Google Research

Advances in Probabilistic Generative Modeling for Scientific Machine Learning

Fei Sha (fsha@google.com)

Google Research

June 11, 2025

http://www.feisha.org

Google Research

Acknowledgement

Multi-disciplinary team of

meteorologists, applied mathematicians, climate scientists, computational physicists and computer scientists

Google Research

Thanks, the organizers

especially Leo, for teaching

a computer scientist like me

PDE, fluids, etc and

allowing me to use some of his slides

Disclaimer. All turbulent drivels are probabilistically mine, not Leo’s faults.

An everyday -life motivation: travel via air

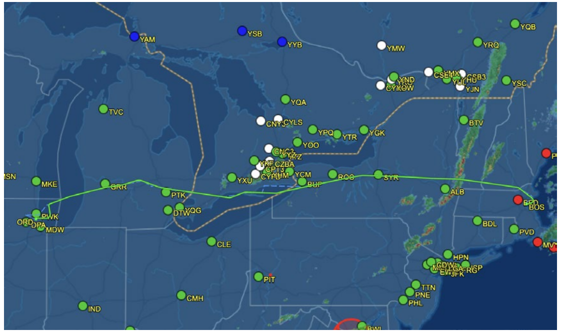
A lot of questions

What makes it fly?

How to make it fly faster, consume less fuel?

How to make it safer, cheaper?

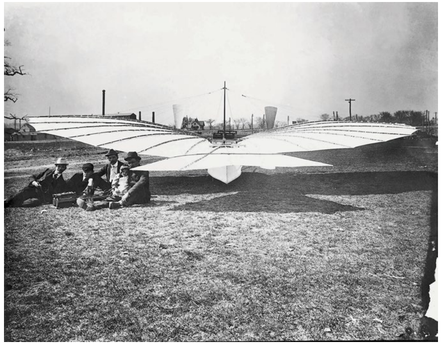
…



[Photo Credits: Flightaware.com]

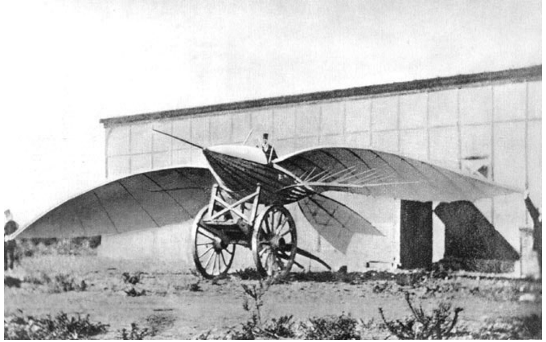
Google Research

Earlier days (1850s - 1900s): build → crash → repeat



Monoplane No 21 by Gustave Weißkopf.
(Wikipedia)

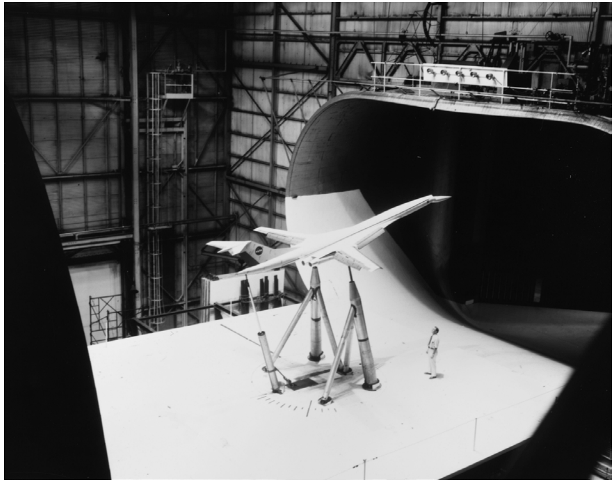
Santos-Dumont 14-bis.
(Wikipedia)



Albatros II by Jean-Marie Le Bris.
(Wikipedia)

1920s-1970s: driven by scientific principles and experiments

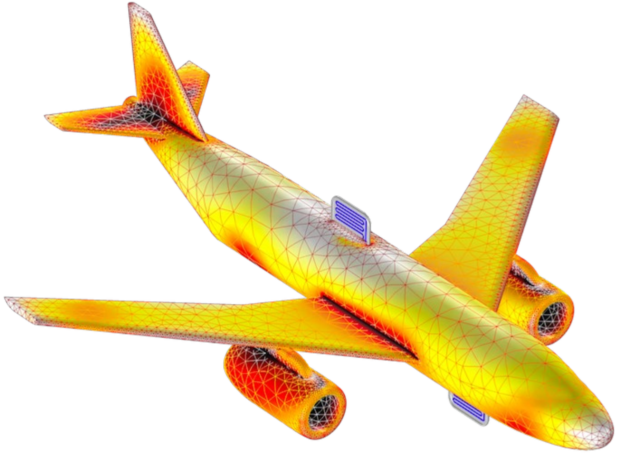
Replica of Wright brothers wind tunnel
(Wikipedia)



Scale model in wind tunnel.
(NASA/JPL)

2000s: in silico design

Aerodynamics simulation
(Simulia)

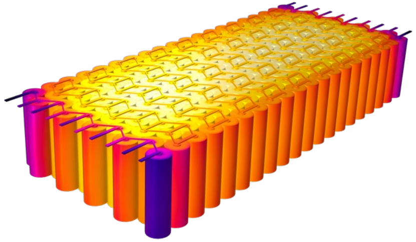


Antenna cross-talk simulation
(Comsol)

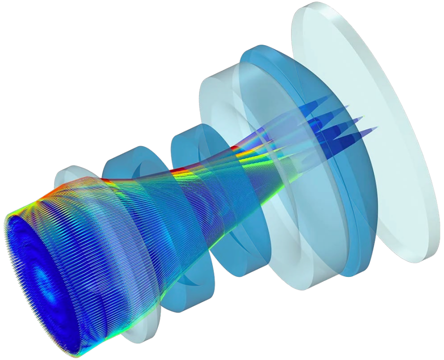
Computation and simulation is ubiquitous

Market for Simulation SW 41.8B in 2033, 10.8% per year

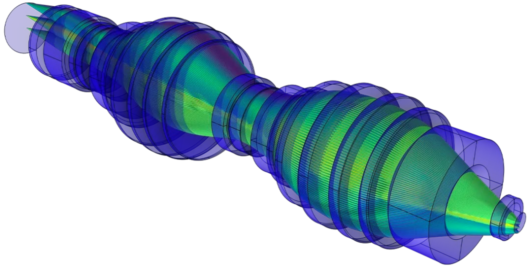
Accelerating design and discovery



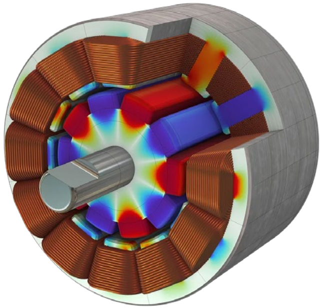
Car battery simulation, COMSOL



Camera lenses simulations (COMSOL)



Microlithography lens system simulation, COMSOL



Electric motor simulation, COMSOL

But, there are many challenges

High fidelity simulations require very fine discretization of space and time

Need to overcome computational cost increase in quadratic to cubic

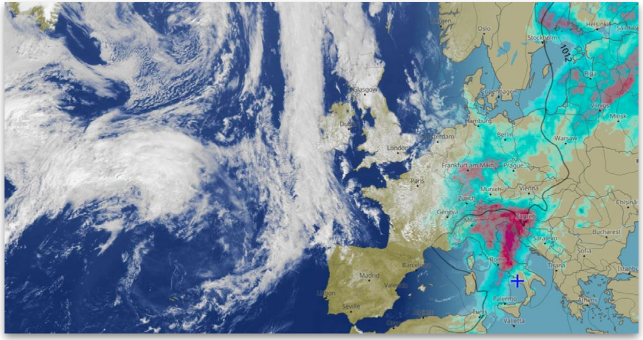
Many systems are highly turbulent/chaotic or operate in unknown conditions

Need uncertainty quantification of their behavior

For optimal design, many configurations need to be run

Need to summarize/survey the whole design space efficiently

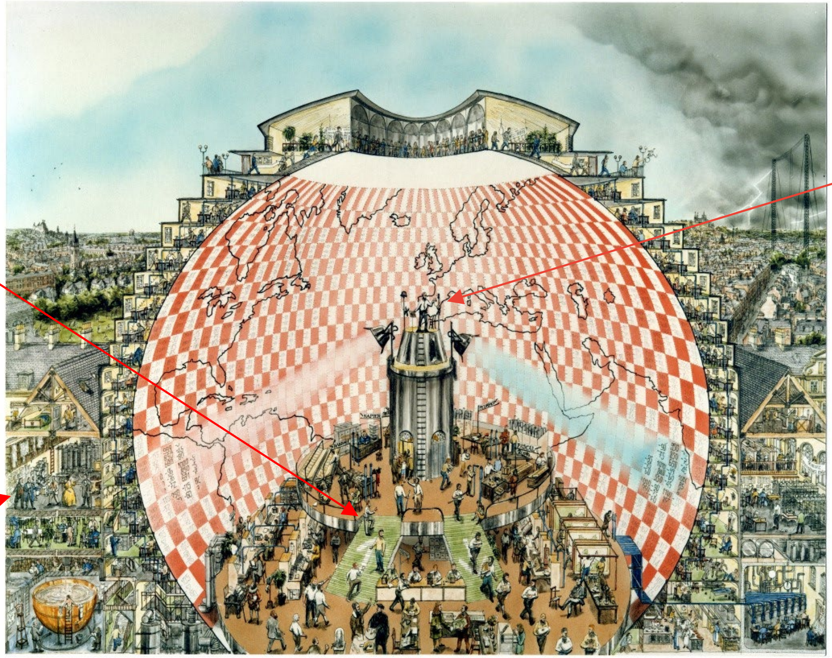
Another real-world example:
understanding and modeling weather/climate systems



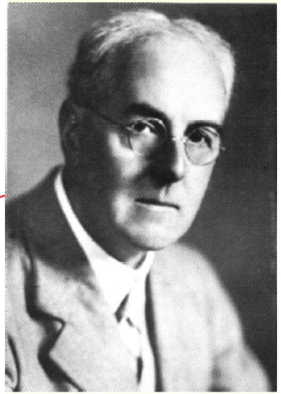
[Photo Credits: MeteoBlue, CNBC, University of Utah]

Richardson's Fantastic Forecast Factory ~ 100 years ago

Charles Babbage



"Algorithm room"
George Boole
Ada Lovelace



Lewis Richardson (1922)
"Weather Prediction by Numerical Process"

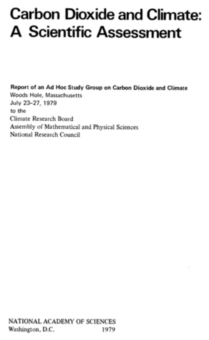
(Credit. https://www.emetsoc.org/resources/rff/)

30 years later, first real computer-generated numerical weather forecast

von Neumann

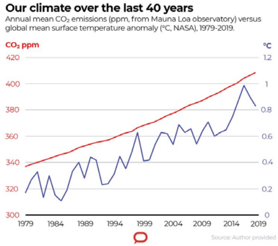
Jules Charney

[First Weather Forecast by Computer, 1950]


Carbon Dioxide and Climate: A Scientific Assessment

Report of an Ad Hoc Study Group on Carbon Dioxide and Climate
Woods Hole, Massachusetts
July 23-27, 1979
to the
Climate Research Board
Assembly of Mathematical and Physical Sciences
National Research Council

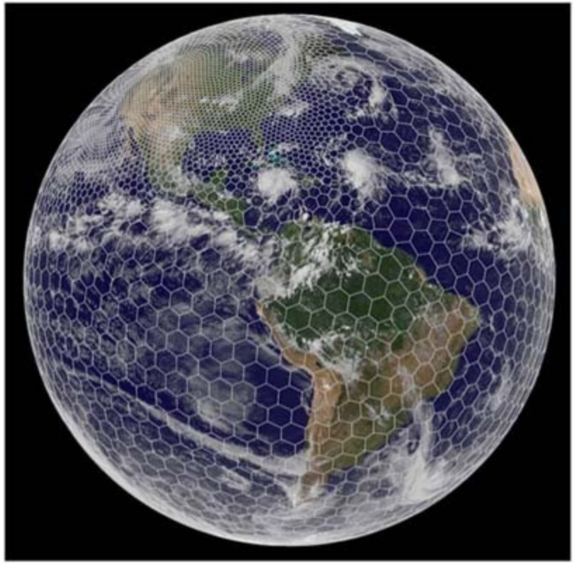
NATIONAL ACADEMY OF SCIENCES
Washington, D.C. 1979



https://phys.org/news/2019-07-charney-years-scientists-accurately-climate.html

Google Research

What is the “scaling” law in numerical weather prediction?



