Institute for Mathematical and Statistical Innovation

Optimal Control and Decision Making Under Uncertainty for Digital twins

Sparse Sensor Optimization Under Uncertainty

Krithika Manohar
Mechanical Engineering

AI INSTITUTE IN DYNAMIC SYSTEMS

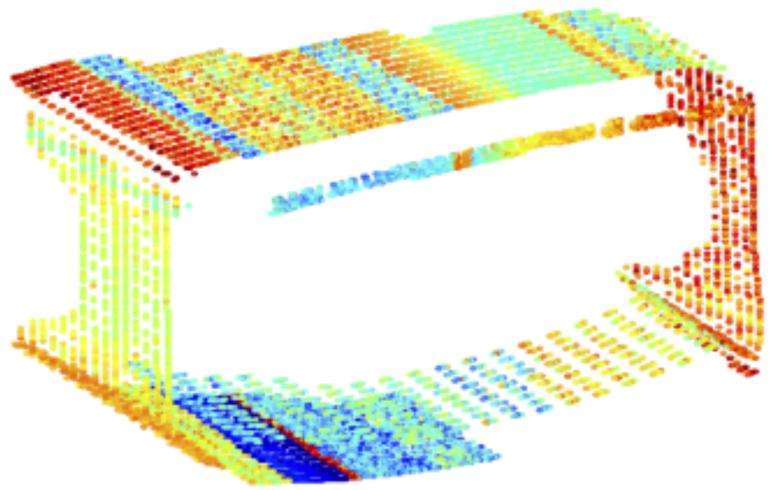
UNIVERSITY of WASHINGTON

Data-Driven Sensing

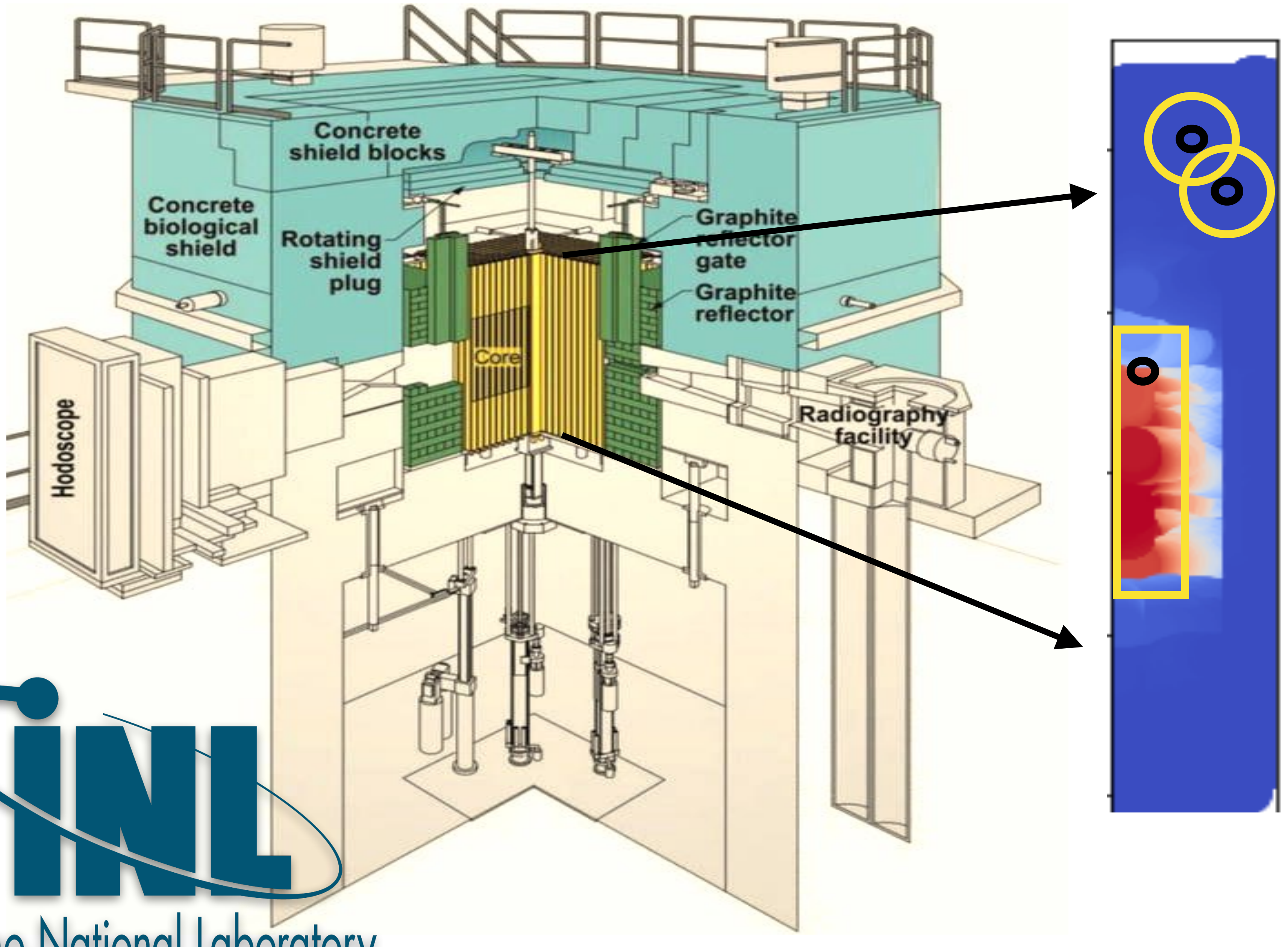
- Data-driven **sensor placement optimization**
  - Choice of measurements crucial for real-time estimation and control
  - Physical resources with placement constraints and deployment cost
  - Uncertainty quantification, interpretability

Large-scale aircraft assembly

Sensor measurements → shims



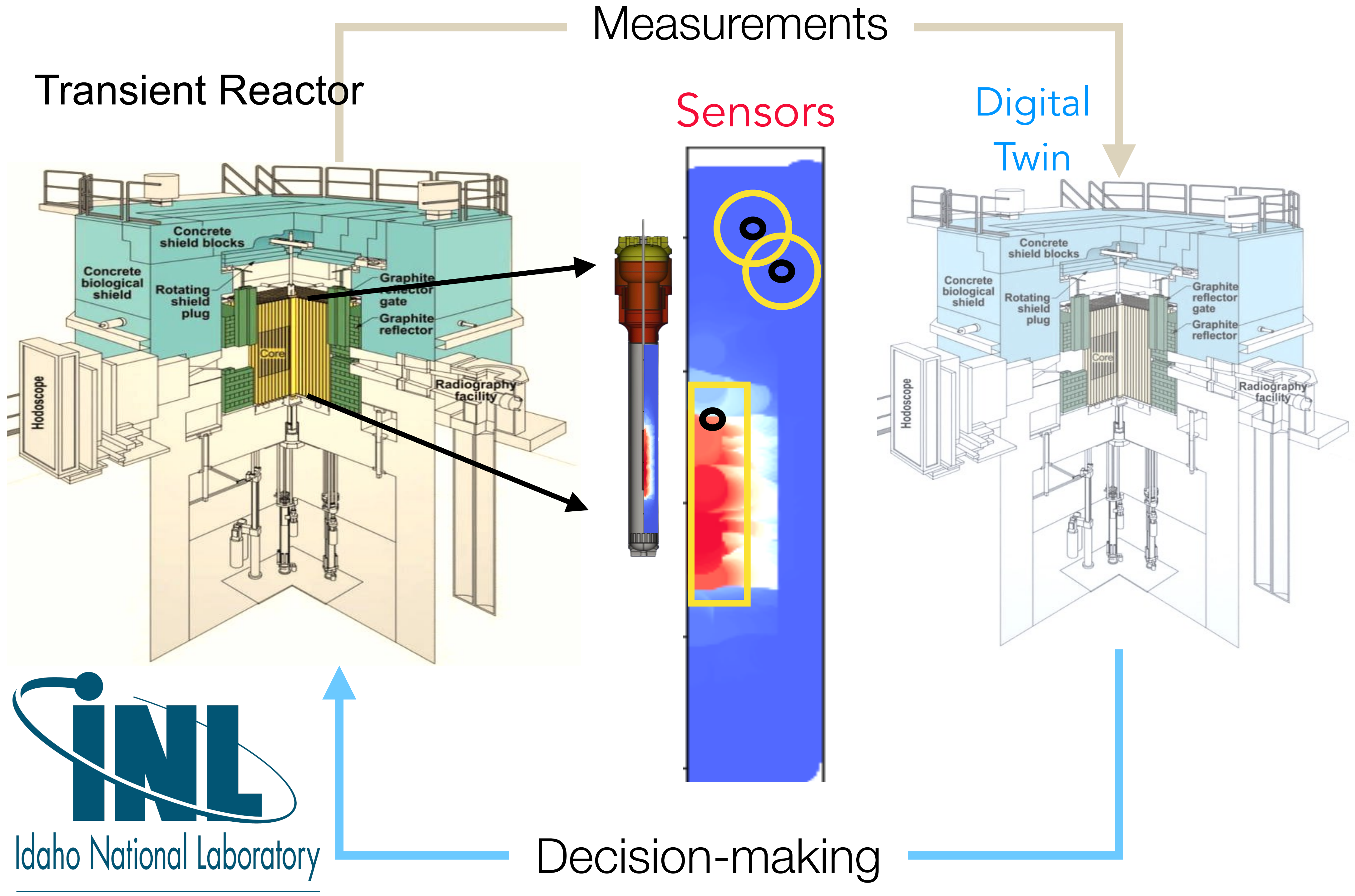
Sparse Sensing in Nuclear Energy Systems



Idaho National Laboratory

Core challenges

- **Detect, characterize sensor failure/perturbation**
- Model discrepancy between virtual & physical spaces
- Uncertainty quantification- expected performance
- Interpretability- non convex, non differentiable optimization landscape



Sparse sensor allocation in nuclear reactors

$$\min_{\mathbb{S},f} \|\mathbf{x} - f_\theta(\mathbb{S}\mathbf{x})\|_2^2$$

State
(Flow field)

Measurement
Operator

Problem setting

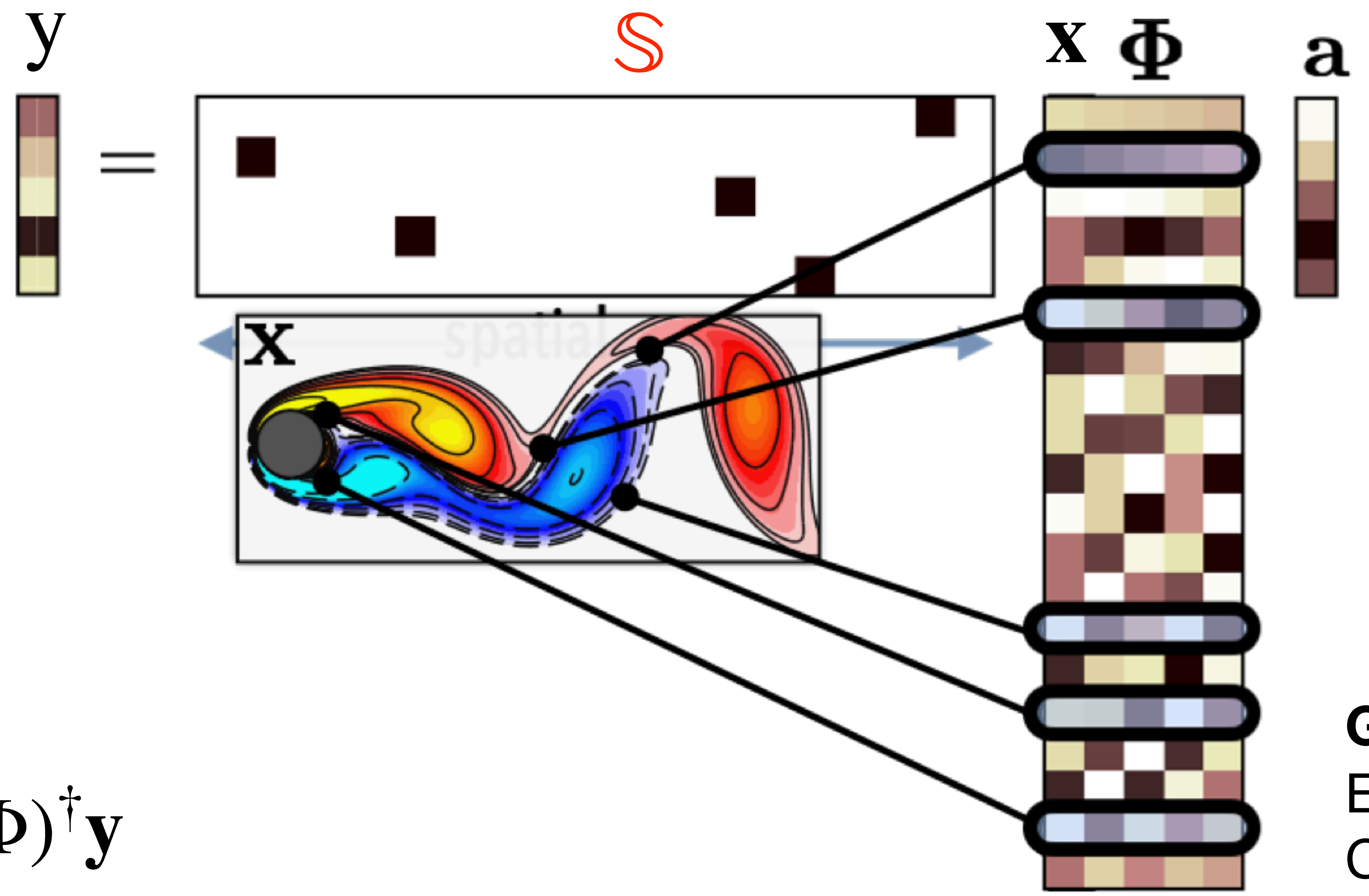
Sensor measurements $\mathbf{y} = \mathbb{S}\mathbf{x} + \xi$

Represent state in a low-dimensional basis $\mathbf{x} \approx \mathbf{\Phi a}$

$\mathbf{x} \in \mathbb{R}^n$

$\mathbf{\Phi} \in \mathbb{R}^{n \times r}$

$\mathbb{S} \in \{0,1\}^{p \times r}$



Reconstruction $\hat{\mathbf{x}} = \Phi(\mathbb{S}\Phi)^{\dagger}\mathbf{y}$

$$\boxed{\mathrm{Var}(\mathbf{a} - \hat{\mathbf{a}}) \propto \det[(\mathbb{S}\Phi_r)^T \mathbb{S}\Phi_r]^{-1}}$$

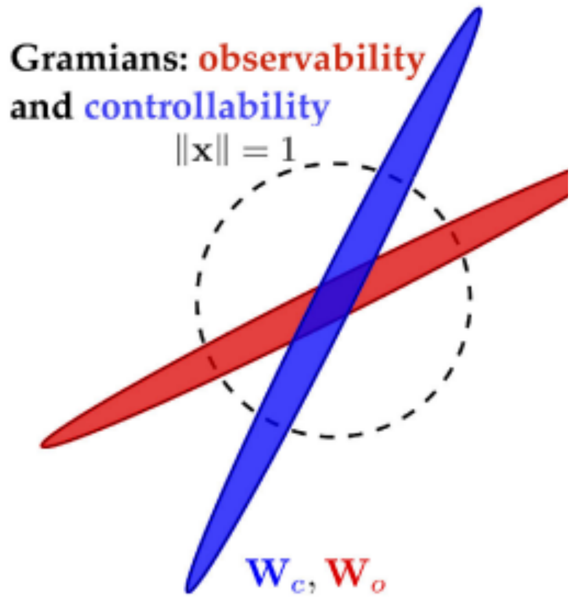
Gappy Reconstruction/Interpolation

Everson & Sirovich 1995 (gappy POD)
Chaturantabut & Sorensen 2010 (DEIM)
Manohar, Brunton, Kutz, Brunton 2018
Manohar, Kutz & Brunton, IEEE TAC 2021
Drmac & Gugercin 2016 (QDEIM)

Objective function at a glance

$$\mathrm{tr}(\Theta^T\Theta)$$

- Observability, H_2 norm optimal
  - Chen & Rowley 2016
  - Manohar et al 2021



Measure	Formula	Geometry
Generalized variance	$\det(\boldsymbol{\Sigma}) = \Pi_i \lambda_i$	area, (hyper)volume
Average variance	$\mathrm{tr}(\boldsymbol{\Sigma}) = \sum_i \lambda_i$	linear sum
Maximal variance	$\lambda_{\max}$	maximum dispersion

$$\frac{1}{\sigma_{\min}(\Theta)}$$

- Empirical Interpolation $p \geq r$
  - Drmac & Gugercin 2016
  - Perherstofer et al 2020

$$\ln\det[\overbrace{(\mathbb{S}\Phi_r)^T\mathbb{S}\Phi_r}^{\Theta^T\Theta}]$$

- When submodular, greedy *near-D-optimal*
- Probabilistic view (max *volume* ellipsoid)
  - Joshi & Boyd 2016
  - Kakasenko, Alexanderian et al, arXiv 2025

Bayesian Inference Setting

- Prior over modal coefficients

$$p(\mathbf{a}) = \exp\left(-\frac{1}{2}\mathbf{a}^T S^{-2}\mathbf{a}\right)$$

- Likelihood

$$\ln p(\mathbf{y}\,|\,\mathbf{a}) = -\frac{1}{2\eta^2}(\mathbf{y} - \Theta\mathbf{a})^T(\mathbf{y} - \Theta\mathbf{a}) - \frac{1}{2}\mathbf{a}^T S^{-2}\mathbf{a}$$

- MAP estimate

$$\mathbf{a} = \underbrace{\left(S^{-2} + \frac{\Theta^T\Theta}{\eta}\right)}_{A}{}^{-1}\frac{\Theta^T\mathbf{y}}{\eta}$$

Klishin, Kutz & KM (2023)
arXiv:2307.11838


Andrei Klishin


Nathan Kutz

D-optimal design

MAP estimate

$$\mathbf{a} = \Bigg(\underbrace{S^{-2} + \frac{\Theta^T\Theta}{\eta}}_{A}\Bigg)^{-1} \frac{\Theta^T\mathbf{y}}{\eta}$$

