Geometry-Aware Post-Hoc Uncertainty Quantification in Operator Learning

Oriol Vendrell-Gallart
PhD Student
University of California, Irvine
ovendrel@uci.edu

Nima Negarandeh
PhD Student
University of California, Irvine
nnegaran@uci.edu

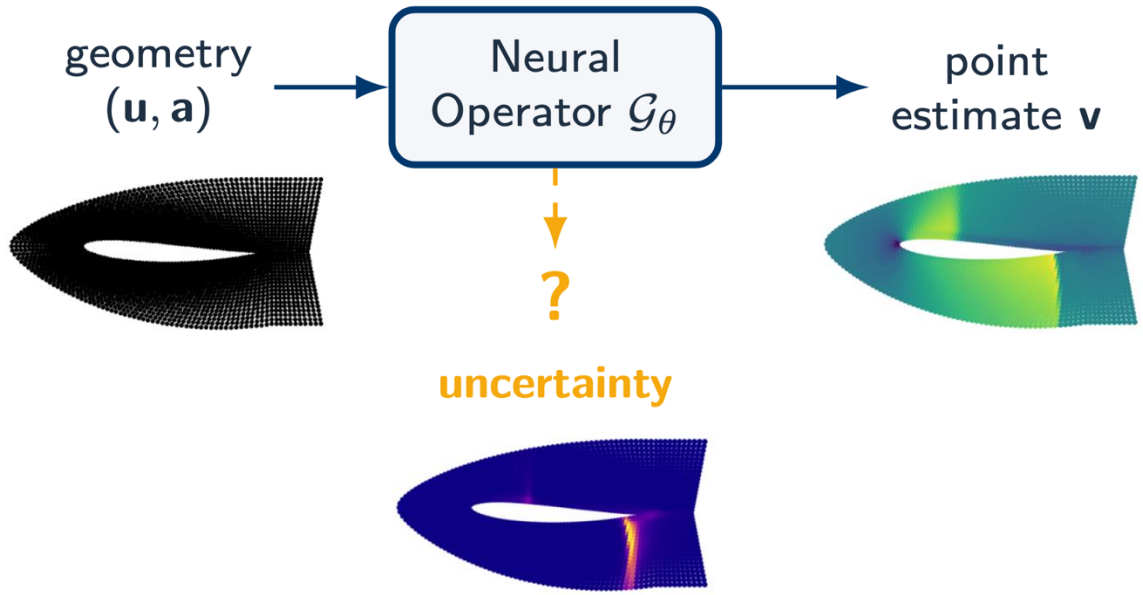
Ramin Bostanabad
Associate Professor
University of California, Irvine
raminb@uci.edu

Sydney, December 2026

NEURAL OPERATORS AS PDE SURROGATES

- Neural operators replace expensive solvers across climate, fluids, plasma and solid mechanics.
- Modern **geometry-aware** backbones (e.g. Transolver) run natively on **point clouds**.
- But they are trained **deterministically**:

  one input → one output



Why this matters?

Scientific use and downstream tasks (active data collection, design, safety) need to know ***when to trust*** the surrogate, especially under ***geometric variability***.

OPERATOR LEARNING WITH GAUSSIAN PROCESSES

Target. Solution operator over varying domains:

$$\mathcal{G}^{\dagger}: \mathcal{U} \times \mathcal{A} \to \mathcal{V}, \quad (\mathbf{u}, \boldsymbol{a}) \to \boldsymbol{v}, \qquad \mathcal{D} = \{(\mathbf{u}_{\mathrm{i}}, \boldsymbol{a}_{\mathrm{i}}, \boldsymbol{v}_{\mathrm{i}})\}_{i=1}^{M}$$

Functional-regression view. Instead of learning an infinite-dimensional output, learn its **pointwise evaluation** observed at finite coordinates.

$$\tilde{\mathcal{G}}^{\dagger}: \mathbb{R}^{\mathrm{d_u}} \times \mathbb{R}^{\mathrm{d_a}} \times \Omega \to \mathbb{R}, \quad (\mathbf{u}, \boldsymbol{a}, \boldsymbol{x}) \to \mathrm{v}(\boldsymbol{x}), \quad \mathcal{D} = \{(\boldsymbol{z}_{\mathrm{i}}, \mathrm{v}_{\mathrm{i}})\}_{i=1}^{MN}, \quad \boldsymbol{z}_{\mathrm{i}} = (\mathbf{u}_{\mathrm{s(i)}}, \boldsymbol{a}_{\mathrm{s(i)}}, \boldsymbol{x}_{\mathrm{i}})$$

Consequence

Operator learning ⇒ **regression** on the augmented input $\boldsymbol{z} = (\boldsymbol{u}, \boldsymbol{a}, \boldsymbol{x})$.
A **GP over $f(z)$** gives pointwise prediction and closed-form uncertainty.

$$\mathrm{f}(\boldsymbol{z}) \sim \mathcal{GP}(\mathrm{m}(\boldsymbol{z}; \boldsymbol{\beta}), \mathrm{k}(\boldsymbol{z}, \boldsymbol{z}'; \phi)).$$

- **Scalability**. M point clouds × N nodes, $MN \sim 10^6$ points, GPs are $\mathcal{O}\left((MN)^3\right)$
- **Kernel for irregular geometries**. How to measure point cloud distances?

