Learning to Emulate Chaos: Adversarial Optimal Transport Regularization

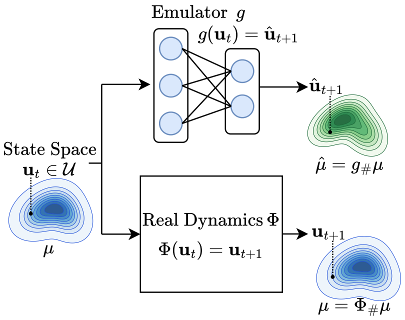
Gabriel Melo[1,2,⋆] Leonardo Santiago[3,4,⋆] Peter Y. Lu[5]

[1]Institut Polytechnique de Paris [2]Télécom Paris [3]NC State University
[4]University of Campinas [5]Tufts University

ICML 2026, Seoul

arXiv:2604.21097

The Challenge: Emulating Chaotic Systems



Chaotic dynamical system Φ

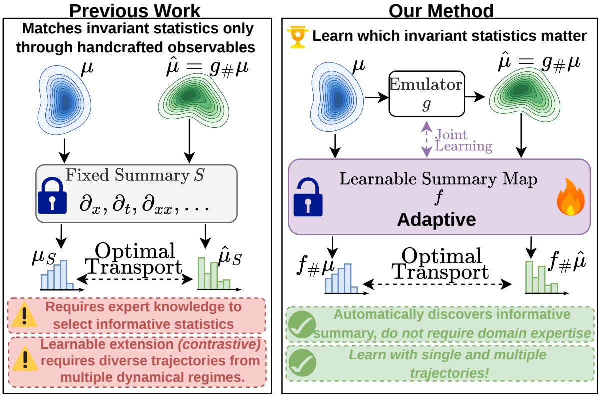
- Small perturbations induce **exponentially growing errors** $\Rightarrow$ MSE $\mathcal{L}_{\text{MSE}}(\mathbf{u}_{t+1}, \hat{\mathbf{u}}_{t+1})$ is ill-suited

Ergodicity: for any observable $f : \mathcal{U} \to \mathcal{S}$,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} f(\mathbf{u}_t) = \mathbb{E}_\mu[f(\mathbf{u})]$$

Goal: preserve *long-term invariant statistics*, not exact trajectories.

How to Emulate Chaos: Preserve Long-Term Statistics



- Noisy $\mathcal{L}_{\text{MSE}}\big(\mathbf{u}_{t+1}, \hat{\mathbf{u}}_{t+1}\big)$ grows **exponentially** with the rollout horizon
- Even a *perfect* emulator $g = \Phi$ suffers MSE blow-up

$\Rightarrow$ **Replace trajectory-level loss with a distributional objective over the invariant measure.**

Our Method: Adversarial OT Loss

Adversarial Optimal Transport objective:

$$\min_{g \in \mathcal{G}} \max_{f \in \mathcal{F}} \underbrace{\mathcal{L}_{\mathrm{MSE}}(\mathbf{u}_{t+1},\, g(\mathbf{u}_t))}_{\text{short-term accuracy}} + \lambda \underbrace{\mathcal{W}_{d_{\mathcal{S}}^p}(f_{\#}\mu,\, f_{\#}\hat{\mu})}_{\text{long-term distributional fidelity}}$$

- $f : \mathcal{U} \to \mathcal{S}$ is a **learnable summary map** adversarially optimized
- μ = true invariant measure; $\hat{\mu}$ = emulator's induced measure
- OT term is non-differentiable in practice $\Rightarrow$ two relaxations

Differentiable Relaxations

Sinkhorn (Primal)

$$\min_{g\in\mathcal{G}} \max_{f\in\mathcal{F}} \; \mathcal{L}_{\mathrm{MSE}} + \lambda\, S^{\varepsilon}_{c_f}(\mu, \hat{\mu})$$

where $c_f(\mathbf{u}, \hat{\mathbf{u}}) := {d_{\mathcal{S}}}^p(f(\mathbf{u}), f(\hat{\mathbf{u}}))$

- Entropic OT (Sinkhorn algorithm)
- $S^{\varepsilon}_c = \mathcal{W}^{\varepsilon}_c(\mu, \nu) - \frac{1}{2}\mathcal{W}^{\varepsilon}_c(\mu, \mu) - \frac{1}{2}\mathcal{W}^{\varepsilon}_c(\nu, \nu)$
- Smooth, fully differentiable

WGAN-style (Dual)

$$\min_{g} \max_{\substack{f\in\mathcal{F}\\ \varphi\in\mathrm{Lip}_1}} \; \mathcal{L}_{\mathrm{MSE}} + \lambda\Big(\mathbb{E}_{\mu}[\varphi \circ f(\mathbf{u})] - \mathbb{E}_{\hat{\mu}}[\varphi \circ f(\hat{\mathbf{u}})]\Big)$$

Critic $\varphi : \mathcal{S} \to \mathbb{R}$

- Dual 1-Wasserstein formulation
- Lightweight MLP critic φ
- Scalable to large systems

Benchmark Systems

System	Dynamics	Eval. Summary Statistics
Lorenz-96	$\dot{u}_i = (u_{i+1} - u_{i-2})u_{i-1} - u_i + F$	$\{\partial_t u_i,\ (u_{i+1} - u_{i-2})u_{i-1},\ u_i\}$
Kuramoto–Sivashinsky	$\partial_t u = -u\partial_x u - \partial_{xx} u - \partial_{xxxx} u$	$\{\partial_t u,\ \partial_x u,\ \partial_{xx} u\}$
Kolmogorov Flow	$\partial_t \omega + \mathbf{v} \cdot \nabla \omega = \frac{1}{Re}\nabla^2 \omega - \chi\omega + f$	$\{\mathbf{v} \cdot \nabla\omega,\ \partial_t \omega,\ \omega\}$

