Heritability: a counterfactual perspective

Hongyuan Cao
hcao@fsu.edu

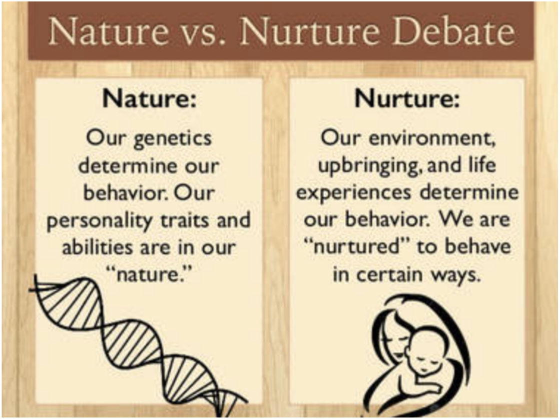
Department of Statistics, Florida State University

Joint work with Haochen Lei, Jieru Shi and Qingyuan Zhao

IMSI long-term program on digital twins
Chicago, IL

December 5, 2025

Outline



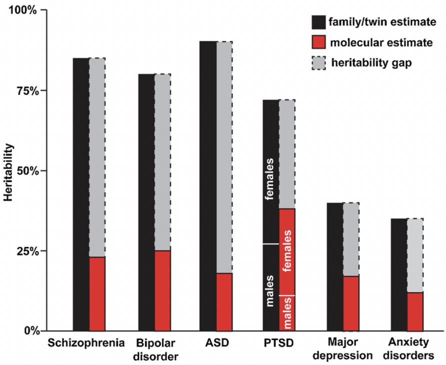
A fundamental question in behavioral, biological, and medical sciences

1.1 Heritability

- A central concept of nature vs nurture is heritability.
- Heritability measures how much genes contribute to complex human traits.
- Parallel to design-based and model-based causal inference: heritability can be derived from twin data and population-based cohort studies.
- Identical twins share 100% of genetics and fraternal twins share 50% of their genetics.
- Twin heritability is defined as twice the difference in the correlation coefficient of identical twins and fraternal twins.
- Population-based narrow-sense heritability accounts for additive genetic effects.
- Population-based broad-sense heritability encompasses both additive and non-additive genetic effects.

1.2 Missing heritability

The plot is adapted from Calker and Serchov (2021).



1.3 Various explanations

Finding the **missing heritability** of complex diseases
TA Manolio, FS Collins, NJ Cox, DB Goldstein… - Nature, 2009 - nature.com
… many to question how the remaining, '**missing**' **heritability** can be explained. Here we examine potential sources of **missing heritability** and propose research strategies, including and …
☆ Save 99 Cite Cited by 9806 Related articles All 33 versions Web of Science: 5614

The mystery of **missing heritability**: Genetic interactions create phantom **heritability**
O Zuk, E Hechter, SR Sunyaev… - Proceedings of the …, 2012 - National Acad Sciences
… The prevailing view has been that the explanation for **missing heritability** lies in the … **missing heritability** could arise from overestimation of the denominator, creating "phantom **heritability**…
☆ Save 99 Cite Cited by 1940 Related articles All 14 versions Web of Science: 980

Missing heritability may be hiding in repeats
M Gymrek, A Goren - Science, 2021 - science.org
… still fail to fully explain the **heritability** of most traits analyzed, even as sample sizes have reached hundreds of thousands of people, a problem dubbed the "**missing heritability**" (5). …
☆ Save 99 Cite Cited by 11 Related articles All 4 versions Web of Science: 5

Genetic nurturing, **missing heritability**, and causal analysis in genetic statistics
H Shen, MW Feldman - … of the national academy of sciences, 2020 - National Acad Sciences
… Both categories can contribute to **missing heritability** and … genetic nurturing to the **missing heritability** problem (27–29). … our understanding of the **missing heritability** problem. We also …
☆ Save 99 Cite Cited by 20 Related articles All 6 versions Web of Science: 12

2.1 Definition

- Let A denote whether certain gene is passed to the children and Y denote the phenotype.
- Denote $P(A = 1) = p$ and $q = 1 - p$.
- Suppose we use *CRISPR* to edit genes.
- $Y(0)$ is the potential outcome of the phenotype when $A = 0$ and $Y(1)$ is the potential outcome of the phenotype when $A = 1$.
- Let A' be a random variable and $A' \overset{d}{=} A$.
- Assume $A \perp\!\!\!\perp A' \perp\!\!\!\perp \{Y(0), Y(1)\}$
- The causal heritability is defined as

$$h^2 = \frac{\mathrm{Var}\{Y(A) - Y(A')\}}{2\mathrm{Var}(Y(A))}. \tag{1}$$

2.2 Interpreting h^2

Numerator

$$\begin{aligned} & \mathrm{Var}\{Y(A) - Y(A')\} = E\{\mathrm{Var}(Y(A) - Y(A') \mid Y(0), Y(1))\} \\ + \quad & \mathrm{Var}\{E(Y(A) - Y(A') \mid Y(0), Y(1))\} \\ = \quad & 2p(1-p)E\{(Y(1) - Y(0))^2\}. \end{aligned}$$

Denominator

$$\begin{aligned} \mathrm{Var}(Y(A)) &= E(\mathrm{Var}(Y(A) \mid Y(0), Y(1))) + \mathrm{Var}(E(Y(A) \mid Y(0), Y(1))) \\ &= p(1-p)E((Y(0) - Y(1))^2) + \mathrm{Var}(pY(1) + (1-p)Y(0)). \end{aligned}$$

- $pY(1) + (1-p)Y(0) = Y(0) + p\tau$, where $Y(1) - Y(0) = \tau$.


Thank You
and
Questions

Digital Twin for Determining the Functional Significance of Coronary Artery Lesions

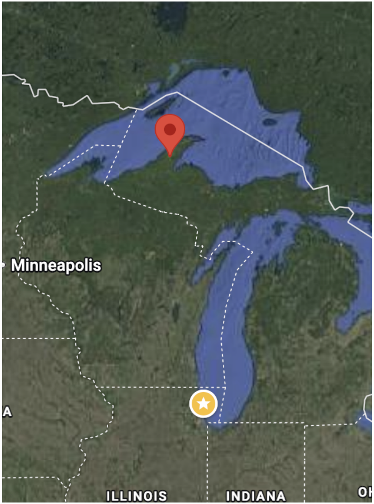
Jiguang Sun

Michigan Technological University

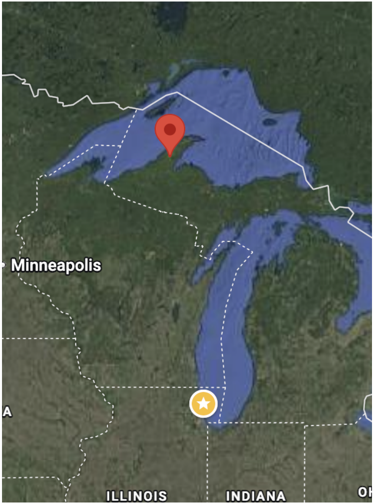
Collaborator: Dr. Weihua Zhou

Digital Twins, IMSI, Fall 2025.

Houghton, MI



Houghton, MI



Research Interests

- Numerical Methods for PDE Eigenvalue Problems
  - Linear Eigenvalue Problems
  - Band Structures of Photonic Crystals
  - Scattering Resonances
- Inverse Problems
  - Inverse Scattering Problems
  - Inverse Spectral Problems
  - Bayesian Inversion
- Through-wall Imaging
- Diagnostic of Coronary Artery Disease (CAD)

Coronary Artery Disease (CAD)

- The gold standard for diagnosing CAD is invasive coronary angiography (ICA) - an X-ray procedure that uses a catheter to diagnose coronary artery disease by visualizing blood flow.

Figure: Coronary revascularization (CR) - Google.

- Fractional flow reserve (FFR) is a medical diagnostic tool used to measure the severity of a blockage in a coronary artery - comparing the blood pressure and flow in a specific part of the artery against the theoretical maximum blood flow of a normal artery. An FFR less than 0.80 is often considered abnormal.

Project Overview

multi-view ICA videos $\longrightarrow$ FFR $\longrightarrow$ clinical decision-making of CR

Goals:

1. understanding the mapping from "contrast dye" to images
   - blood flow
   - artery boundary - blockage/narrowing
   - dynamic motion of the heart
   - angles of the images
2. reconstructing FFR from the images.
   - boundaries
   - initial data
3. making decision based on FFR + ?
   - patient specific parameters

Tools

1. Finite element method for NS equations [Canic et al. 25]:

$$
\begin{aligned}
&\frac{\partial A}{\partial t} + \frac{\partial Q}{\partial z} = 0,\\
&\frac{\partial Q}{\partial t} + \frac{\partial}{\partial z}\left[(a+\alpha_c)\frac{Q^2}{A}\right] + \frac{A}{\rho_f}\frac{\partial p}{\partial z} = -2(\gamma+2)\nu\frac{Q}{A} + \frac{Q^2}{A}\frac{\partial \alpha_c}{\partial z},\\
&p = p_{ext} + \frac{hE}{R_0^2(1-\sigma^2)}(R-R_0) + (\gamma+2)\rho_f\nu\frac{Q}{A}\frac{\partial \ln R}{\partial z}.
\end{aligned}
$$

2. Operator learning for parametrized NS equations, e.g., boundary to pressure.
3. Reduced order methods.
4. Bayesian inverse problems, e.g., image to boundary

A Data-Driven Digital Twin Framework for Non-Stationary Complex Systems

Ivan Sudakow (The Open University, UK)

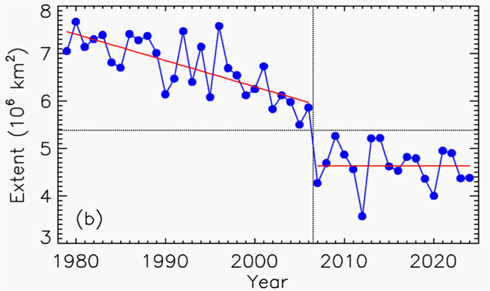
Dmitri Kondrashov (UCLA)

The Challenge: Modeling Systems Under Regime Shift

The Complex System:

- Arctic Sea Ice is characterized by high-dimensional chaos and strong atmosphere-ocean coupling.
- **The Problem:** A distinct dynamical **Regime Shift** occurred ≈ 2007.