Mora, C., Yousefpour, A., Hosseinmardi, S., Owhadi, H., & Bostanabad, R. (2025). Operator learning with Gaussian processes. Computer Methods in Applied Mechanics and Engineering, 434, 117581.

THE SCALABLE UQ LANDSCAPE IS IN WEIGHT SPACE

Train-time UQ (modify training):

- Deep ensembles
- MC Dropout, PNO (2025)

- Rich epistemic exploration
- Prohibitively expensive for large scale meshes (needs retraining models)

Post-hoc UQ (freeze the model):

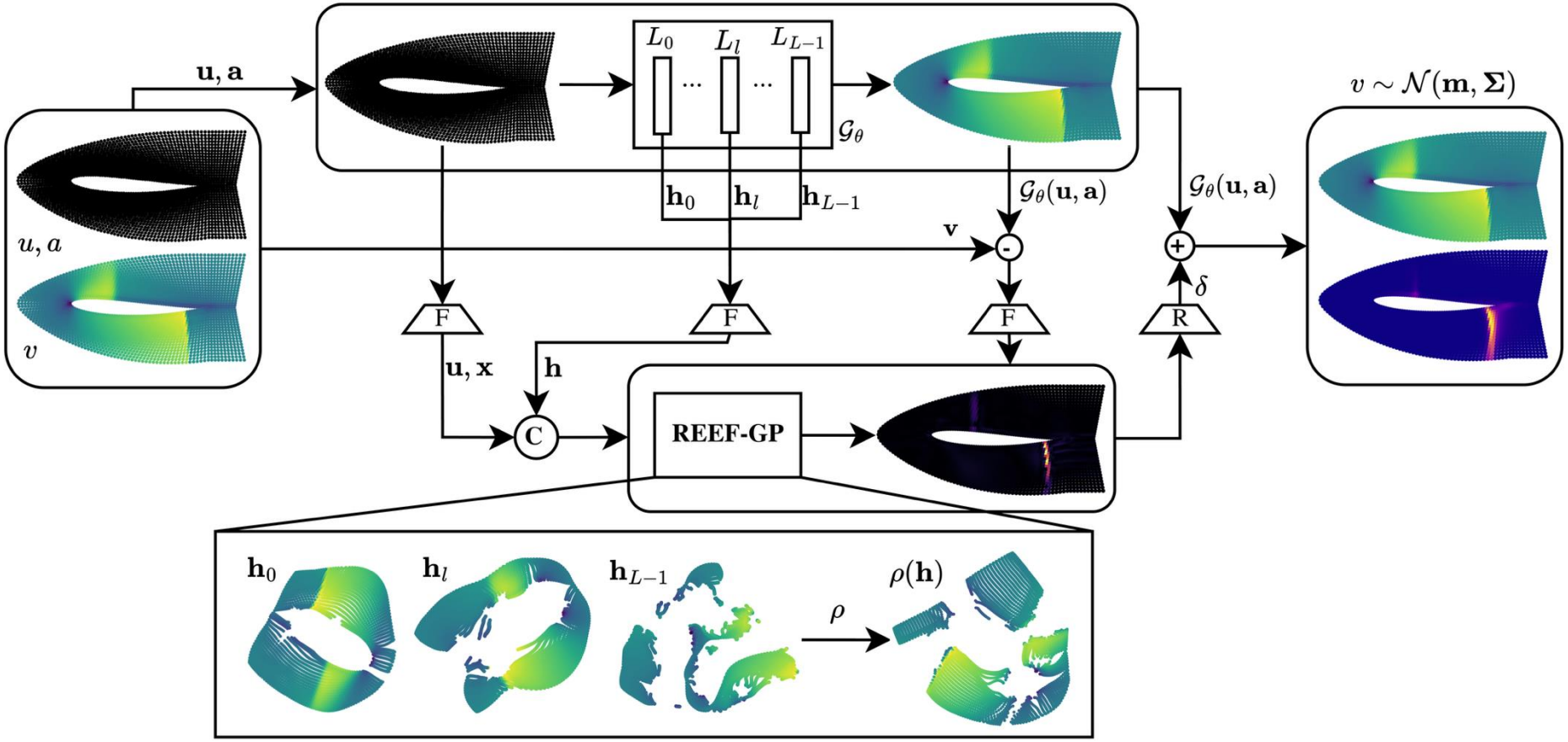
- Laplace approximation, LUNO (2025).
- Deep Vecchia Ensembles (2025)

- No need to modify training (cheaper calibration)

The common limitation

Almost all methods model uncertainty in **network parameters.**
The learnt **geometry-aware representations** are left unused.