Hamiltonian of a given sensor set γ

$$\boxed{\mathcal{H}(\gamma) \equiv -\ln\det A = E_b - \operatorname{tr}\ln\left(\mathbf{I} + \frac{\Theta S^2\Theta^T}{\eta^2}\right)}$$

Decompose into sums over 1pt, 2pt, ... s pt interactions:

$$\mathcal{H}(\gamma) = E_b - \operatorname{tr}\ln\left(\mathbf{I} + \frac{1}{\eta^2}(\mathbf{D} + \mathbf{R})\right)$$

$$= E_b - \operatorname{tr}\ln(\mathbf{I} + \eta^{-1}\mathbf{D})$$

$$+ \sum_{s=1}^{\infty} \frac{(-1)^s}{s!} \operatorname{tr}\left(\left[\eta^{-2}\mathbf{R}(\mathbf{I} + \eta^{-2}\mathbf{D})^{-1}\right]^s\right)$$

Klishin, Kutz & KM (2023)
arXiv:2307.11838

Two-point Energy Approximation

$$\boxed{\begin{aligned}\mathscr{H}_{2pt}(\gamma) \equiv & -\operatorname{tr}\ln(\mathbf{I} + \eta^{-2}\mathbf{D}) \\ & + \frac{1}{2}\operatorname{tr}\left(\left[\eta^{-2}\mathbf{R}(\mathbf{I} + \eta^{-2}\mathbf{D})^{-1}\right]^2\right)\end{aligned}}$$

Express as dot products of *sensing vectors* (columns of measurement-basis product)

$$G \equiv \Theta S$$

$$\mathscr{H}_{2pt}(\gamma) \equiv \sum_i h_i + \sum_{i,j} J_{ij}$$

Sensor 1-pt energy

$$h_i = -\ln\left(1 + \mathbf{g}_i \cdot \mathbf{g}_i/\eta^2\right) \leq 0$$

Attracted to high-variance locations

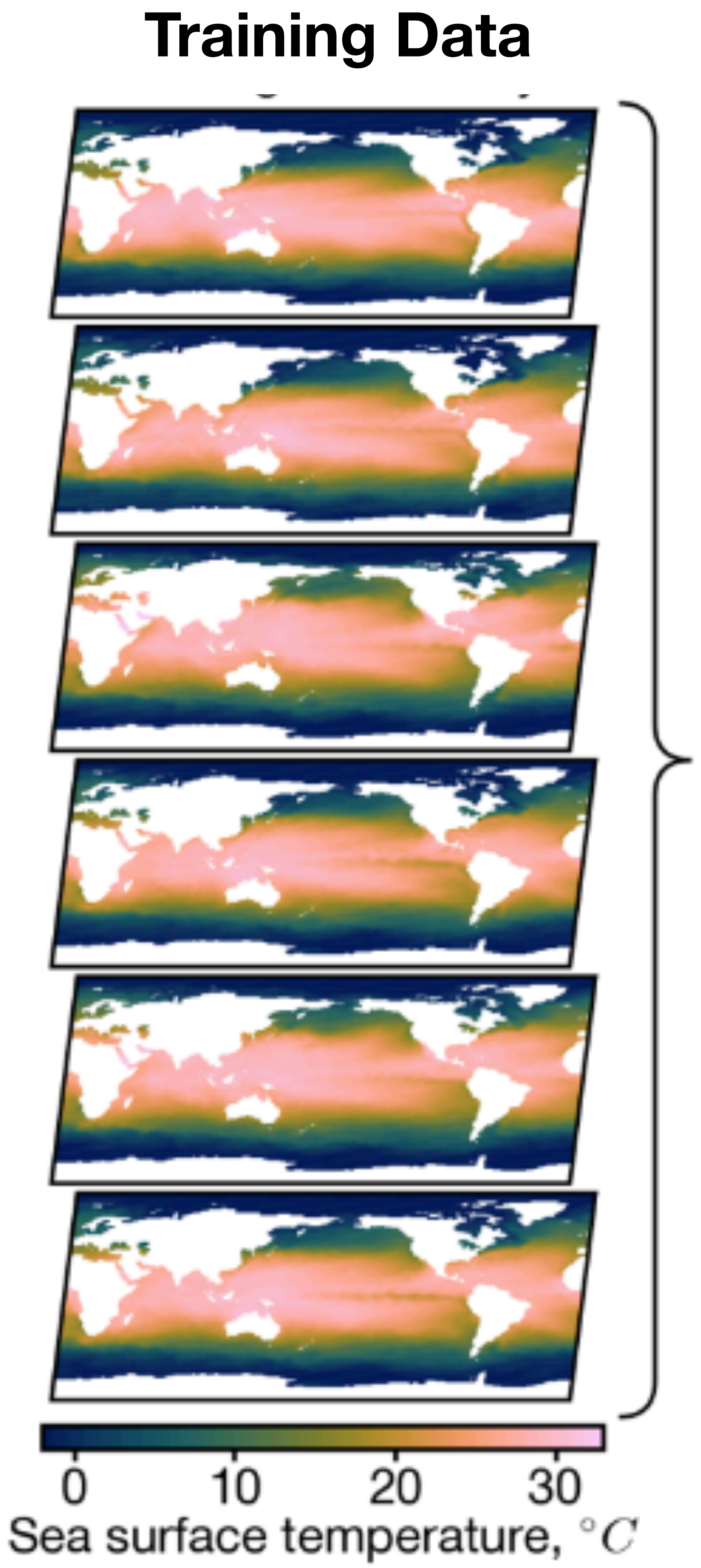
Sensor 2-pt energy (crosstalk)

$$J_{ij} = \frac{(\mathbf{g}_i \cdot \mathbf{g}_j/\eta^2)^2}{\left(1 + \mathbf{g}_i \cdot \mathbf{g}_i/\eta^2\right)\left(1 + \mathbf{g}_j \cdot \mathbf{g}_j/\eta^2\right)} \geq 0$$

Repelled from correlated sensors

Klishin, Kutz & KM (2023)
arXiv:2307.11838

Two-point greedy sensor placement

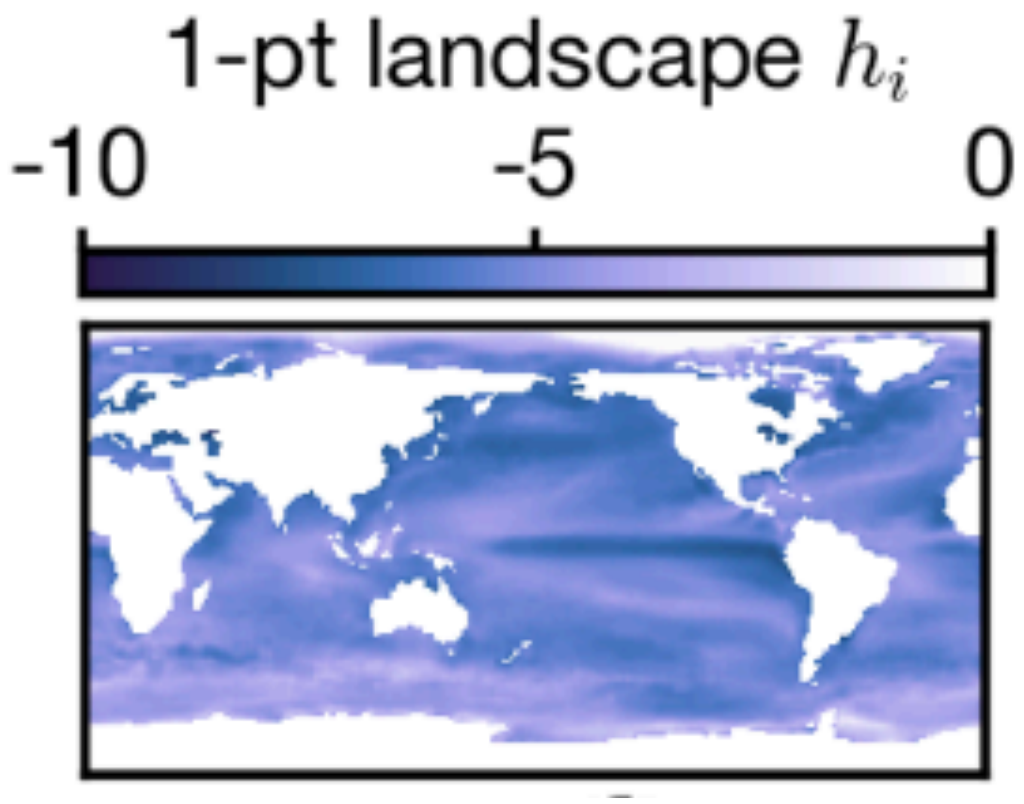


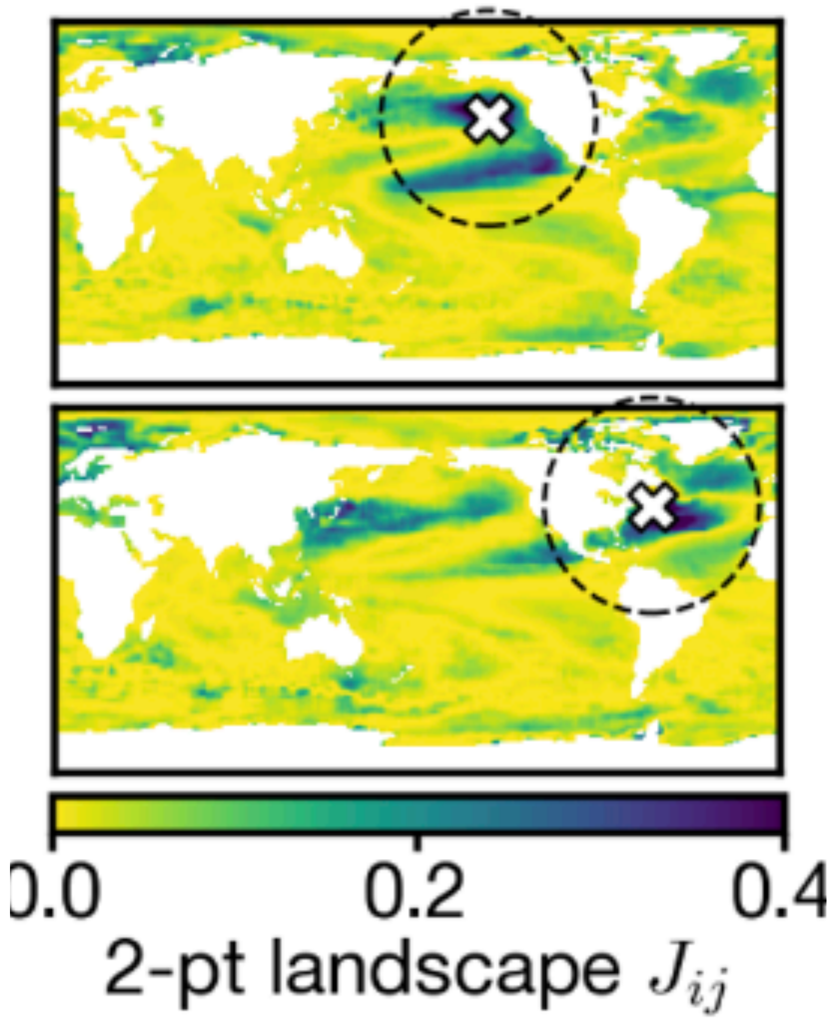
Greedy strategy for selection of the next sensor aims to minimize sensor 1- and 2- body interactions (crosstalk) $O(nrp)$

$$q = \arg\min_{q \notin \gamma} \Big(h_q + 2 \sum_{i \in \gamma} J_{iq} \Big); \;\; \gamma \leftarrow q$$

Two-point interactions provide a sensor optimization *landscape*

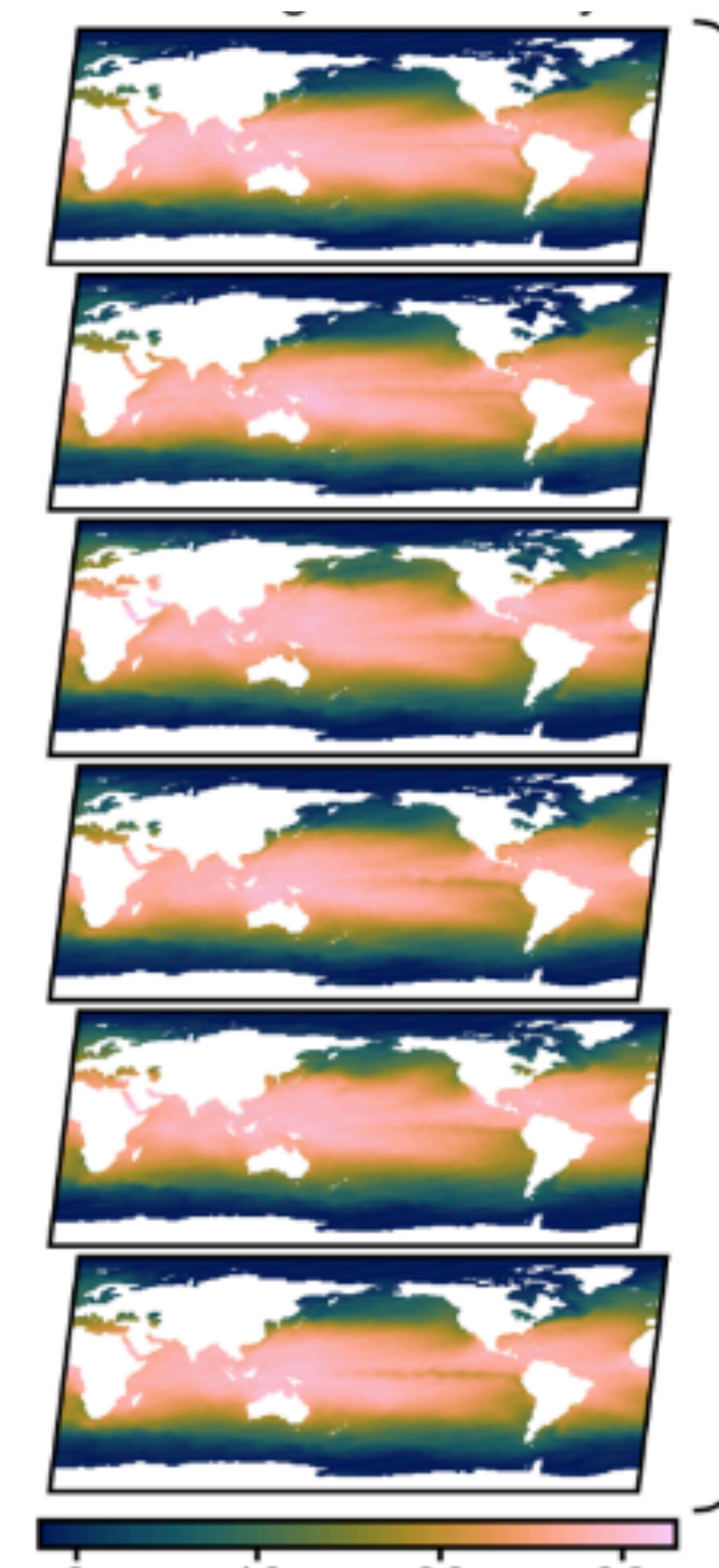
→ Alternatives to the minimizer for emerging design constraints





Two-point Energy Approximation

Training Data

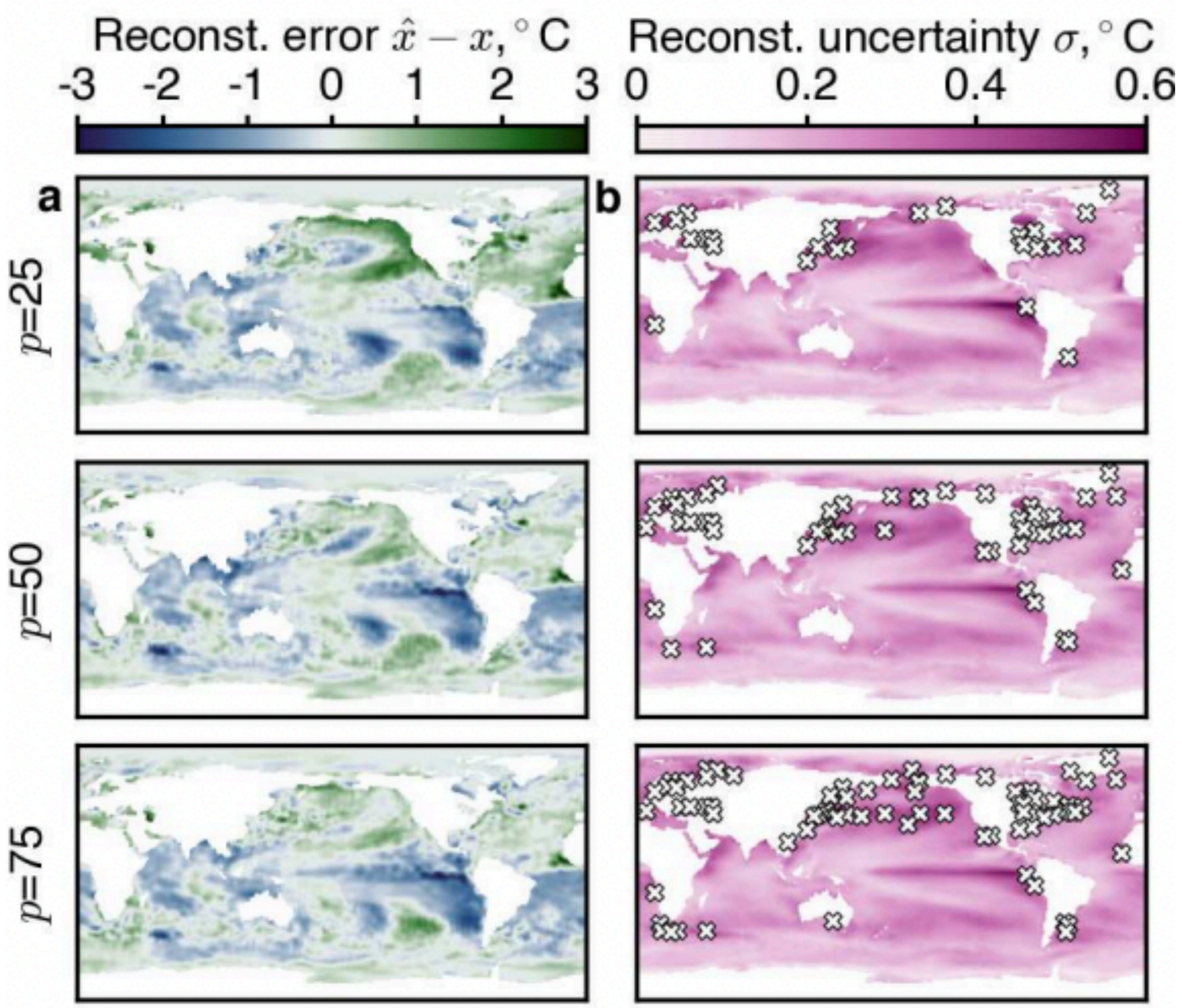


State reconstruction fluctuation propagated from sensor reading

Compute only diagonal part of the covariance matrix $\langle \Delta \mathbf{x} \Delta \mathbf{x}^T \rangle$