September Sea Ice Extent



Digital Twin Requirements:

- Physics-based models (GCMs) struggle with sub-grid parameterization and computational cost.

The Encoder: Data-Adaptive Harmonic Decomposition (DAHD)

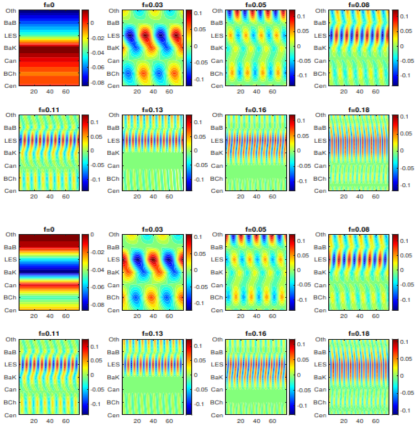
Spectral Decomposition of the State Space:

- We utilize time-embedding to compute a spectral decomposition of the system's correlation matrix.
- This provides a finite-dimensional approximation of the **Koopman Operator** for nonlinear systems.

Feature Extraction:

- The system is decomposed into d pairs of **Data-Adaptive Harmonic Modes (DAHMs)**.
- Modes appear in **phase-quadrature** (sin/cos pairs), acting as narrow-band filters isolating specific temporal scales.

The Encoder: Data-Adaptive Harmonic Decomposition (DAHD)



Space-time patterns of data-adaptive harmonic mode pair for largest λ at selected frequencies f (cycle/week), see left, x-axis – time (weeks), y-axis – Arctic regions: units – non-dimensional.

The Generative Engine: Coupled Stochastic Oscillators

Reduced Order Modeling (ROM): We model the evolution of the mode coefficients $z(t)$ using coupled **Stuart-Landau (SL)** oscillators.

The Stochastic Generative Model

$$\dot{z} = (\mu + i\gamma)z - (1 + i\beta)|z|^2 z + \eta_t$$

- **Interpretable Physics:**
  - μ, γ: Linear growth rate and frequency.
  - β: Nonlinear amplitude saturation (anharmonicity).
- **The "Twin" Aspect:**
  - Parameters are estimated via inverse modeling (regression with linear constraints).
  - η_t (Noise): Captures **unresolved sub-grid scales** and maintains variance.

Simulation: Trajectory Reconstruction & Uncertainty

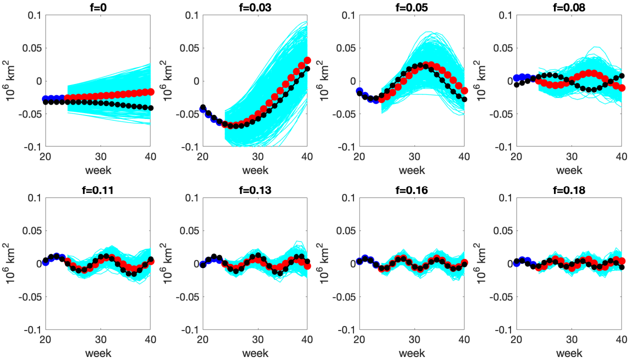
Probabilistic Digital Twin Simulation (2024 Case Study):

(blue) observed trajectory (ground truth).

(black) observed trajectory at the end of summer

(cyan) the digital twin's stochastic ensemble.

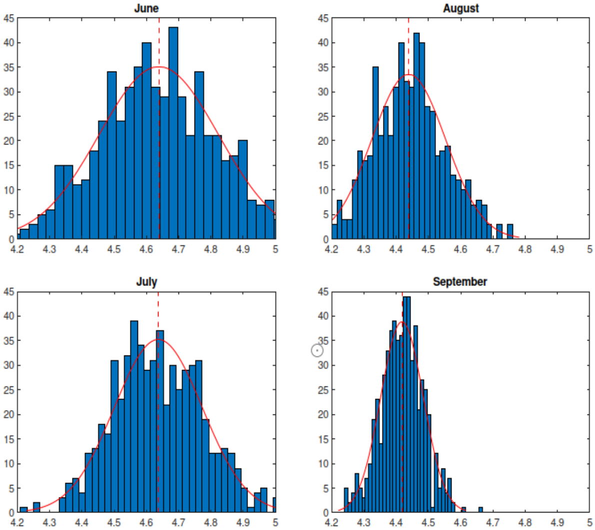
(red) ensemble mean.



Key Result: The Twin accurately reproduces the **phase-space trajectory** and **amplitude modulations** of the chaotic system, providing rigorous uncertainty quantification.

Blind Prediction Performance:

- Target: September 2024 Pan-Arctic Sea Ice Extent.
- Accuracy: Ensemble mean aligns closely with observation.
- Error: ≈ 0.3 Mkm^2 (Competitive with physics-based GCMs).



Summary:

- DAHD-SL constitutes a rigorous, lightweight **Digital Twin**.
- Efficiently bridges Data

IMSI Long Program Report-Out: Digital Twins, Housing Analytics, and Collaborations

Nancy L. Glenn Griesinger, Ph.D.
Research Professor and Statistician
Huston–Tillotson University
Austin, Texas

December 5, 2025

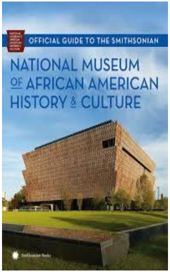
Background

N. L. Glenn Griesinger, Ph.D.

HUSTON-TILLOTSON UNIVERSITY
1875 AUSTIN, TEXAS

TEXAS COMMISSION ON ENVIRONMENTAL QUALITY

- **Research Expertise: Statistical Methodologist**
- **Nonparametric Statistics**
- Nuclear Physics, Brookhaven National Laboratory (2024)
- Computational Chemistry Software Development, Ames National Laboratory (2025)
- U. S. Department of Housing and Urban Development
- **Active Grant – HUD** Artificial Intelligence for Housing Infrastructure (2024 -)
- **Grant –** Texas Commission on Environmental Quality (2019 - 2024)
- **Elementary Statistics: A Guide to Data Analysis Using R** (2023)



Concept

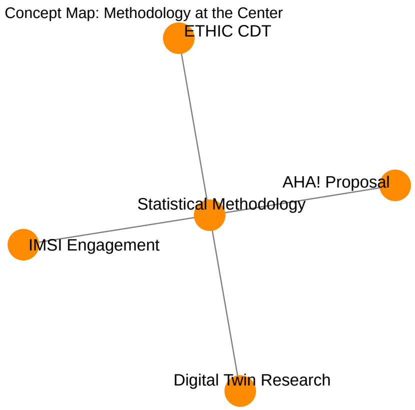


Figure: Concept Map for IMSI Projects

Overview

- ETHIC (Enabling Technologies for Housing Innovation Center) CDT architecture and collaborations
- Digital Twin theoretical research
- AHA! Affordable Housing Analytics proposal
- IMSI workshops, seminars, and standards work

KPI 1: AHA! Affordable Housing Analytics

- Proposal integrating statistical modeling, geospatial ML, digital twins
- Dynamic Opportunity Index using real-time data assimilation
- Developed during IMSI Digital Twin Workshop 1

Workshop contributions:

- Functional data analysis
- Scenario-based forecasting
- Spatial–temporal modeling

KPI 2: Digital Twin Research Supporting DOE Missions

- Real-time modeling & uncertainty quantification
- Scalable statistical DT architecture
- Integration of AI with physical systems
- Supports nomination for 2026 Enrico Fermi Presidential Award

KPI 3: IMSI Research Seminars

- Omar Ghattas: GPU Bayesian inversion for real-time prediction
- Georg Stadler: Extreme-event modeling via PDE optimization

Applications:

- Urban resilience modeling
- Sensor placement optimization
- Housing/infrastructure forecasting

KPI 4: ETHIC Digital Twin

- Student contractor supporting ML/AI for ETHIC RQ1
- Reproducible CDT architecture (R, Python, SQL, Docker)
- Dynamic systems: Lorenz96 + Kalman filtering

KPI 4: Prototype v0.1.8 System Architecture

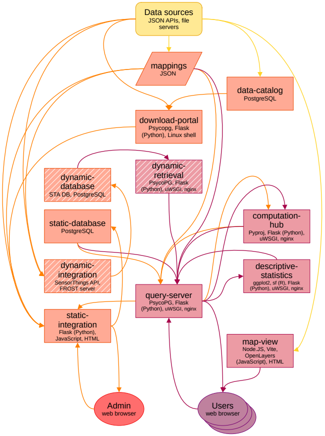
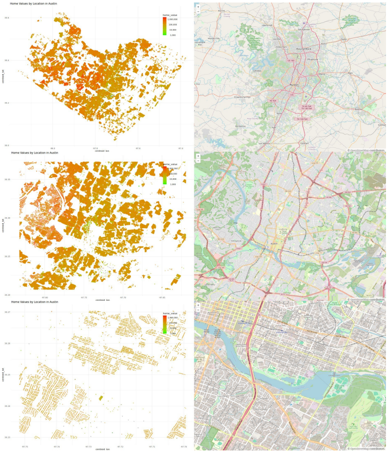