REEF-GP AT A GLANCE



Parameter-centric
uncertainty on θ

shift

REEF-GP
GP on learned features **h**

Frozen $\mathcal{G}_\theta$ maps $(\mathbf{u}, \mathbf{a}) \rightarrow \mathbf{v}$. REEF-GP models its **residual**: **transforms** (ρ) the operator's geometry-aware embeddings $(\mathbf{h_0}, \mathbf{h_l}, \mathbf{h_{L-1}})$ into a kernel space, and adds the discrepancy $\boldsymbol{\delta}$ to yield an output distribution. Note how the kernel space **tears exactly at the shock fronts**.

MODELING THE DISCREPANCY

Treat **frozen** $\mathcal{G}_\theta$ as a **computer model**, the gap to truth is **model inadequacy** δ:

$$v(\mathbf{x}) = \mathcal{G}_\theta(\mathbf{u}, \mathbf{a})(\mathbf{x}) + \delta\big(\mathbf{u}, \mathbf{x}, \mathbf{h}(\mathbf{u}, \mathbf{a}, \mathbf{x})\big) + \epsilon, \qquad \bar{\mathbf{z}} = (\mathbf{u}, \mathbf{x}, \mathbf{h}(\mathbf{u}, \mathbf{a}, \mathbf{x}))$$

Unlike classic KOH, we **open the black box**: errors accumulate through the representational limits of the layers composition $\mathcal{L}_\mathcal{L} \circ \cdots \circ \mathcal{L}_1$, so we condition δ on internal features $\mathbf{h}$.

Probabilistic split.

$$\delta = f(\bar{\mathbf{z}}) + \varepsilon(\bar{\mathbf{z}})$$

$$\mathrm{f}(\bar{\mathbf{z}}) \sim \mathcal{GP}(0, \mathrm{k}_\phi(\bar{\mathbf{z}}, \bar{\mathbf{z}}'))$$

$$\varepsilon + \epsilon \sim \mathcal{N}(0, {\sigma_\psi}^2(\bar{\mathbf{z}}) + \sigma_n^2)$$

ε : **heteroscedastic noise**

${\sigma_\psi}^2$: **spectral-normalized MLP**

Internal features.

$$\mathbf{h}(\mathbf{u}, \mathbf{a}, \mathbf{x}) = \oplus_{\mathrm{l}=1}^{\mathrm{L}'} \boldsymbol{h}_\mathrm{l}(\mathbf{u}, \mathbf{a}, \mathbf{x}), \qquad \mathrm{L}' < \mathrm{L}$$

Marginal prior at query point.

$$\mathrm{v}(\boldsymbol{x}) \sim \mathcal{N}(\mathcal{G}_\theta(\mathbf{u}, \mathbf{a})(\mathbf{x}), \mathrm{k}_\phi(\bar{\mathbf{z}}, \bar{\mathbf{z}}') + {\sigma_\psi}^2(\bar{\mathbf{z}}) + \sigma_n^2)$$

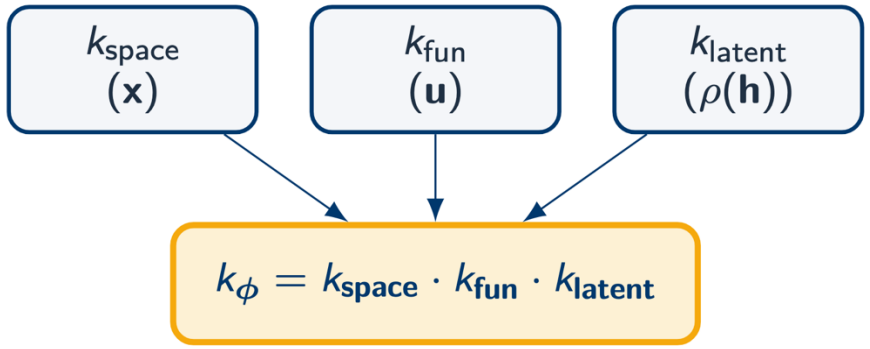
Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 63(3):425–464, 01 2002.

A GEOMETRY-AWARE DEEP KERNEL

Respect the distinct roles of **space**, **input function** and **latent state** via a product kernel:

$$\mathrm{k}_{\phi}(\bar{\boldsymbol{z}}, \bar{\boldsymbol{z}}') \;=\; \sigma^2\, \mathrm{k}_{\mathrm{space}}(\boldsymbol{x}, \boldsymbol{x}') \cdot \mathrm{k}_{\mathrm{fun}}(\boldsymbol{u}, \boldsymbol{u}') \cdot \mathrm{k}_{\mathrm{latent}}(\rho(\boldsymbol{h}), \rho(\boldsymbol{h}'))$$

- ρ: spectral-normalized MLP warping the operator embeddings into a compact latent space.
- k_{ϕ} is a **non-stationary** prior: spatial correlations are modulated by latent geometry-aware correlations.



Why not just a fixed prior?