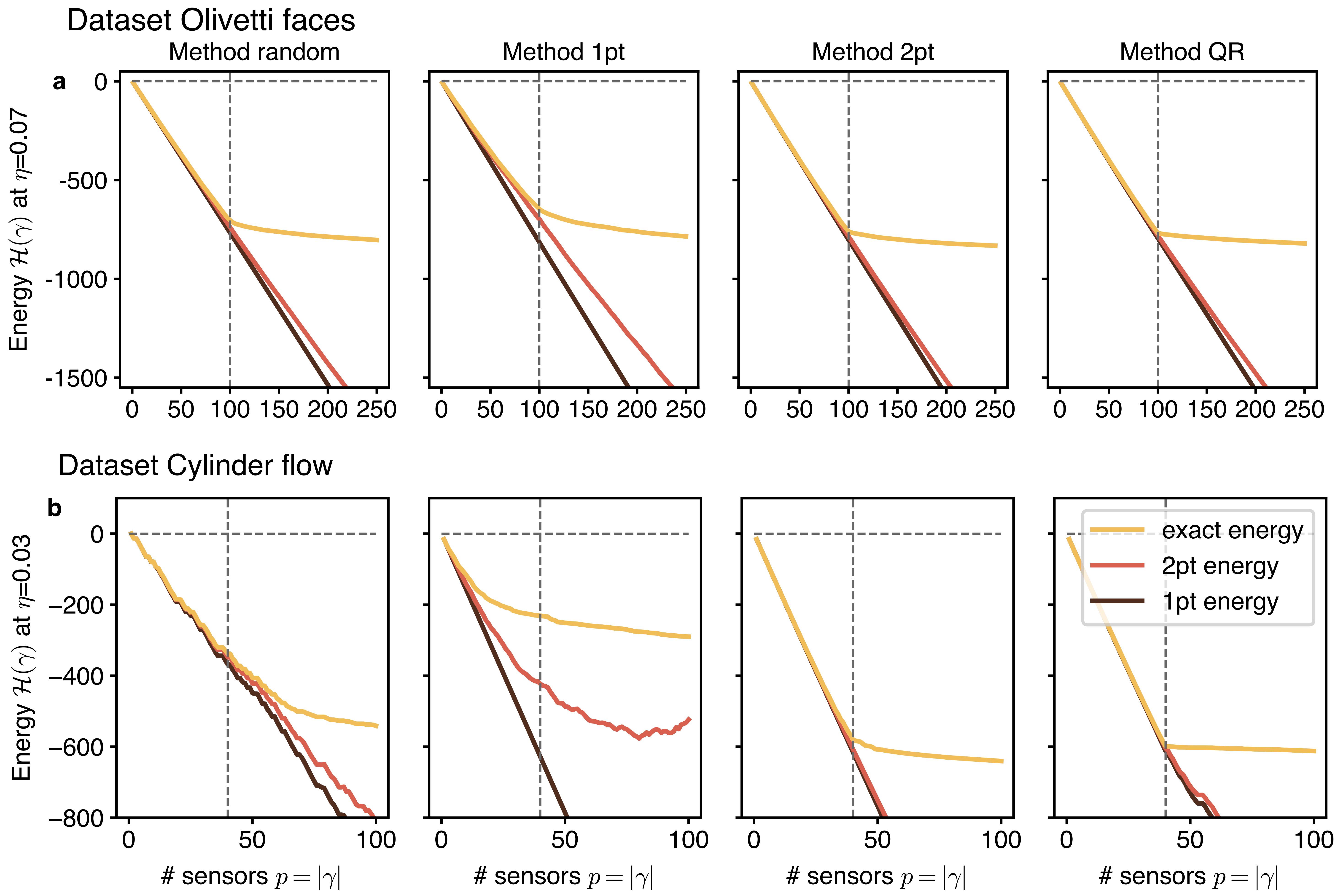
$$B = \Phi_r A^{-1} \frac{\Theta^T}{\eta^2}$$

$$\sigma_i = \eta \sqrt{\sum_j B_{ij}^2}$$

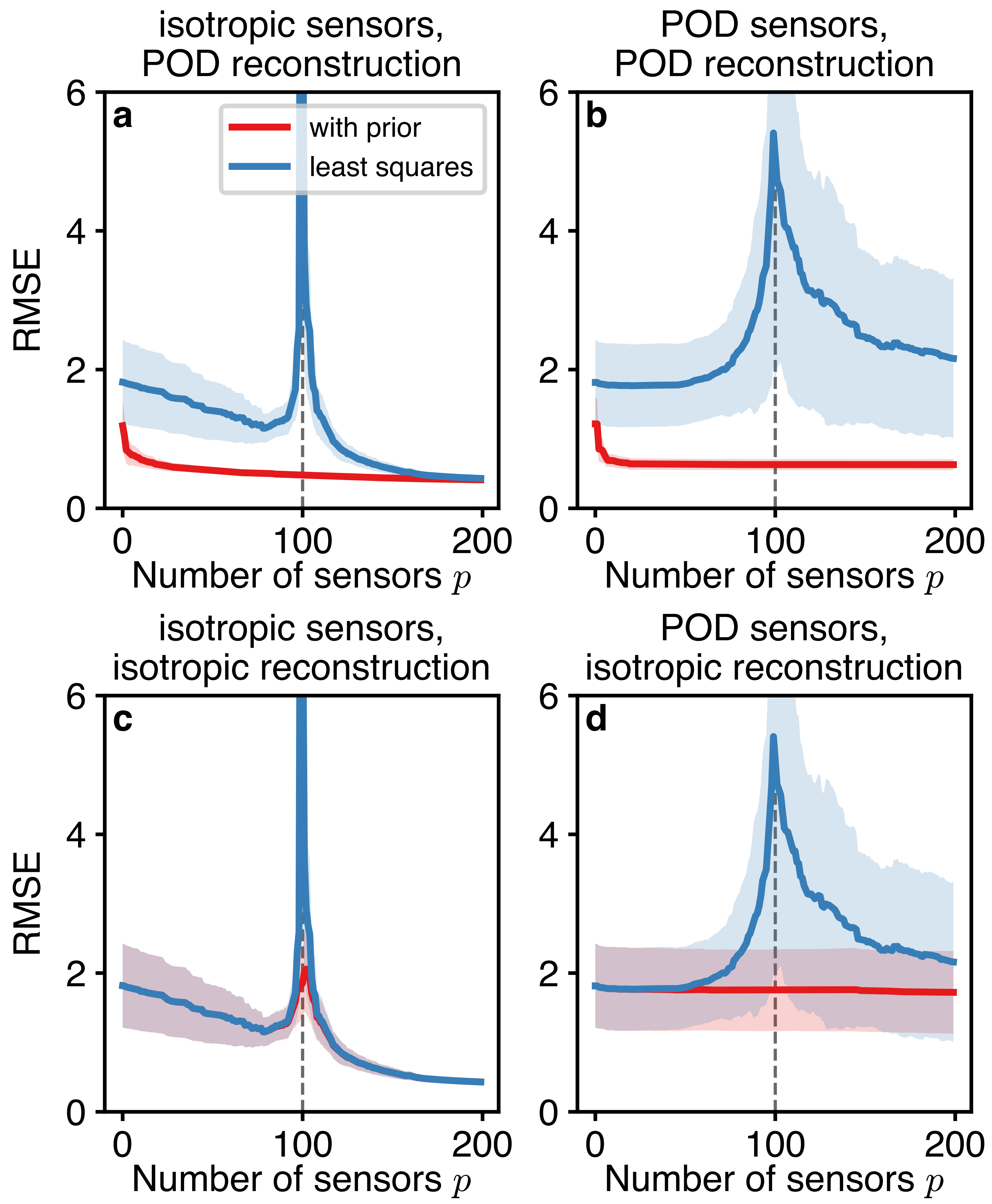


Validity of approximation

For small p (no. sensors) 2pt energy is a good approximation



Effect of Data-Driven Prior

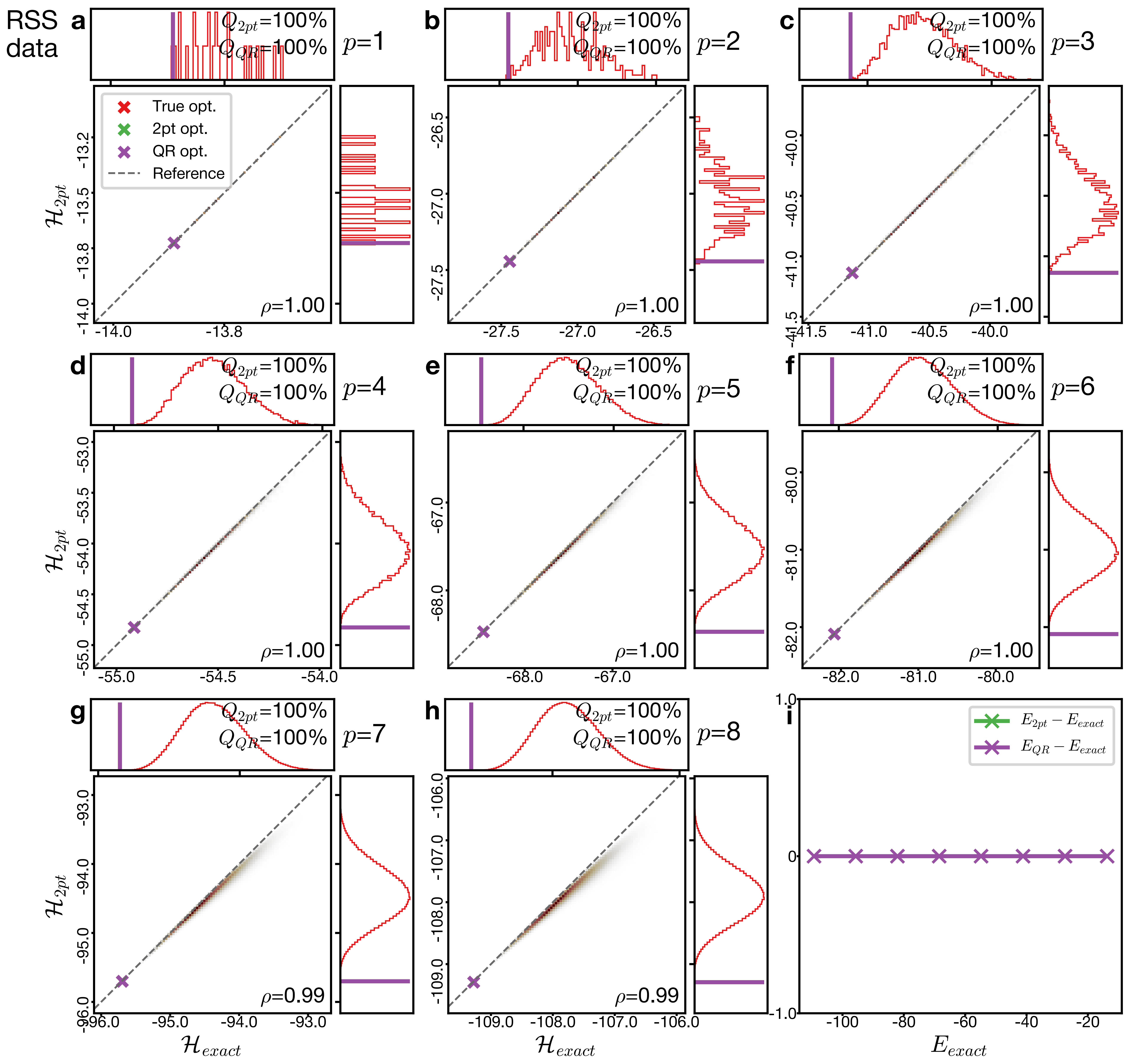


- Data-driven prior $S = \Sigma^2/\sqrt{n-1}$
- Isotropic prior $S = \sigma_{prior} I$
- Mitigates CPQR instability when $p = r$
- Outlook: Error covariance decomposition

Outlook: Analytic expression for RMSE curve

$$\mathbf{K} = \underbrace{\mathbf{\Psi}_c \mathbf{B}_0 \mathbf{B}_0^T \mathbf{\Psi}_c^T}_{\text{subleading modes}} + \underbrace{\mathbf{\Psi}_r \mathbf{B}_1 \mathbf{B}_1^T \mathbf{\Psi}_r^T}_{\text{leading modes}} + \underbrace{\mathbf{\Psi}_r \mathbf{B}_2 \mathbf{B}_2^T \mathbf{\Psi}_r^T}_{\text{contamination}} + \underbrace{\mathbf{\Psi}_r \mathbf{B}_3 \mathbf{B}_3^T \mathbf{\Psi}_r^T}_{\text{noise}}$$

Comparison to exhaustive search

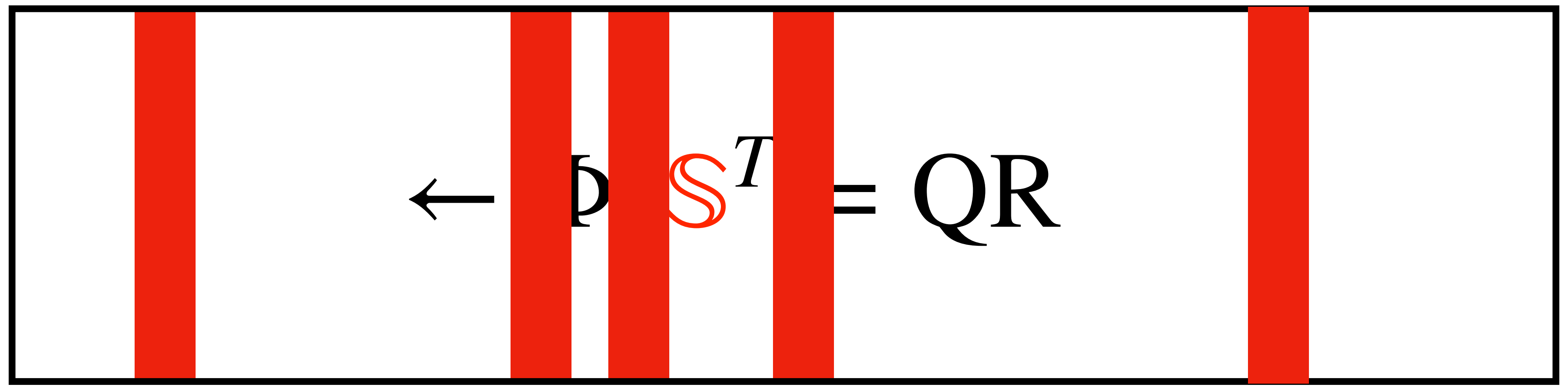


Column-Pivoted QR

Minimize error covariance of estimation $\mathbb{S}_\star = \underset{\mathbb{S}}{\arg\max}\, \sigma((\mathbb{S}\Phi_r)^T \mathbb{S}\Phi_r)$

$$\mathrm{Var}(\mathbf{a} - \hat{\mathbf{a}}) \propto \det[(\mathbb{S}\Phi_r)^T \mathbb{S}\Phi_r]^{-1}$$

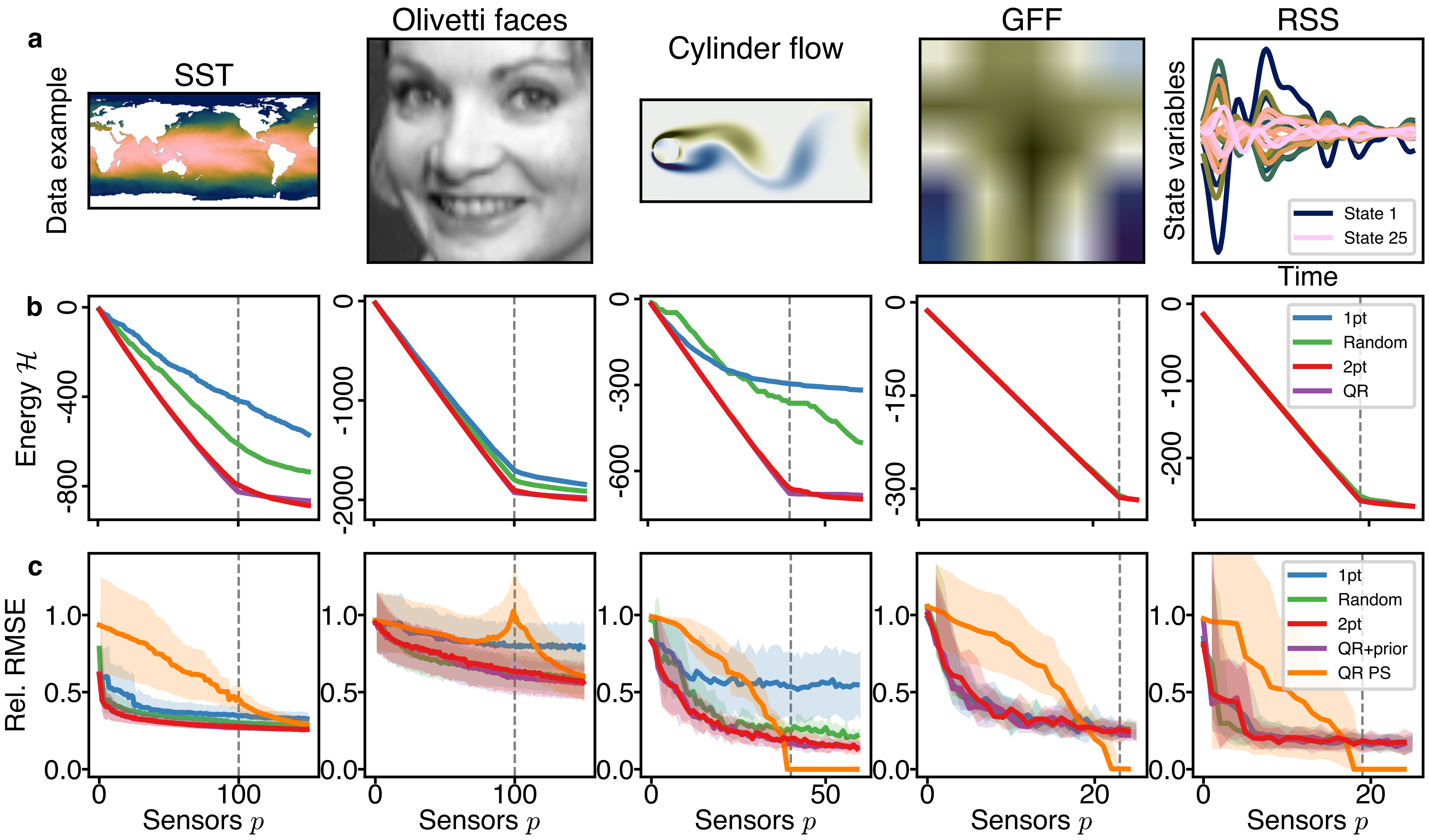
$$\Phi^T$$



$$\det \Theta^T \Theta = \prod_i R_{ii}^2$$

Efficient interpolation points $O(nr^2)$
Drmac and Gugercin 2016
KM, Brunton, Kutz & Brunton, IEEE CSM 2018
KM, Kutz & Brunton, IEEE TAC 2021
Peherstofer et al, 2020 (ODEIM $p > r$)
....