Figure: System Architecture

KPI 4: Prototype

Scatter plot



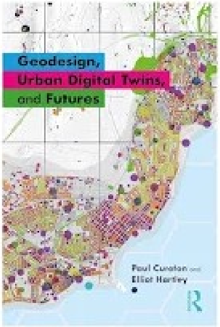
KPI 4A: ETHIC Q3 Report

- Updated H.AI (Housing Access Index)
- Added scenario-based forecasting
- Functional data analysis + data assimilation
- Started Geodesign & Ethics DT group

KPI 5: National Digital Twin Standards

- Joined Digital Twin Consortium Working Group 1
- Contributing to:
  - Data requirements
  - Interoperability standards
  - Built-environment use cases

KPI 6: Geodesign Discussion Group



- Reading *Geodesign, Urban Digital Twins and Futures*
- Professor Gerard Awanou, University of Illinois – Chicago
- Integrating ethics + governance perspectives
- Strengthened CDT community learning

KPIs 7–15: Additional Achievements

- New collaborations (e.g., Sina Ansari, DePaul)
- DOE computational chemistry work with Siwei Luo
- AAAS S&T Policy Fellowship (HUD track)
- Connections: City of Chicago Housing, DePaul, Mansueto
- Attended all IMSI workshops + NISS Conference
- Presented DT updates to HUD ETHIC team (Oct 16)

Summary

- Digital Twin architecture advancements
- Housing analytics innovation (H.AI + AHA! proposal)
- Real-time modeling and extreme-event methods
- ML + geospatial integration
- National DT standards contributions

Acknowledgments

- Work performed while visiting IMSI, supported by NSF DMS-2425650
- Supported by HUD HSI Center of Excellence (ETHIC)

Some Thoughts on Scalable Migration Digital Twin Modeling

Li-Hsiang Lin
Georgia State University

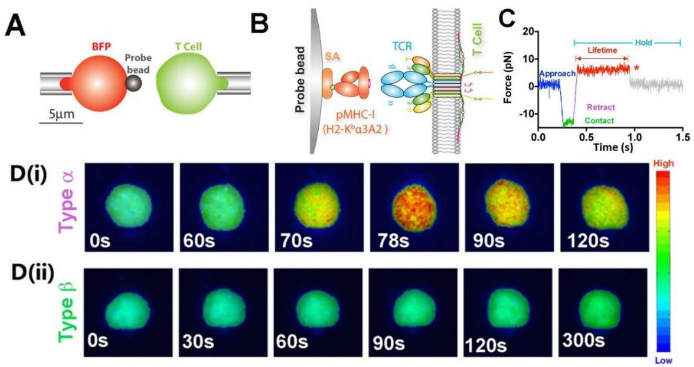
December 5 2025

Long Program on Digital Twin

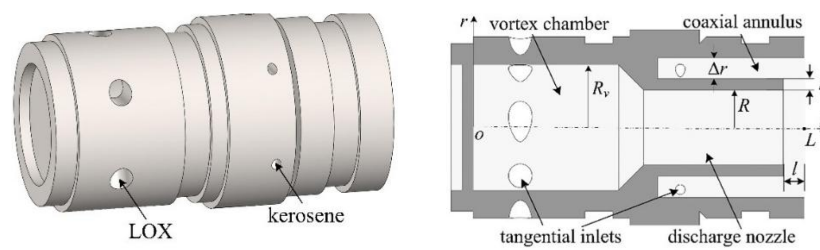
Institute of Mathematical and Statistical Innovation

Georgia State University

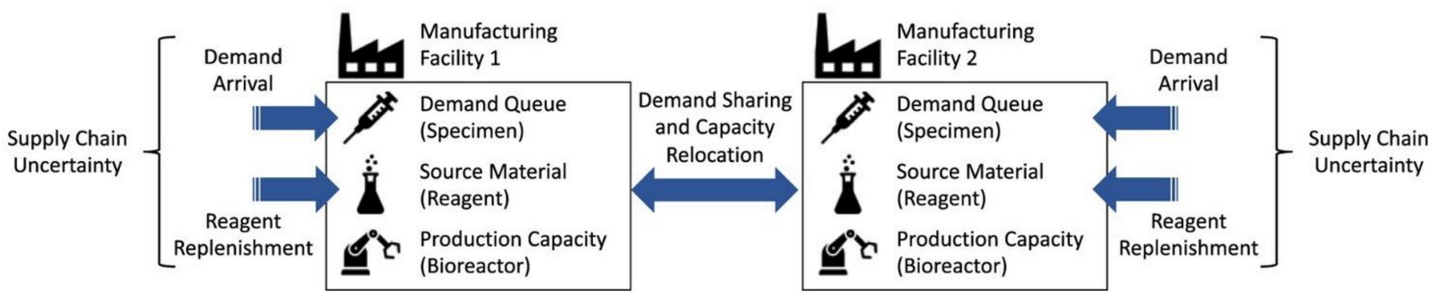
My Previous Research



Single Molecular Experiments



Rocket Engine Design Experiments

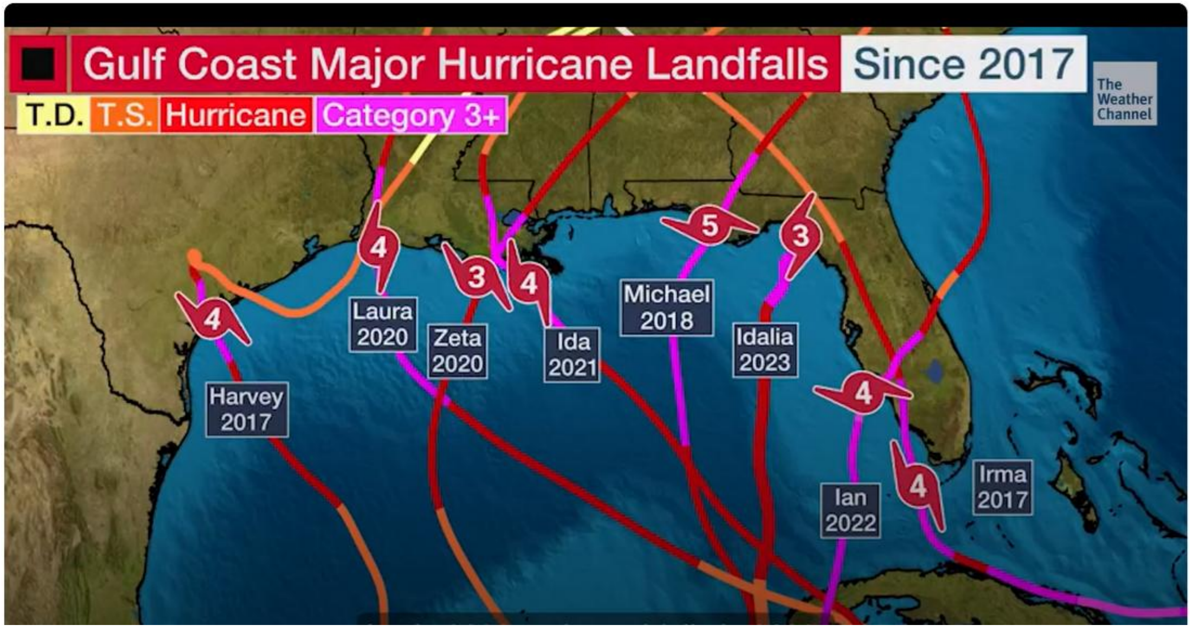


Medical Supply Chain Experiments

My Previous Research (Cont.)

- **Computer Experiments (CEs)** are expensive and/or requires quantify prediction uncertainty
  - ➢Surrogate Models: For example, Transformation and Additive Gaussian Process (Lin and Joseph, Technometrics, 2020)
- **Physical Experiments** are often integrated with CEs to interact and ` ` calibrate" computer models
  - ➢ Simulator Selection: Hung, Lin, and Wu, JASA, 2023

Motivation of Dynamic Modeling



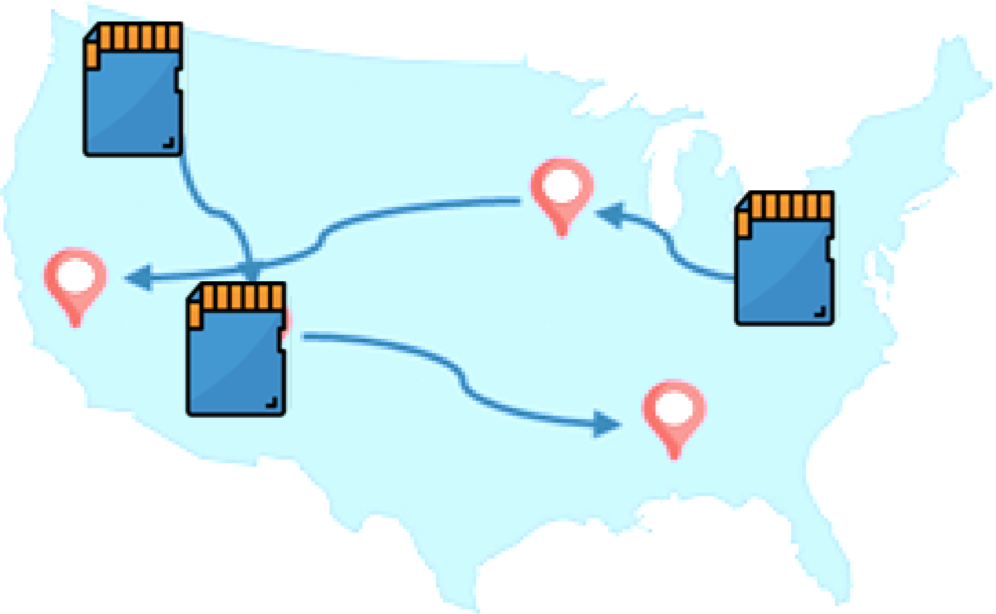
Picture from The Weather Channel https://weather.com

↓Picture from National Weather Services https://www.weather.gov

Georgia State University

Real Migration Experiments

- Communication Constraints



Scalable and Communication-Efficient Varying Coefficient Mixed Effect Models

- Find a statistical efficient estimator under “morden data constraint”, like the communication constraint

$$\max_{\boldsymbol{\beta}(\mathbf{h}),\boldsymbol{\alpha},\boldsymbol{\eta}} \quad \ell_{\text{joint}}(\boldsymbol{\beta}(\mathbf{h}),\boldsymbol{\alpha},\boldsymbol{\eta}) \quad \text{subject to} \quad \text{total bits communicated} \leq cdk,$$

- Show a proposed estimator

$$\boldsymbol{\theta}^{(1)} = \boldsymbol{\theta}_0 - \mathbf{K}_1^{-1}\mathbf{g}(\boldsymbol{\theta}_0).$$

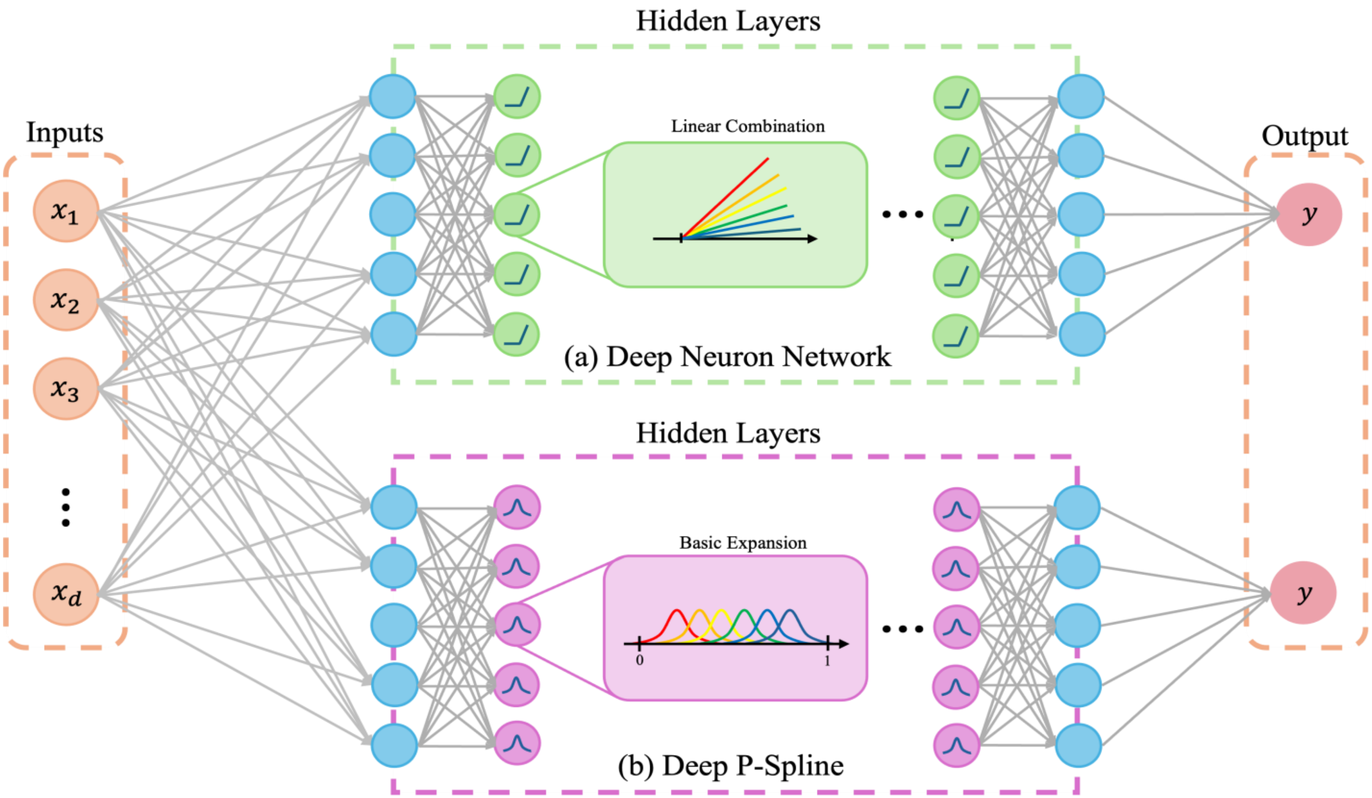
Pilot Estimator, such as Estimator from One Node

Achieve the same first-order asymptotic efficiency as the maximum likelihood estimator with no constraint.

- More details can be found in Chalangar Jalili Dehkharghani, L., and Lin (2025+)

Chalangar Jalili Dehkharghani, L., & Lin, L.-H. (2025). *Scalable and Communication-Efficient Varying Coefficient Mixed Effect Models: Methodology, Theory, and Applications*. arXiv:2511.12732. https://doi.org/10.48550/arXiv.2511.12732

Deep P-Spline: Surrogate Model and Fast Tuning



Objective Function:

$$\sum_{i=1}^{n}\left(y_i - f_{\mathbf{w}_{(1)},\cdots,\mathbf{w}_{(L-1)},\mathbf{w}_{(L)}}(\mathbf{x}_i)\right)^2 + \sum_{\ell=2}^{L-1}\lambda_{(\ell)}\left(\sum_{j=1}^{P(\ell)}||\mathbf{D}_{(\ell),r}\boldsymbol{\omega}_{j,(\ell)}||^2\right)$$

Difference Penalty:

$$\mathbf{D}_{(\ell),1} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ & \ddots & & & \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}; \quad \mathbf{D}_{(\ell),2} = \begin{bmatrix} -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ & \ddots & & & & \\ 0 & \cdots & 0 & -1 & 2 & -1 \end{bmatrix}$$

Deep P-Spline: Surrogate Model and Fast Tuning (II)

A Fast Tuning Method:

- Find priors for each layer weight coefficients
- Connect with *Expected Maximum Likelihood* Algorithm for Structure Selection and Network Fitting

$$\mathbf{y}_{i,(\ell)} = h_{(\ell)}\left(\mathbf{z}_{(\ell-1)}; \mathbf{W}_{(\ell)}\right) + \epsilon_{i,(\ell)}, \quad \ell = 1, 2, \ldots, L-1,$$

$$\text{where} \quad \mathbf{z}_{(\ell-1)} = h_{(\ell-1)}\left(\mathbf{z}_{(\ell-2)}; \mathbf{W}_{(\ell-1)}\right), \quad \mathbf{z}_{(0)} = \{\mathbf{x}_i\}_{i=1}^{n},$$

$$\mathbf{y}_{(L)} = h_{(L)}\left(\mathbf{z}_{(L-1)}; \mathbf{W}_{(L)}\right) + \epsilon_{(L)},$$

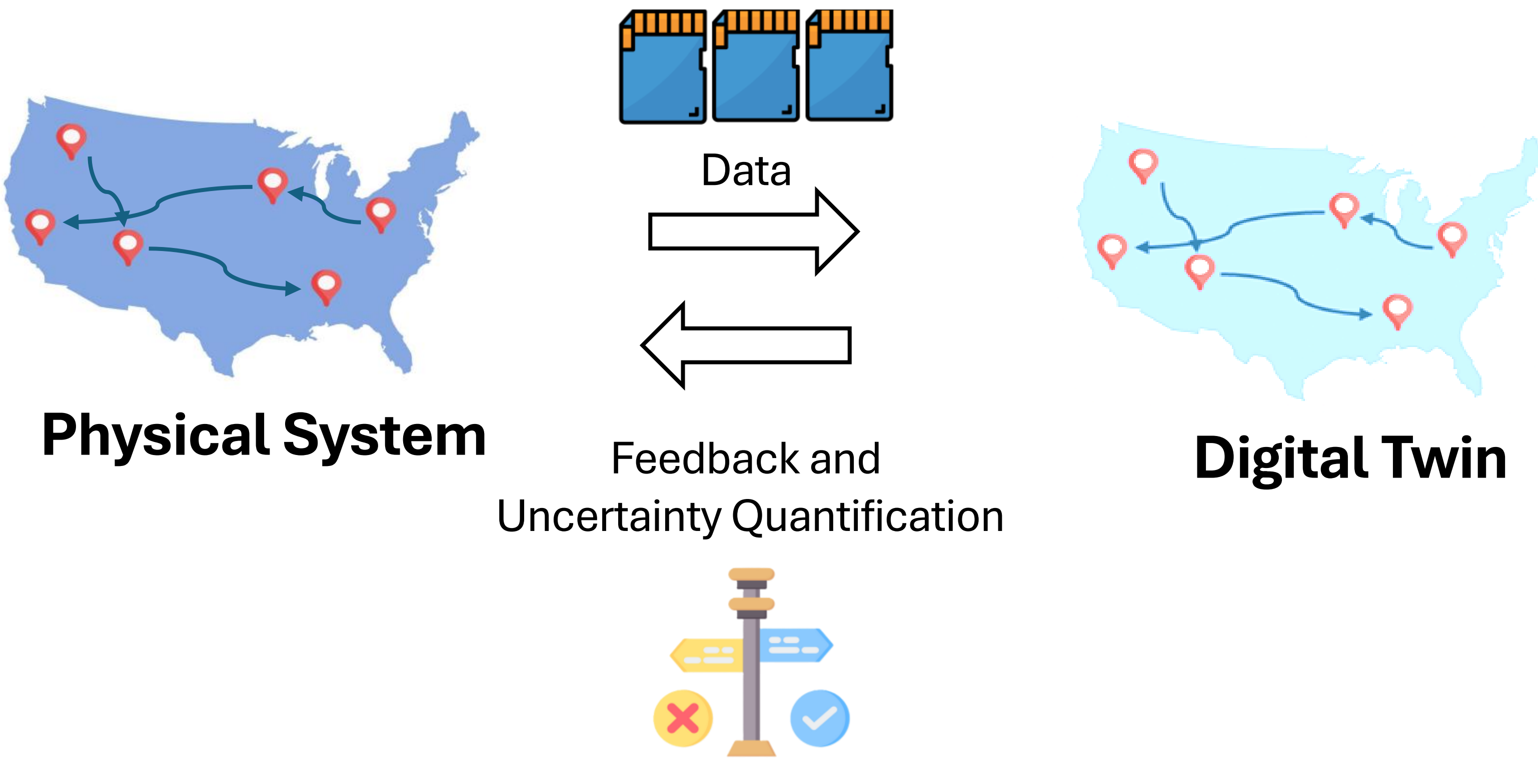
$$\mathbf{W}_{(\ell)} = \begin{pmatrix} \mathbf{w}_{1,(\ell)} \\ \vdots \\ \mathbf{w}_{p_{(\ell)},(\ell)} \end{pmatrix} \sim \text{Prior}, \quad \ell = 1, 2, \ldots, L.$$

ls can be found in Hung, Lin, Calhoun (2025)

Hung, N. Y. T., Lin, L. H., & Calhoun, V. D. (2025). Deep P-Spline: Theory, Fast Tuning, and Application. *arXiv preprint arXiv:2501.01376*.

Future Research

- Select important spatial temporal migration events



- Many thanks to IMSI for enabling me and my students to deepen our understanding of advanced techniques for digital modeling for migration modeling.

Bayesian inverse problems: well-posedness, optimal experimental design, and computation

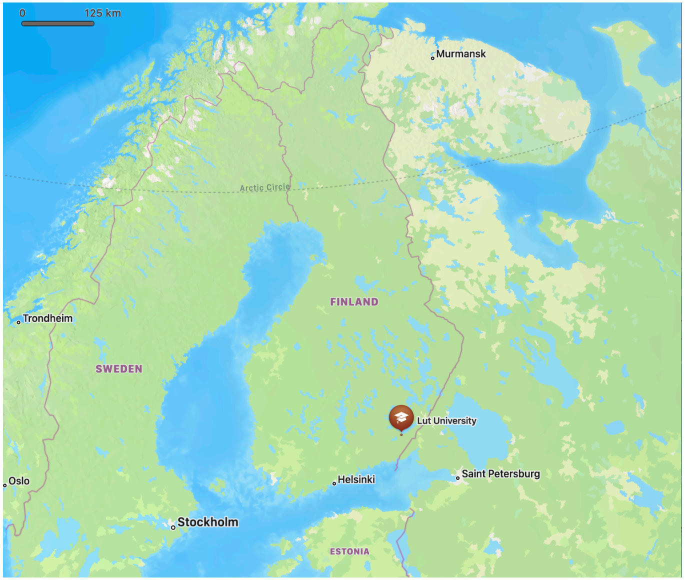
Duc-Lam Duong
duc-lam.duong@lut.fi

LUT University, Finland

IMSI Digital Twins Program
University of Chicago
December 5, 2025

Briefly about LUT/my background

- Lappeenranta-Lahti University of Technology (LUT)



- My research interests: Bayesian inverse problems, nonlinear PDEs, mathematical modelling, optimal experimental design, uncertainty quantification and machine learning

Bayesian inverse problems (BIPs)

Identify $x \in \mathcal{X}$ (Hilbert/Banach) from noisy data $y \in \mathbb{R}^n$,

$$y = \mathcal{G}(x) + \xi,$$

with a measurement map $\mathcal{G} : \mathcal{X} \to \mathbb{R}^n$ and observational noise ξ.

$$\text{prior } \mu_0 \xrightarrow{\text{measurement data } y} \text{posterior } \mu^y$$

Topics of interest include

- Well-posedness of μ^y: would μ^y be stable under perturbation in $\mathcal{G}$, especially when $\mathcal{G}$ is governed by a PDE?
- Optimal experimental design: how to design data acquisition to maximize 'information gain' and reduce uncertainty?
- Sampling from the posterior: how to sample from μ^y efficiently in complex, large-scale problems?

Uncertainty Quantification

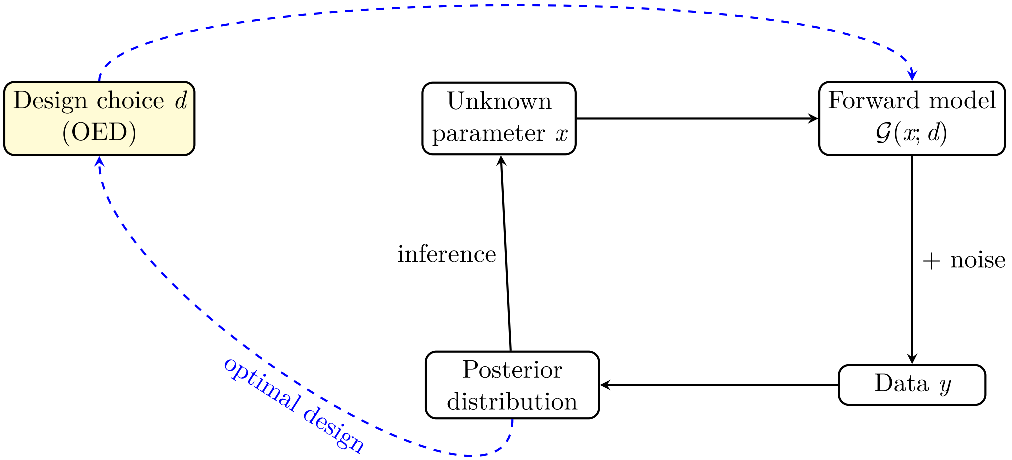
Optimal experimental design (OED)

The inverse problem can be reformulated as

$$y = \mathcal{O}_d \circ \mathcal{F}(x) + \xi,$$

i.e. $\mathcal{G} = \mathcal{O}_d \circ \mathcal{F}$, where $\mathcal{F} : \mathcal{X} \to \mathcal{X}'$ is the forward operator and $\mathcal{O}$ is the measurement operator, which depends on a design variable d.

OED problem: how to 'design' d to get the 'most informative' y and reduce uncertainty in x at the same time?



OED criteria and their robustness

'Informative data'? $\to$ OED *criteria*, defined via an *expected utility* U.

- D-optimality:

$$U(d) = \mathbb{E}^{y|x}\mathbb{E}^{\mu_0}\left[D_{\mathrm{KL}}\left(\mu^y \parallel \mu_0\right)\right], \quad d^* = \arg\max_d U(d).$$

- A-optimality:

$$U(d) = \mathbb{E}^{y|x}\mathbb{E}^{\mu_0}\Big[-\|x - \hat{x}(y;d)\|^2\Big], \quad d^* = \arg\max_d U(d),$$

where $\hat{x}(y;d)$ is the posterior mean/MAP estimator.

Challenges: expensive to compute, requires approximations.

But, *how reliable are OED criteria under approximations of model or likelihood?* Our answers[1] [2]: Yes, under 'reasonable' assumptions.

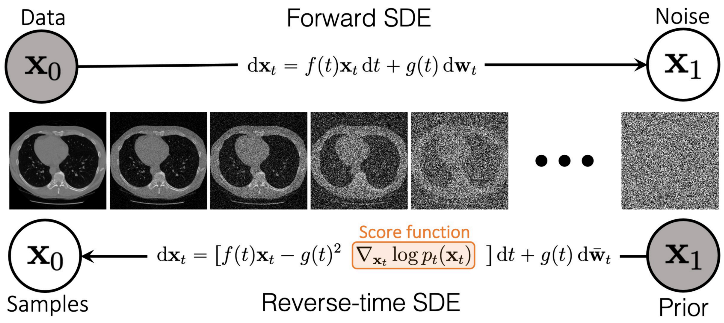
---

[1] D.-L. Duong, T. Helin, R. Rojo-Garcia, *Stability estimates for the expected utility in Bayesian optimal experimental design*, IPs 2023

[2] D.-L. Duong, T. Helin, ongoing

Posterior sampling: tools from ML

Diffusion models: learn/generate samples from data.



Challenges:

(i) expensive to run, especially large-scale IPs

(ii) may require access to a lot of data

(iii) infinite dimension? Needs to adapt the score function

Our works[3] [4] overcome some of these with a new formulation: UCoS

[3] Schneider, DLD, Lassas, de Hoop, Helin, *An unconditional representation of the conditional score in infinite-dimensional linear inverse problems*, 2025

[4] Mozumder et al., *Diffusion models for diffuse optical tomography*, ongoing

Other activities at IMSI

- Reading groups:
  1. (1) M Heinkenschloss and D P Kouri, Optimization problems governed by systems of PDEs with uncertainties, *Acta Numerica* 2025
  2. (2) R Nickl, Bayesian non-linear statistical inverse problems, EMS 2023
- Research discussions with other participants: Georg Stadler, Jiguang Sun, Ruiyi Yang, Hassan Iqbal, ...
- What I take from here into future:
  1. (1) Statistical aspects of inverse problems: RY
  2. (2) Inverse scattering: JS
  3. (3) Disign and control with mobile sensors: HI ...

Special thanks to IMSI, the DT long program organizers, and IMSI staff for making all this possible!

Surrogate modeling for adaptive predictive control over parameter spaces

Hassan Iqbal[1]

Joint work with:
Xingjian Li[1], Tyler Ingebrand[1], Adam Thorpe[1], Krishna Kumar[1], Ufuk Topcu[1], Ján Drgoňa[2]

[1]The University of Texas at Austin
[2]Johns Hopkins University

Iqbal, H. et al. (2025). "Zero-Shot Function Encoder-Based Differentiable Predictive Control". In: *arXiv preprint arXiv:2511.05757*

Dec 5, 2025
IMSI, University of Chicago
Workshop on Application of Digital Twins to Large-Scale Complex Systems

Model Predictive Control (MPC)

MPC solves online the following discrete-time optimal control problem (OCP),

$$
\begin{aligned}
\min_{\mathbf{u}_0,\dots,\mathbf{u}_{N-1}} \; & \sum_{k=0}^{N-1} \ell(\mathbf{x}_k, \mathbf{u}_k) + p_N(\mathbf{x}_N) \\
\text{s.t. } & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \;\; k \in \mathbb{N}_0^{N-1} \\
& h(\mathbf{x}_k) \leq 0 \\
& g(\mathbf{u}_k) \leq 0 \\
& \mathbf{x}_0 = \mathbf{x}(t)
\end{aligned}
$$

Model Predictive Control (MPC)

MPC solves online the following discrete-time optimal control problem (OCP),

$$\begin{aligned} \min_{\mathbf{u}_0,\ldots,\mathbf{u}_{N-1}} \; & \sum_{k=0}^{N-1} \ell(\mathbf{x}_k, \mathbf{u}_k) + p_N(\mathbf{x}_N) \\ \text{s.t. } \; & \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \;\; k \in \mathbb{N}_0^{N-1} \\ & h(\mathbf{x}_k) \leq 0 \\ & g(\mathbf{u}_k) \leq 0 \\ & \mathbf{x}_0 = \mathbf{x}(t) \end{aligned}$$

A typical reference-tracking objective with a reference $\mathbf{r}_k$ and control action penalty,

$$\ell(\mathbf{x}_k, \mathbf{u}_k, \mathbf{r}_k) = ||\mathbf{x}_k - \mathbf{r}_k||_2^2 + ||\mathbf{u}_k||_2^2$$

The system dynamics are assumed to be differentiable,

$$\mathbf{x}_{k+1} = A\mathbf{x}_k + B\mathbf{u}_k, \quad \text{or} \quad \mathbf{x}_{k+1} = \text{ODESolve}(\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k)).$$

This is the discretize-then-optimize approach to OCP.

Differentiable Predictive Control (DPC)

DPC (Drgoňa et al. 2024) is sampling-based strategy with expected risk minimization loss for approximately solving the following OCP,

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}_{\mathbf{x}_0 \sim P_{\mathbf{x}_0}, \boldsymbol{\xi} \sim P_{\boldsymbol{\xi}}, \boldsymbol{\nu} \sim P_{\boldsymbol{\nu}}} \left(\int_0^T \ell(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\xi}) \mathrm{d}t + p_T(\mathbf{x}(T)) \right) \\
\text{s.t.} \quad & \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\nu}), \\
& \mathbf{u}(t) = \pi(\mathbf{x}(t); \boldsymbol{\xi}, \boldsymbol{\nu}), \\
& h(\mathbf{x}(t); \boldsymbol{\xi}) \leq 0, \\
& g(\mathbf{u}(t); \boldsymbol{\xi}) \leq 0,
\end{aligned}
$$

Differentiable Predictive Control (DPC)

DPC (Drgoňa et al. 2024) is sampling-based strategy with expected risk minimization loss for approximately solving the following OCP,

$$\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}_{\mathbf{x}_0 \sim P_{\mathbf{x}_0}, \boldsymbol{\xi} \sim P_{\boldsymbol{\xi}}, \boldsymbol{\nu} \sim P_{\boldsymbol{\nu}}} \left(\int_0^T \ell(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\xi}) \mathrm{d}t + p_T(\mathbf{x}(T)) \right) \\
\text{s.t.} \quad & \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\nu}), \\
& \mathbf{u}(t) = \pi(\mathbf{x}(t); \boldsymbol{\xi}, \boldsymbol{\nu}), \\
& h(\mathbf{x}(t); \boldsymbol{\xi}) \leq 0, \\
& g(\mathbf{u}(t); \boldsymbol{\xi}) \leq 0,
\end{aligned}$$

We use penalty method to include the constraints in the loss function,

$$\begin{aligned}
\min_{\mathbf{W}} \quad & \mathbb{E}_{\mathbf{x}_0 \sim P_{\mathbf{x}_0}, \boldsymbol{\xi} \sim P_{\boldsymbol{\xi}}, \boldsymbol{\nu} \sim P_{\boldsymbol{\nu}}} \left[\sum_{k=0}^{N-1} \Big(\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) + p_h(h(\mathbf{x}_k, \boldsymbol{\xi})) + p_g(g(\mathbf{u}_k, \boldsymbol{\xi})) \Big) + p_N(\mathbf{x}_N) \right] \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{x}_k + \int_{t_k}^{t_{k+1}} \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\nu}) \mathrm{d}t, \quad \mathbf{u}_k = \pi_{\mathbf{W}}(\mathbf{x}_k; \boldsymbol{\xi}, \boldsymbol{\nu}).
\end{aligned}$$

- Assumes known differentiable dynamics and objectives for policy optimization.
- This can be considered as discretize-then-sample-and-learn approach.

Model, data-driven or learning-based control: comparison

Method	**Online cost**	**Offline cost**	**Constraint-aware?**	**Handles unknown dynamics?**
Classical MPC	✗ high online cost	✓ zero offline cost	✓ industry standard	✗ needs recursive system ID*
Approx. MPC (supervised)	✓	✗ requires labelled dataset	✓ postprocessing & verification	✗ requires online correction
DPC (self-supervised)	✓	✓ low cost: no labelled data needed	✓ constraints embedded during training	✗ needs known model
Goal	✓	✓	✓	✓

Modeling the Dynamics Using Function Encoders

- FE (Ingebrand et al. 2024) learns a set of NODE basis functions $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_B\}$ that span a subspace $\hat{\mathcal{F}} = \mathrm{span}\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_B\}$ supported by the data.
- Given some $\boldsymbol{\nu}$, the dynamics are approximated as $\mathbf{f} \approx \hat{\mathbf{f}} \in \hat{\mathcal{F}}$, which takes the form of a linear combination of the learned basis functions under time discretization,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \int_{t_k}^{t_{k+1}} \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\nu})\, \mathrm{d}t \approx \mathbf{x}_k + \int_{t_k}^{t_{k+1}} \sum_{j=1}^{B} c_j(\boldsymbol{\nu})\, \mathbf{g}_j(\mathbf{x}(t), \mathbf{u}_k; \boldsymbol{\theta}_j)\, \mathrm{d}t$$

$$\approx \mathbf{x}_k + \sum_{j=1}^{B} c_j(\boldsymbol{\nu}) \int_{t_k}^{t_{k+1}} \mathbf{g}_j(\mathbf{x}(t), \mathbf{u}_k; \boldsymbol{\theta}_j)\, \mathrm{d}t \quad (1)$$

  where $\boldsymbol{\theta}_j$ are the network parameters of the basis network $\mathbf{g}_j$.
- Importantly, the NODE basis functions $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_B\}$ do not depend explicitly on $\boldsymbol{\nu}$; as such, the dynamics function is uniquely determined by the coefficients $\mathbf{c} \in \mathbb{R}^B$.

- The coefficients $\mathbf{c}$ to some $\mathbf{f} \in \mathcal{F}$ can be computed in closed-form via the normal equation as $(\mathbf{G} + \lambda \mathbf{I})\mathbf{c} = \mathbf{F}$.
- Here, $\mathbf{G}_{ij} = \langle \mathbf{g}_i, \mathbf{g}_j \rangle$ and $\mathbf{F}_i = \langle \mathbf{f}, \mathbf{g}_i \rangle$ can both be easily computed using Monte Carlo integration from a small amount of input-output data collected online.

Offline Learning of FE Basis Functions for System Dynamics

- The coefficients $\mathbf{c}$ to some $\mathbf{f} \in \mathcal{F}$ can be computed in closed-form via the normal equation as $(\mathbf{G} + \lambda \mathbf{I})\mathbf{c} = \mathbf{F}$.
- Here, $\mathbf{G}_{ij} = \langle \mathbf{g}_i, \mathbf{g}_j \rangle$ and $\mathbf{F}_i = \langle \mathbf{f}, \mathbf{g}_i \rangle$ can both be easily computed using Monte Carlo integration from a small amount of input-output data collected online.

Algorithm 2 (using $\mathbf{x}_{k+1} = \mathbf{x}_k + \Delta t \cdot \mathbf{f}(\mathbf{x_k}, \mathbf{u_k}; \boldsymbol{\nu})$ for simplicity).

Input: set of datasets $\mathcal{D}$ collected from dynamics $\mathcal{F}$, learning rate α
Initialize basis functions $\mathbf{g}_1, \ldots, \mathbf{g}_B$ with trainable parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_B$
while not converged **do**
    for all $\mathcal{D}_l \in \mathcal{D}$ **do**
        reset loss $L = 0$
        for all $(\mathbf{x}_k, \mathbf{u}_k, \mathbf{x}_{k+1}) \in \mathcal{D}_i$ **do**
            $\mathbf{c} \leftarrow (\mathbf{G} + \lambda \mathbf{I})^{-1}\mathbf{F}$
            $\hat{\mathbf{x}}_{k+1} \leftarrow \hat{\mathbf{x}}_{k+1}$ from (1)
            $L \leftarrow L + \|\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1}\|_2^2$
        end for
        $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \nabla_{\boldsymbol{\theta}} L$
    end for
end while
Output: trained basis functions $\mathbf{g}_1, \ldots, \mathbf{g}_B$

DPC with FE-NODE Dynamics

$$\min_{\mathbf{W}} \quad \mathbb{E}_{\mathbf{x}_0 \sim P_{\mathbf{x}_0},\, \boldsymbol{\xi} \sim P_{\boldsymbol{\xi}},\, \mathbf{c} \sim P_{\mathbf{c}}} \left[\sum_{k=0}^{N-1} \Big(\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) + p_h(h(\mathbf{x}_k, \boldsymbol{\xi})) + p_g(g(\mathbf{u}_k, \boldsymbol{\xi})) \Big) + p_N(\mathbf{x}_N) \right] \tag{2}$$

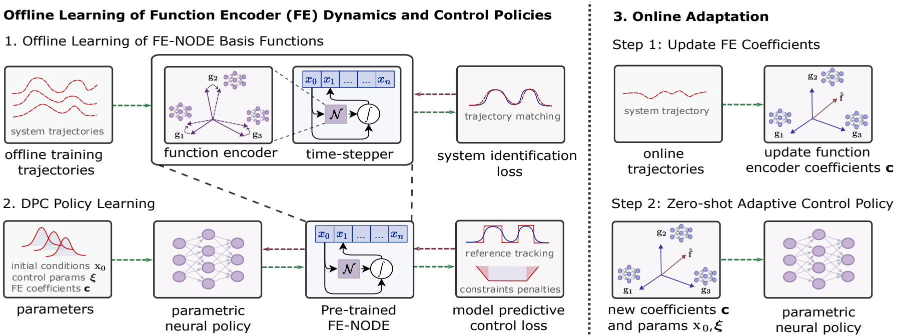
$$\text{s.t.} \quad \mathbf{x}_{k+1} = \mathbf{x}_k + \int_{t_k}^{t_{k+1}} \sum_{j=1}^{B} c_j\, \mathbf{g}_j(\mathbf{x}(t), \mathbf{u}_k; \boldsymbol{\theta}_j)\, \mathrm{d}t, \tag{3}$$

$$\mathbf{u}_k = \pi_{\mathbf{W}}(\mathbf{x}_k; \boldsymbol{\xi}, \mathbf{c}). \tag{4}$$

Algorithm 3
Offline learning of parametric neural policies via DPC

Input: pre-trained NODE basis functions $\mathbf{g}_1, \ldots, \mathbf{g}_B$, learning rate β
Initialize policy network $\pi_{\mathbf{W}}$ with parameters $\mathbf{W}$
while not converged **do**
  for all $\mathbf{x}_0 \sim P_{\mathbf{x}_0}$, $\boldsymbol{\xi} \sim P_{\boldsymbol{\xi}}$, $\mathbf{c} \sim P_{\mathbf{c}}$ **do**
    loss $L \leftarrow 0$
    for $k = 0, 1, N-1$ **do**
      $\mathbf{u}_k \leftarrow \pi_{\mathbf{W}}(\mathbf{x}_k; \boldsymbol{\xi}, \mathbf{c})$
      $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_{k+1}$ from (3)
    end for
    $L \leftarrow L$ from (2)
    $\mathbf{W} \leftarrow \mathbf{W} - \beta \nabla_{\mathbf{W}} L$
  end for
end while
Output: trained policy network $\pi_{\mathbf{W}}$

DPC with FE-NODE Dynamics and Neural Policies



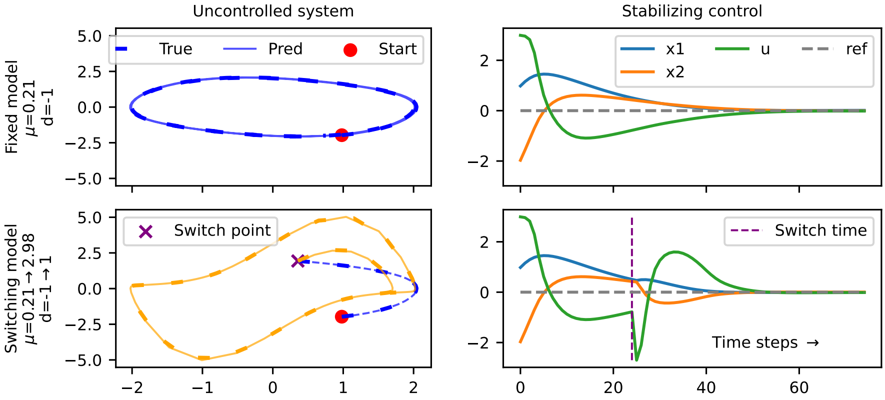
Example 1: Stabilizing a Van der Pol Oscillator

$$\begin{cases} \dot{x}_1 &= d \cdot x_2, \\ \dot{x}_2 &= \mu(1 - x_1^2)x_2 - x_1 + u, \end{cases}$$

where $[x_1, x_2]^\top \in [-2, 2] \times [-5, 5]$, $u \in [-3.0, 3.0]$, and parameters $\mu \sim \mathcal{U}[0.1, 3.0]$ and $d \in \{-1, 1\}$ determine the dynamics. The objective is, $p_N(\mathbf{x}_N) = \|\mathbf{x}_N\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{u}_k\|^2$.

$$\begin{cases} \dot{x}_1 &= d \cdot x_2, \\ \dot{x}_2 &= \mu(1 - x_1^2)x_2 - x_1 + u, \end{cases}$$

where $[x_1, x_2]^\top \in [-2, 2] \times [-5, 5]$, $u \in [-3.0,\ 3.0]$, and parameters $\mu \sim \mathcal{U}[0.1, 3.0]$ and $d \in \{-1, 1\}$ determine the dynamics. The objective is, $p_N(\mathbf{x}_N) = \|\mathbf{x}_N\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{u}_k\|^2$.



Example 2: Reference Tracking of a Two-tank System

$$\begin{cases} \dot{x}_1 &= d_1(1-v)p - d_2\sqrt{x_1}, \\ \dot{x}_2 &= d_1 vp + d_2\sqrt{x_1} - d_2\sqrt{x_2}. \end{cases}$$

where $[x_1, x_2]^\top \in [0,1]^2$, control inputs $[p, v]^\top \in [0,1]^2$, $d_1 \sim \mathcal{U}[0.06,\ 0.1]$ and $d_2 \sim \mathcal{U}[0.01,\ 0.06]$. The objective is $p_N(\mathbf{x}_N) = \|\mathbf{x}_N - \mathbf{x}_{\text{ref}}(\boldsymbol{\xi})\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{x}_k - \mathbf{x}_{\text{ref}}(\boldsymbol{\xi})\|^2 + \|\mathbf{u}_k\|^2$.

Example 2: Reference Tracking of a Two-tank System

$$\begin{cases} \dot{x}_1 &= d_1(1-v)p - d_2\sqrt{x_1}, \\ \dot{x}_2 &= d_1 vp + d_2\sqrt{x_1} - d_2\sqrt{x_2}. \end{cases}$$

where $[x_1, x_2]^\top \in [0,1]^2$, control inputs $[p, v]^\top \in [0,1]^2$, $d_1 \sim \mathcal{U}[0.06,\ 0.1]$ and $d_2 \sim \mathcal{U}[0.01,\ 0.06]$. The objective is $p_N(\mathbf{x}_N) = \|\mathbf{x}_N - \mathbf{x}_{\text{ref}}(\boldsymbol{\xi})\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{x}_k - \mathbf{x}_{\text{ref}}(\boldsymbol{\xi})\|^2 + \|\mathbf{u}_k\|^2$.

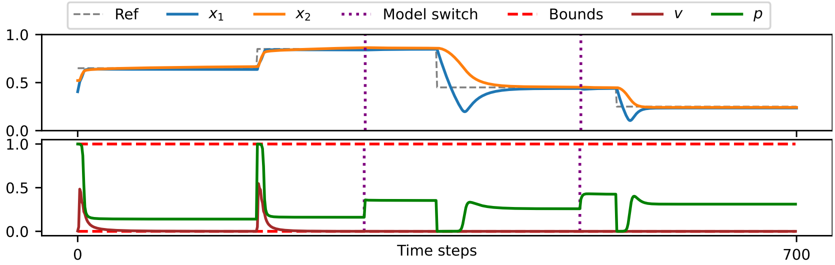


Figure: Two-tank system reference tracking under multiple system switches using FE-DPC.

Example 3: Reference Tracking of a Glycolytic Oscillator (GO)

$$
\begin{cases}
\dot{x_1} &= J_0 - \frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} + u, \text{(no control in below figure)} \\
\dot{x_2} &= 2\frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} - k_2 x_2 (N - x_5) - k_6 x_2 x_5, \\
\dot{x_3} &= k_2 x_2 (N - x_5) - k_3 x_3 (A - x_6), \\
\dot{x_4} &= k_3 x_3 (A - x_6) - k_4 x_4 x_5 - \kappa (x_4 - x_7), \\
\dot{x_5} &= k_2 x_2 (N - x_5) - k_4 x_4 x_5 - k_6 x_2 x_5, \\
\dot{x_6} &= -2\frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} + 2 k_3 x_3 (A - x_6) - k_5 x_6, \\
\dot{x_7} &= \psi \kappa (x_4 - x_7) - k x_7.
\end{cases}
$$

where $k_1 \in \{80, 90, 100\}$ and $K_1 \in \{0.5, 0.75\}$.

Example 3: Reference Tracking of a Glycolytic Oscillator (GO)

$$\begin{cases} \dot{x}_1 &= J_0 - \frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} + u, \text{(no control in below figure)} \\ \dot{x}_2 &= 2\frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} - k_2 x_2 (N - x_5) - k_6 x_2 x_5, \\ \dot{x}_3 &= k_2 x_2 (N - x_5) - k_3 x_3 (A - x_6), \\ \dot{x}_4 &= k_3 x_3 (A - x_6) - k_4 x_4 x_5 - \kappa(x_4 - x_7), \\ \dot{x}_5 &= k_2 x_2 (N - x_5) - k_4 x_4 x_5 - k_6 x_2 x_5, \\ \dot{x}_6 &= -2\frac{k_1 x_1 x_6}{1+(x_6/K_1)^4} + 2k_3 x_3 (A - x_6) - k_5 x_6, \\ \dot{x}_7 &= \psi\kappa(x_4 - x_7) - k x_7. \end{cases}$$

where $k_1 \in \{80, 90, 100\}$ and $K_1 \in \{0.5, 0.75\}$.

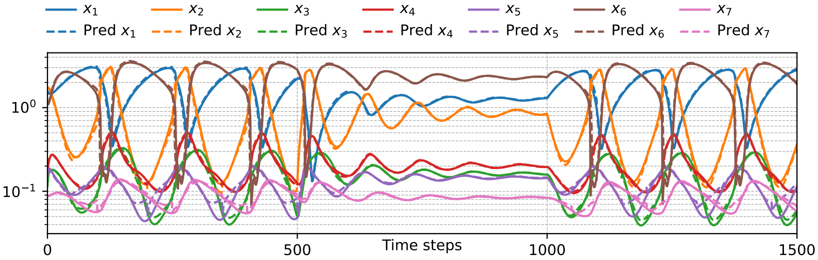


Figure: Parameters switch every 500 step, predictions calibrated against true states every 50 step.

Example 3: Reference Tracking of a Glycolytic Oscillator (G0)

Control objective is defined by $p_N(\mathbf{x}_N) = \|\mathbf{x}_{1,N} - \mathbf{x}_{1,\text{ref}}(\boldsymbol{\xi})\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{x}_{1,k} - \mathbf{x}_{1,\text{ref}}(\boldsymbol{\xi})\|^2 + \|\mathbf{u}_k\|^2$.

Example 3: Reference Tracking of a Glycolytic Oscillator (G0)

Control objective is defined by $p_N(\mathbf{x}_N) = \|\mathbf{x}_{1,N} - \mathbf{x}_{1,\text{ref}}(\boldsymbol{\xi})\|^2$ and $\ell(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\xi}) = \|\mathbf{x}_{1,k} - \mathbf{x}_{1,\text{ref}}(\boldsymbol{\xi})\|^2 + \|\mathbf{u}_k\|^2$.

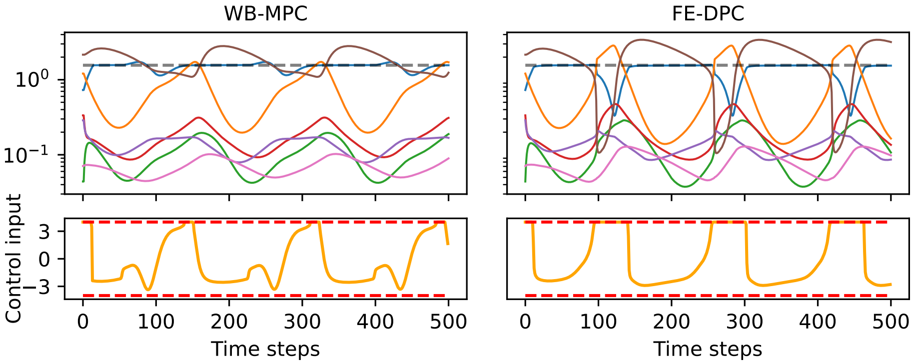


Figure: WB-MPC (left) and FE-DPC (right) based reference tracking of x_1 (blue) state.

Example 4: Controlling a Quadrotor

$$\begin{cases} \dot{p}_n = \cos\theta\cos\psi\, u \\ \quad + (\sin\phi\sin\theta\cos\psi - \cos\phi\sin\psi)\, v \\ \quad + (\cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi)\, w, \\ \dot{p}_e = \cos\theta\sin\psi\, u \\ \quad + (\sin\phi\sin\theta\sin\psi + \cos\phi\cos\psi)\, v \\ \quad + (\cos\phi\sin\theta\sin\psi - \sin\phi\cos\psi)\, w, \\ \dot{h} = \sin\theta\, u - \sin\phi\cos\theta\, v - \cos\phi\cos\theta\, w, \\ \dot{u} = rv - qw - g\sin\theta, \\ \dot{v} = pw - ru + g\cos\theta\sin\phi, \\ \dot{w} = qu - pv + g\cos\theta\cos\phi - (F/m), \\ \tau_\psi = 0., \end{cases}$$

$$\begin{aligned} F &= mg - 10(h - \alpha_1) + 3w, \\ \tau_\phi &= -(\phi - \alpha_2) - p, \\ \tau_\theta &= -(\theta - \alpha_3) - q, \end{aligned}$$

$$\begin{aligned} \dot{\phi} &= p + \sin\phi\tan\theta\, q + \cos\phi\tan\theta\, r, \\ \dot{\theta} &= \cos\phi\, q - \sin\phi\, r, \\ \dot{\psi} &= (\sin\phi/\cos\theta)\, q + (\cos\phi/\cos\theta)\, r, \\ \dot{p} &= ((J_y - J_z)/J_x)\, qr + (1/J_x)\, \tau_\phi, \\ \dot{q} &= ((J_z - J_x)/J_y)\, pr + (1/J_y)\, \tau_\theta, \\ \dot{r} &= ((J_x - J_y)/J_z)\, pq + (1/J_z)\, \tau_\psi, \end{aligned}$$

$$[\alpha_1, \alpha_2, \alpha_3]^\top \in [\boldsymbol{\alpha}_{\min}, \boldsymbol{\alpha}_{\max}] = \begin{bmatrix} 0, & -0.524, & -0.524 \\ 1.5, & 0.524, & 0.524 \end{bmatrix}^\top \subset \mathbb{R}^3,$$

$$m \sim \mathcal{U}[1.2,\ 1.6], \quad J_x = J_y \sim \mathcal{U}[0.050,\ 0.058], \quad J_z \sim \mathcal{U}[0.090,\ 0.110].$$

Objective: maintain an altitude of 0.4m with zero linear and angular velocities.

Example 4: Controlling a Quadrotor

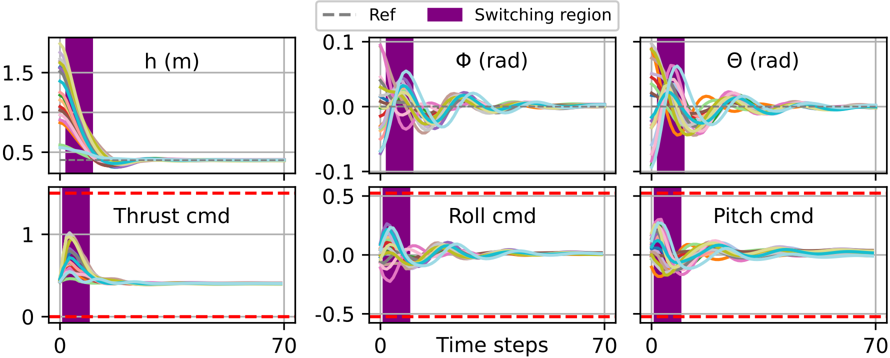


Figure: 20 quadrotor models with distinct dynamics parameterization are randomly initialized within state bounds. Each experiences a random dynamics switch between 2–20 s.

Algorithm	Metric	Van der Pol	Two Tank	GO (7D)	Quad (12D)
FE-DPC	Error (MSE)	0.0027	0.0085	0.1803	0.0220
	Time (s)	0.53	1.13	5.89	1.93
WB-MPC*	Error (MSE)	0.0027	0.0042	0.0323	0.0242
	Time (s)	1.21	6.75	136.07	155.85

Table: Comparison of error (MSE) and computation time (s) for each benchmark.

* Assumed no plant-model mismatch.

Conclusion

We introduced FE-DPC, a framework for zero-shot adaptive predictive control:

- FE-NODE to identify unknown dynamics online from a few data points.
- DPC to learn a parametric policy offline conditioned on identified dynamics.
- Instantly adapts to unknown dynamics without retraining.
- Examples shown for high-dimensional (12D) and stiff (7D) systems.
- Achieves accuracy competitive with MPC at fraction of computational cost.

Conclusion

We introduced FE-DPC, a framework for zero-shot adaptive predictive control:

- FE-NODE to identify unknown dynamics online from a few data points.
- DPC to learn a parametric policy offline conditioned on identified dynamics.
- Instantly adapts to unknown dynamics without retraining.
- Examples shown for high-dimensional (12D) and stiff (7D) systems.
- Achieves accuracy competitive with MPC at fraction of computational cost.

Some remarks and future work:

- Complimentary to model-based approaches, potential benefits by combining e.g. parameter estimation, warm starting, stochastic processes etc.
- Closed-loop guarantees.
- Surrogate modeling for complex engineering systems (PGD etc.).

References I

Drgoňa, J. et al. (2024). "Learning constrained parametric differentiable predictive control policies with guarantees". In: *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 54.6, pp. 3596–3607.

Ingebrand, T. et al. (2024). "Zero-shot transfer of neural ODEs". In: *Advances in Neural Information Processing Systems* 37, pp. 67604–67626.

Iqbal, H. et al. (2025). "Zero-Shot Function Encoder-Based Differentiable Predictive Control". In: *arXiv preprint arXiv:2511.05757*.

Low, S. A. et al. (2025). "Function Spaces Without Kernels: Learning Compact Hilbert Space Representations". In: *arXiv preprint arXiv:2509.20605*.