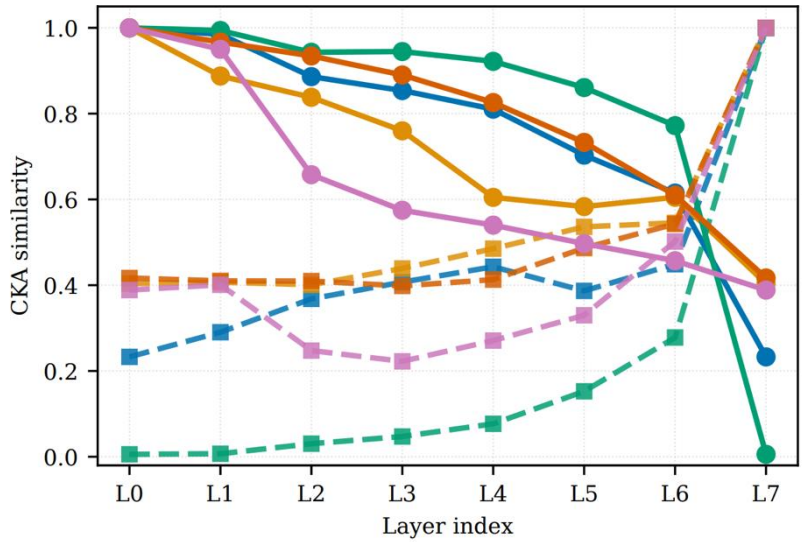
Bayesian inference in **infinite dimensions is brittle**: a slightly misspecified prior can converge **confidently to the wrong answer**.

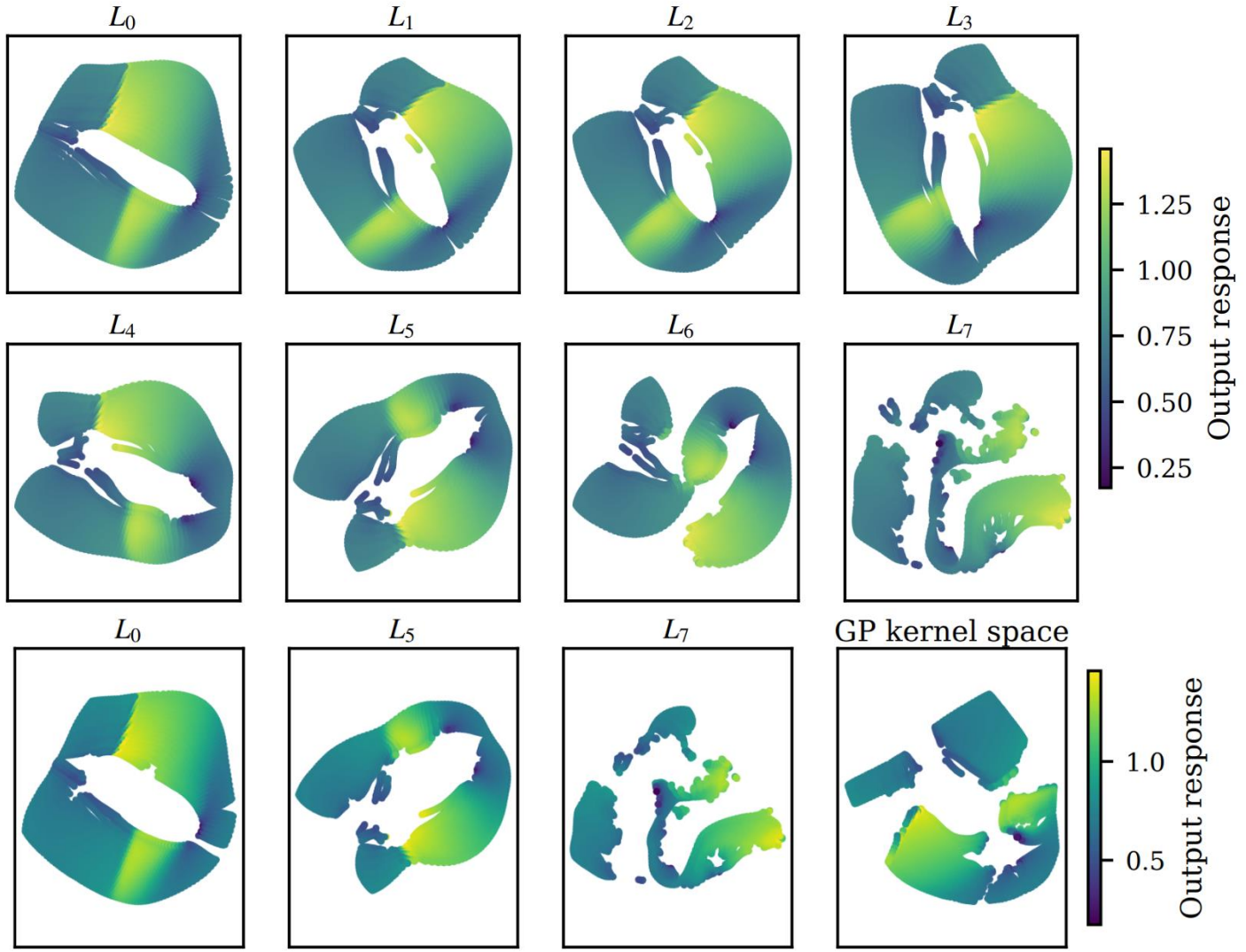
We reuse the surrogate learned representation instead of learning a manifold from scratch.