Training

- FNO emulator on *noisy* observations
- MSE warm-up, then OT regularization

Statistical metrics (full roll-out)

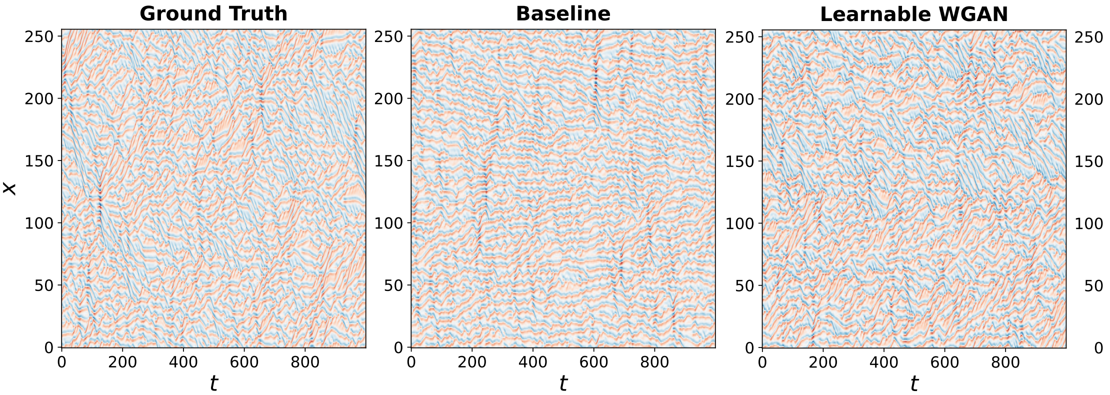
- L^1 **Histogram:** distributional fidelity of eval. summary statistics
- **Spectral distance:** Relative error of the energy spectrum

Results: Statistical Metrics

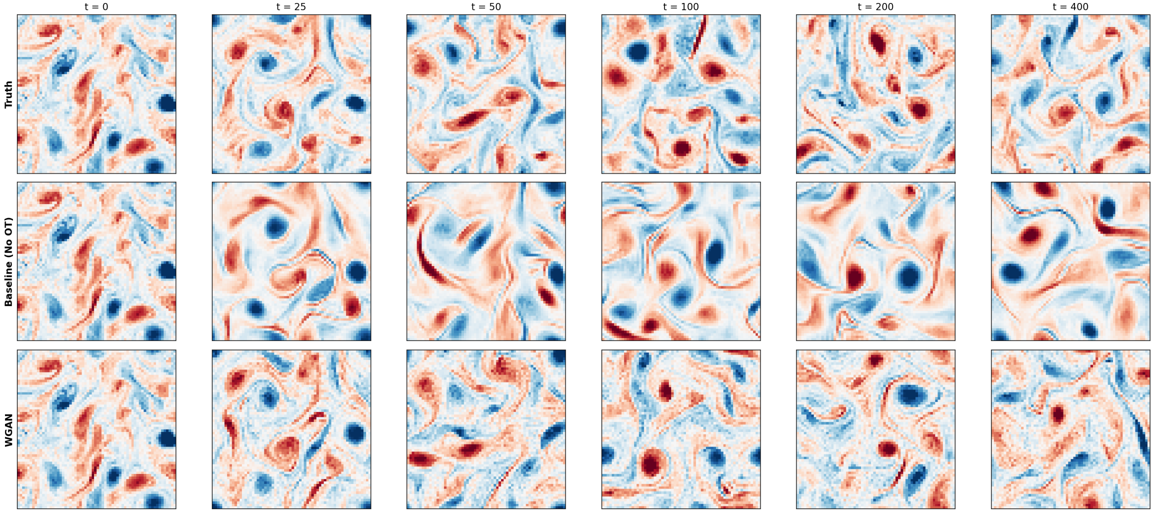
Table: Statistical metrics evaluated on noisy data ($\sigma = 0.3$). Shaded = our methods.

	Method	$L^1 \downarrow$	Spec. $\downarrow$
Lorenz-96	No OT (baseline)	0.348	0.321
	Fixed OT	0.175	0.141
	Sinkhorn (MLP)	0.175	0.112
	WGAN (MLP)	**0.149**	**0.049**
Kuramoto–Sivashinsky	No OT (baseline)	0.454	0.351
	Fixed OT	0.290	0.349
	Sinkhorn (MLP)	0.435	0.219
	WGAN (MLP)	0.339	**0.148**
	WGAN (1D-Conv)	**0.259**	0.198
Kolmogorov Flow	No OT (baseline)	0.660	0.249
	Fixed OT	0.614	0.168
	Sinkhorn (2D-Conv)	0.360	**0.144**
	WGAN (2D-Conv)	**0.273**	0.157

Results: Kuramoto–Sivashinsky Roll-out Visualization



Results: Kolmogorov Flow Roll-out Visualization



Summary

Contributions

1. **Problem:** MSE is ill-suited for chaotic emulation at long horizons
2. **Method:** Adversarial OT regularization with a *learnable* summary map f
3. **Relaxations:** Sinkhorn (primal) and WGAN-style (dual) formulations
4. **Theory:**
   - MSE bounds OT $\Rightarrow$ warm-up theoretically justified
   - MSE blows up exponentially even for a perfect emulator
   - Wasserstein is robust to IC noise via ergodic mixing
5. **Experiments:** Consistent distributional improvements on L96, KS, KF; WGAN recovers true Lyapunov exponent

`arXiv:2604.21097`