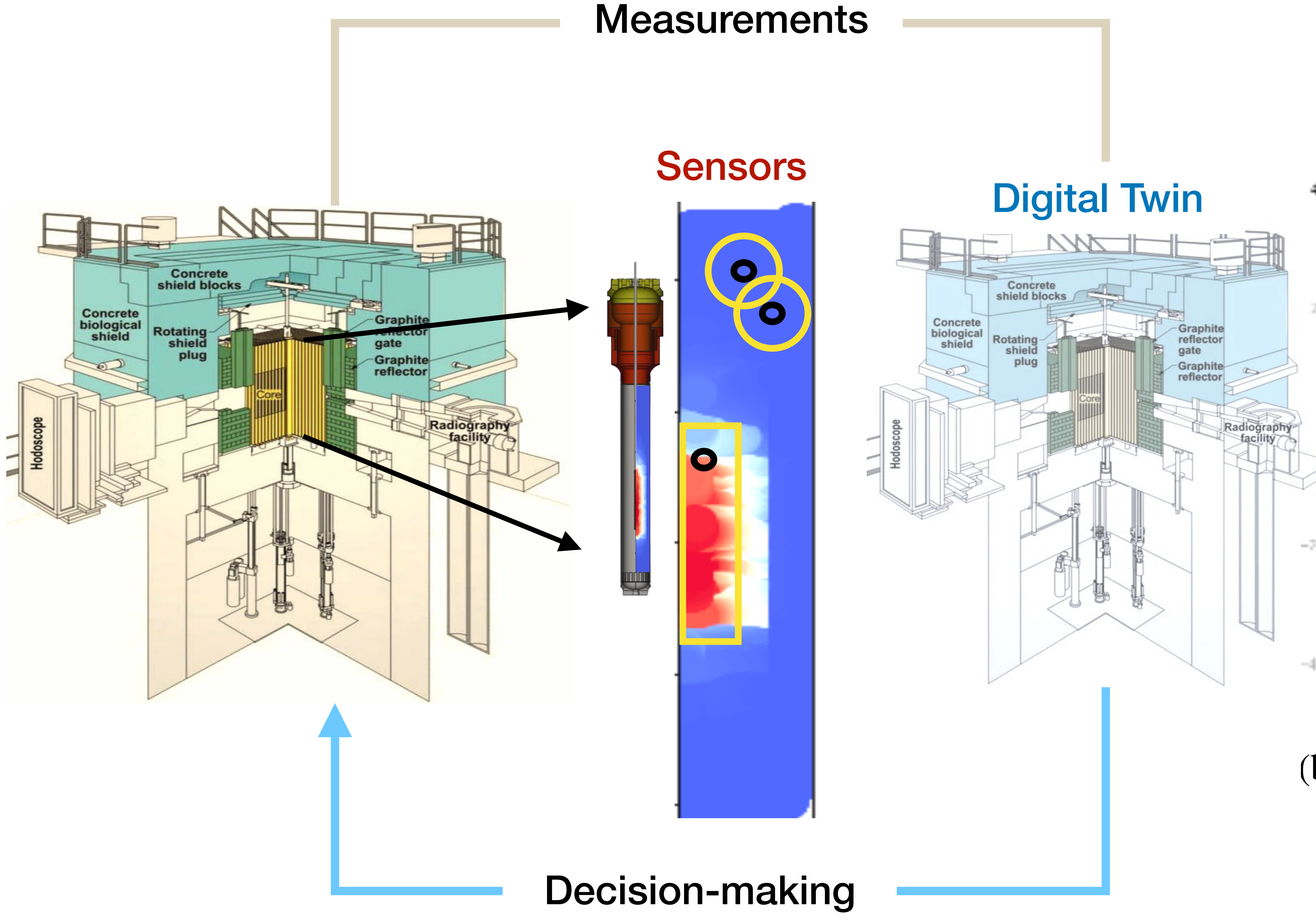
Reconstruction results

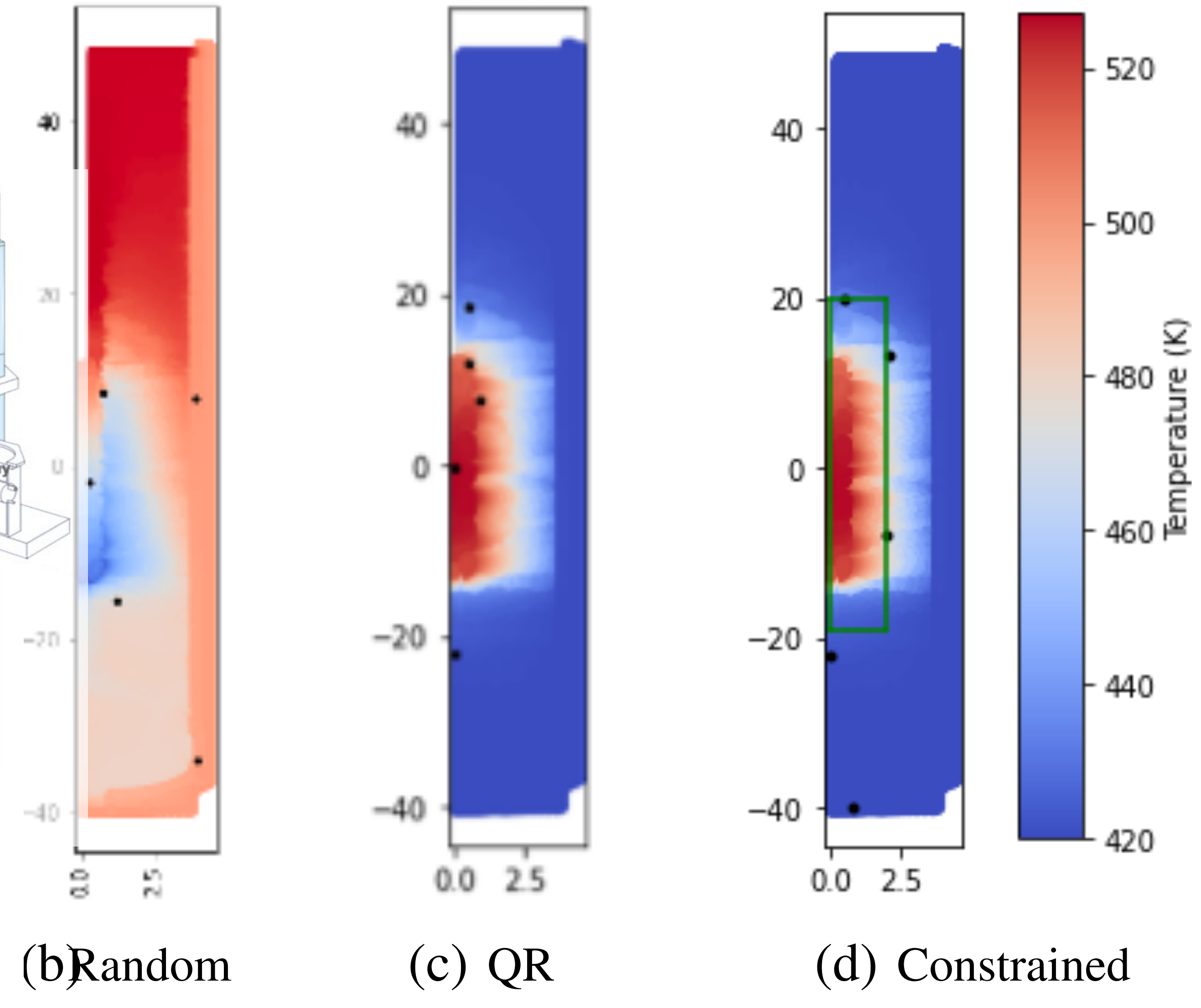


CPQR and 1pt consistently underperform 2pt greedy and data-driven prior

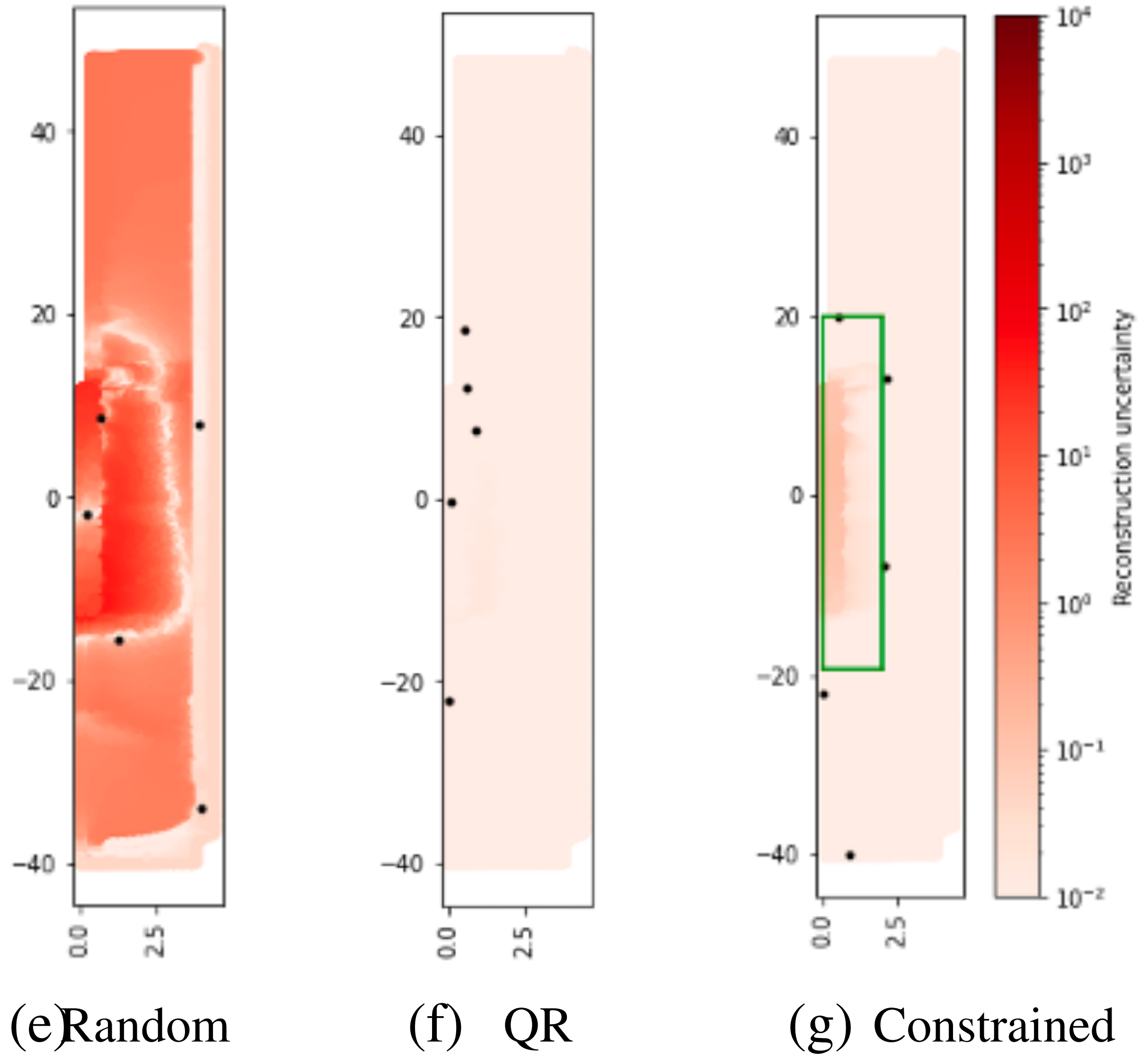
Constrained Sensing for Nuclear DTs

Reconstruct temperature field from optimized, spatially constrained measurements
Heater adjacent sensors optimize reconstruction

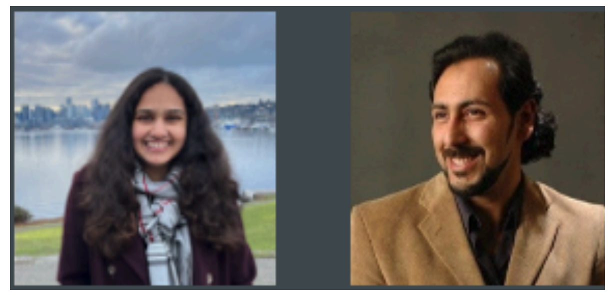




Sparse sensor reconstruction



Uncertainty heatmaps



Karnik, Abdo, Perez, Yoo, Cogliati, Skifton, Calderoni, Brunton & KM,
arXiv:2306.13637 *IEEE Sensors Journal* (2024)

Outlook: sensor information criteria

- *A set function submodular* if $f(\gamma \cup \{s\}) - f(\gamma)$ is monotone decreasing $\forall s \in V$ *(Diminishing returns)*

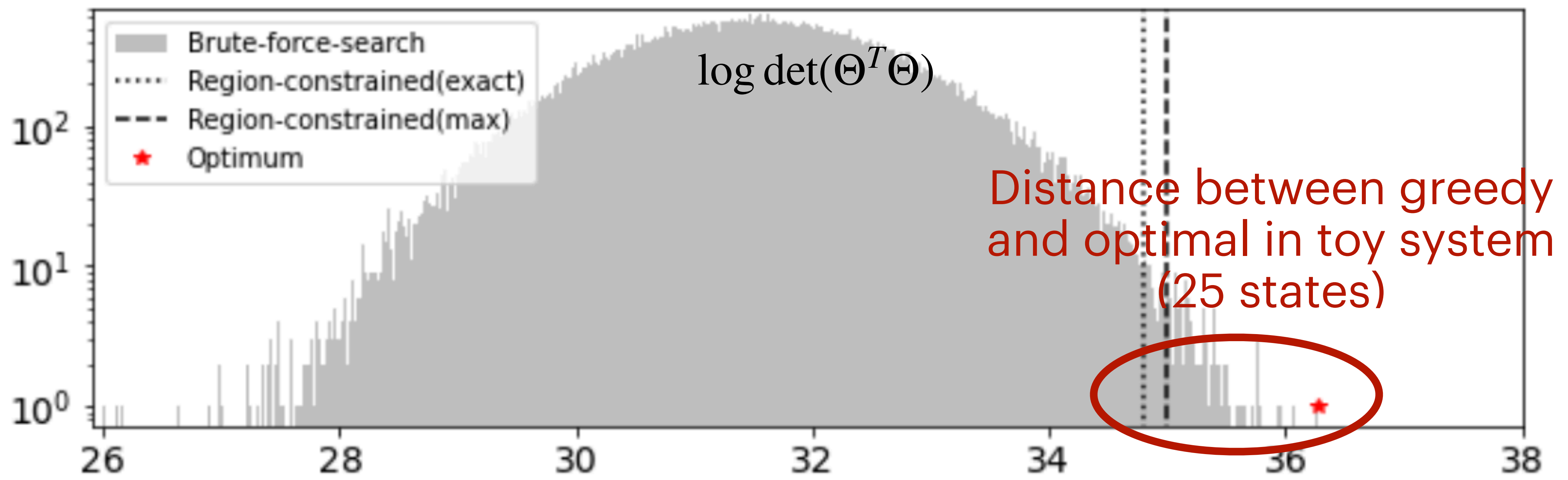
$$\ln \det \mathbb{S} W_c \mathbb{S}^T$$