Houman Owhadi, Clint Scovel, and Tim Sullivan. On the brittleness of bayesian inference. SIAM Review, 57(4):566–582, 2015.

WHY INTERNAL LAYERS?

- Transolver's blocks act like a **discrete flow** (connection to Kernel Flows).
- Intermediate layers carry **richer** information than the final readout. We pick **3 CKA-distinct layers**.





Houman Owhadi and Gene Ryan Yoo. Kernel flows: From learning kernels from data into the abyss. Journal of Computational Physics, 389:22–47, 2019.

Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries, 2024.

MAKING IT SCALE

Training (KF-inspired).

At each step, draw a random subset of $N_s \ll MN$ points and minimize the mini-batch NLL:

$$\mathcal{L} = \frac{1}{2}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{K}^{-1}\boldsymbol{\delta} + \frac{1}{2}\log|\boldsymbol{K}| + \frac{N_s}{2}\log 2\pi$$

Decouples memory from mesh resolution N, no restrictive low-rank/inducing point assumptions.

Inference (gPoE).

K experts on random subsets, aggregated in precision space:

$$\frac{1}{\sigma_*^2} = \frac{1}{K}\sum_{k}\frac{1}{\sigma_k^2}, \qquad \bar{\delta}_* = \sigma_*^2\frac{1}{K}\sum_{k}\frac{\bar{\delta}_k}{\sigma_k^2}$$

Test-time memory bounded to $\mathcal{O}(N_S^2)$ per expert, independent of MN.

Takeaway

Stochastic subsets + gPoE give enable scaling GPs at **millions of points**.
Empirically stable training with spectral normalization and heteroscedastic noise.

EXPERIMENTAL SETUP

Base neural operator.

- **Transolver** (frozen)

Metrics.

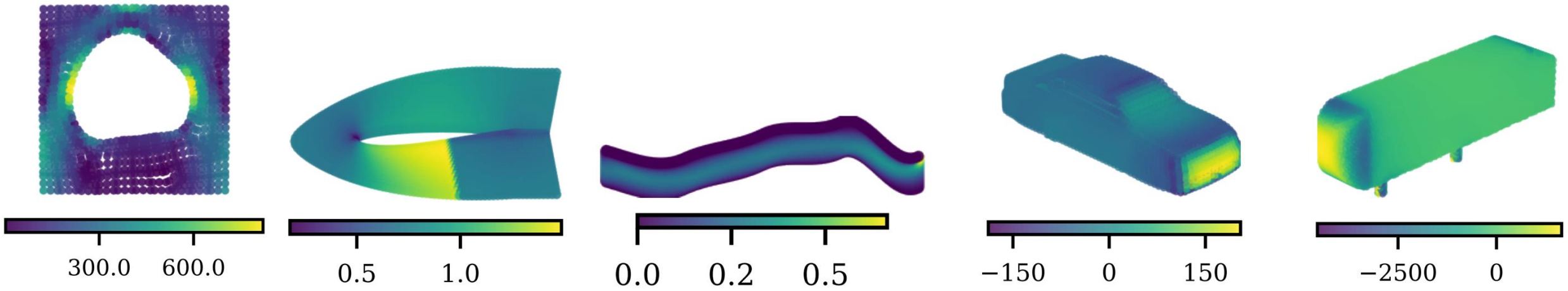
- rL2, NLL, CRPS, NIS, ES

5 benchmarks.

- **2D**: Elasticity, Airfoil, Pipe
- **3D**: ShapeNet Car, Ahmed Body

6 baselines.

- Deep ensemble, MC dropout, PNO
- Input perturbation, LUNO-LA (Laplace), DVE-spatial.

