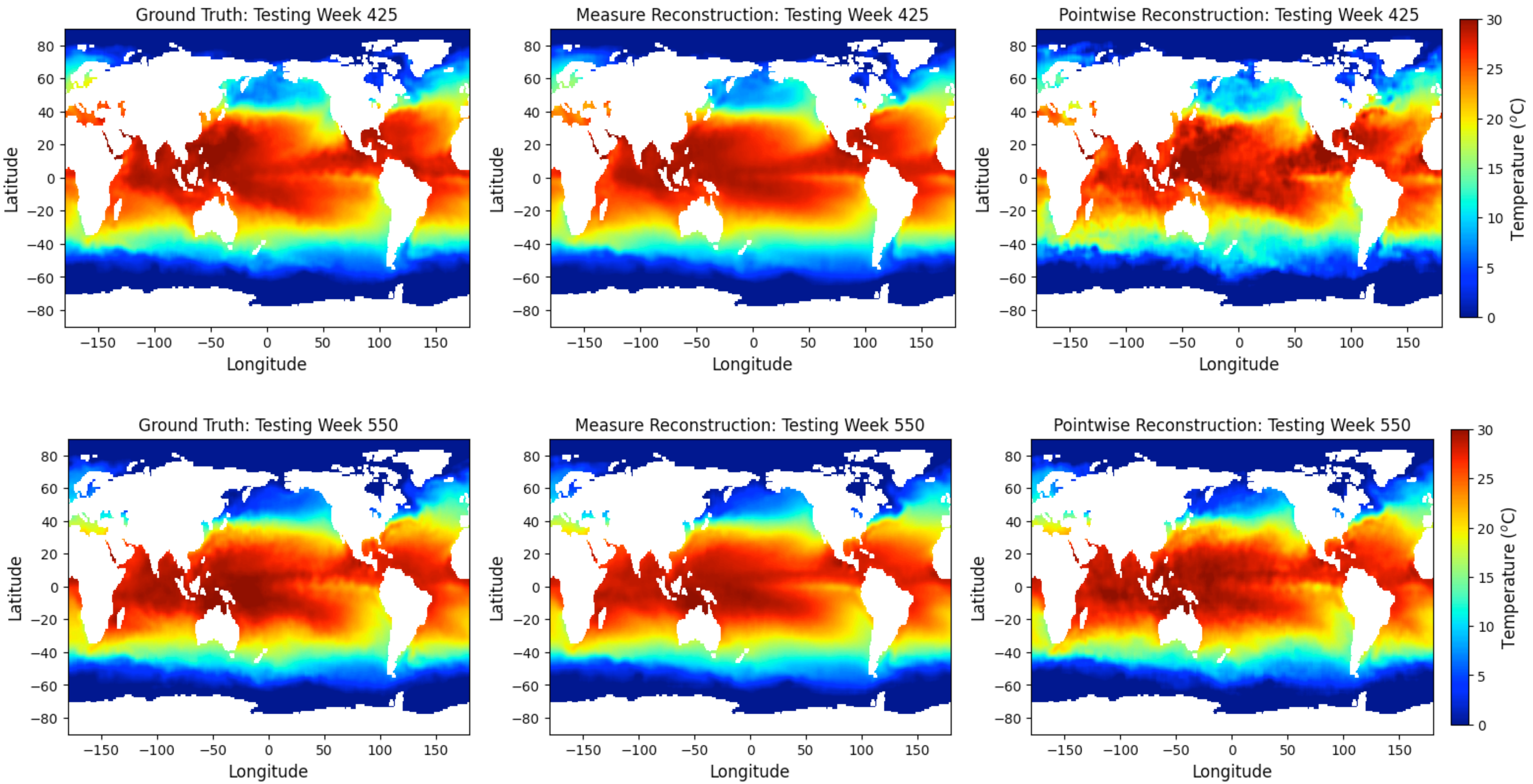
Numerical Results

Lorentz system



Roessler system




Ground Truth: Testing Week 425
Measure Reconstruction: Testing Week 425
Pointwise Reconstruction: Testing Week 425
Ground Truth: Testing Week 550
Measure Reconstruction: Testing Week 550
Pointwise Reconstruction: Testing Week 550
Latitude
Longitude
Temperature (°C)

Thank you for your attention!

Contact me: mao237@cornell.edu

References: [1] Floris Takens. Detecting strange attractors in turbulence. In Dynamical Systems and Turbulence, Warwick 1980, pages 366–381. Springer, 1981

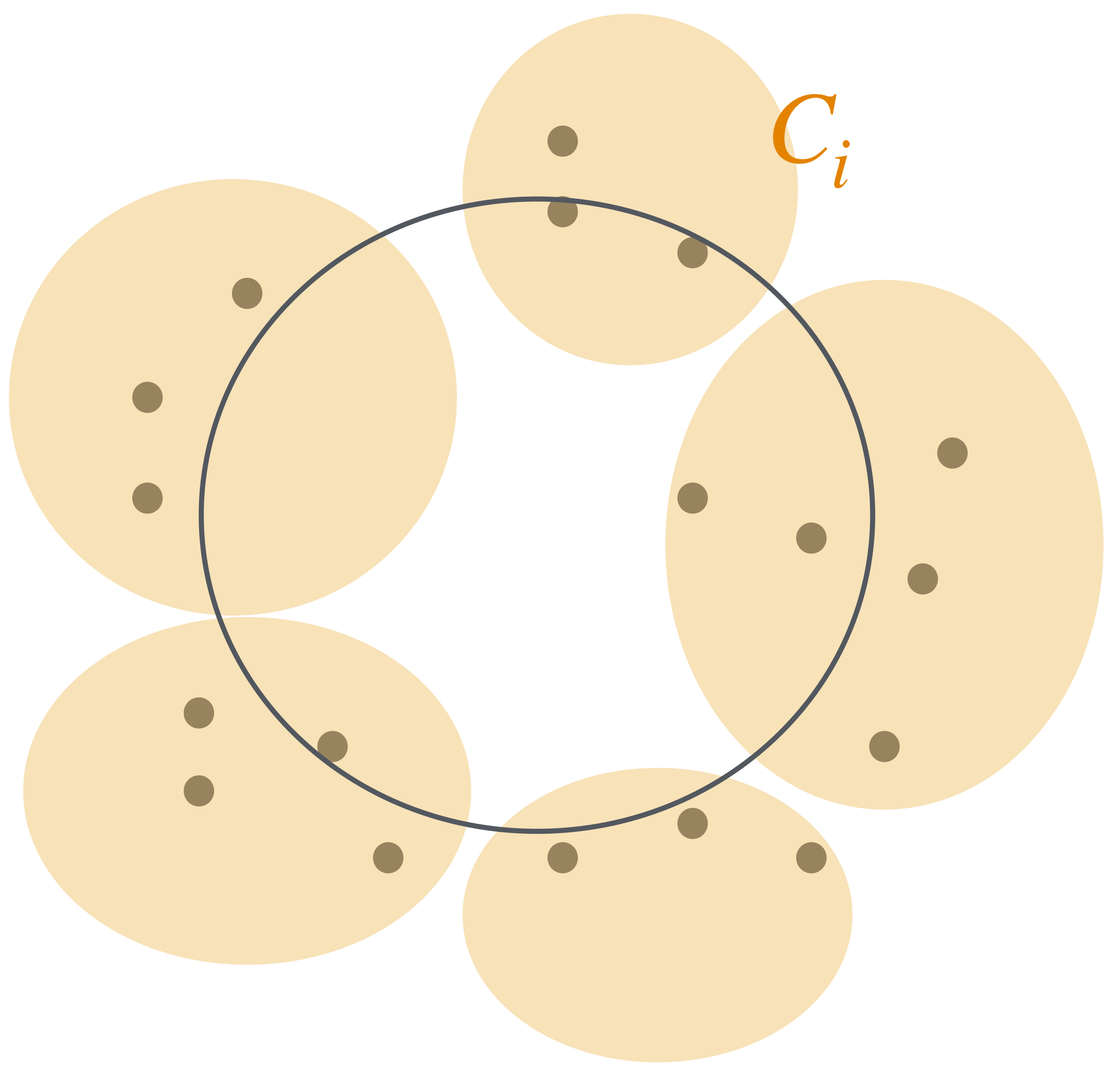
[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savare. Gradient Flows in Metric Spaces and in the Space of Probability Measures. Birkhaeuser Verlag, 2005.

[3] Jonah Botvinick-Greenhouse, Maria Oprea, Yunan Yang and Romit Maulik, Measure-Theoretic Takens' Time-Delay Embedding:Analysis and Application, in preparation, 2024

Supplemental slides

Transform trajectory data into measure data

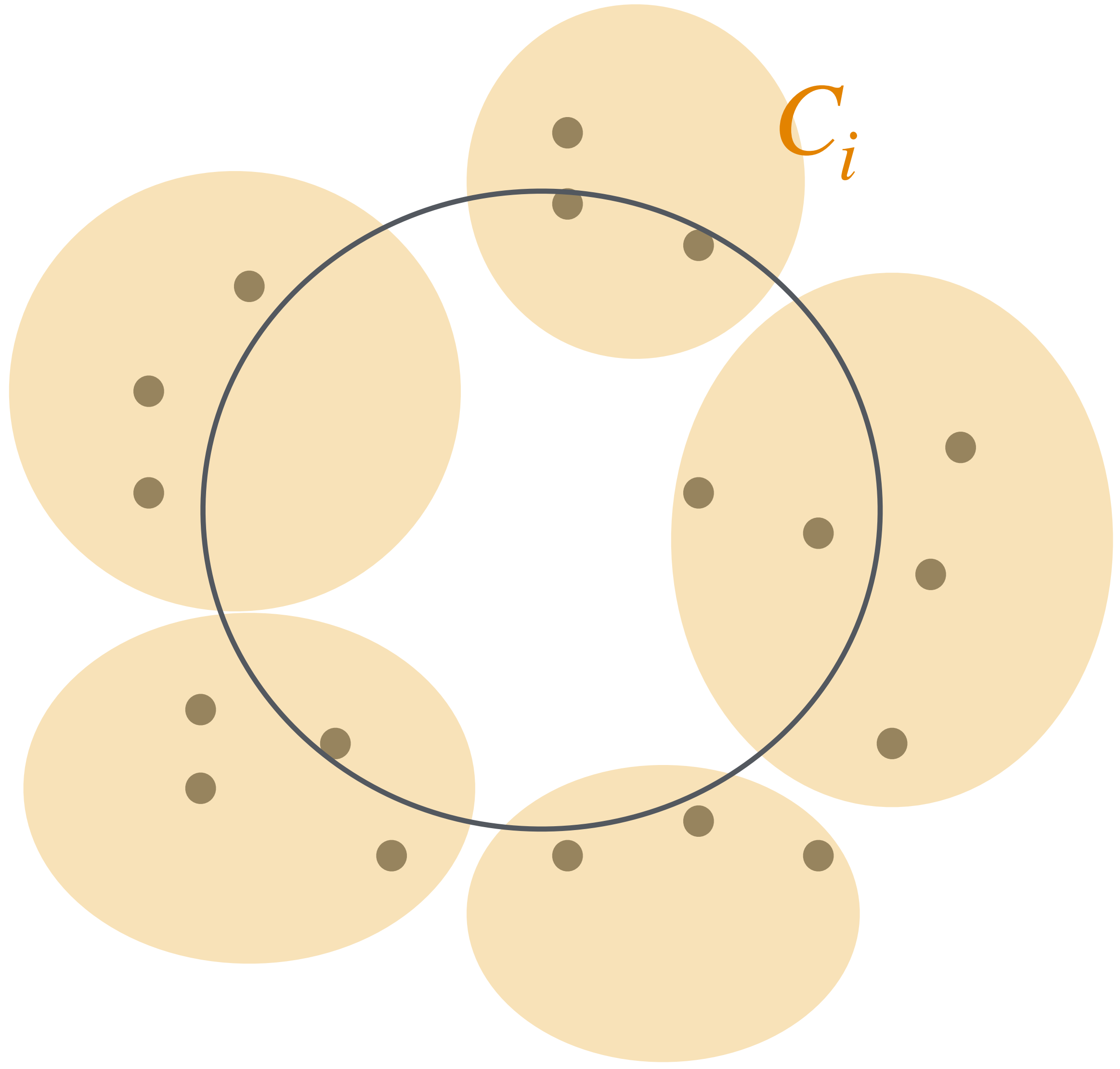
- Partition domain into $\{C_i\}_{i=1}^{M}$
-
-
-

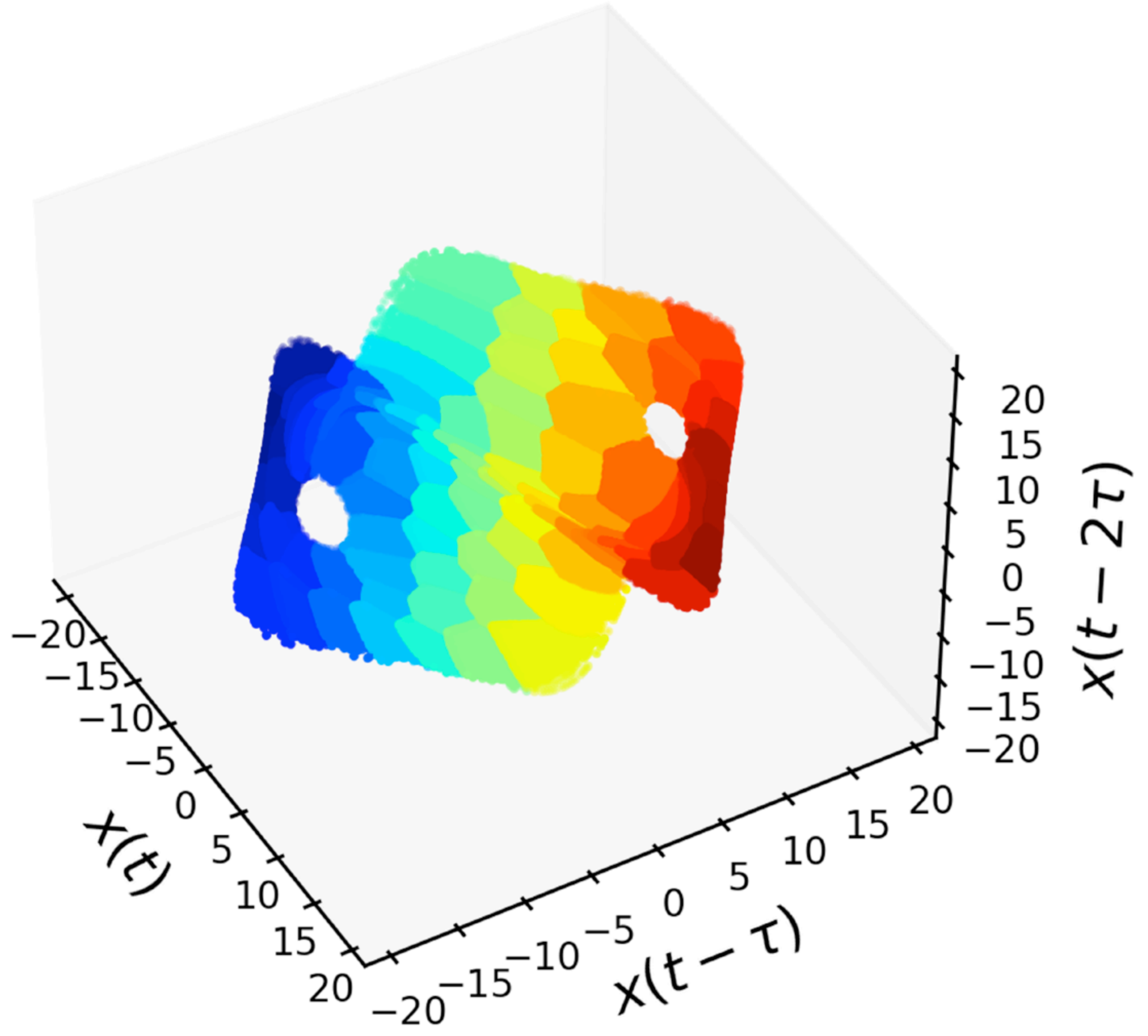


Supplemental slides

Transform trajectory data into measure data

- Partition domain into $\{C_i\}_{i=1}^{M}$
- Empirical distribution conditioned on the cell $\mu_i \sim \sum_{x_j \in C_i} \delta_{x_j}$
- As diameter of $C_i \to 0$ we obtain point wise reconstruction
-





Supplemental slides

Transform trajectory data into measure data

- Partition domain into $\{C_i\}_{i=1}^{M}$
- Empirical distribution conditioned on the cell $\mu_i \sim \sum_{x_j \in C_i} \delta_{x_j}$
- As diameter of $C_i \to 0$ we obtain point wise reconstruction
- Works with different data assumptions