Summers et al, 2016

- Ex. Observability/controllability metrics, mutual information, neural estimation
  Belghazi, Bengio, et al., ICML (2018)

$$\arg\min_{\gamma} I(X_\gamma; X_{S \setminus \gamma})$$

Krause et al, JMLR, 2008
Gaussian Processes

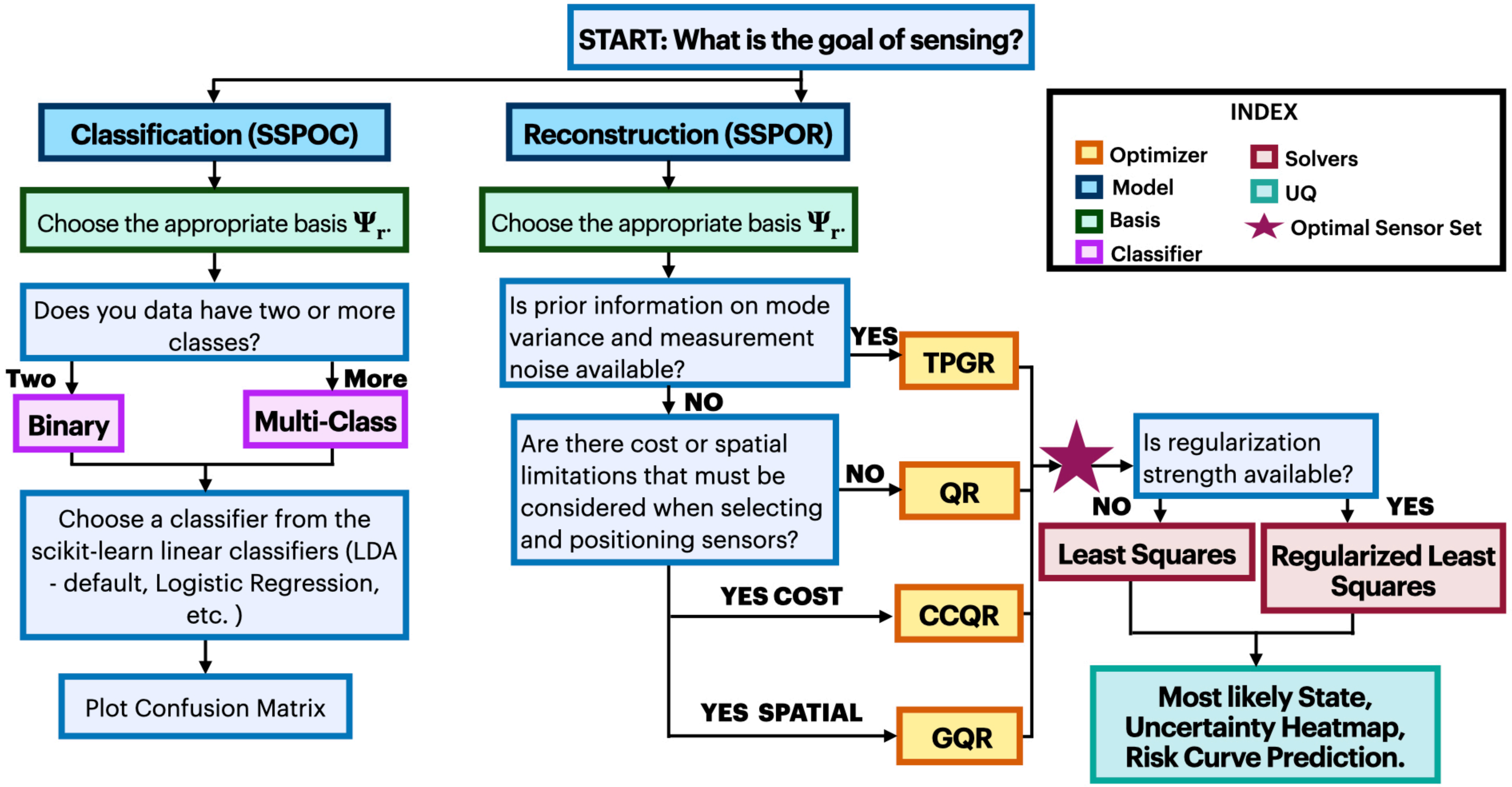


PySensors 2.0

Open-source code and benchmark datasets

Code 2: Implementation of the Two Point Greedy Optimizer

```
basis = ps.basis.SVD(n_basis_modes=r)
optimizer = ps.optimizers.TPGR(n_sensors, noise, prior)
model = ps.SSPOR(basis=basis, optimizer=optimizer)
model.fit(data)
sensors=model.get_selected_sensors()
one_pt_landscape = model.one_pt_energy_landscape(prior, noise)
two_pt_landscape = model.two_pt_energy_landscape(prior, noise, sensors)
```



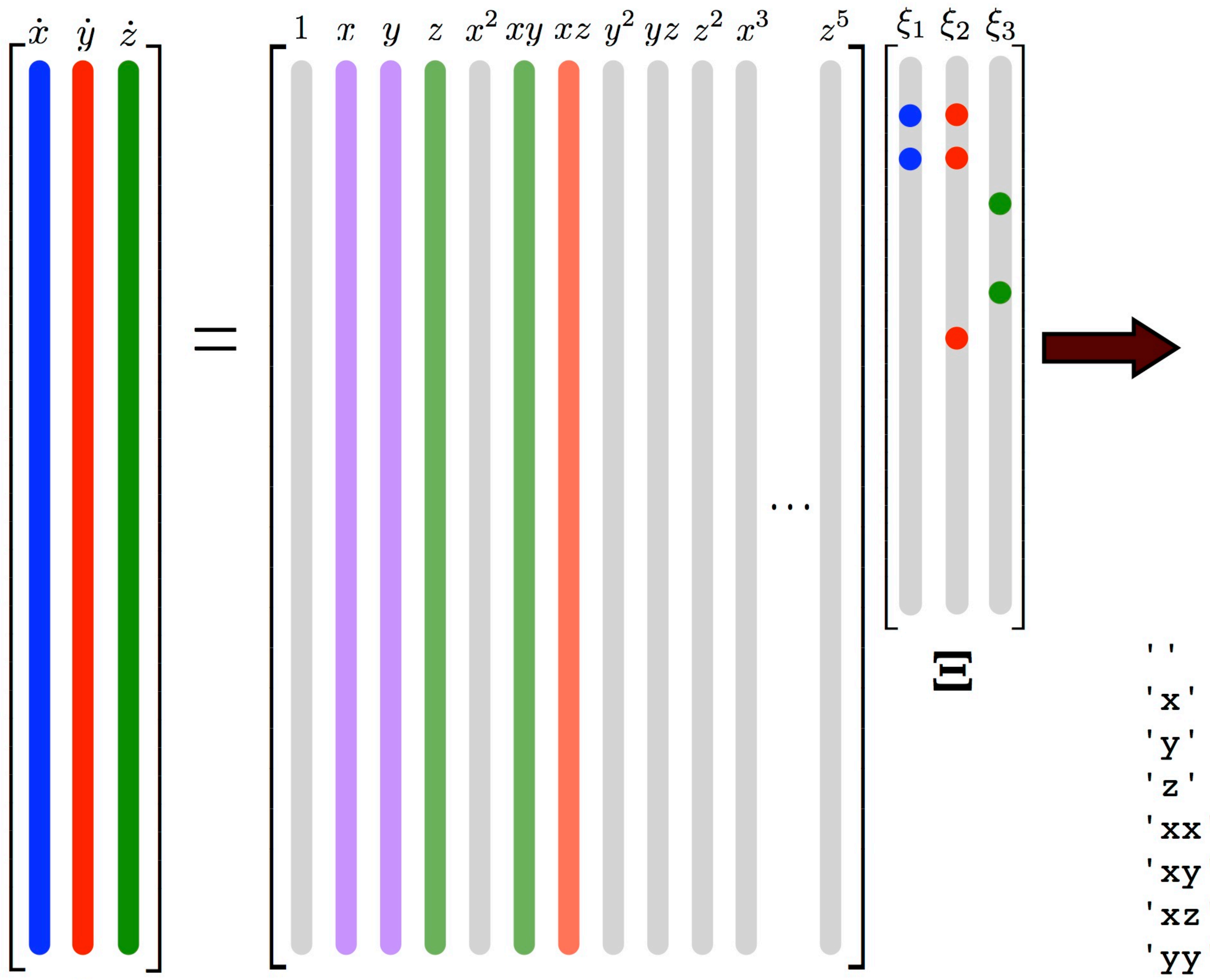
pysensors

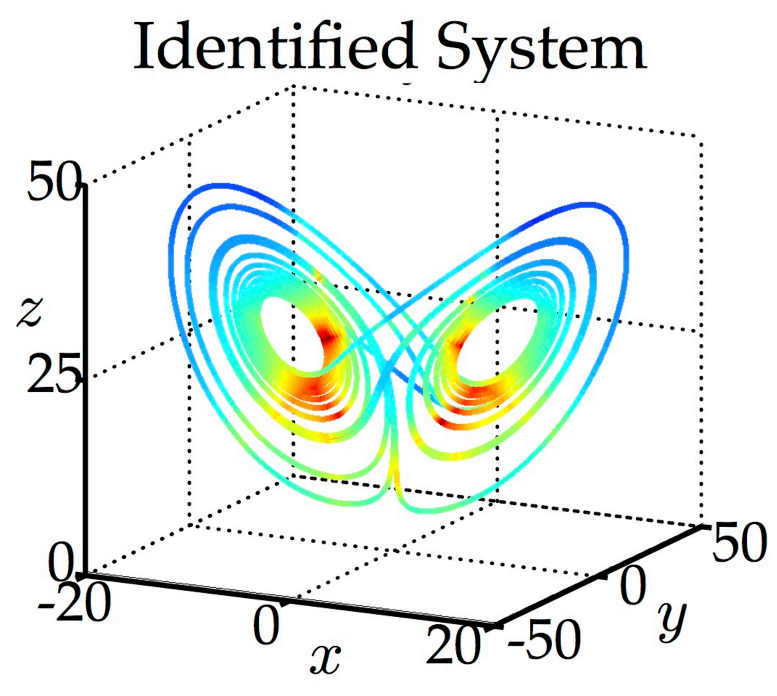
Karnik, Bhangale, Abdo, Klishin, Cogliati, Brunton, Kutz, Brunton & Manohar, arXiv 2509.08017 (2025)

Nonlinear System Identification

- **Goal: learn ODEs, laws of motion from trajectory data**
- Nonconvex optimization = many solutions
- When is recovery possible: trajectory length n, Δt , noise η
- Hyperparameters: sparsity λ, resolution ρ

Sparse regression in constrained library: SINDy





''	'xi_1'	'xi_2'	'xi_3'
'x'	[-9.9996]	[27.9980]	[0]
'y'	[9.9998]	[-0.9997]	[0]
'z'	[0]	[0]	[-2.6665]
'xx'	[0]	[0]	[0]
'xy'	[0]	[0]	[1.0000]
'xz'	[0]	[-0.9999]	[0]
'yy'	[0]	[0]	[0]

Brunton, Proctor, Kutz, *PNAS (2016)*

SINDy sparse regression loss function

$$\Xi_l = \arg\min_{\Xi} \frac{1}{2\rho^2} \sum_t (\dot{x}_l - \Theta^T(x) \cdot \Xi_l)^2 + \underbrace{\lambda \|\Xi_l\|_p}_{\text{Sparsifying for } p \leq 1}$$

Statistical mechanics for SINDy


Andrei Klishin


Joseph Bakarji

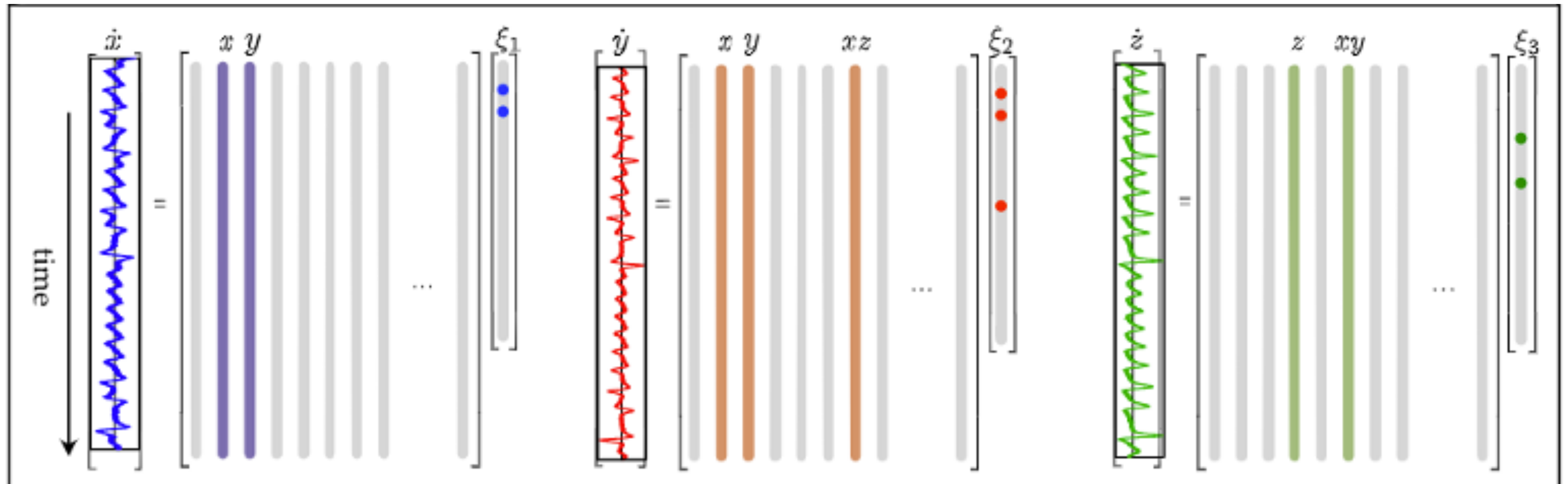

Nathan Kutz

Sparse Identification of Nonlinear Dynamics

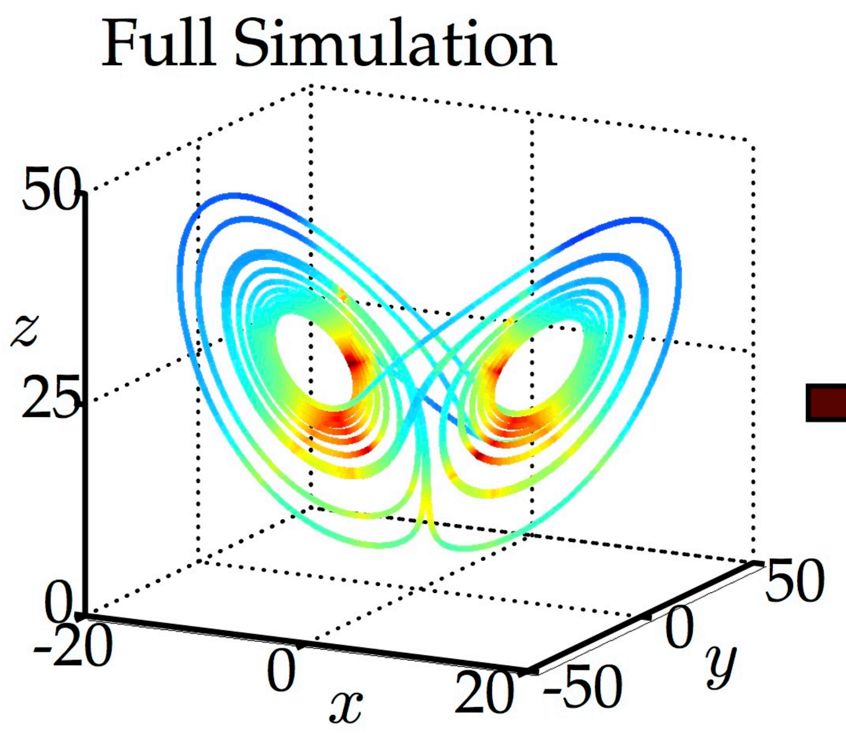
Identify ODE from trajectory data

$$\dot{x}_l = f_l(\overrightarrow{x}) \approx \sum_{i=1}^{N} \Theta_i(\overrightarrow{x})\Xi_{il}$$

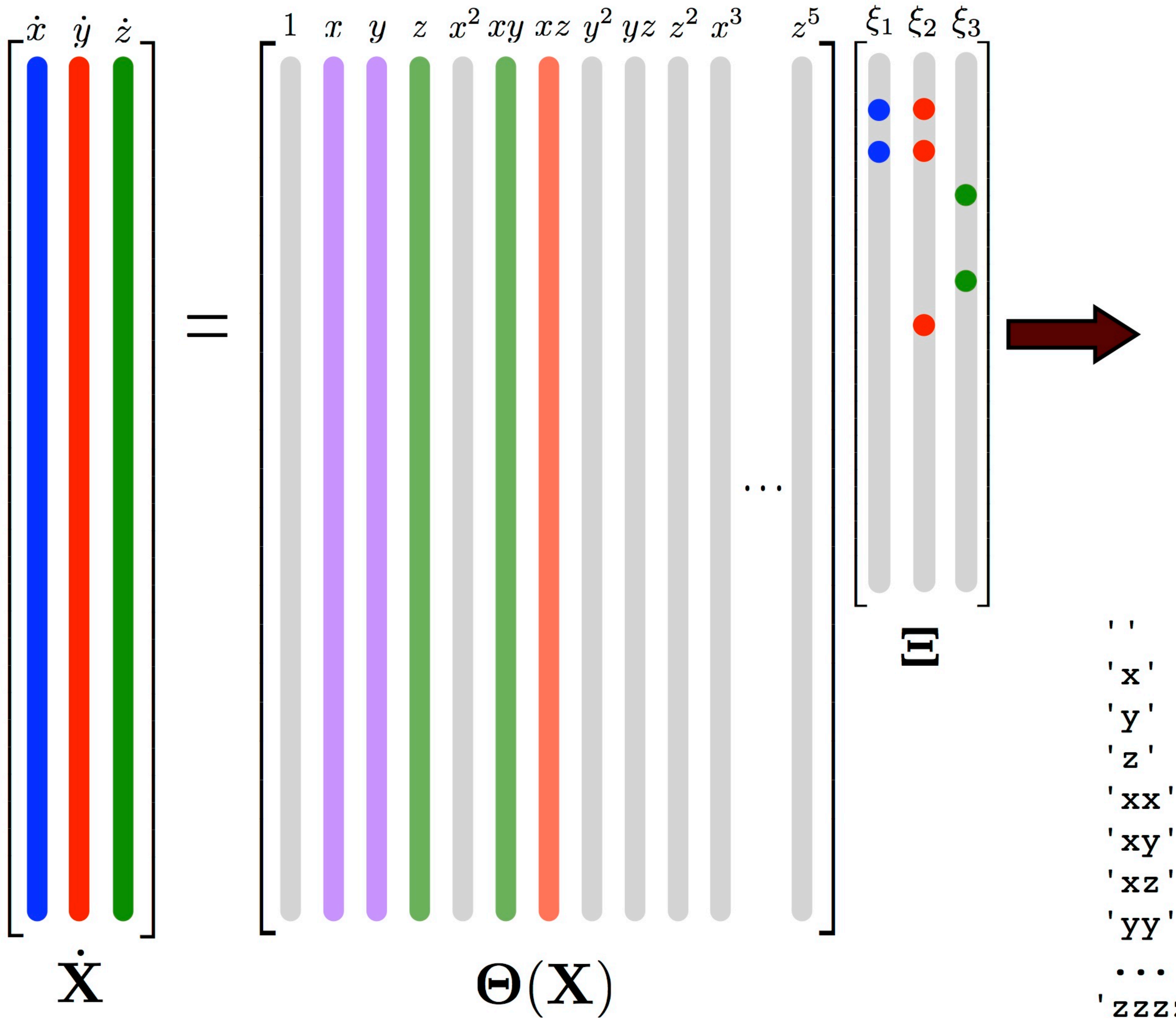
where $\Theta_i(\overrightarrow{x})$ is an overcomplete nonlinear library and Ξ_{il} is *sparse*

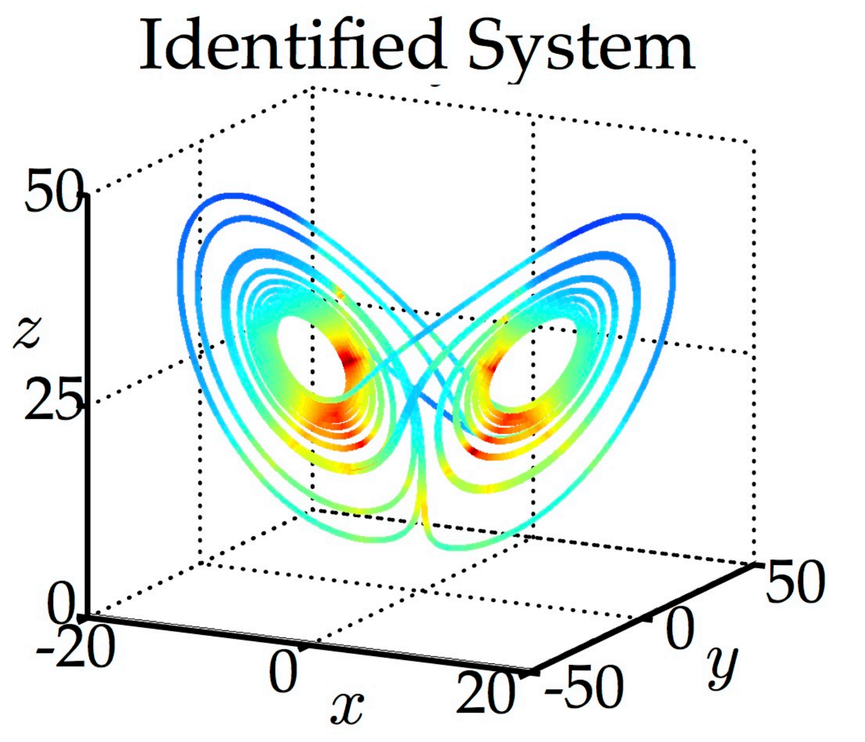


II. Sparse Regression to Solve for Active Terms in the Dynamics



$$\begin{aligned}\dot{x} &= \sigma(y-x)\\ \dot{y} &= x(\rho - z) - y\\ \dot{z} &= xy - \beta z.\end{aligned}$$



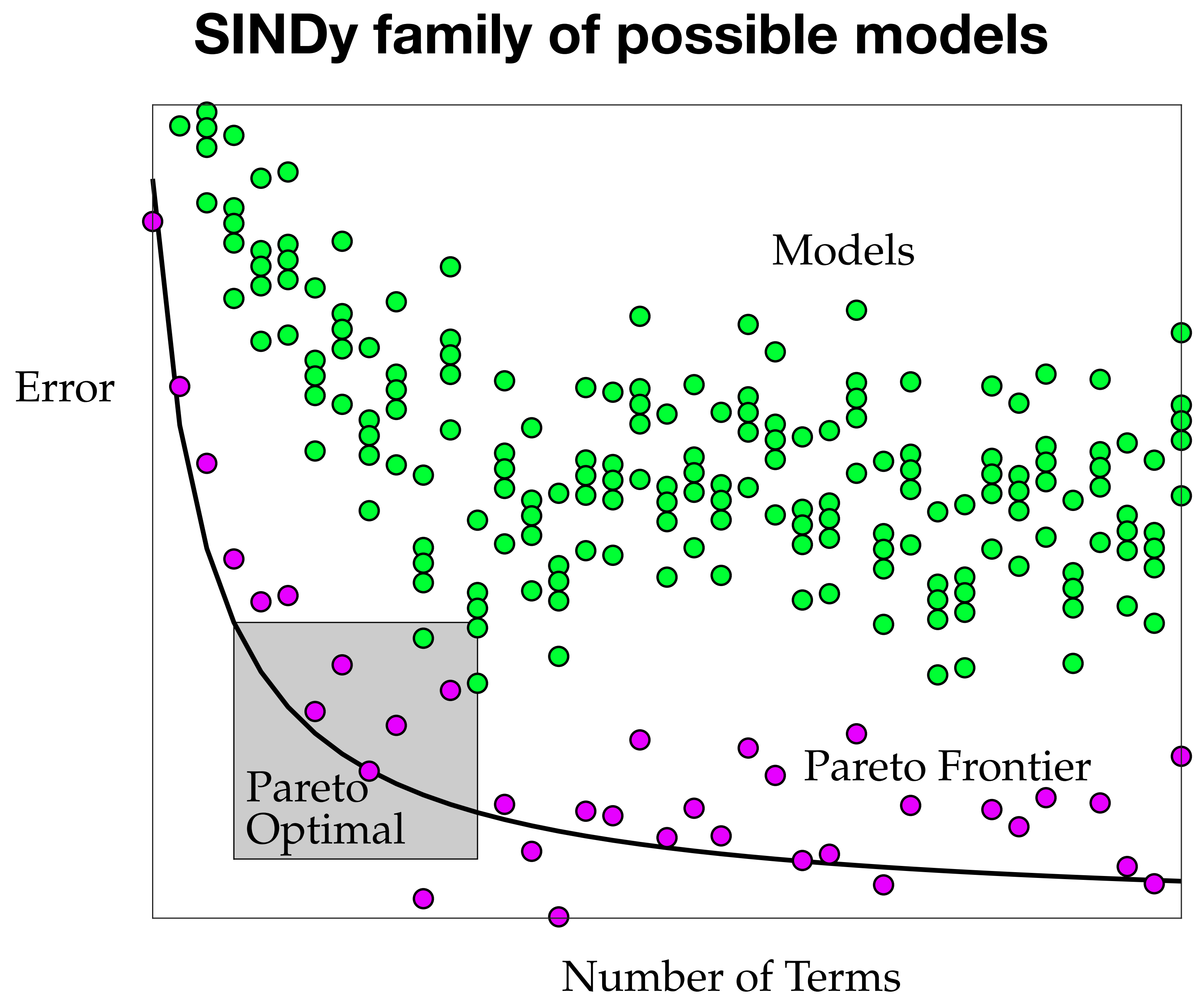


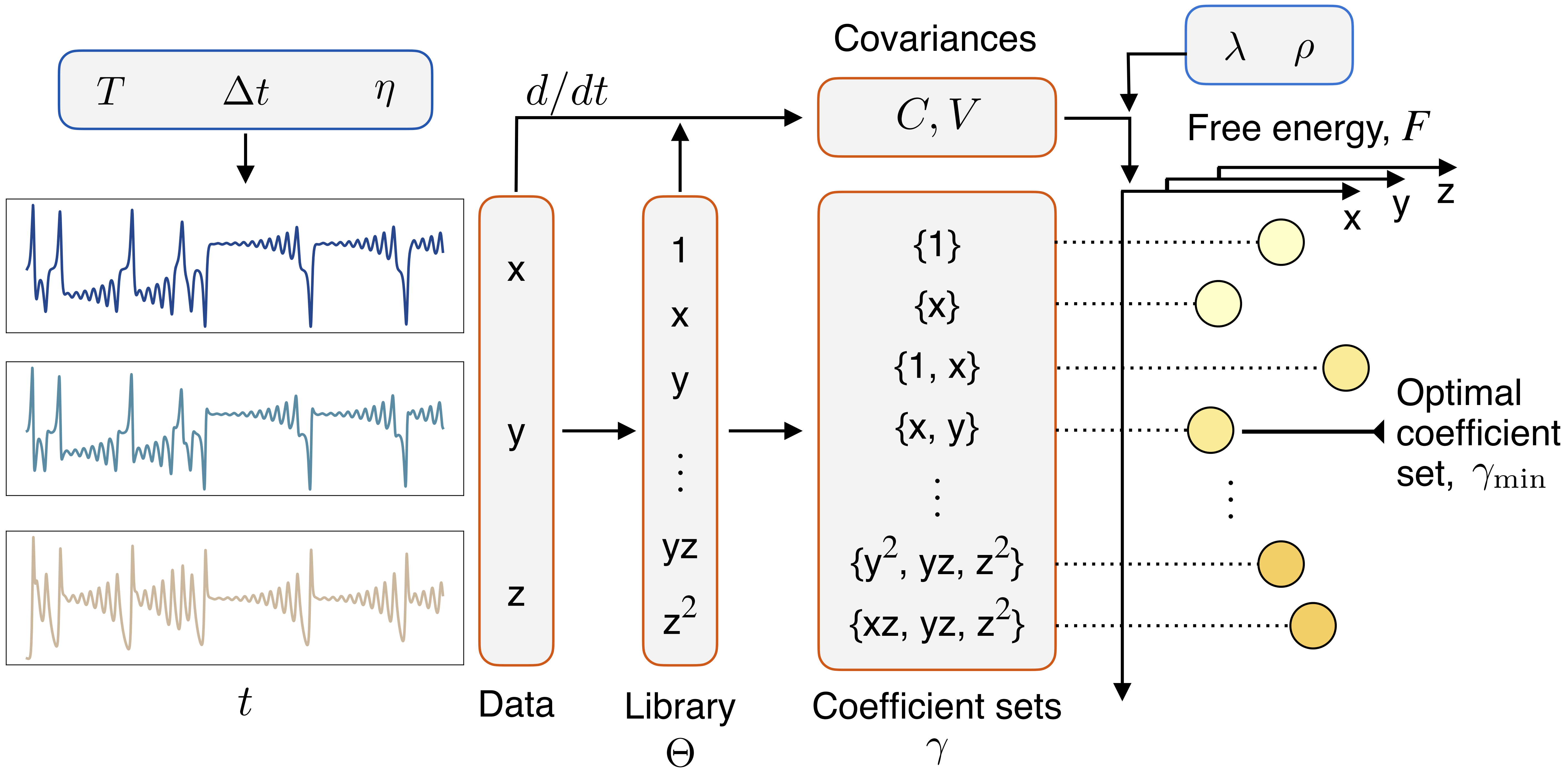
''	'xi_1'	'xi_2'	'xi_3'
'x'	[-9.9996]	[27.9980]	[0]
'y'	[9.9998]	[-0.9997]	[0]
'z'	[0]	[0]	[-2.6665]
'xx'	[0]	[0]	[0]
'xy'	[0]	[0]	[1.0000]
'xz'	[0]	[-0.9999]	[0]
'yy'	[0]	[0]	[0]
...	...	...	...
'zzzzz'	[0]	[0]	[0]

Brunton, Proctor, Kutz, *PNAS (2016)*

Nonlinear System Identification - ZSINDy

- *Closed form* posterior for l_0 SINDy: $$p(\Xi\,|\,x) = \frac{1}{Z}\left(-\frac{1}{2\rho^2}\sum_t (\dot{x} - \Theta^T\Xi)^2 - \Lambda\|\Xi\|_0\right)$$
- Free energy criterion *ranks models*
- Predicts model transition behavior with changing Λ
- Identifies Lotka-Volterra from Mahaffy population dataset





Nonlinear System Identification - ZSINDy

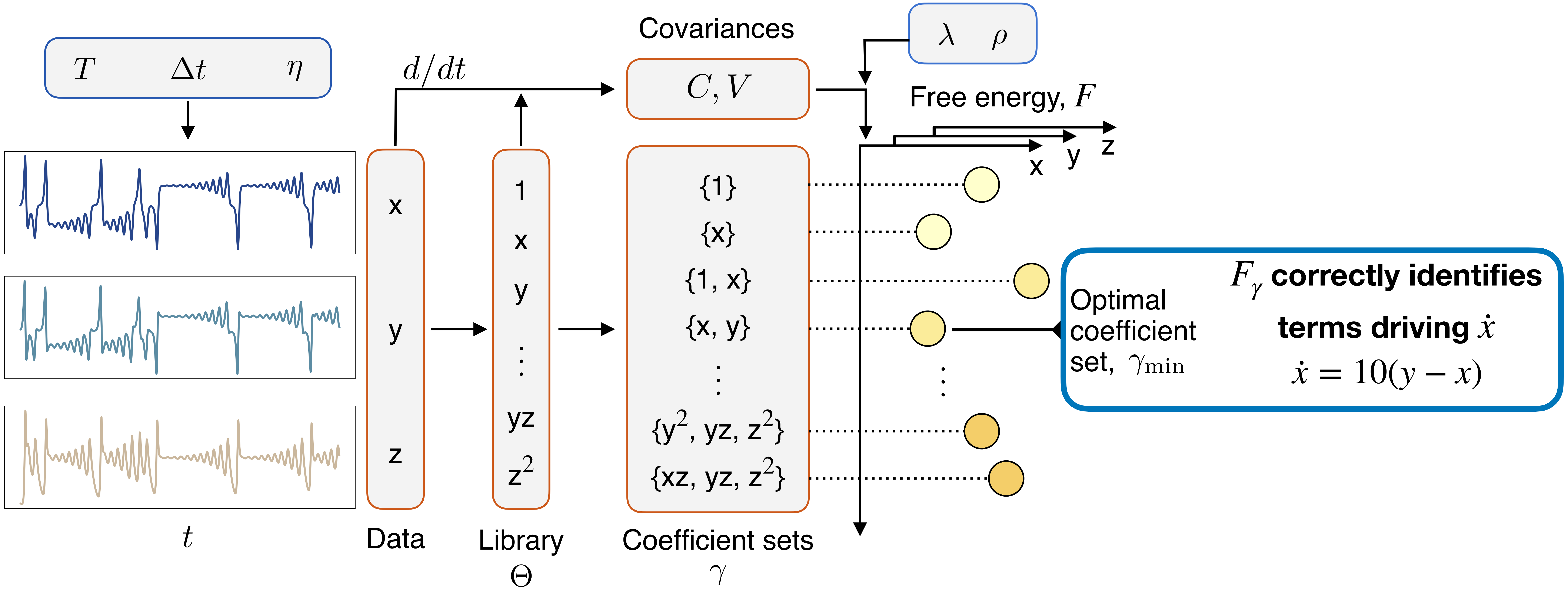
Recast SINDy as Bayesian inference problem, characterize optimization landscape of different coefficient sets γ using statistical mechanics

Statistical "free energy" characterizes optimal set $\gamma_{\min}$

$$F_\gamma = \frac{1}{2\rho^2}\sum_t \dot{x}^2 \underbrace{- \frac{|\gamma|}{2}\ln(2\pi\rho^2) + \frac{1}{2}\ln\det C_\gamma}_{\textit{natural sparsity}} \underbrace{- \frac{1}{2\rho^2} V_{l,\gamma}^T C_\gamma^{-1} V_{l,\gamma}}_{\textit{O(n) LS fit}} + \underbrace{|\gamma| n \cdot \lambda}_{\textit{explicit}}$$

$$p(\Xi\,|\,x) = \frac{\sum_\gamma p(\Xi\,|\,\gamma, x) Z_\gamma}{\sum_\gamma Z_\gamma}$$

Factorized posterior over γ weighted by $Z_\gamma = e^{-F_\gamma}$



Nonlinear System Identification - ZSINDy

Factorized posterior over coefficient sets γ weighted by $Z_\gamma = e^{-F_\gamma}$

$$p(\Xi | x) = \frac{\sum_\gamma p(\Xi | \gamma, x) Z_\gamma}{\sum_\gamma Z_\gamma}$$

Bernoulli-Gaussian prior

$$p(\Xi) = \prod_i \frac{1}{1 + e^{-\Lambda}} (\delta(\Xi_i) + w(\Xi_i) e^{-\Lambda})$$

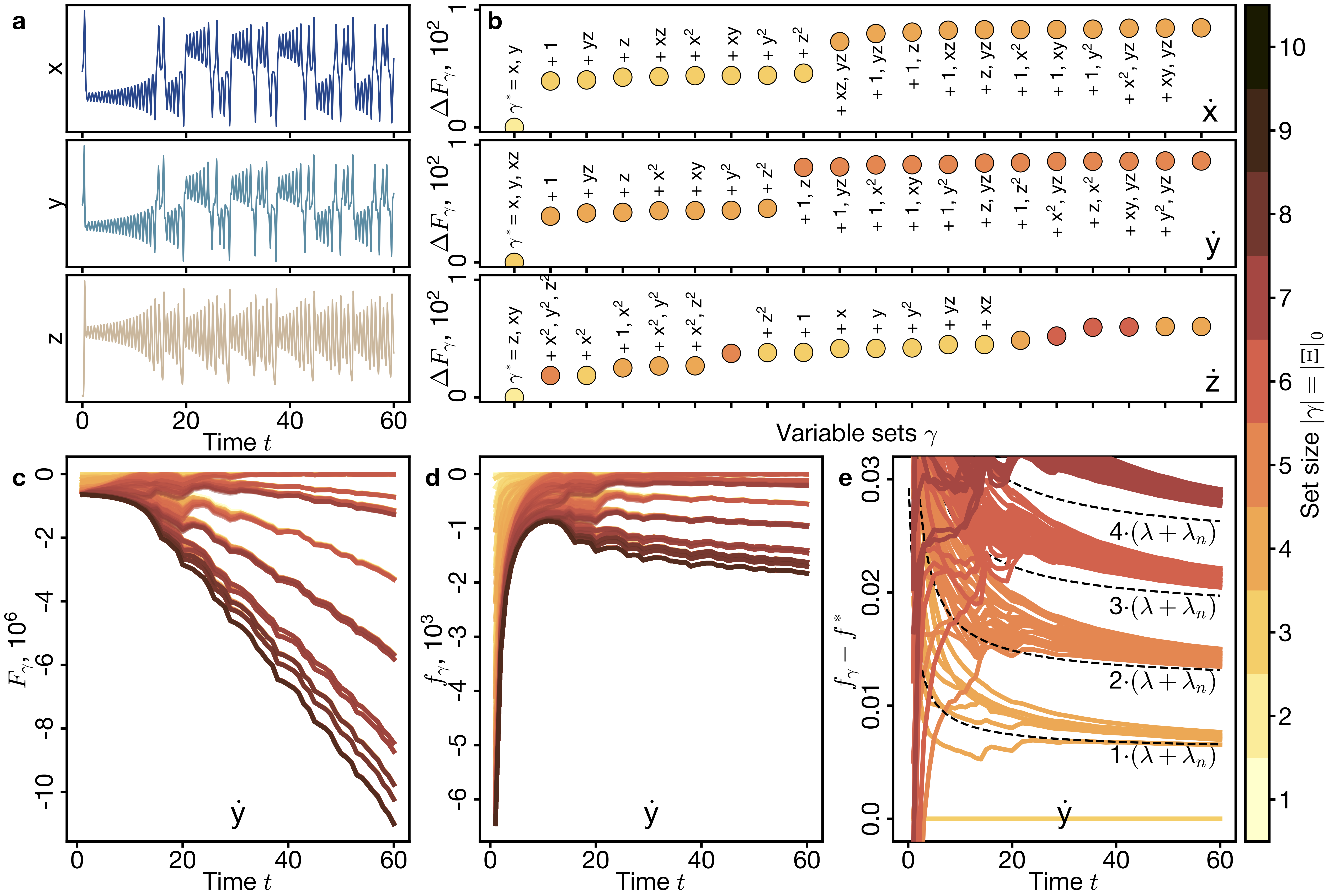
$F_\gamma = -\ln Z_\gamma$**: Minimizer set identifies** l_0 **optimal solution**

$$\boxed{F_\gamma \propto \underbrace{-\frac{|\gamma|}{2} \ln(2\pi\rho^2) + \frac{1}{2} \ln \det C_\gamma}_{natural\ sparsity} - \underbrace{\frac{1}{2\rho^2} V_{l,\gamma}^T C_\gamma^{-1} V_{l,\gamma}}_{O(n)\ LS\ fit} + \underbrace{|\gamma| n \cdot \lambda}_{explicit}}$$

Klishin, Bakarji, Kutz & Manohar, *Phys. Rev. Research 2025*

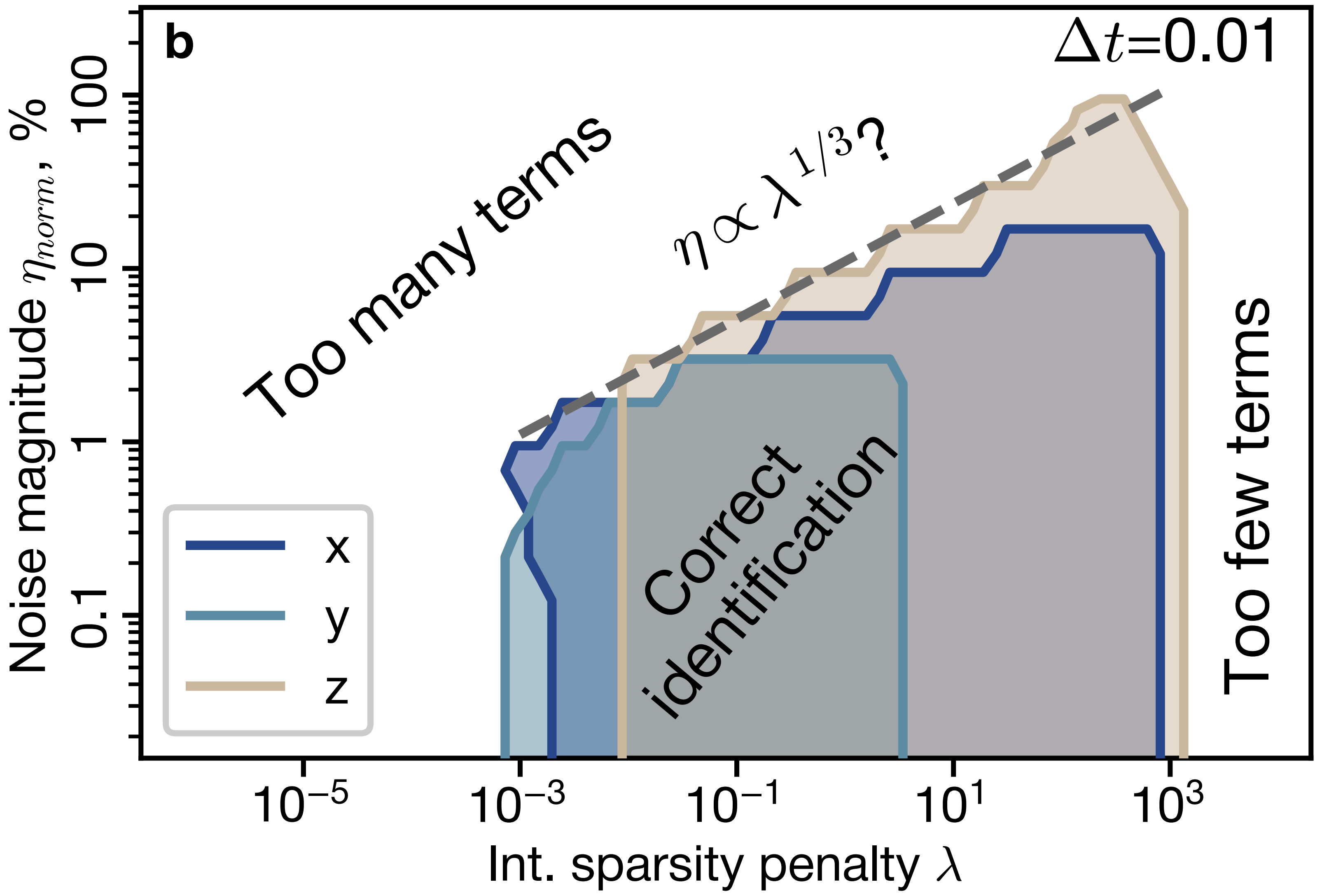
Identification of optimal coefficient set

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$

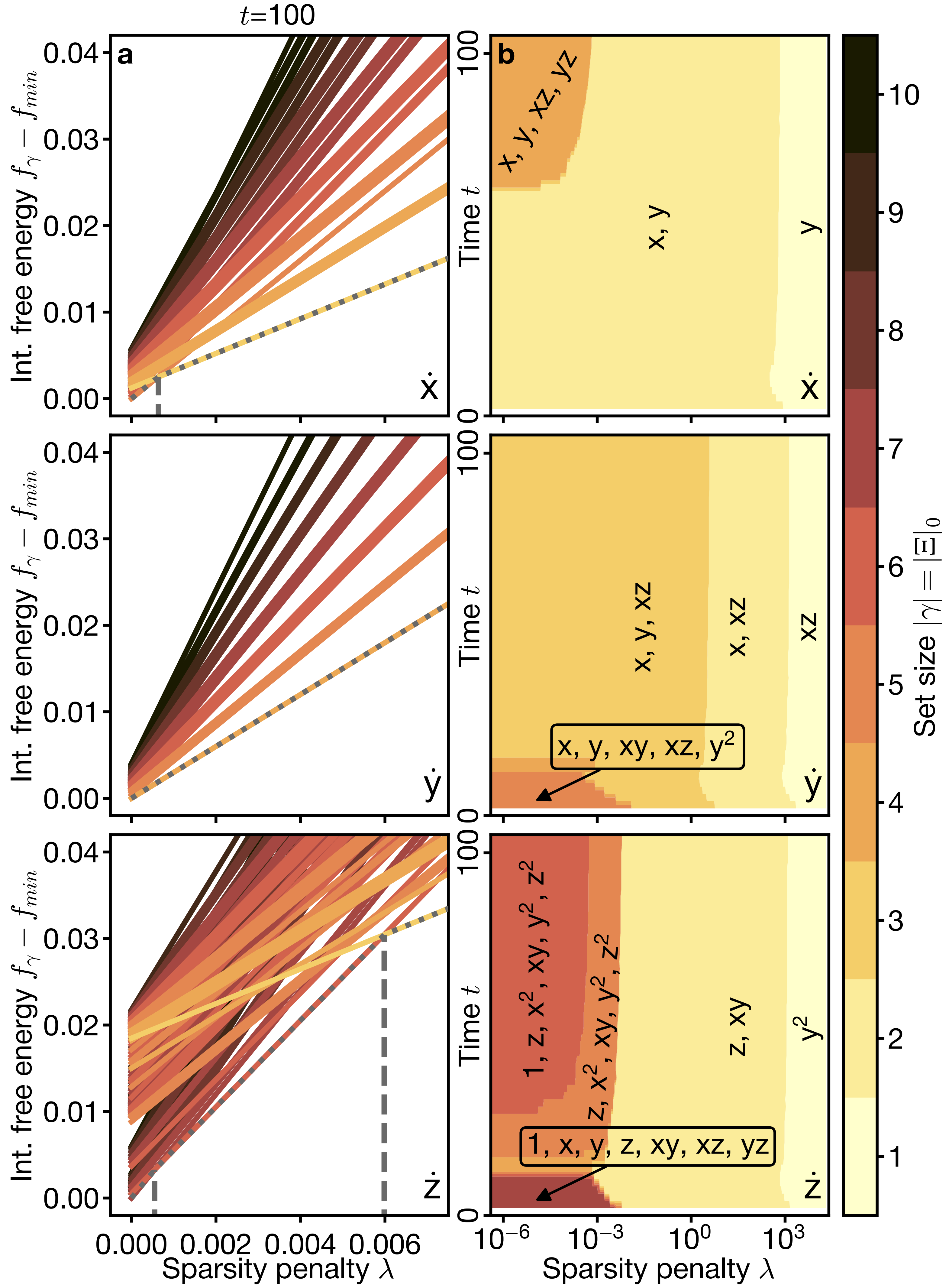


When is inference possible

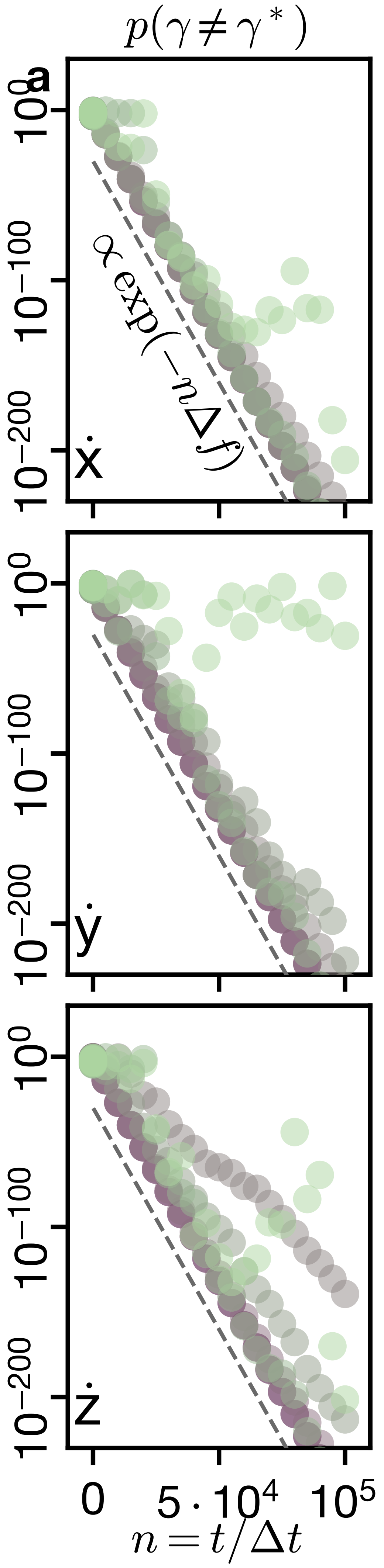
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = x(\rho - z) - y$$
$$\dot{z} = xy - \beta z$$



Critical noise levels for inference



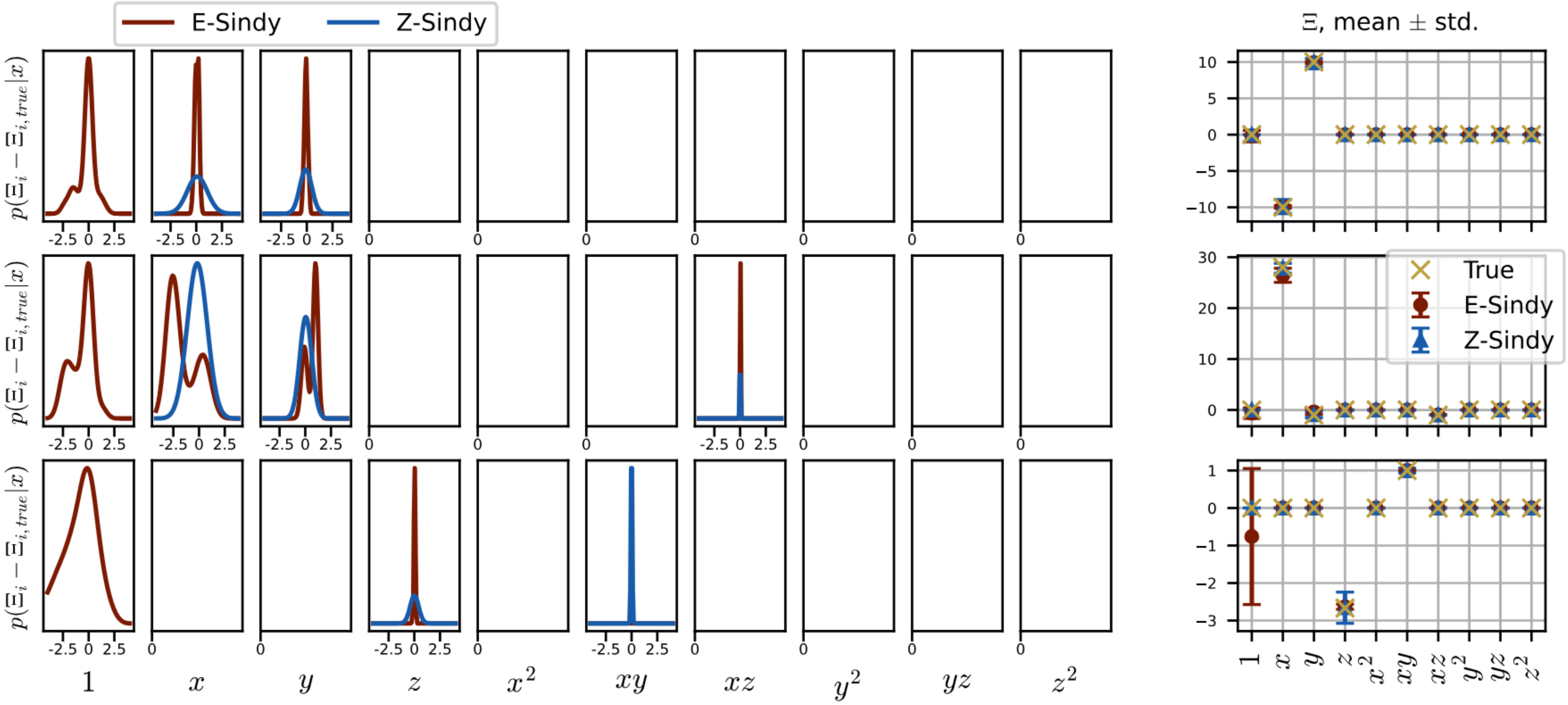
Sharp transitions of coefficient sets with increasing sparsity penalty



Probability of selecting suboptimal set decays *exponentially* w/ trajectory length $p(\gamma^*) \approx 1 - e^{-n\Delta f}$

Comparison with Ensemble SINDy

- E-SINDy fits distributions to bootstrapped ensemble models
- Z-SINDy consistently identifies sparser solutions



Klishin, Bakarji, Kutz & Manohar, *Phys. Rev. Research 2025*

Long-Term Forecasting

- Williams, J. P., Kutz, J. N., & Manohar, K. (2024). Reservoir computing for system identification and predictive control with limited data. *arXiv preprint arXiv:2411.05016*.

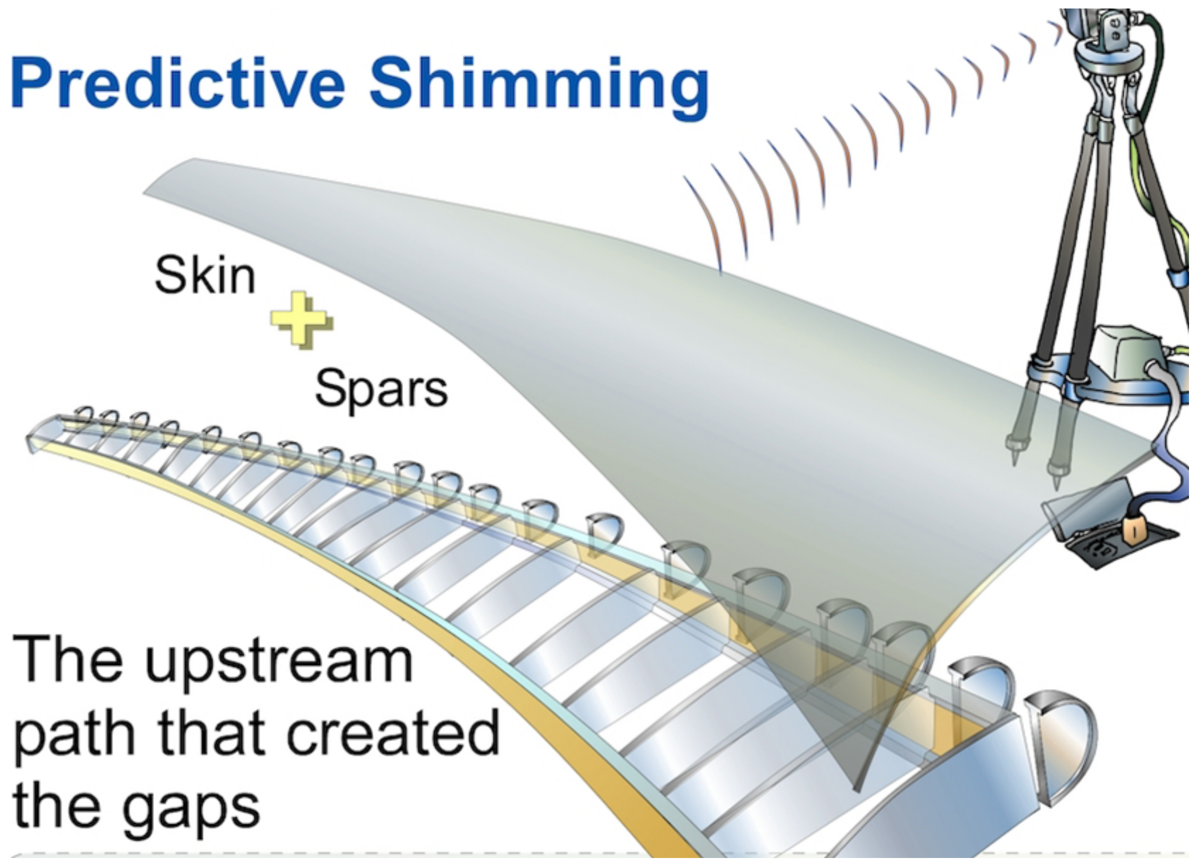
- Klishin, A. A., Bakarji, J., Kutz, J. N., & Manohar, K. (2025). Statistical mechanics of dynamical system identification. *Physical Review Research, 7(3), 033181*.

- Peng, M., Kaptanoglu, A. A., Hansen, C. J., Stevens-Haas, J., Manohar, K., & Brunton, S. L. *(2025)*. Extending the trapping theorem to provide local stability guarantees for quadratically nonlinear models. *Physics of Fluids, 37(10)*.

- Salazar, Manohar, & Banerjee. Online Kernel Dynamic Mode Decomposition for Real-Time Time Series Forecasting with Adaptive Windowing *(In preparation)*

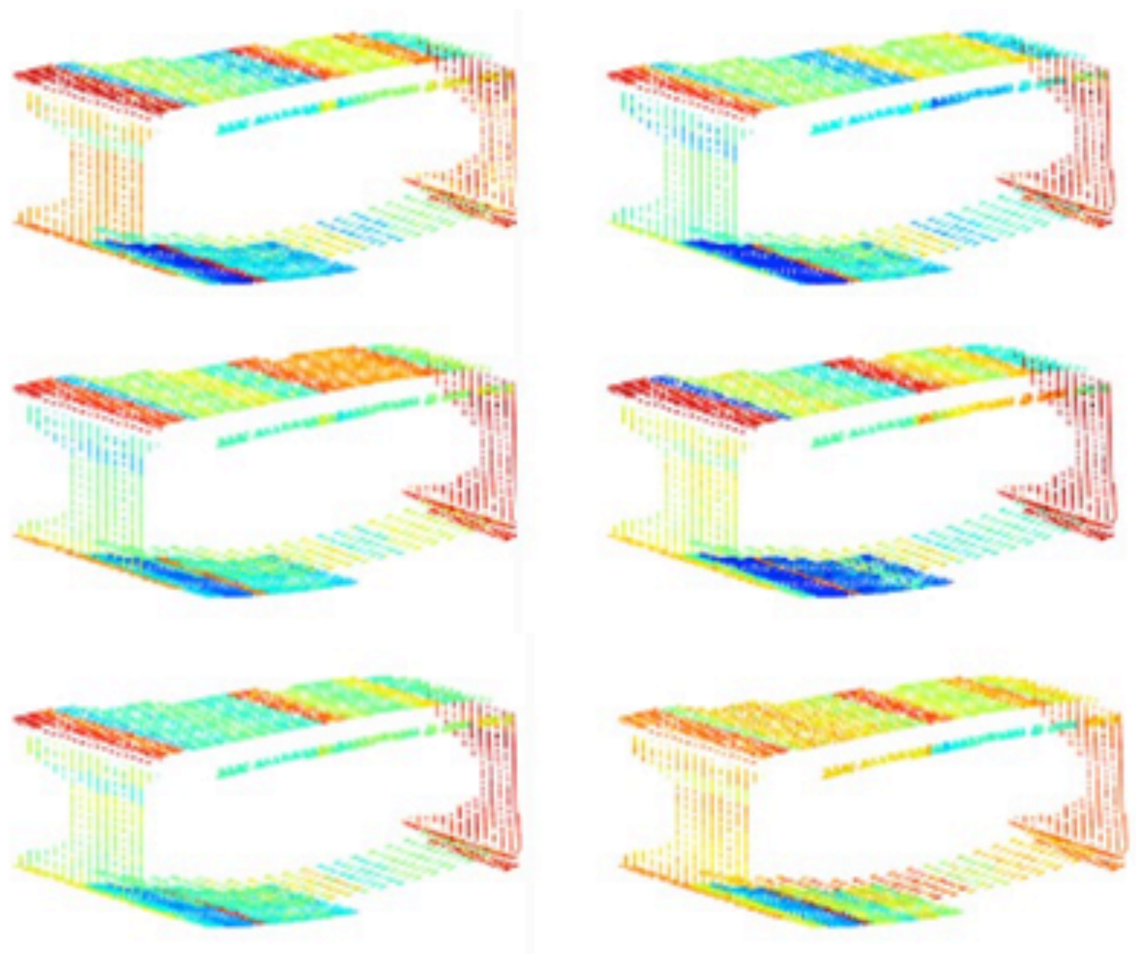
Predictive shimming in large-scale assembly

787 Wing-to-Body Join

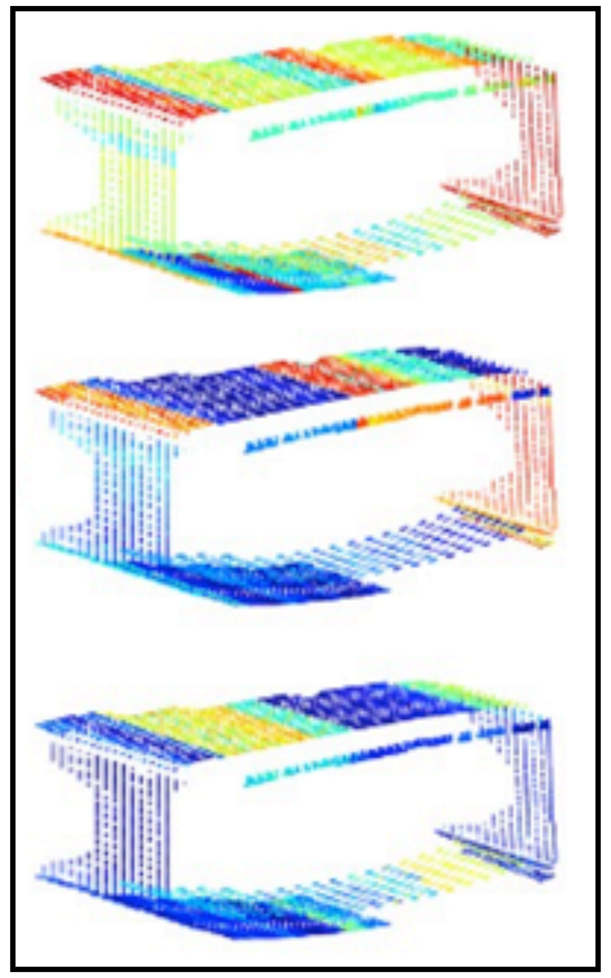


Training data: 10,000 gap measurements (truth data) for 50 aircraft

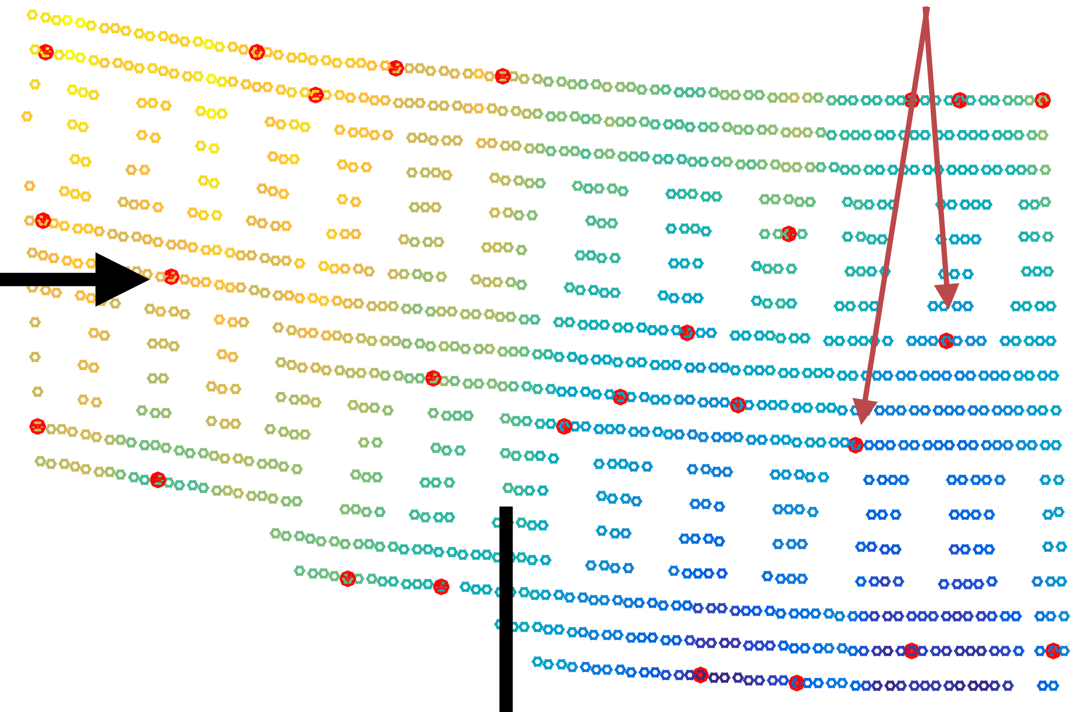
Data from individual aircraft



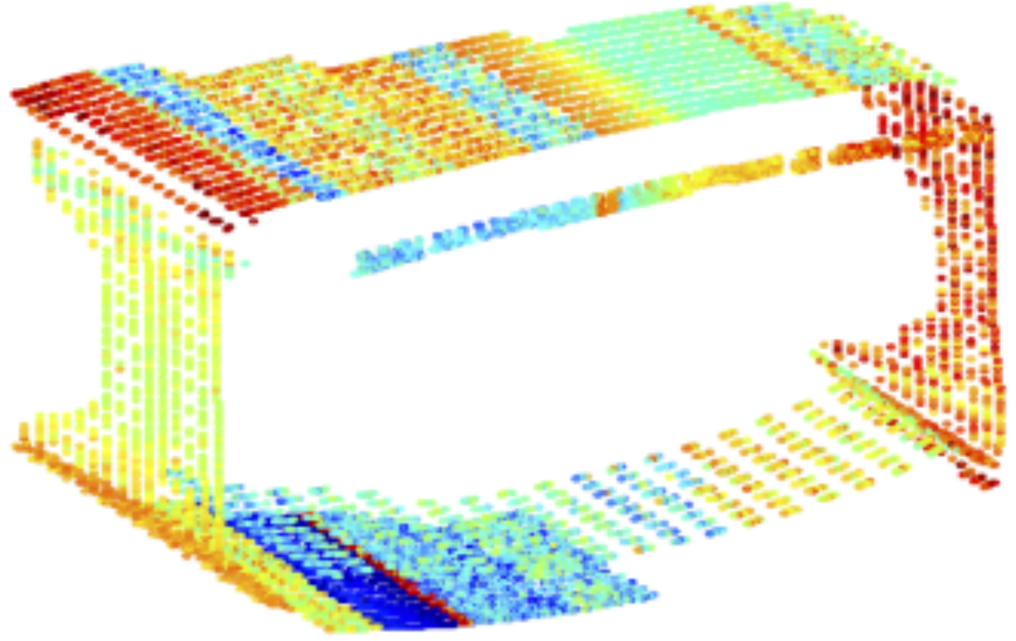
Dominant patterns
Modes



Sensors



Predicted shims

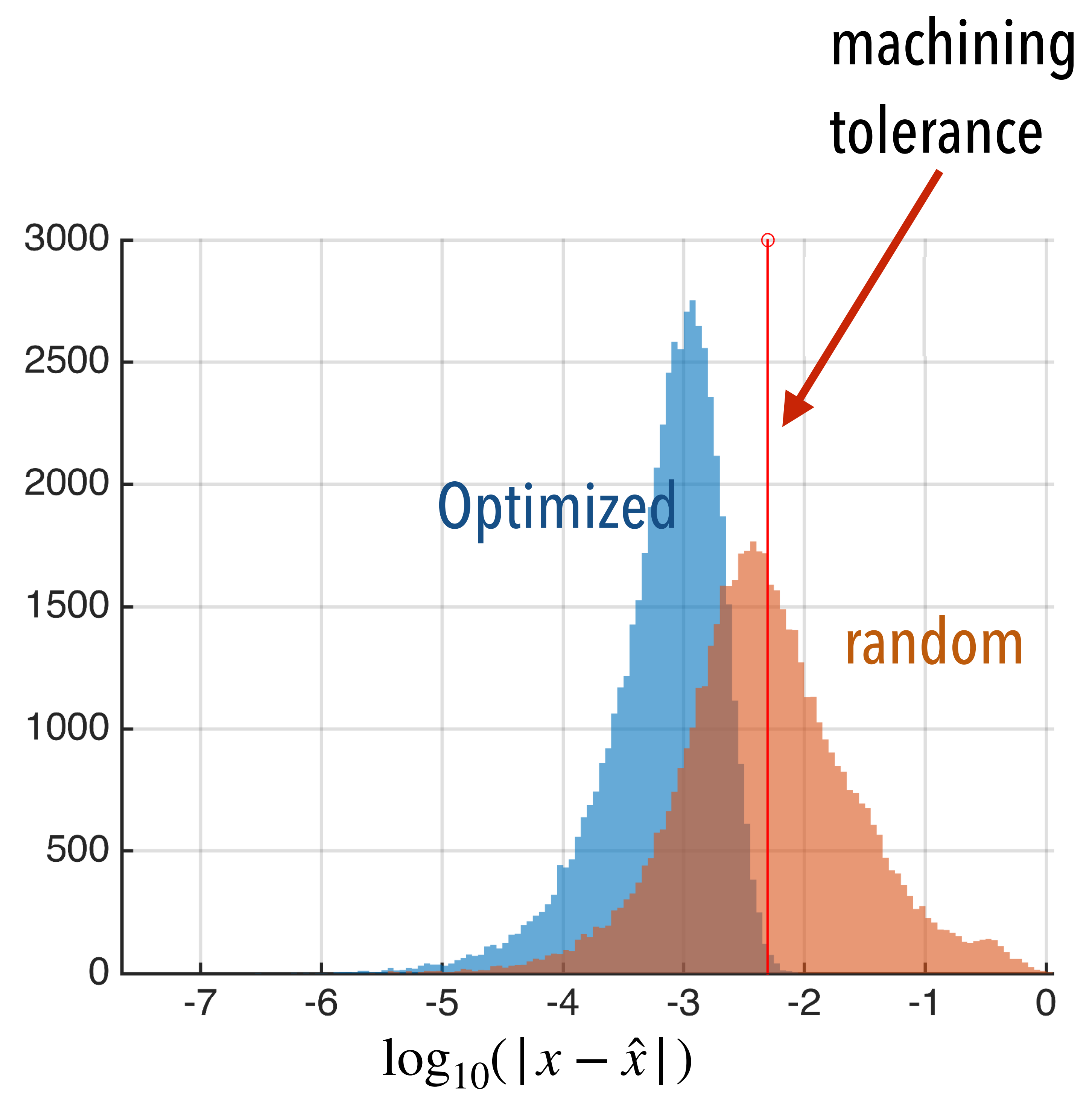


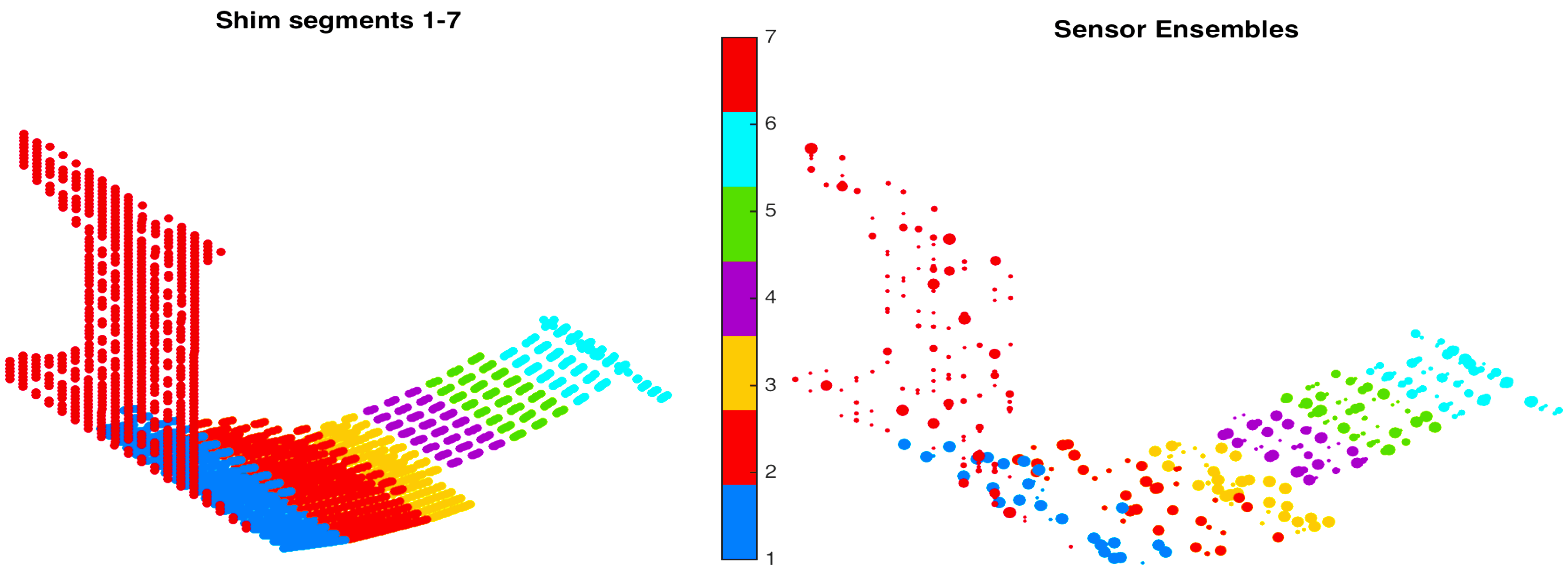
KM, Hogan, Buttrick, Banerjee, Kutz & Brunton, *JMS* (2018)

Predictive shimming in large-scale assembly

99% unknown gaps within 0.005″ with 30X less measurement

Shim No.	1	2	3	4	5	6	7
Percent Accurate	97.90	98.05	99.82	99.94	99.99	99.03	99.97
Optimal sensors (avg)	26	26	25	26	25	26	25
Total points	1003	1116	453	692	709	768	664





KM, Hogan, Buttrick, Banerjee, Kutz & Brunton, *JMS* (2018)