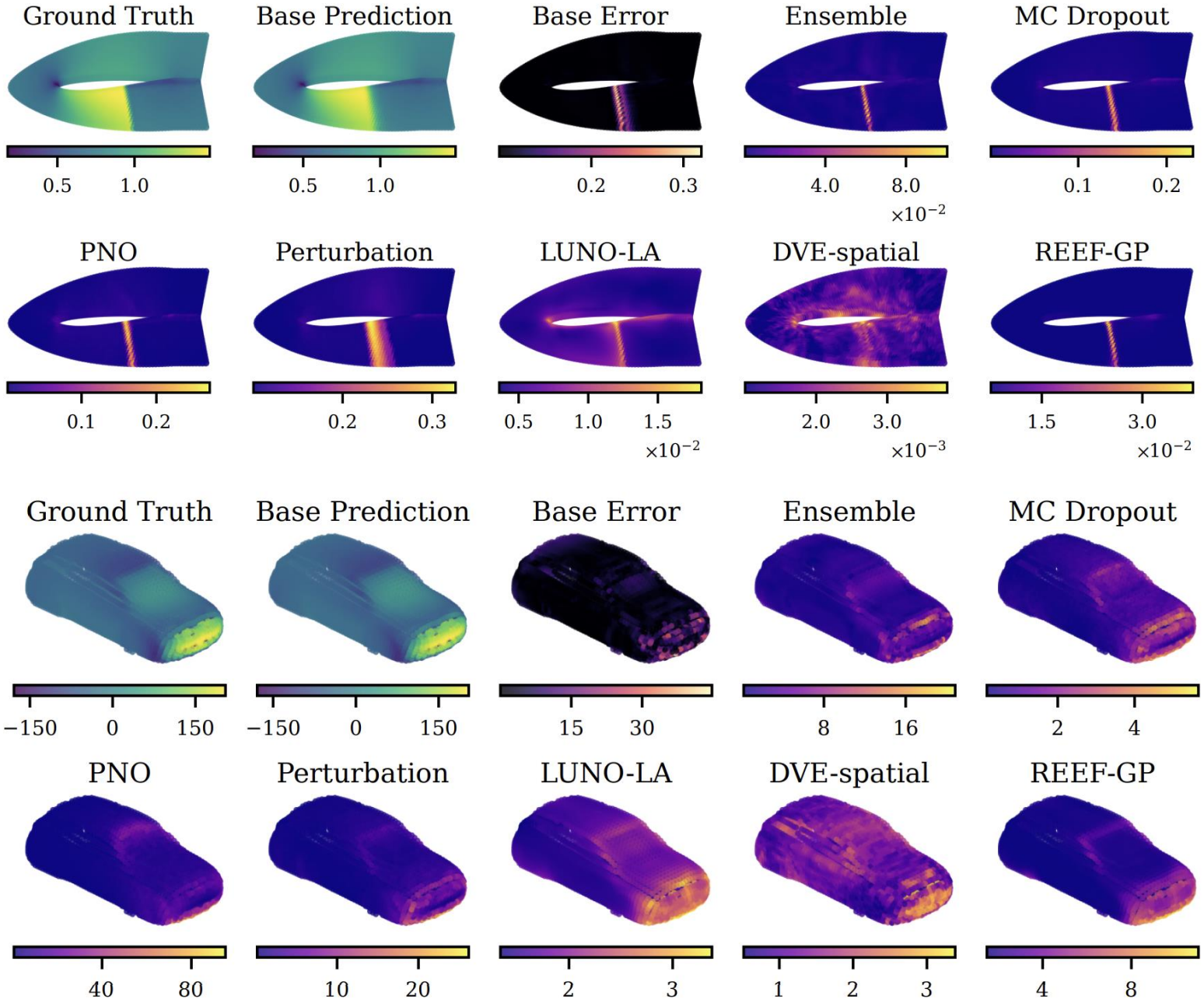


RESULTS

Method	rL2 ↓	NLL ↓	CRPS ↓	NIS ↓	ES ↓
Airfoil (2D)	(%)		($\times 10^{-4}$)	($\times 10^{-2}$)	
Base (no UQ)	1.24 ±0.12	–	–	–	–
Ensemble	**1.06**	−3.46	**28.63**	**3.89**	**0.35**
MC Dropout	1.58 ±0.21	−3.38 ±0.15	44.88 ±5.66	5.55 ±0.91	0.55 ±0.08
PNO	2.32 ±1.40	−3.28 ±0.65	74.27 ±56.12	8.26 ±6.39	0.79 ±0.47
Perturbation	3.33 ±0.19	−3.32 ±0.05	81.05 ±5.67	9.71 ±0.66	1.08 ±0.06
LUNO-LA	1.24 ±0.12	−3.07 ±0.82	38.70 ±2.50	6.19 ±0.54	0.47 ±0.04
DVE-spatial	2.09 ±0.25	19.98 ±7.32	88.41 ±12.53	25.89 ±3.09	0.92 ±0.10
REEF-GP	1.24 ±0.12	**−3.51** ±0.06	39.80 ±2.35	5.69 ±0.47	0.45 ±0.05
ShapeNet Car (3D)	(%)				
Base (no UQ)	8.62 ±0.13	–	–	–	–
Ensemble	**7.73**	4.16	**1.88**	30.28	**194.57**
MC Dropout	9.04 ±0.59	17.21 ±4.21	2.52 ±0.26	67.96 ±9.04	267.68 ±17.66
PNO	13.76 ±1.74	4.98 ±1.52	3.53 ±0.50	51.10 ±12.17	337.16 ±43.04
Perturbation	8.62 ±0.13	13.44 ±1.14	2.28 ±0.03	50.82 ±1.10	224.73 ±3.97
LUNO-LA	8.62 ±0.13	33.55 ±23.43	2.47 ±0.13	73.47 ±14.91	271.32 ±12.10
DVE-spatial	13.08 ±0.26	11.22 ±3.03	3.76 ±0.10	96.17 ±10.43	396.70 ±6.79
REEF-GP	8.52 ±0.14	**2.84** ±0.09	2.05 ±0.03	**28.81** ±1.65	224.31 ±4.18