[Source. ***https://www.ecmwf.int/en/newsletter/172/editorial/towards-greater-resolution***]

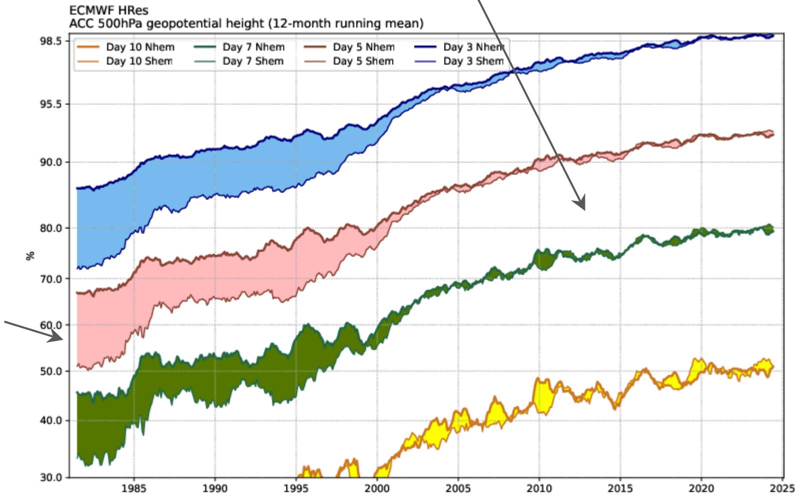
Year	Resolution	
1950	270km	
1979	200km	~1.5 Chicago metro
1991	60km	~1.5 Cook county
2006	25km	~1 Chicago city
2022	9km	~
2025	3km (exp)	~
2030?	3km	
2035+	1km (cloud resolving)	~ 0.25 Hyde Park

"Headline" news: AI for NWP has been accelerating the process

Before 2022: Quiet (r) evolution

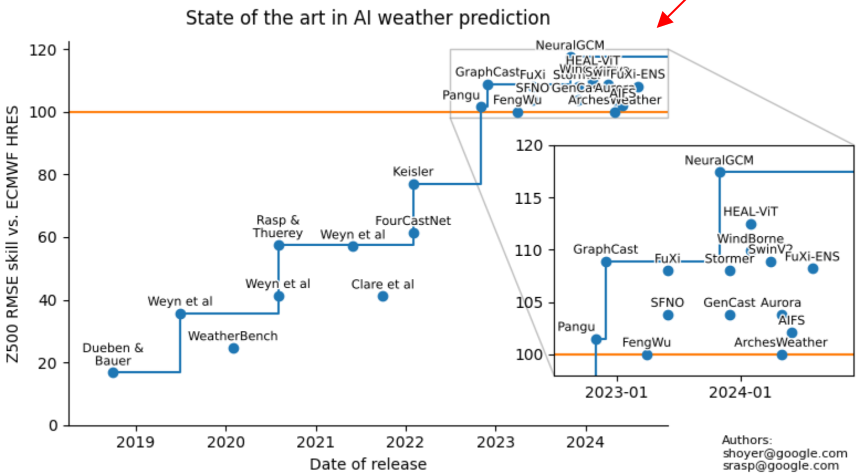
Improvement: 1 day / decade

Useful (60%)



[Plot Credits. ECMWF (https://arxiv.org/pdf/2407.03787)]

After 2022: Roaring revolution led by *AI-based* numerical weather prediction



WeatherBench 2

AI surpassing operational system w/ caveat

Scientific ML: develop ML technology to tackle those challenges

High fidelity simulations require very fine discretization of space and time

Need to overcome computational cost increase in quadratic to cubic

Many systems are highly turbulent/chaotic or operate in unknown conditions

Need uncertainty quantification of their behavior

For optimal design, many configurations need to be run

Need to summarize/survey the whole design space efficiently

Talk based on a subset of our work in this space

Representation and dynamics learning

A. Boral, Z. Y. Wan, L Zepeda-Núñez, J. Lottes, Q. Wang, Y. Chen, J. Anderson, F Sha. Neural Ideal Large Eddy Simulation: Modeling Turbulence with Neural Stochastic Differential Equations, NeurIPS 2023.

Z. Y. Wan, L. Zepeda-Núñez, A. Boral, F. Sha. Evolve Smoothly, Fit Consistently: Learning Smooth Latent Dynamics For Advection-Dominated Systems, ICLR 2023

Probabilistic generative modelling

M. A. Finzi, A. Boral, A. G. Wilson, F. Sha, L. Zepeda-Núñez, User-defined Event Sampling and Uncertainty Quantification in Diffusion Models for Physical Dynamical Systems, ICML 2023

L. Li, R. Carver, I. Lopez-Gomez, F. Sha, J. Anderson. SEEDS: Emulation of Weather Forecast Ensembles with Diffusion Models. Sciences Advances 2024.

Y. Schiff, Z. Y. Wan, J. B. Parker, S. Hoyer, V. Kuleshov, F. Sha, L. Zepeda-Núñez. DySLIM: Dynamics Stable Learning by Invariant Measure for Chaotic Systems. ICML 2024.

Z. Y. Wan, R. Baptista, Y. Chen, J. Anderson, A. Boral, F. Sha, L. Zepeda-Núñez. Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models, NeurIPS 2023.

I. Lopez-Gomez, Z. Y. Wan, L. Zepeda-Núñez, T. Schneider, J. Anderson, F. Sha. Dynamical-generative downscaling of climate model ensembles. 2024

R. Molinaro, S. Lanthaler, B. Raonić, T. Rohner, V. Armegioiu, Z. Y. Wan, F. Sha, S. Mishra, L. Zepeda-Núñez. Generative AI for fast and accurate Statistical Computation of Fluids. 2024

Open source code

https://github.com/google-research/swirl-dynamics

Prelude

Vignette 1: Methods

Vignette 2: Application

Vignette 3: Theory

Final thoughts

Prelude

Modeling dynamical systems

State variables (and/or observations) evolve, in discrete time steps

$$\mathbf{u}_k = \mathcal{S}(\mathbf{u}_{k-1}) = \ldots = \mathcal{S}^k(\mathbf{u}_0) \qquad \mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathcal{U}$$

Sometimes with known governing equations (ex: Navier-Stokes equations)

$$\partial_t u + u \cdot \nabla u = -\frac{1}{\rho}\nabla P + \nu^2 \nabla^2 u$$

Those variables are vector-valued functions in both space and time.
thus, infinite-dimensional/high-dimensional objects.

Learning tasks

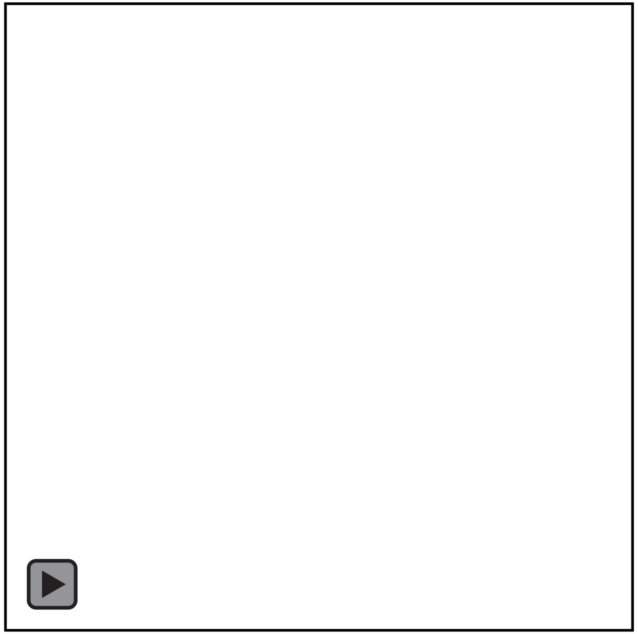
Data

trajectories, from observations or simulation

Goals: learn a model that mimics system dynamics

Reduce computation : evolving the learned model efficiently

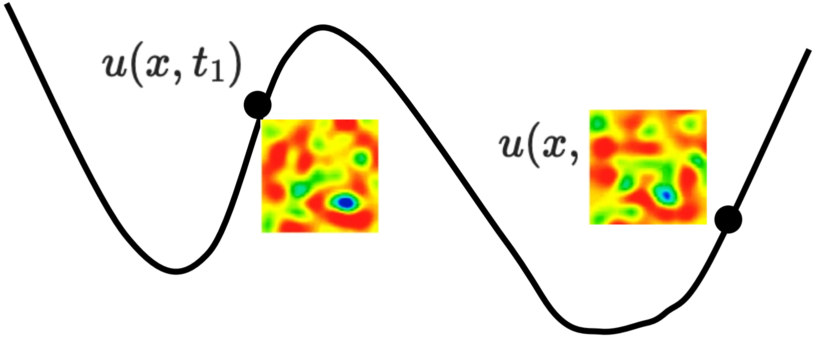
Maintain fidelity : matching trajectories or their (large - scale) spatiotemporal patterns



Is not this just a supervised learning of predicting “video”?

On a high-level, right. But there are **nuances**.

observing trajectory τ



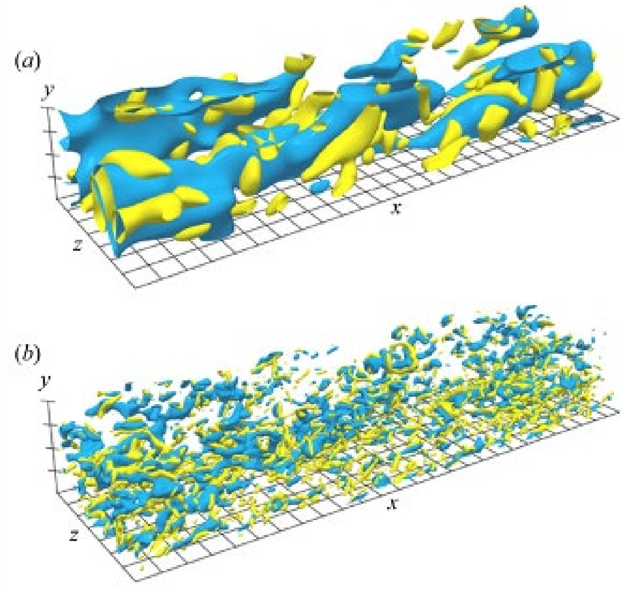
learning to predict

$$\min \ \mathbb{E}_\tau \sum_{(u_i,u_j)\in\tau} \|\mathcal{S}_\theta(u_i) - u_j\|_2^2$$

Multi - scale structures are common in turbulent flows

coarse resolution

high resolution



Big whirls have little whirls,
That feed on their velocity;
And little whirls have lesser whirls,
And so on to viscosity.

(Lewis Fry Richardson)

(Credit. https://www.youtube.com/watch?v=IwAoNha2Jpc&ab_channel=AmericanPhysicalSociety)

Less poetic translation

At larger (spatial) scales, we see well

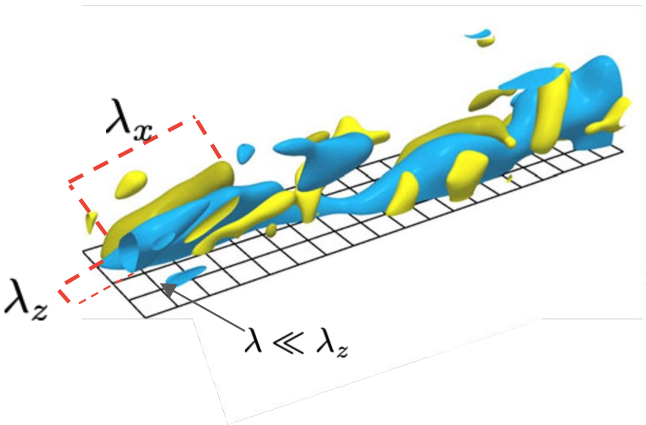
thus, physics is resolved .

At smaller scales, we cannot afford the compute,

thus, physics is unresolved .

Yet, their interaction is the troublemaker

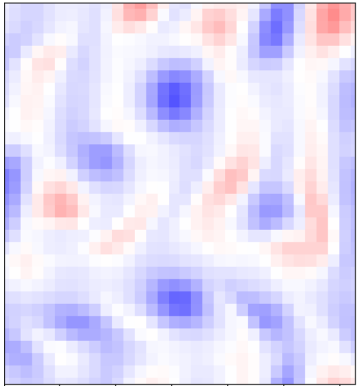
as the nonlinearity is the culprit .



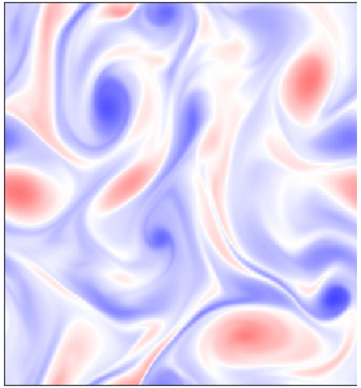
Challenge: cost of reducing computation

Bias: how do we bridge the gap?

low-fidelity, biased simulation
cheap to compute



high-fidelity simulation
costly to compute



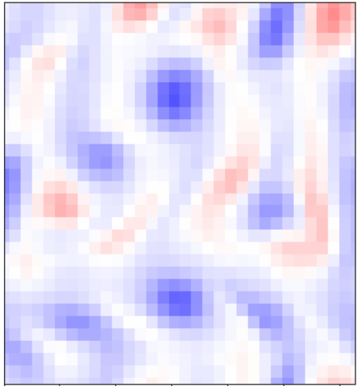
Vignette #1: Methods

generative modeling for
coarse - grained modeling

Google Research

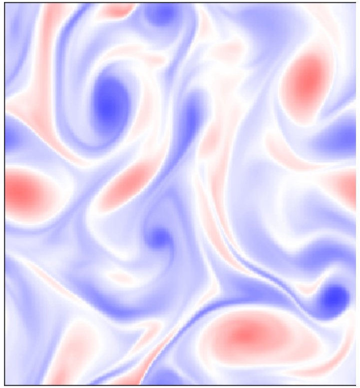
Increasing resolution via statistical models

low resolution simulation
cheap to compute



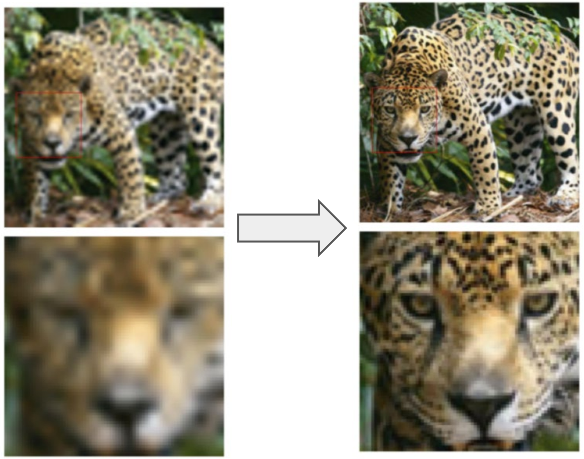
AI/ML methods?

high resolution simulation
costly to compute



Can we just super-resolution?

classical supervised learning setup in computer vision

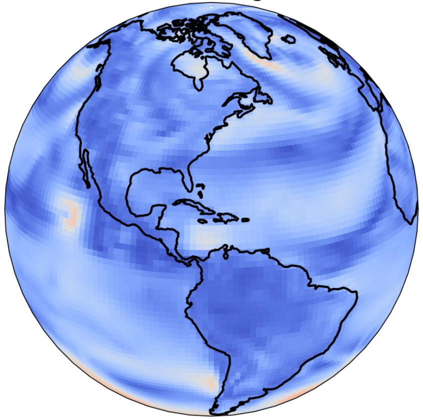


Google Research

Snag #1: bias in real world problem

coarse climate simulation

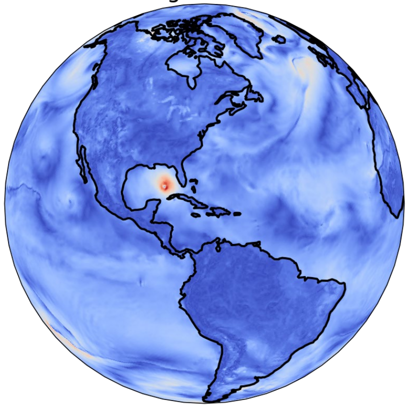
LENS2 (member 1) August 28th 2005



Hurricane Katrina is absent

high-resolution weather

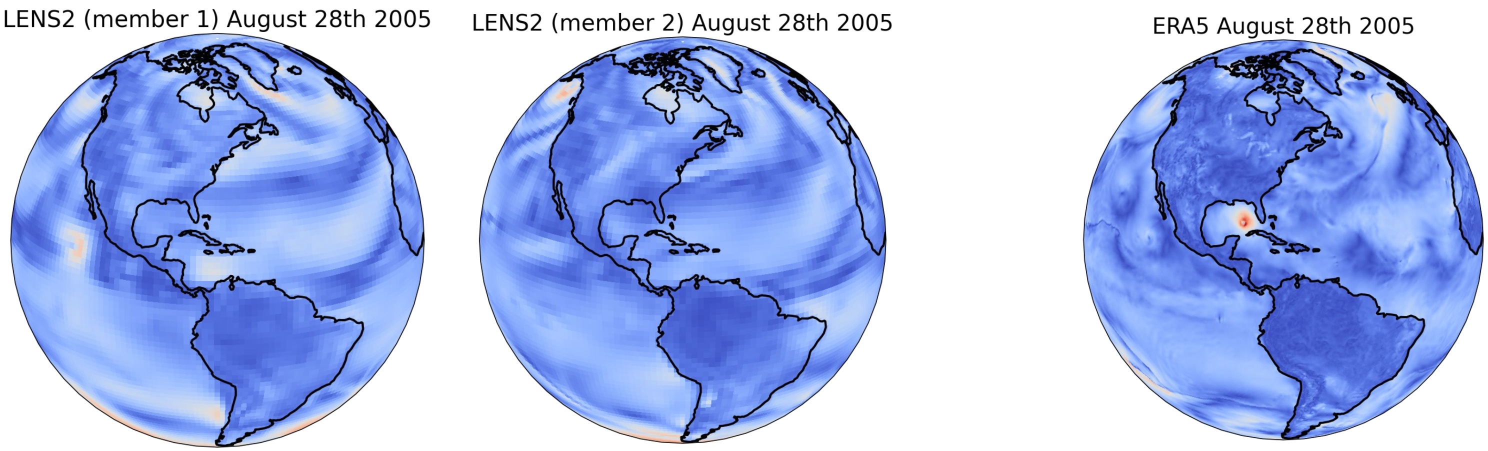
ERA5 August 28th 2005



Snag #2: intertwined bias and lack of correspondence

from the same period (decade), but do not match at the same time point

not a supervised learning problem where exact one to one mapping exists



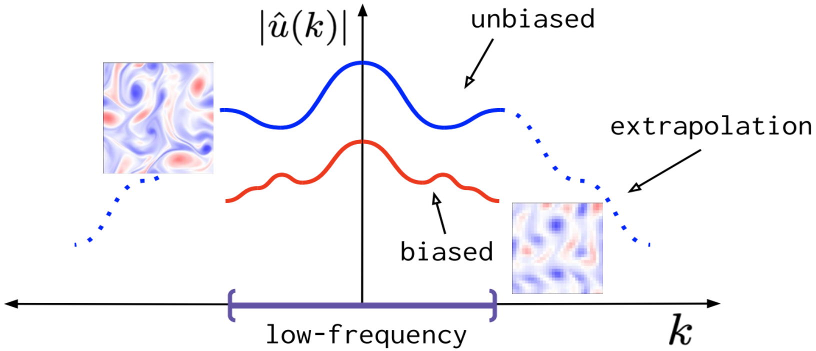
coarse climate simulation

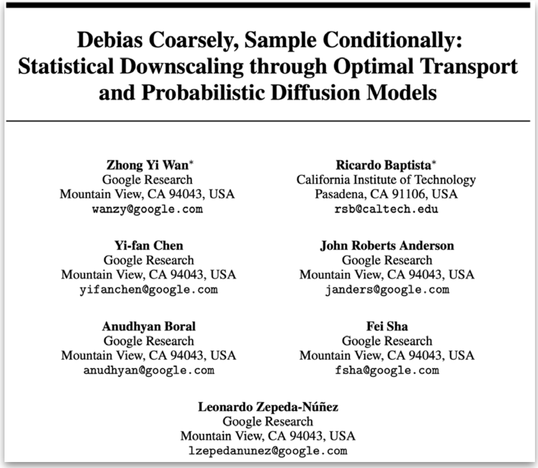
high - resolution weather

Two goals in one task

Correcting bias

Adding details





Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models

Zhong Yi Wan* Google Research Mountain View, CA 94043, USA wanzy@google.com

Ricardo Baptista* California Institute of Technology Pasadena, CA 91106, USA rsb@caltech.edu

Yi-fan Chen Google Research Mountain View, CA 94043, USA yifanchen@google.com

John Roberts Anderson Google Research Mountain View, CA 94043, USA janders@google.com

Anudhyan Boral Google Research Mountain View, CA 94043, USA anudhyan@google.com

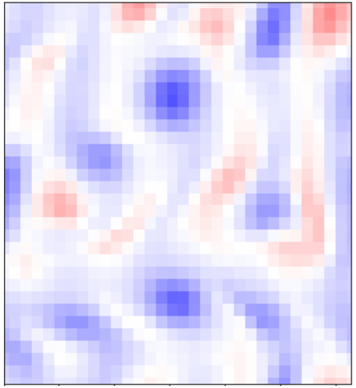
Fei Sha Google Research Mountain View, CA 94043, USA fsha@google.com

Leonardo Zepeda-Núñez Google Research Mountain View, CA 94043, USA lzepedanunez@google.com

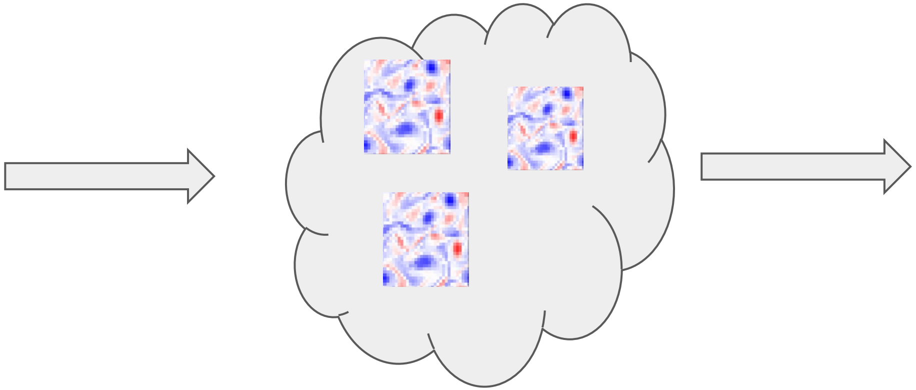
[NeurIPS 2023]

Our approach: latent variable modeling

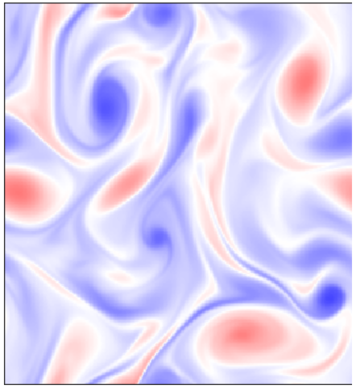
biased low resolution *y*



latent unbiased low - resolution *y'*



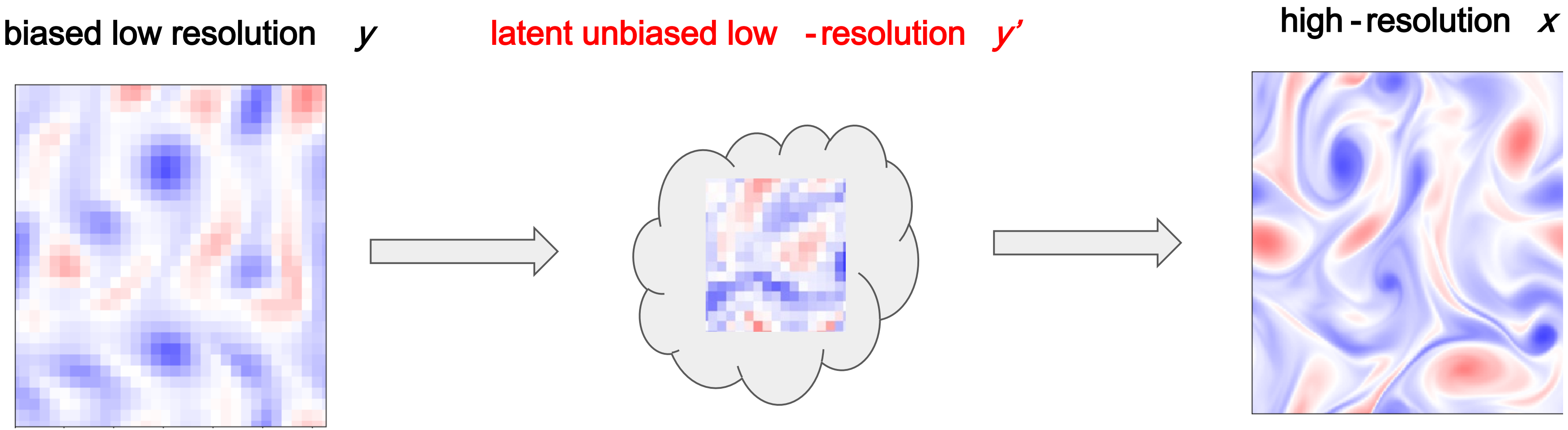
high - resolution *x*



$$p(x|y) = \int_{Y'} p(x|y')p(y'|y)dy'$$

Avoid costly posterior inference via “clamping” the latent variable

Often referred as “deterministic” EM



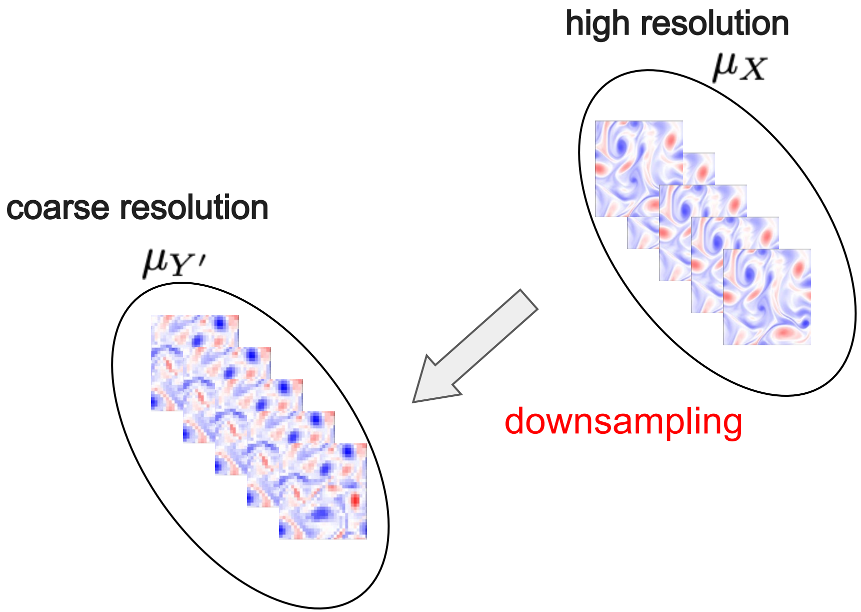
$$p(x|y) = \int_{Y'} p(x|y')p(y'|y)dy'$$
$$= p(x|y')\delta(y' = T(y))$$

The key is to select a good candidate latent variable

Downsample high-resolution training data

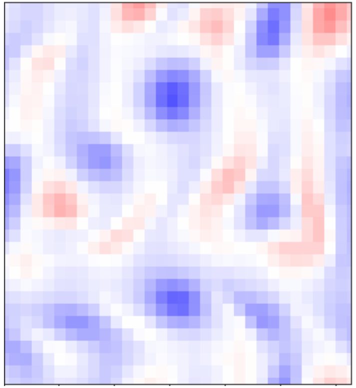
Pretend it as the mode of the posterior

$$p(x|y) = \int_{Y'} p(x|y')p(y'|y)dy'$$
$$= p(x|y')\delta(y' = T(y))$$



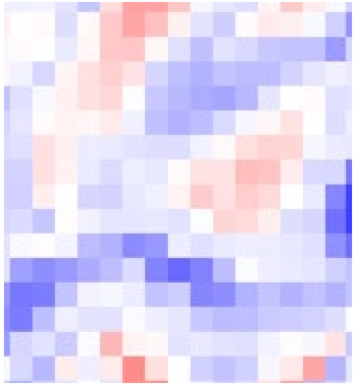
Deterministic EM decouples the learning into two stages

biased low resolution y



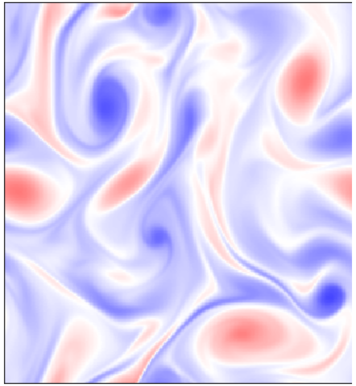
$T(\cdot)$

latent unbiased low-resolution y'



invert downsampling

high-resolution x



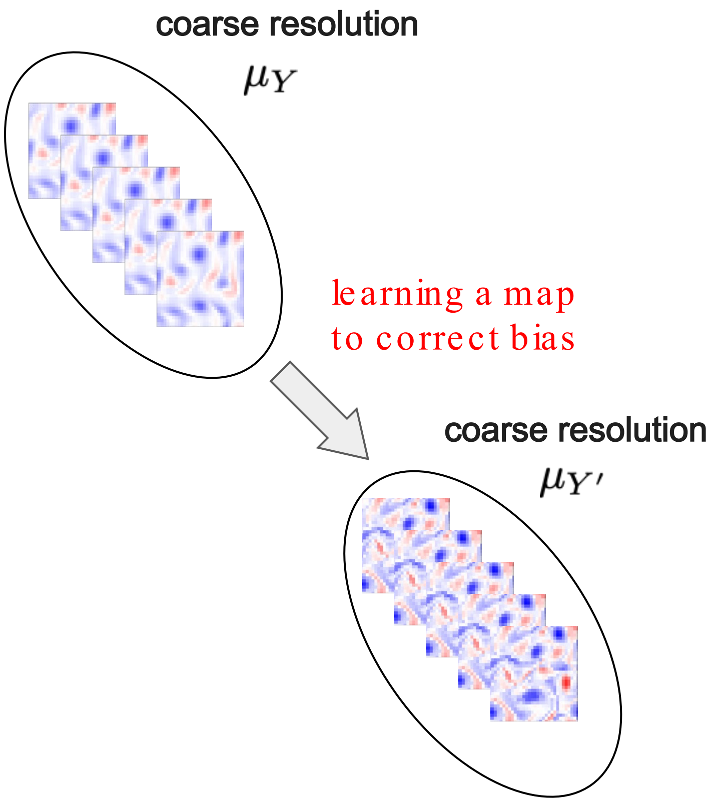
$$p(x|y) = \int_{Y'} p(x|y')p(y'|y)dy'$$
$$= p(x|y')\delta(y' = T(y))$$

Stage 1: align manifolds and match distributions

Learning via optimal transport

$$\min_T \left\{ \int c(y, T(y))\, d\mu_Y(y) : T_{\#}\mu_Y = \mu_{Y'} \right\}$$

- many scalable optimization algorithms have been developed
- connected to generative modeling methodologies such as flow matching.



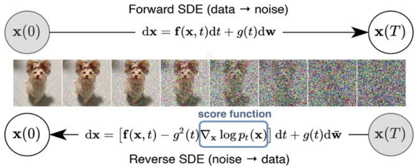
Stage 2: super-resolution via denoising diffusion model

Learn to invert the downsampling

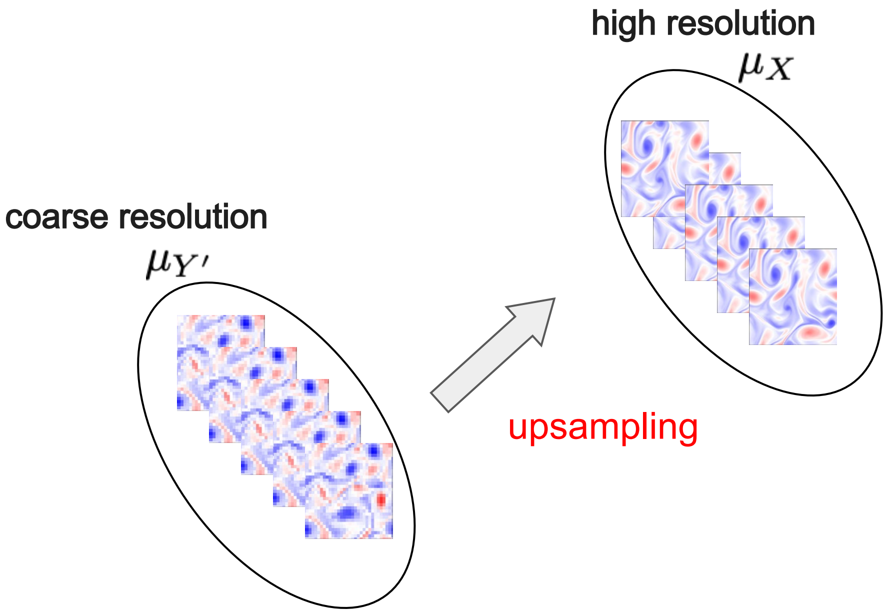
Supervised learning with paired data

Well studied, standard recipes

[Saharia et al, 2020]

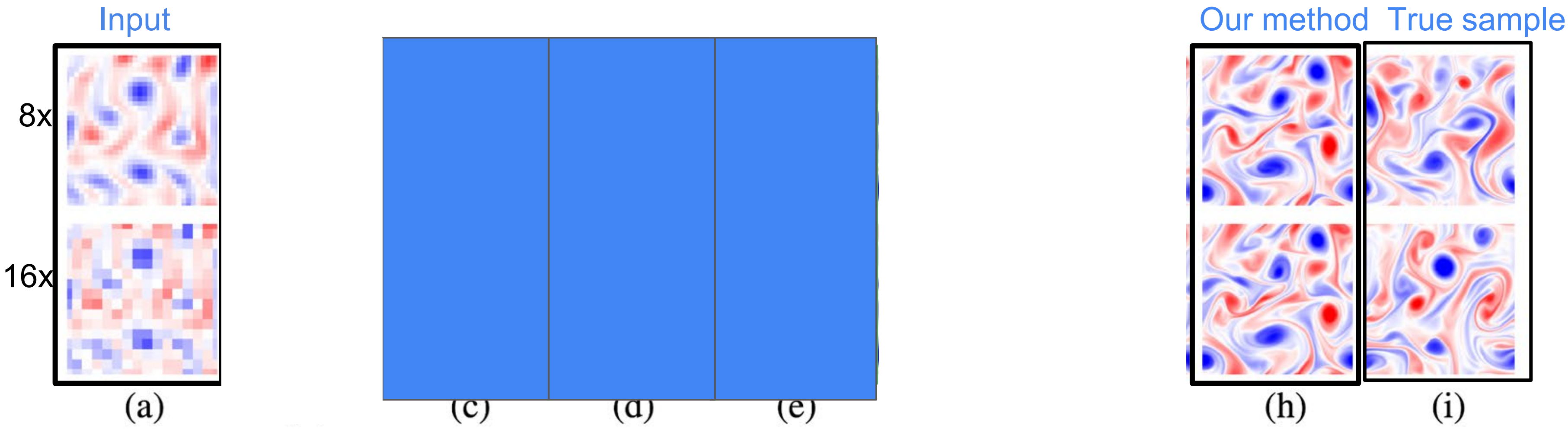


[Song et al, 2020]



Generative downscaling a low-resolution Kolmogorov flow

Optimal transport corrects bias; denoising diffusion model fills details



a) low resolution

c) BCSD

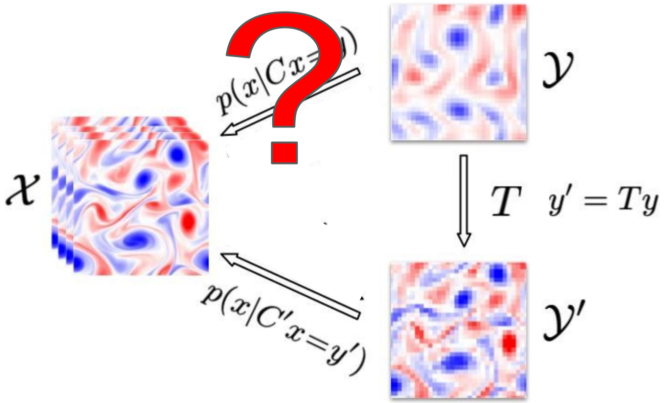
f) OT + cubic interpolation

g) OT + ViT
h) Ours
i) true samples

Why our method works better

Model multivariate distributions

Maintain spatiotemporal patterns

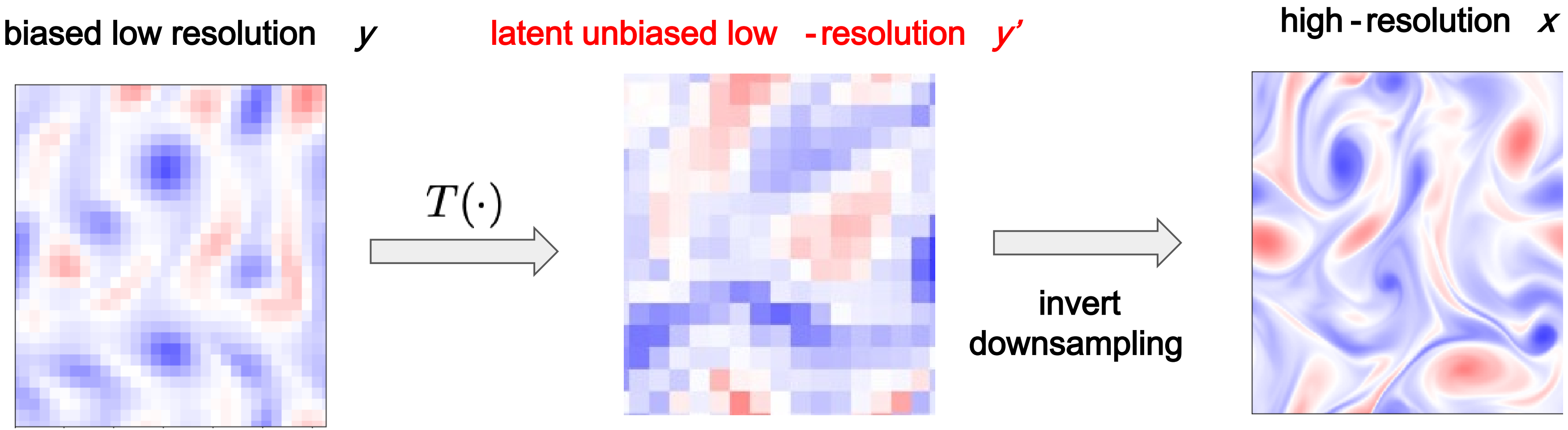


more quantitative measures

Model	Var	covRMSE↓	MELRu↓	MELRw↓	KLD↓	Wass1↓	MMD↓
8× downscale							
BCSD	0	0.31	0.67	0.25	2.19	**0.23**	0.10
cycGAN	0	0.15	0.08	0.05	1.62	0.32	0.08
ClimAlign	0	2.19	0.64	0.45	64.37	2.77	0.53
Raw+cDfn	0.27	0.46	0.79	0.37	73.16	1.04	0.42
OT+Cubic	0	**0.12**	0.52	0.06	1.46	0.42	0.10
OT+ViT	0	0.43	0.38	0.18	1.72	1.11	0.31
(ours) OT+cDfn	0.36	**0.12**	**0.06**	**0.02**	**1.40**	0.26	**0.07**
16× downscale							
BCSD	0	0.34	0.67	0.25	2.17	**0.21**	0.11
cycGAN	0	0.32	1.14	0.28	2.05	0.48	0.13
ClimAlign	0	2.53	0.81	0.50	77.51	3.15	0.55
Raw+cDfn	1.07	0.46	0.54	0.30	93.87	0.99	0.39
OT+Cubic	0	0.25	0.55	0.13	7.30	0.85	0.20
OT+ViT	0	0.14	1.38	0.09	1.67	0.32	**0.07**
(ours) OT+cDfn	1.56	**0.12**	**0.05**	**0.02**	**0.83**	0.29	**0.07**

But, if we want to get "real" unbiased coarse resolution?

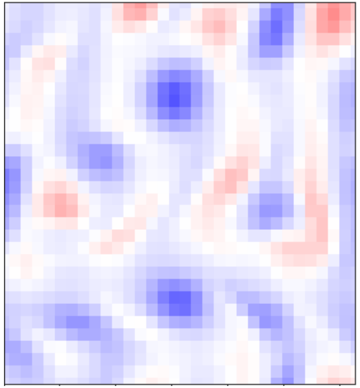
Distribution matching is a weaker notion (of correspondence)



Incorporating physics into generative modeling

Statistics work with stronger prior knowledge

low resolution simulation
cheap to compute



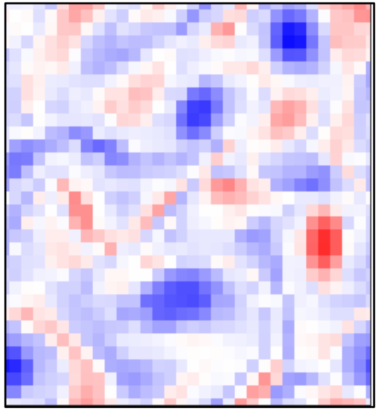
adding physics

+

statistical tooling

→

unbiased coarse resolution



Closure modeling with generative models

Classical setup

$$\partial_t u = \mathcal{R}^{\mathrm{NS}}(u;\nu)$$
$$\bar{u} = G \star u$$
$$\partial_t \bar{u} = \mathcal{R}^{\mathrm{NS}}(\bar{u};\nu) + \mathcal{R}^{\mathrm{closure}}(\bar{u}, u)$$

Supervised learning of a closure model

$$\partial_t \tilde{u} = \mathcal{R}_c^{\mathrm{NS}}(\tilde{u};\nu) + \mathcal{M}(\tilde{u};\theta)$$
$$\theta^* = \arg\min_\theta \sum_i \|\tilde{u}_i - \bar{u}_i\|_2^2$$

Neural Ideal Large Eddy Simulation: Modeling Turbulence with Neural Stochastic Differential Equations

Anudhyan Boral
Google Research
Mountain View, CA 94043, USA
anudhyan@google.com

Zhong Yi Wan
Google Research
Mountain View, CA 94043, USA
wanzy@google.com

Leonardo Zepeda-Núñez
Google Research
Mountain View, CA 94043, USA
lzepedanunez@google.com

James Lottes
Google Research
Mountain View, CA 94043, USA
jlottes@google.com

Qing Wang
Google Research
Mountain View, CA 94043, USA
wqing@google.com

Yi-fan Chen
Google Research
Mountain View, CA 94043, USA
yifanchen@google.com

John Roberts Anderson
Google Research
Mountain View, CA 94043, USA
janders@google.com

Fei Sha
Google Research
Mountain View, CA 94043, USA
fsha@google.com

[NeurIPS 2023]

A Probabilistic Perspective

Ideal LES field

$$\frac{\partial v}{\partial t} = \mathbb{E}_{\pi_t}\left[\overline{\frac{\partial u}{\partial t}} \,\middle|\, \bar{u} = v \right]$$

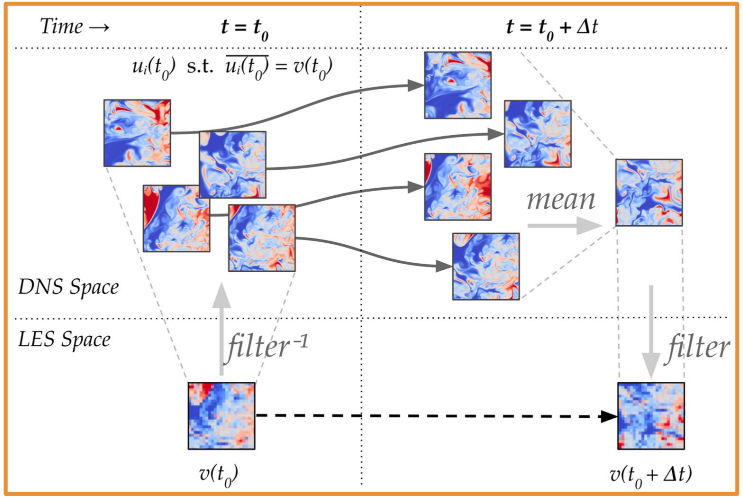
Corresponding closure model

$$\partial_t v = \mathcal{R}_c^{\mathrm{NS}}(v) + \mathcal{M}(v)$$
$$\mathcal{M}(v) = \mathbb{E}_{\pi_t}[\overline{\partial_t u} | \bar{u} = v] - \mathcal{R}_c^{\mathrm{NS}}(v)$$

Challenge

Analytically intractable

Unknown distribution

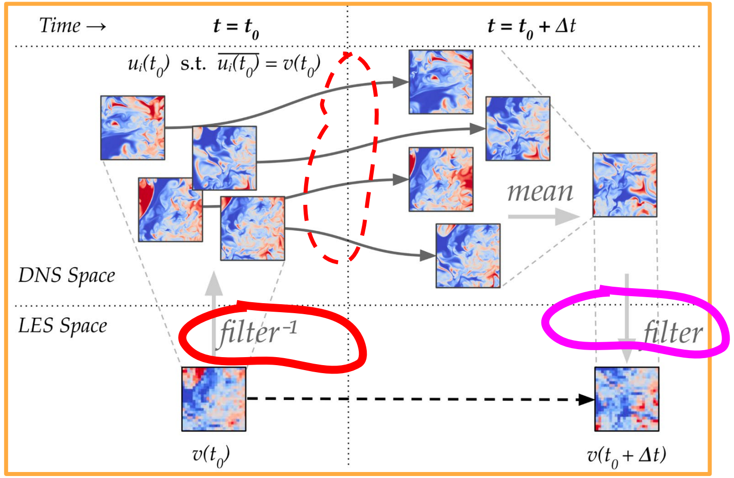


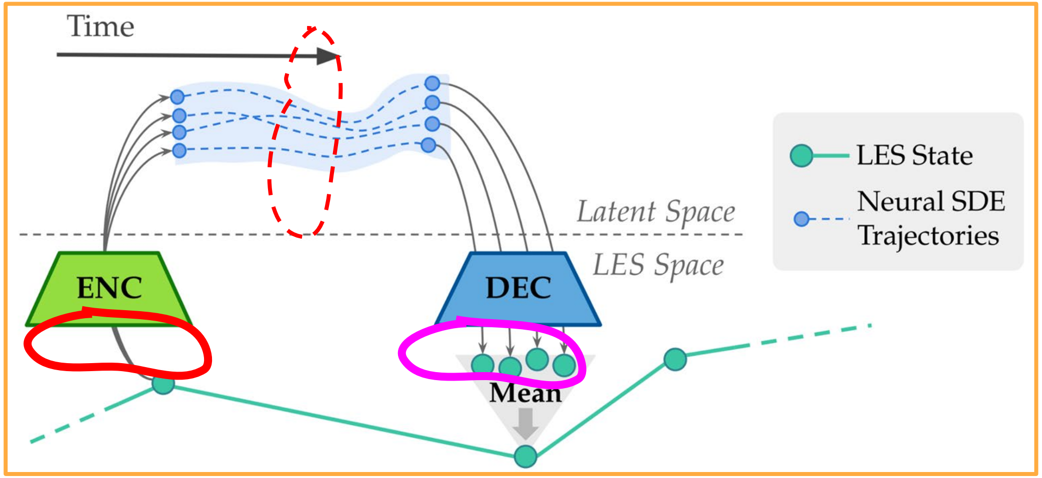
LES field is the ensemble mean of its corresponding DNS fields

[Langford and Moser, 1999]

Neural LES: learning the latent distribution

Intuitively, the latent space is our imaginary “DNS” (high resolution) space





$\mathcal{M}(v) = \mathbb{E}_{\pi_t}[\overline{\partial_t u}|\bar{u} = v] - \mathcal{R}_c^{\mathrm{NS}}(v)$ is approximated by simulation inside the latent space.

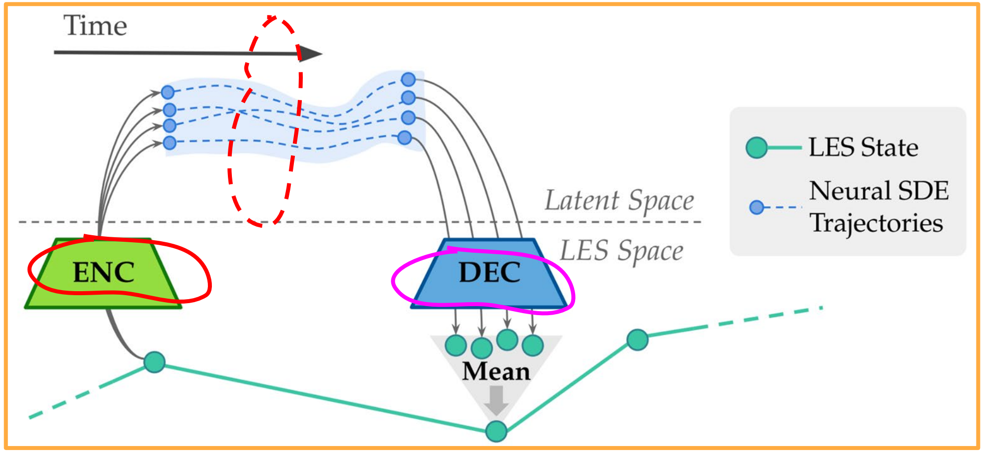
Modeling latent distribution with neural parameterized SDE

latent representation of DNS fields

neural nets for drifting and diffusion

Wiener process for stochasticity

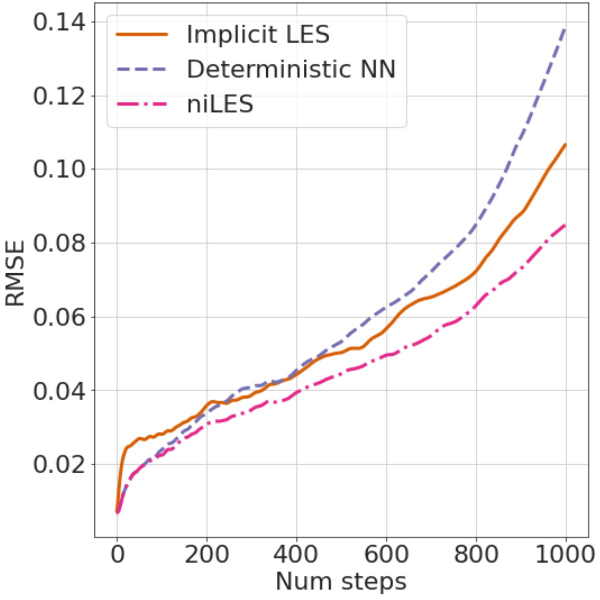
$$dZ_t = h_\theta(Z_t, t)dt + g_\phi(Z_t, t) \circ dW_t$$

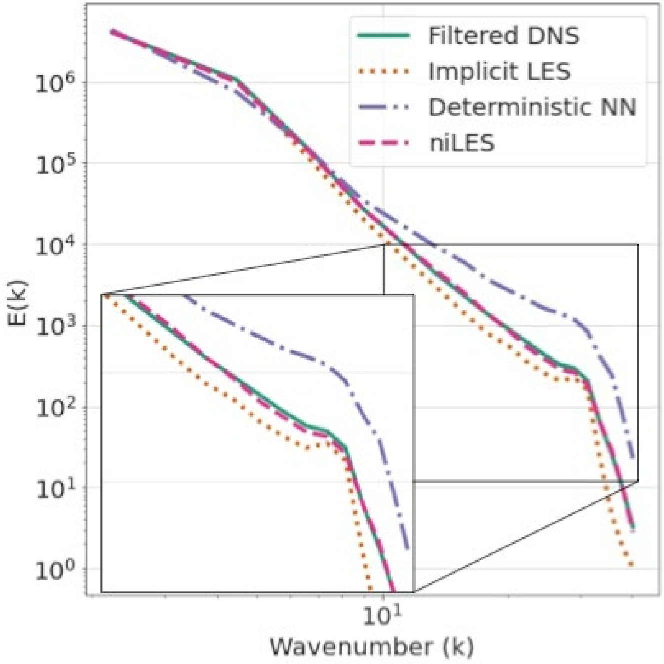


[Li et al, AISTATS, 2020]

Numerical study on Kolmogorov flow

Our neural ideal LES (niLES) is more stable, and get the energy spectra right





Prelude

Vignette 1: Methods

Vignette 2: Application

Vignette 3: Theory

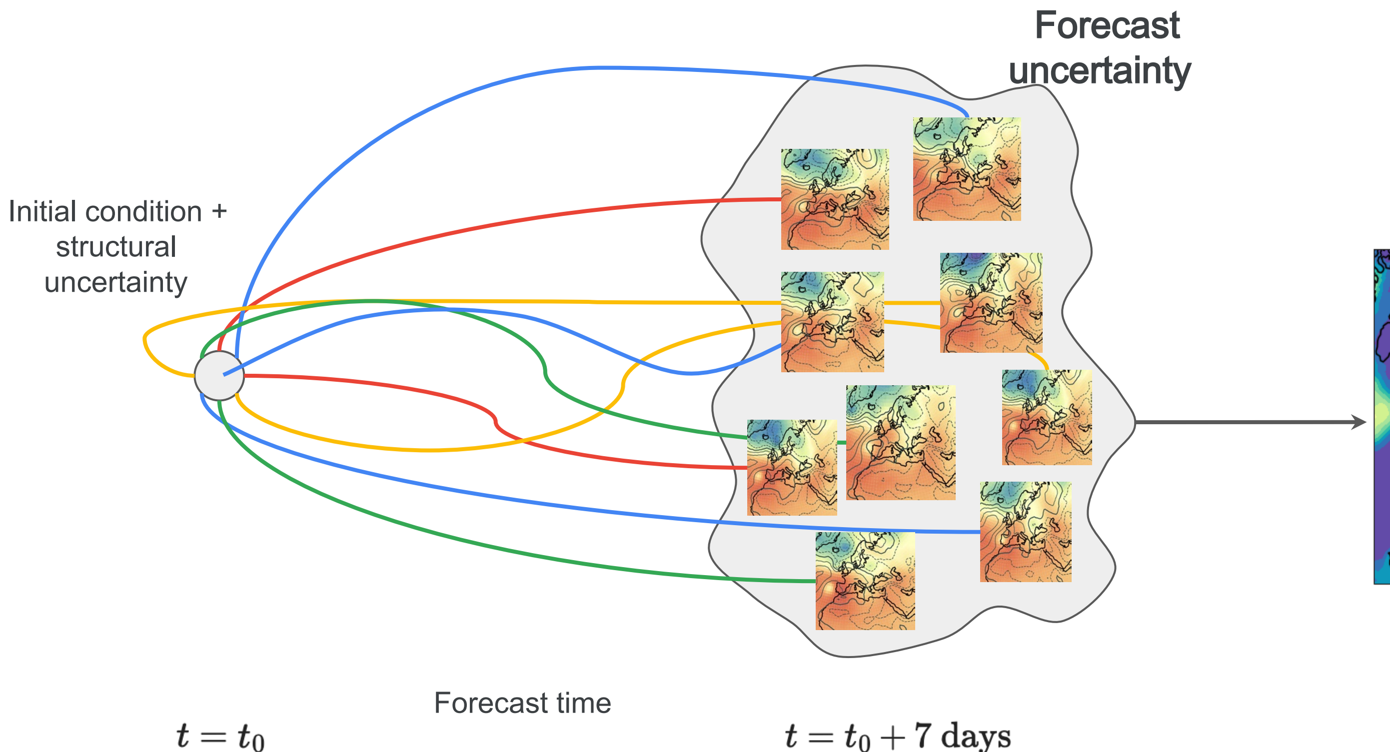
Final thoughts

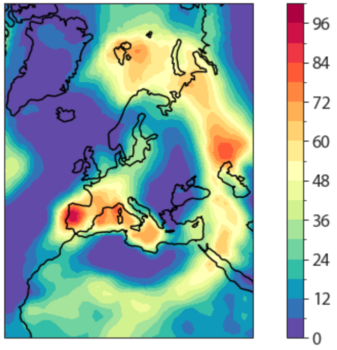
Vignette #2: Application

generative modeling for weather and climate

High - resolution probabilistic forecast

Ensemble forecast is computationally very expensive

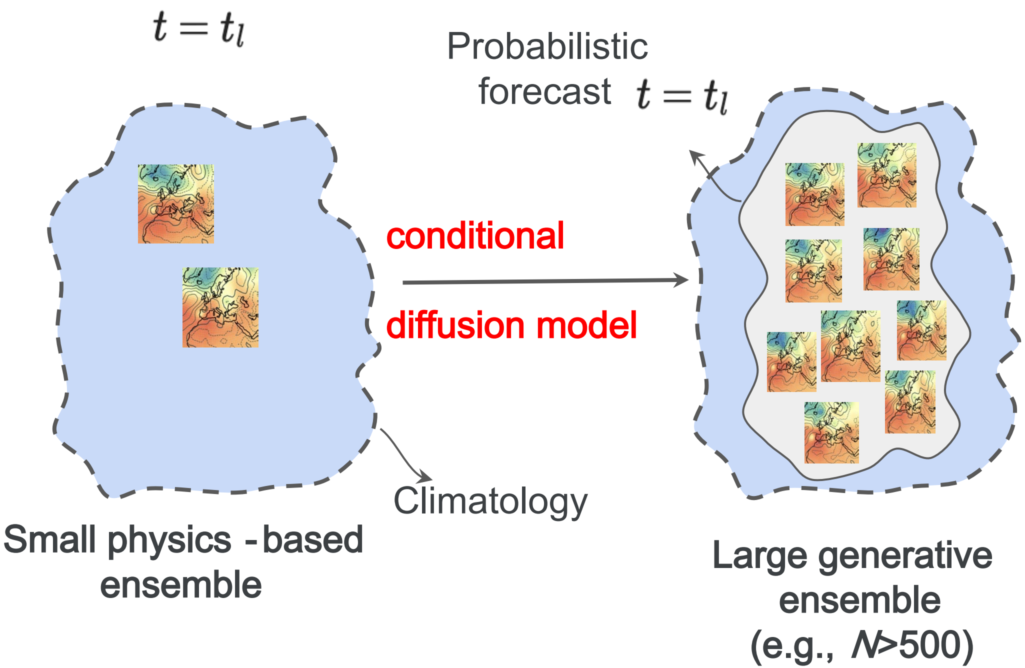


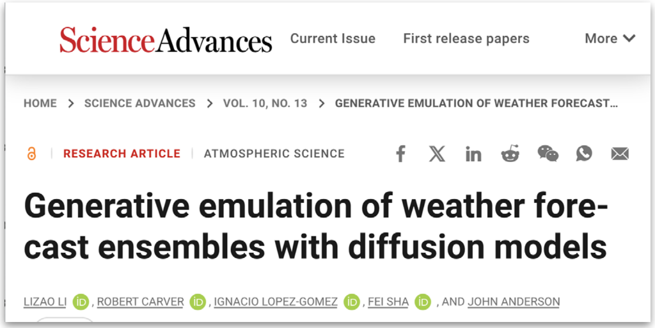


SEEDS: generating conditional distribution

Use diffusion models for density estimate of high resolution data

$$p(x|x_1, x_2)$$





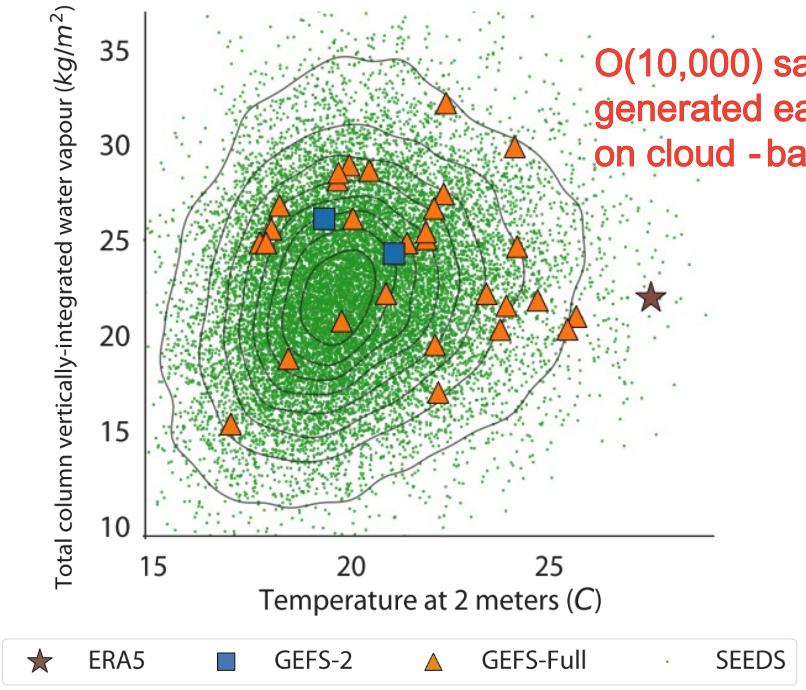
Generative emulation of weather fore-cast ensembles with diffusion models

LIZAO LI, ROBERT CARVER, IGNACIO LOPEZ-GOMEZ, FEI SHA, AND JOHN ANDERSON

[Science Advances 2024]

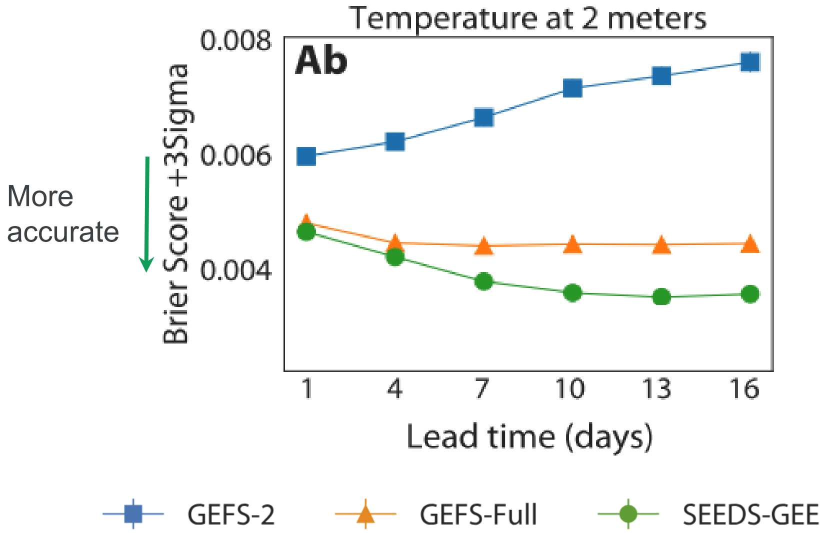
Large ensembles to characterize likelihood of extreme weather

Case study: July 2022 Portugal heatwave (7-day forecast)



O(10,000) samples generated easily in 30min on cloud-based TPUs

Accurate estimation of extreme events

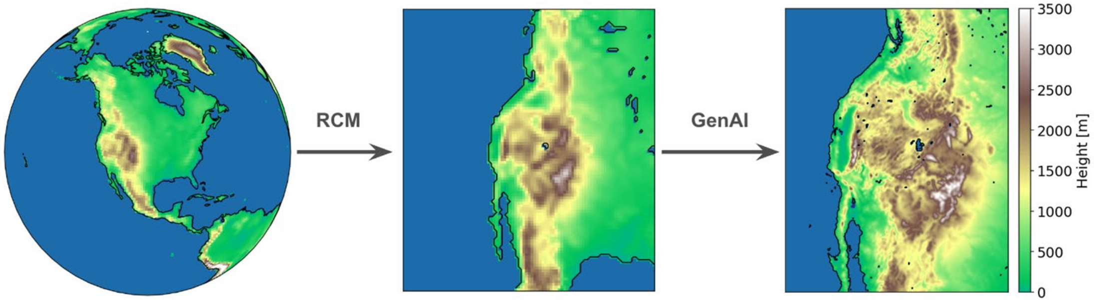


Downscaling future climate projection to meteorological variables

Supervised learning of super - resolution

Traditional approaches create data at **regional** level.

Data production is **limited and costly** .

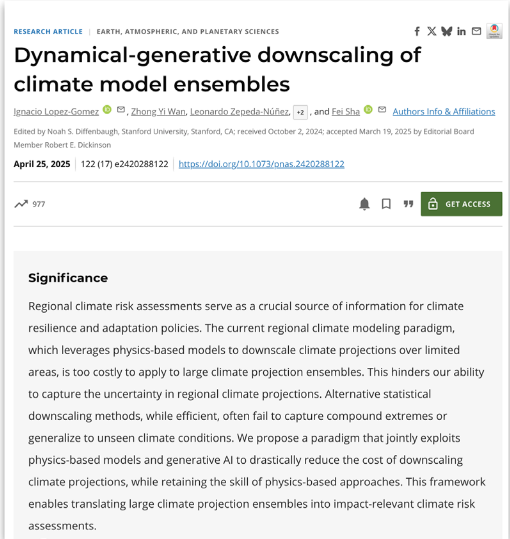


100km resolution

45km resolution

9km resolution

Expensive training data



RESEARCH ARTICLE | EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES

Dynamical-generative downscaling of climate model ensembles

Ignacio Lopez-Gomez, Zhong Yi Wan, Leonardo Zepeda-Núñez, +2 , and Fei Sha Authors Info & Affiliations

Edited by Noah S. Diffenbaugh, Stanford University, Stanford, CA; received October 2, 2024; accepted March 19, 2025 by Editorial Board Member Robert E. Dickinson

April 25, 2025 | 122 (17) e2420288122 | https://doi.org/10.1073/pnas.2420288122

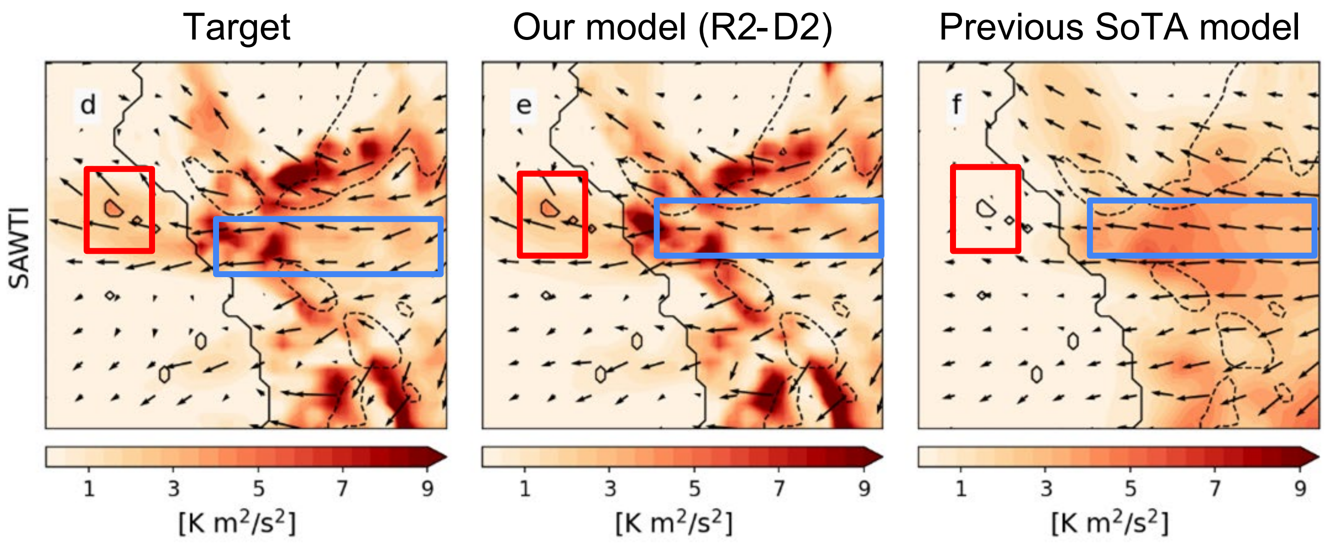
Significance

Regional climate risk assessments serve as a crucial source of information for climate resilience and adaptation policies. The current regional climate modeling paradigm, which leverages physics-based models to downscale climate projections over limited areas, is too costly to apply to large climate projection ensembles. This hinders our ability to capture the uncertainty in regional climate projections. Alternative statistical downscaling methods, while efficient, often fail to capture compound extremes or generalize to unseen climate conditions. We propose a paradigm that jointly exploits physics-based models and generative AI to drastically reduce the cost of downscaling climate projections, while retaining the skill of physics-based approaches. This framework enables translating large climate projection ensembles into impact-relevant climate risk assessments.

[PNAS 2025]

Capture extreme events (Santa Ana wildfire threat index)

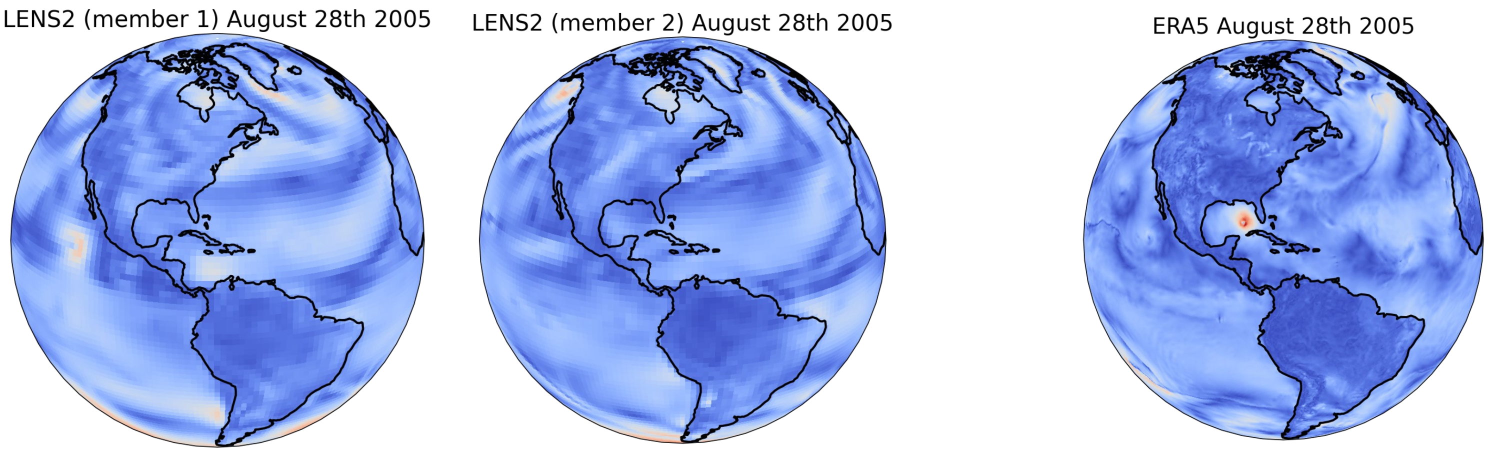
wildfire -producing Santa Ana winds in Southern California, speed and directions accurately predicted, highly similar to physics -based systems



Global downscaling from climate to weather

from the **same period** (decade), but **do not match** at the same **time** point

not a supervised learning problem where exact one to one mapping exists

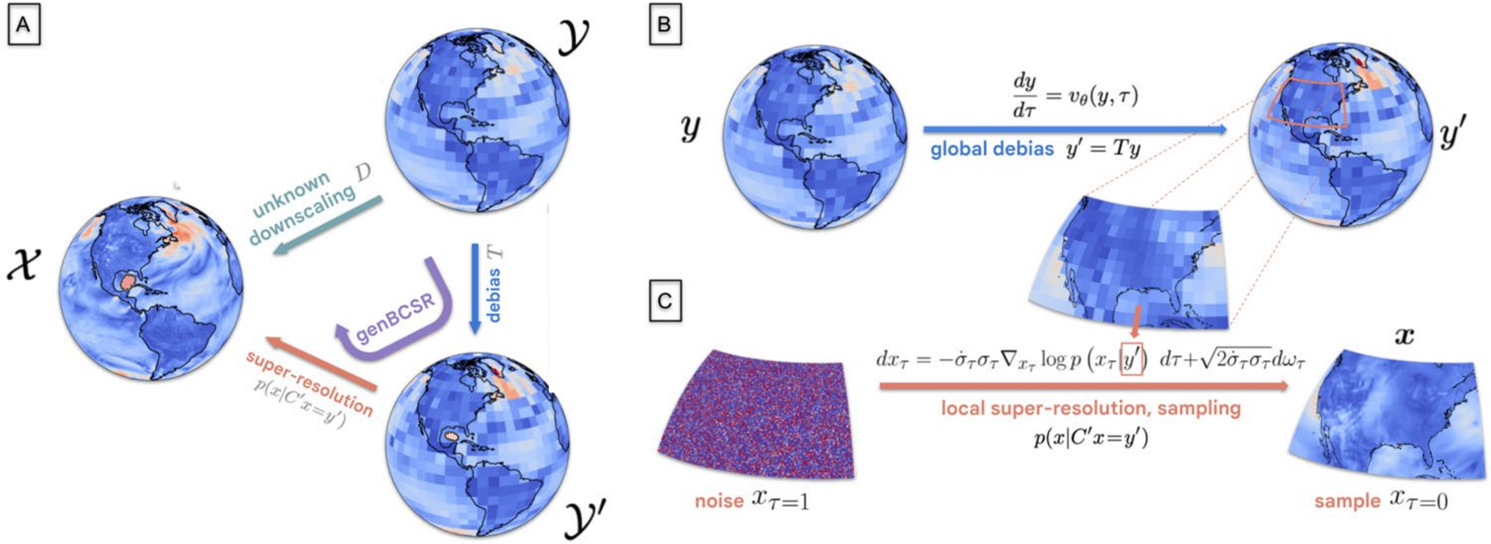


coarse climate simulation

high - resolution weather

Google Research

Methodology adapted to real-world applications

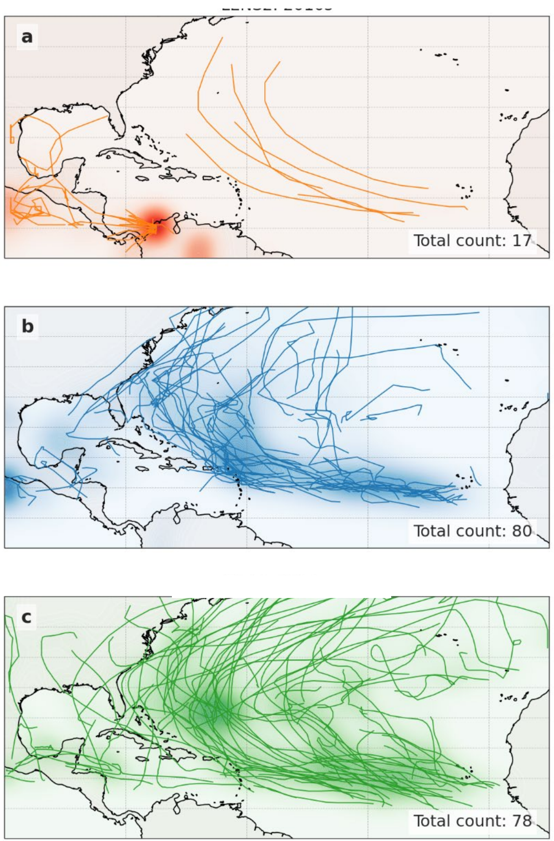


Teaser

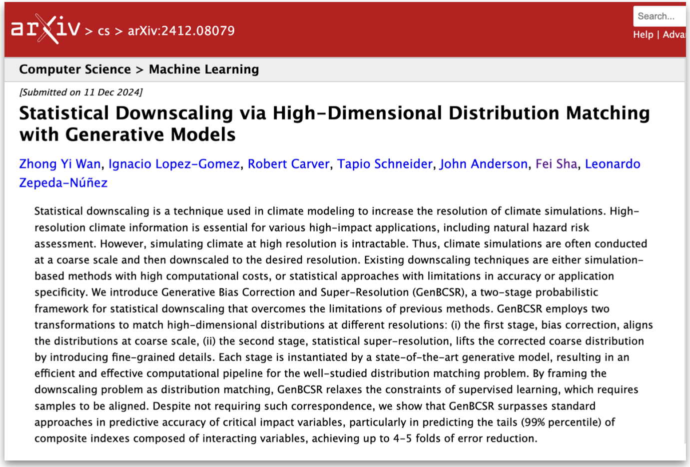
climate simulation

downscaled by our approach

reference



Our approach generates realistic tropical cyclones and accurate statistics.



arXiv > cs > arXiv:2412.08079

Search...

Help | Adva

Computer Science > Machine Learning

[Submitted on 11 Dec 2024]

Statistical Downscaling via High-Dimensional Distribution Matching with Generative Models

Zhong Yi Wan, Ignacio Lopez-Gomez, Robert Carver, Tapio Schneider, John Anderson, Fei Sha, Leonardo Zepeda-Núñez

Statistical downscaling is a technique used in climate modeling to increase the resolution of climate simulations. High-resolution climate information is essential for various high-impact applications, including natural hazard risk assessment. However, simulating climate at high resolution is intractable. Thus, climate simulations are often conducted at a coarse scale and then downscaled to the desired resolution. Existing downscaling techniques are either simulation-based methods with high computational costs, or statistical approaches with limitations in accuracy or application specificity. We introduce Generative Bias Correction and Super-Resolution (GenBCSR), a two-stage probabilistic framework for statistical downscaling that overcomes the limitations of previous methods. GenBCSR employs two transformations to match high-dimensional distributions at different resolutions: (i) the first stage, bias correction, aligns the distributions at coarse scale, (ii) the second stage, statistical super-resolution, lifts the corrected coarse distribution by introducing fine-grained details. Each stage is instantiated by a state-of-the-art generative model, resulting in an efficient and effective computational pipeline for the well-studied distribution matching problem. By framing the downscaling problem as distribution matching, GenBCSR relaxes the constraints of supervised learning, which requires samples to be aligned. Despite not requiring such correspondence, we show that GenBCSR surpasses standard approaches in predictive accuracy of critical impact variables, particularly in predicting the tails (99% percentile) of composite indexes composed of interacting variables, achieving up to 4-5 folds of error reduction.

[Being updated, forthcoming]

Prelude

Vignette 1: Methods

Vignette 2: Application

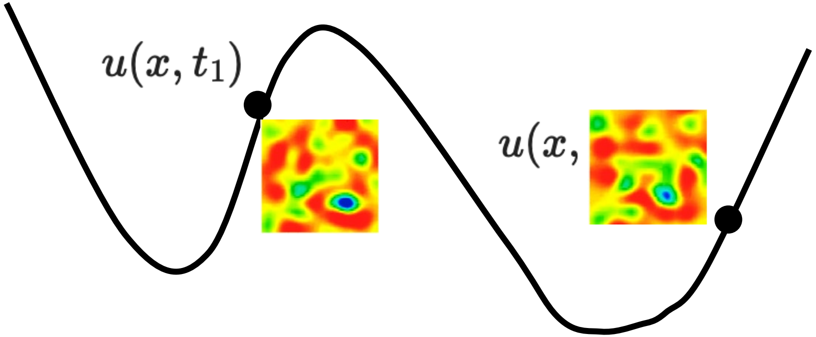
Vignette 3: Theory

Final thoughts

Why these are not supervised learning of predicting “video”

On a high-level, right. But there are **nuances**

observing trajectory τ



learning to predict

$$\min \ \mathbb{E}_{\tau} \sum_{(u_i, u_j) \in \tau} \|\mathcal{S}_\theta(u_i) - u_j\|_2^2$$

Nuances: challenges and lessons from our earlier attempts

Trajectory matching is not necessarily a good or easy to attain learning goal

$$\min \ \mathbb{E}_{\tau} \sum_{(u_i, u_j) \in \tau} \|\mathcal{S}_{\theta}(u_i) - u_j\|_2^2$$

- Trajectories after long -horizon are divergent for chaotic system
- Short-term trajectory matching leads to unstable rollout (“blow-up”, “unphysical”)

arXiv:2301.10391 (cs)

[Submitted on 25 Jan 2023 (v1), last revised 6 Feb 2023 (this version, v3)]

Evolve Smoothly, Fit Consistently: Learning Smooth Latent Dynamics For Advection-Dominated Systems

arXiv:2402.04467 (cs)

[Submitted on 6 Feb 2024 (v1), last revised 5 Jun 2024 (this version, v2)]

DySLIM: Dynamics Stable Learning by Invariant Measure for Chaotic Systems

Yair Schiff, Zhong Yi Wan, Jeffrey B. Parker, Stephan Hoyer, Volodymyr Kuleshov, Fei Sha, Leonardo Zepeda-Núñez

Lessons

- Sometimes, statistical characterization is more useful
- Discover and exploit unknown latent structures

[ICLR 2023, ICML 2024]

Vignette #3: Theory

why does generative modeling work so well?

- Comprehensive empirical evaluation of data-driven probabilistic modeling of 3D turbulent flows
- Rigorous analysis of the mechanism of generative modeling

arXiv > cs > arXiv:2409.18359

Search...

Help | Ad

Computer Science > Machine Learning

[Submitted on 27 Sep 2024]

Generative AI for fast and accurate Statistical Computation of Fluids

Roberto Molinaro, Samuel Lanthaler, Bogdan Raonić, Tobias Rohner, Victor Armegioiu, Zhong Yi Wan, Fei Sha, Siddhartha Mishra, Leonardo Zepeda-Núñez

We present a generative AI algorithm for addressing the challenging task of fast, accurate and robust statistical computation of three-dimensional turbulent fluid flows. Our algorithm, termed as GenCFD, is based on a conditional score-based diffusion model. Through extensive numerical experimentation with both incompressible and compressible fluid flows, we demonstrate that GenCFD provides very accurate approximation of statistical quantities of interest such as mean, variance, point pdfs, higher-order moments, while also generating high quality realistic samples of turbulent fluid flows and ensuring excellent spectral resolution. In contrast, ensembles of operator learning baselines which are trained to minimize mean (absolute) square errors regress to the mean flow. We present rigorous theoretical results uncovering the surprising mechanisms through which diffusion models accurately generate fluid flows. These mechanisms are illustrated with solvable toy models that exhibit the relevant features of turbulent fluid flows while being amenable to explicit analytical formulas.

Google Research

Google Research

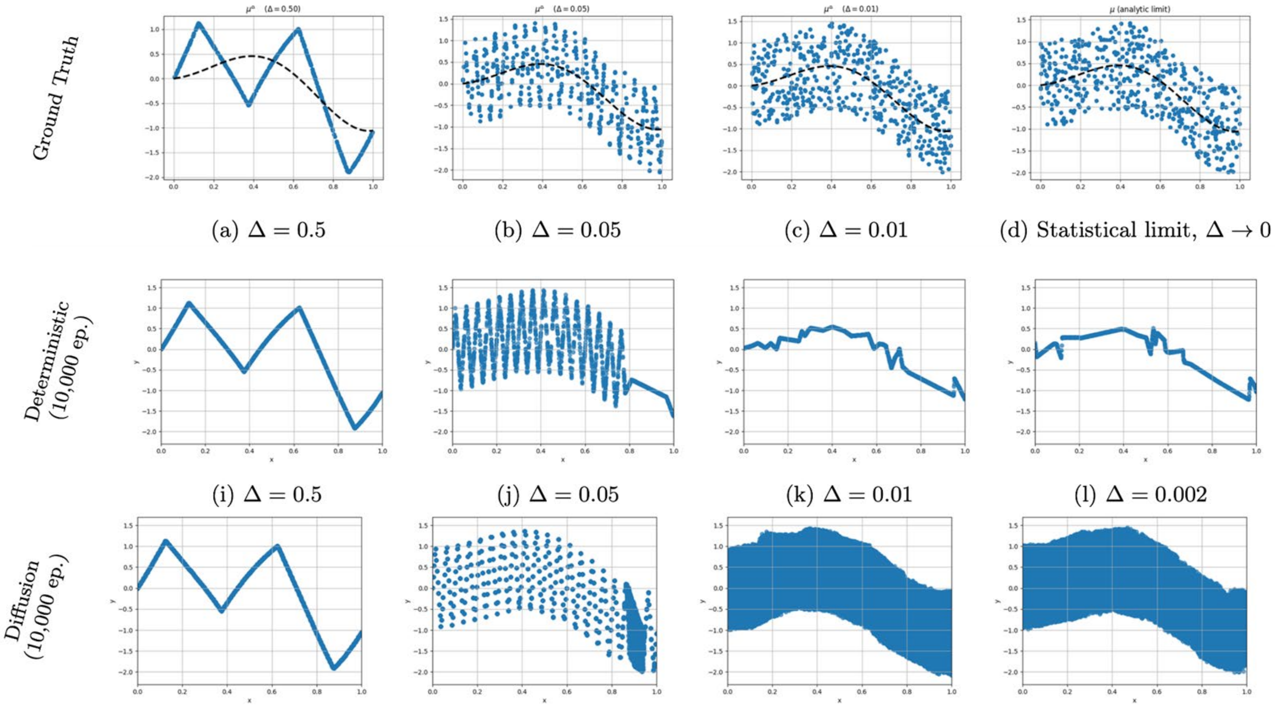
Analytically tractable toy example

$$S^{\Delta}(\bar{u}) = m(\bar{u}) + \Lambda(N\bar{\mu})$$

Λ: bounded hat function
m: mean function
N: # of grids

fitting to trajectory, collapse to mean (expected)

fitting to distribution, recover the underlying measure (expected)



(a) $\Delta = 0.5$ (b) $\Delta = 0.05$ (c) $\Delta = 0.01$ (d) Statistical limit, $\Delta \to 0$

(i) $\Delta = 0.5$ (j) $\Delta = 0.05$ (k) $\Delta = 0.01$ (l) $\Delta = 0.002$

(q) $\Delta = 0.5$ (r) $\Delta = 0.05$ (s) $\Delta = 0.01$ (t) $\Delta = 0.002$

Formal statement

Theorem C.5 (Constrained probabilistic approximation is tractable). *Assume that the optimal conditional denoiser* $D_{\text{opt}}(u; \bar{u}, \sigma)$ *for the statistical limit* μ*, with density* $p(u, \bar{u}) = p(u \mid \bar{u}) p_{\text{prior}}(\bar{u})$*, is* L^**-Lipschitz continuous. Assume that* μ *and* μ^Δ *are supported on* $B_M = \{\|u\| \leq M\}$*. Then, the optimal constrained denoiser* D_θ^Δ *trained on the numerical distribution* μ^Δ*, corresponding to* $p^\Delta(u, \bar{u}) = \delta(u - \mathcal{S}^\Delta(\bar{u}))\, p_{\text{prior}}(\bar{u})$*,*

$$D_\theta^\Delta(u; \bar{u}, \sigma) = \underset{\text{Lip}(D_\theta) \leq L^*}{\text{argmin}} \; \mathcal{J}^\Delta(D_\theta, \sigma),$$

satisfies

$$\mathcal{J}(D_\theta^\Delta, \sigma) \leq \mathcal{J}(D_{\text{opt}}, \sigma) + C L^* W_1(\mu^\Delta, \mu), \quad \forall \sigma > 0, \tag{61}$$

with constant C *independent of* Δ*,* L^* *and* σ*.*

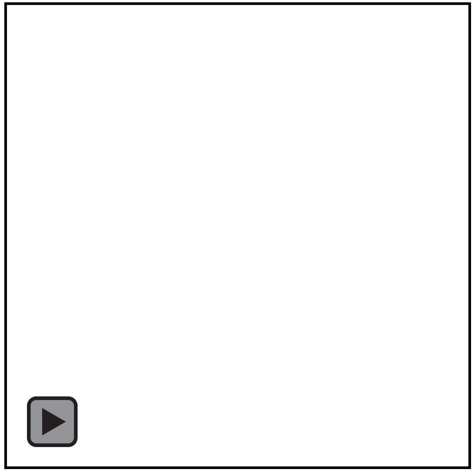
In English,

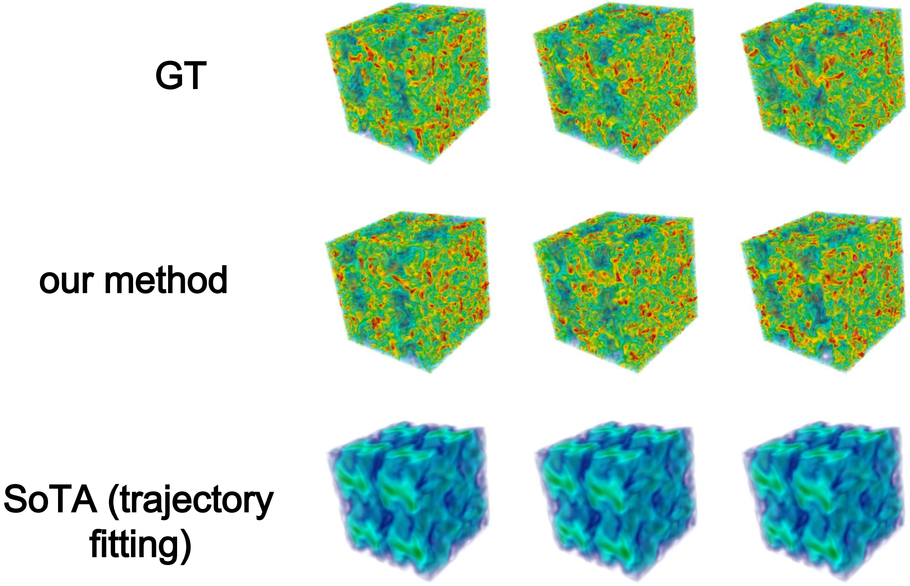
if the underlying probabilistic measure under the continuous chaotic system exists, the denoiser of the optimally trained diffusion model from data converges as the discretization increases, and recovers the underlying measure.

Example: Taylor - Green Vortex

Our method converges to the right statistical characterization

SoTA method collapses to the mean





Prelude

Vignette 1: Methods

Vignette 2: Application

Vignette 3: Theory

Final thoughts

Google Research

Looking ahead

Exciting time for inventing new ways of computing and doing science

AI/ML has shown impacts on advancing scientific computing at **unprecedented** pace

Still **nascent**, we have a lot **unknown operating conditions** of different paradigms

Opportunities to advance foundational AI/ML with **unfamiliar and new challenges**

- How to physics-prior into statistical modeling choices
- How to generate statistical outputs that are physically plausible
- How to reduce sample complexity when acquiring high-resolution simulation is the primary data source
- …

Acknowledgement

Multi-disciplinary team of

meteorologists, applied mathematicians, climate scientists, computational physicists and computer scientists

Thank you!