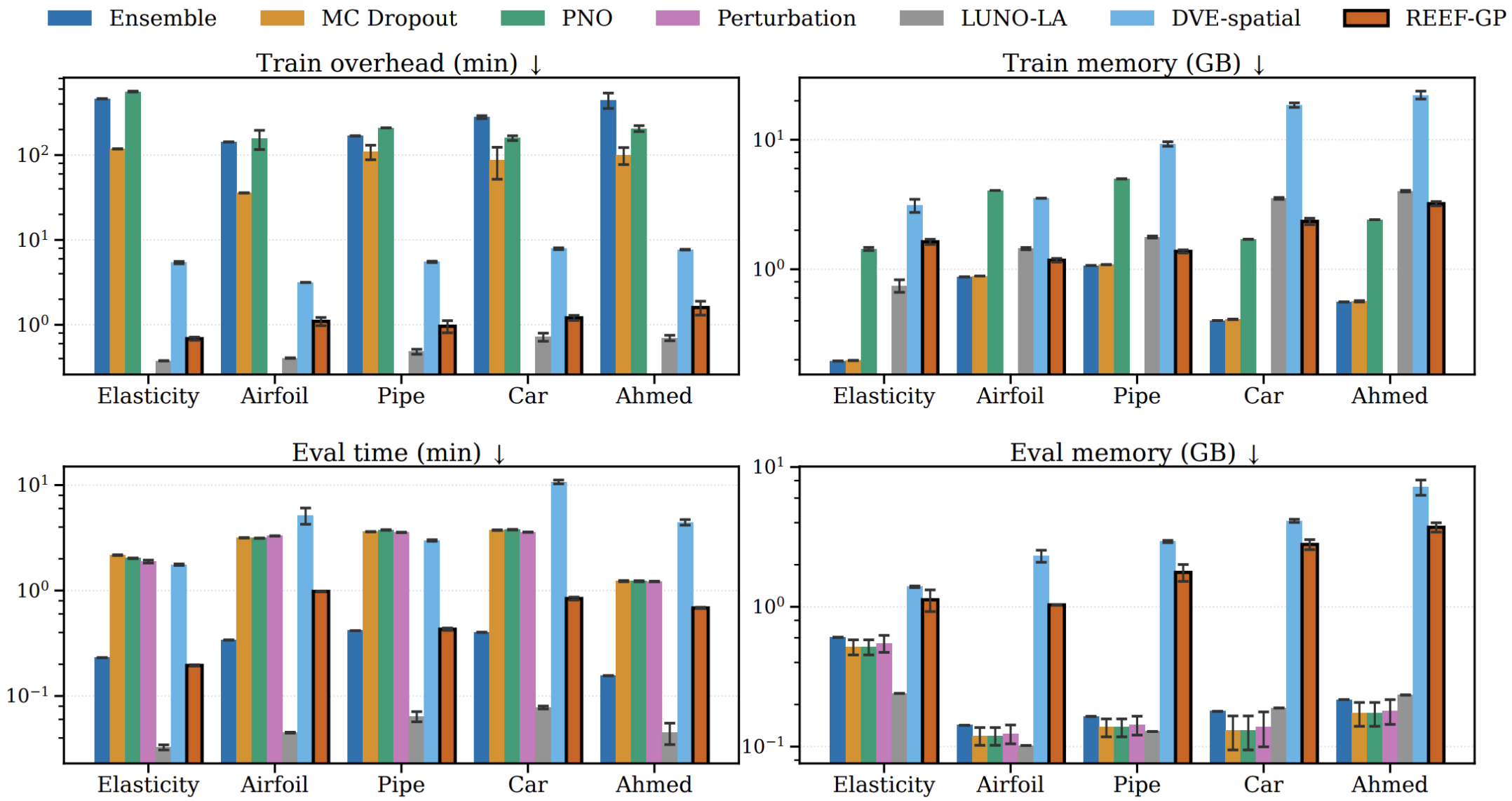
- No accuracy degradation (rL2).
- Competitive to Deep ensembles on UQ metrics.

UNCERTAINTY IS SPATIALLY MEANINGFUL



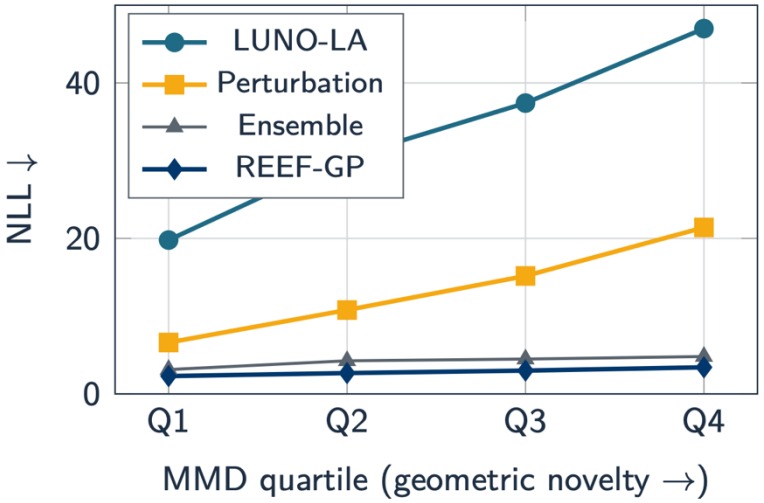
- Predicted std tracks where the **base error** is larger.
- Higher uncertainty concentrates at **shock fronts** (Airfoil) and the **front bumper** (Car).
- The distribution is **physically interpretable**.

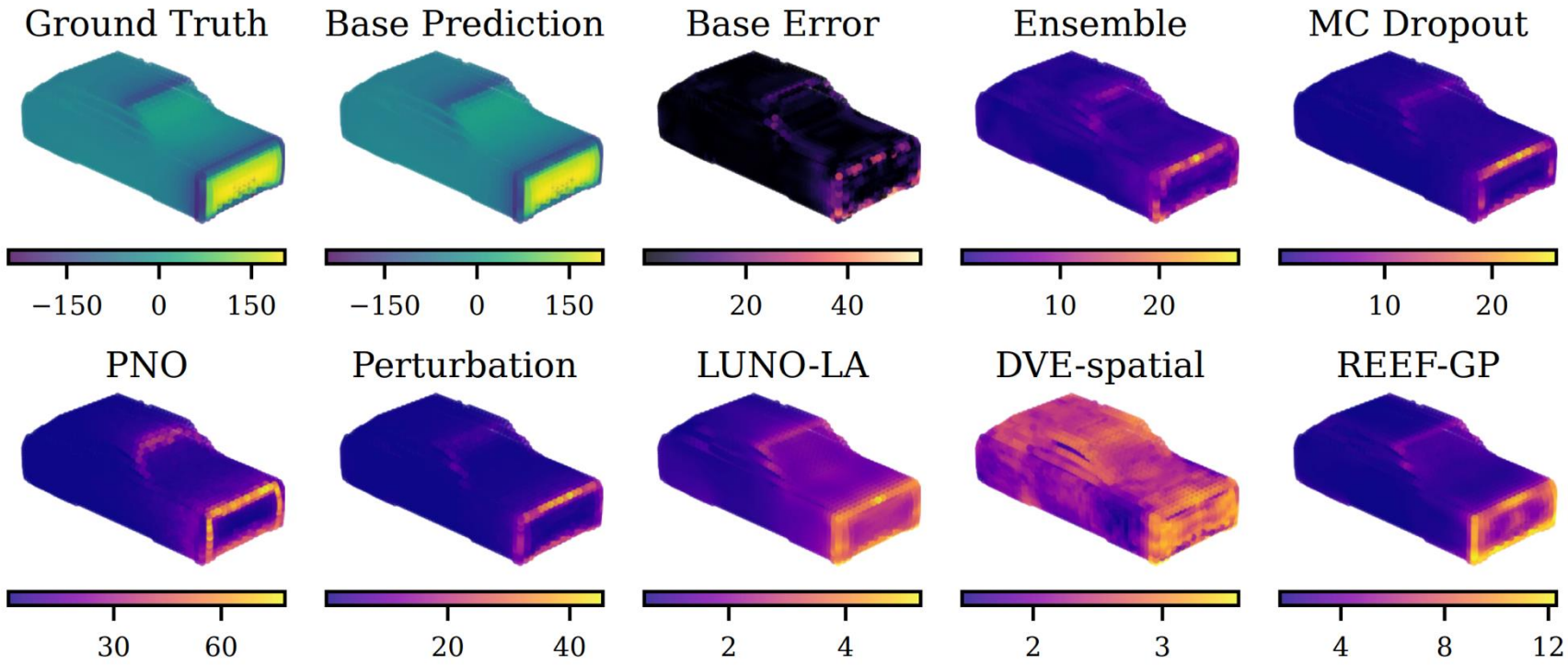
COMPUTATIONAL COST



ROBUST UNDER GEOMETRIC DISTRIBUTION SHIFT

- Split each test set into **MMD quartiles**: Q1 (closest) to Q4 (more novel).
- For 3D cases, REEF-GP's NLL is **stable** in all quartiles.





CONCLUSIONS

Takeaways

- The **coordinate-feature** space of a frozen neural operator is enough for calibrated post-hoc UQ.
- REEF-GP **matches Deep Ensembles** across 5 benchmarks at a fraction of the cost, with **no accuracy loss**.
- Uncertainty is **spatially interpretable** (shock fronts) and stable under **geometric shift**.

Paper

Code

ACKNOWLEDGEMENT



NSF



ONR

Thank you!

California·Catalonia Program
BALSELLS FELLOWSHIP

Balsells Fellowship

NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS