Learning hybrid models combining scientific models and machine learning

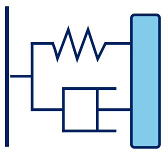
Naoya Takeishi (UTokyo)

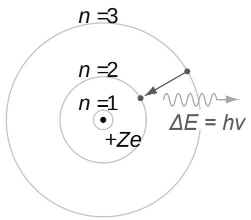
Workshop on Functional Inference and Machine Intelligence 2025 @ Naha

Two worlds

Scientific models

- Few parameters
- Scarce data
- Extrapolation
- Domain knowledge

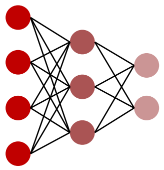


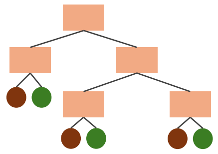


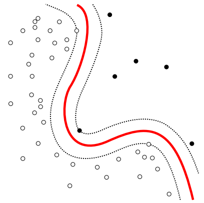
$$\frac{\partial u}{\partial t} = f\left(\frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}, \dots\right)$$

Machine learning models

- Many parameters
- “Big” data
- High adaptivity
- “Inductive bias”







Hybrid modeling

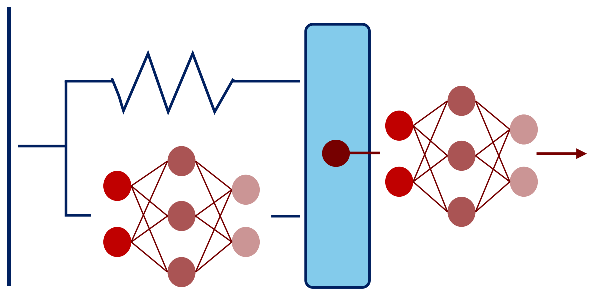
Scientific models

- Few parameters
- Scarce data
- Extrapolation
- Domain knowledge

Machine learning models

- Many parameters
- “Big” data
- High adaptivity
- “Inductive bias”

- Take the best of both worlds
  - for better generalization
  - for extrapolation
  - for (partial) interpretability
  - for inference with misspecified simulators
  - hopefully…



Hamiltonian neural nets [Greydanus+ 19]

- Hamiltonian mechanics: $\dfrac{d\boldsymbol{q}(t,x)}{dt} = \dfrac{\partial H(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{p}}, \ \dfrac{d\boldsymbol{p}(t,x)}{dt} = -\dfrac{\partial H(\boldsymbol{q},\boldsymbol{p})}{\partial \boldsymbol{q}}$
- Learn Hamiltonian H with neural nets

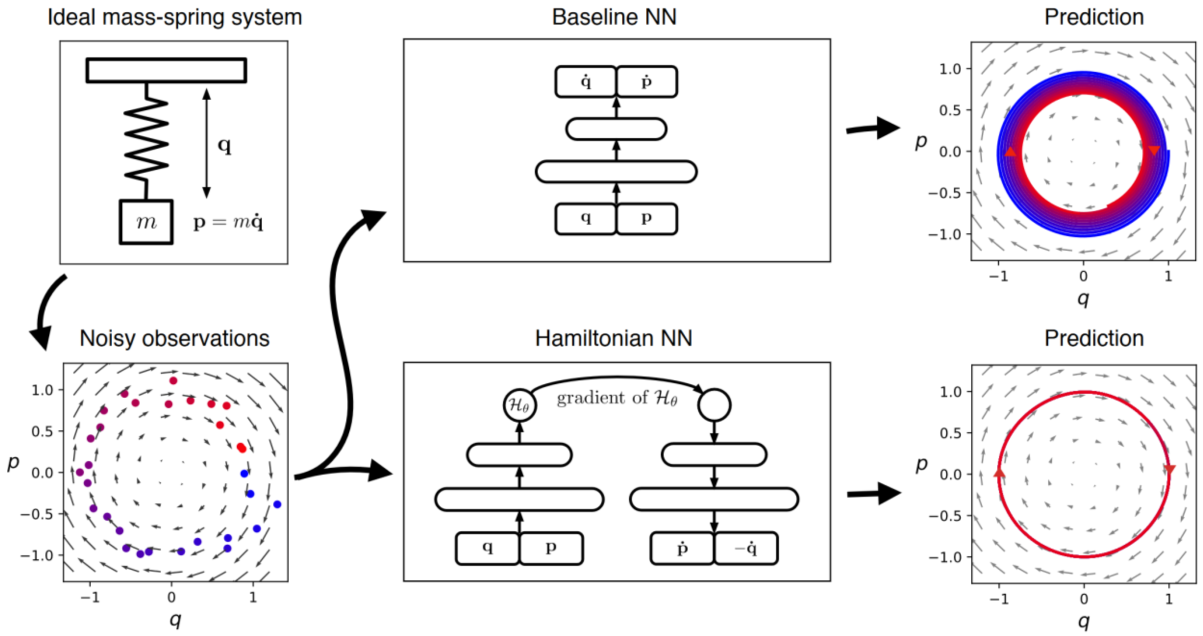


figure from Greydanus+ 19

Greydanus+, Hamiltonian neural networks, NeurIPS 2019.

Hybrid disease progression model [Qian+ 21]

- Pharmacological model of disease progression

$$\dot{z}_1 = k_{IR} \cdot z_4 + k_{PF} \cdot z_4 \cdot z_1 - k_O \cdot z_1 + \frac{E_{max} \cdot z_1^{h_P}}{EC_{50}^{h_P} + z_1^{h_P}} - k_{Dex} \cdot z_1 \cdot z_2$$
$$\dot{z}_4 = k_{DP} \cdot z_4 - k_{IIR} \cdot z_4 \cdot z_1 - k_{DC} \cdot z_4 \cdot z_5^{h_C}$$
$$\dot{z}_5 = k_1 \cdot z_1$$
$$\dot{z}_2 = -k_2 \cdot z_2 + k_3 \cdot z_3$$
$$\dot{z}_3 = -k_3 \cdot z_3$$

- Combined with neural ODEs

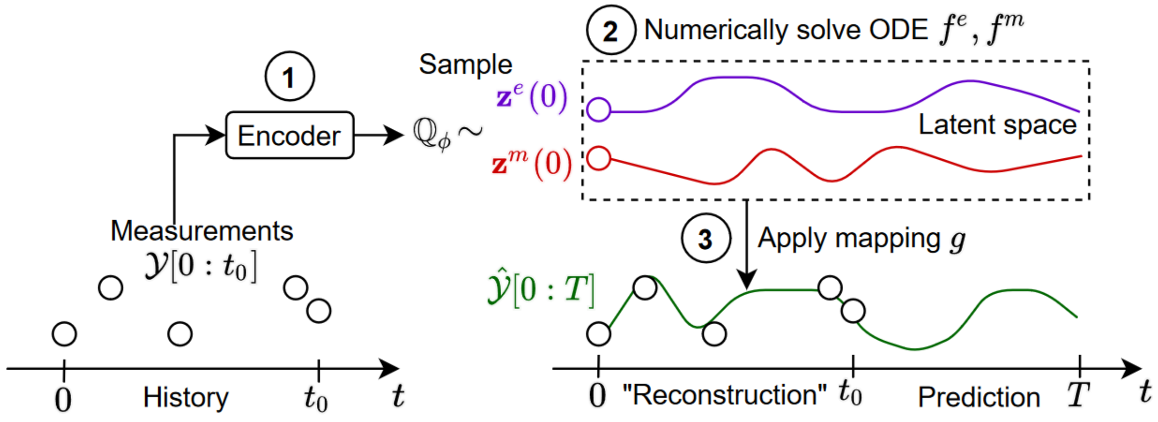
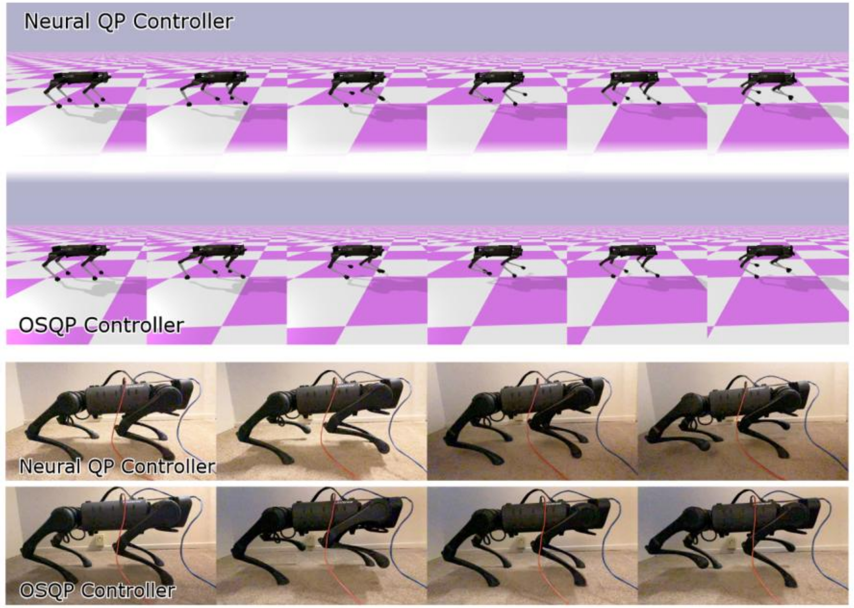


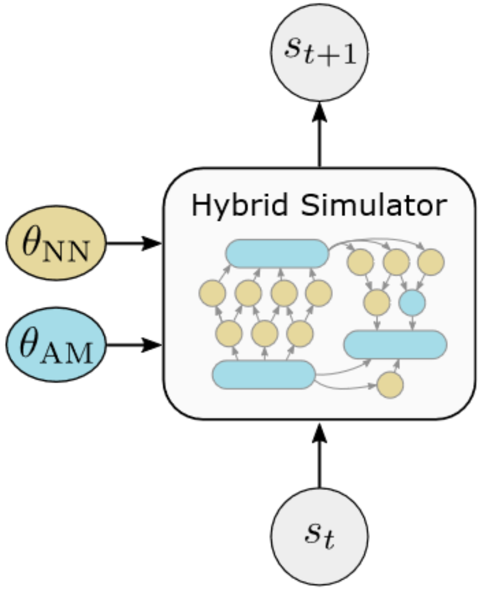
figure from Qian+ 21

Qian+, Integrating expert ODEs into neural ODEs: Pharmacology and disease progression, NeurIPS 2021.

Neural rigid body dynamics [Heiden+ 21]

- Rigid body dynamics with contacts
- Replace friction model by neural nets



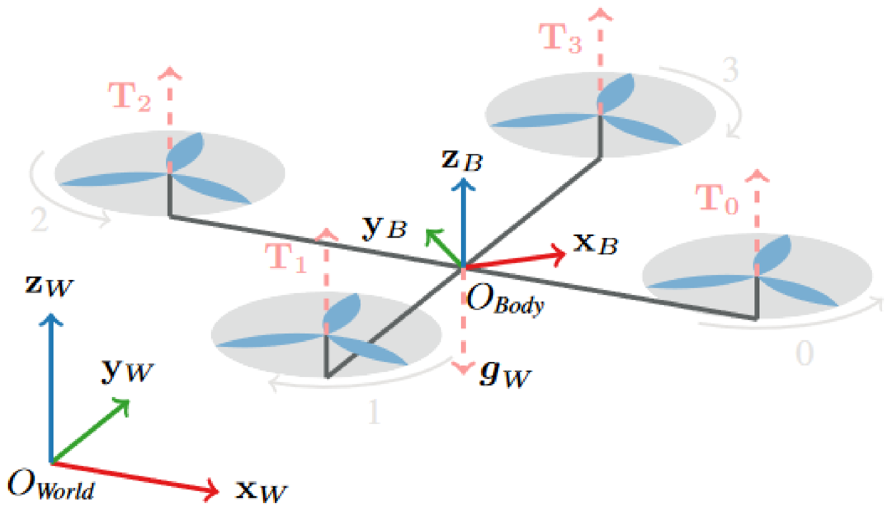


figures from Heiden+ 21

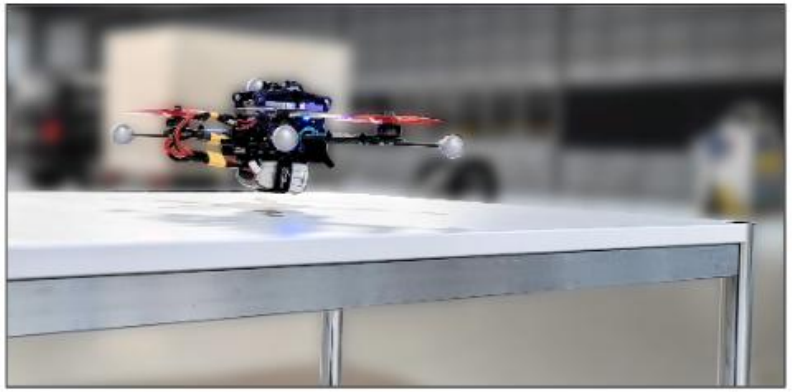
Heiden+, NeuralSim: Augmenting differentiable simulators with neural networks, ICRA 2021.

MPC with hybrid models [Salzmann+ 23]

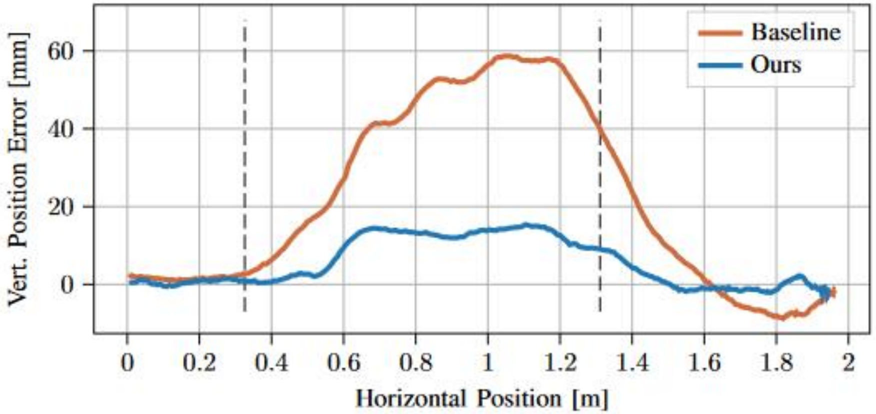
- Rigid body dynamics of drone
- and neural net to predict aerodynamic effects
- Model predictive control improved



figures from Salzmann+ 23



(a)



(b)

Salzmann+, Real-time neural MPC: Deep learning model predictive control for quadrotors and agile robotic platforms, IEEE RA-L, 2023.

Traffic queue length prediction [Shirakami+ 23]

- Feature extraction by GNN
- Prediction of queue length using "sandglass model"

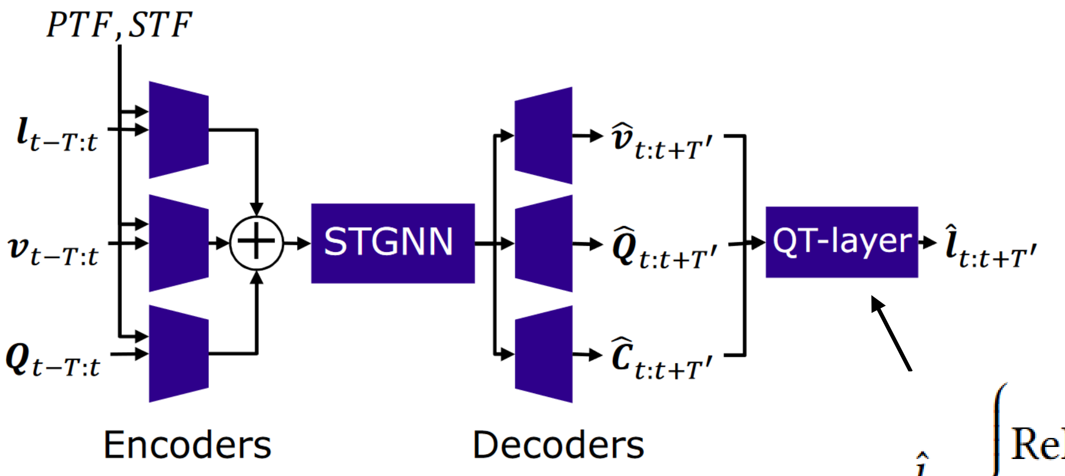


figure from Shirakami+ 23

sandglass model

$$\hat{l} = \begin{cases} \mathrm{ReLU}\left(\dfrac{\hat{d}}{\frac{\rho_0 - a\hat{Q}}{\hat{Q}} - \frac{1}{v_0}} + \hat{C} \right) & \text{if } \hat{d} \geq 0 \text{ and } \frac{\rho_0 - a\hat{Q}}{\hat{Q}} - \frac{1}{v_0} > 0, \\ 0 & \text{otherwise,} \end{cases}$$

Shirakami+, QTNet: Theory-based queue length prediction for urban traffic, KDD 2023.

Neural advection PDEs [Verma+ 24]

- Transport & compression of air: $$\frac{\partial u(t,\boldsymbol{x})}{\partial t} = -\boldsymbol{v}(t,\boldsymbol{x}) \cdot \nabla u(t,\boldsymbol{x}) - u(t,\boldsymbol{x}) \nabla \cdot \boldsymbol{v}(t,\boldsymbol{x})$$
    - u: some physical quantity (e.g., temperature)
    - $\boldsymbol{v}$: flow's velocity
- Learn dynamics of $\boldsymbol{v}$ with neural nets

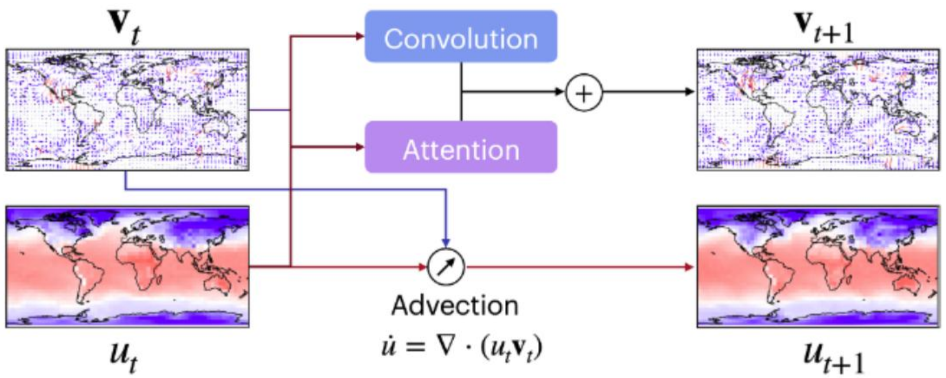
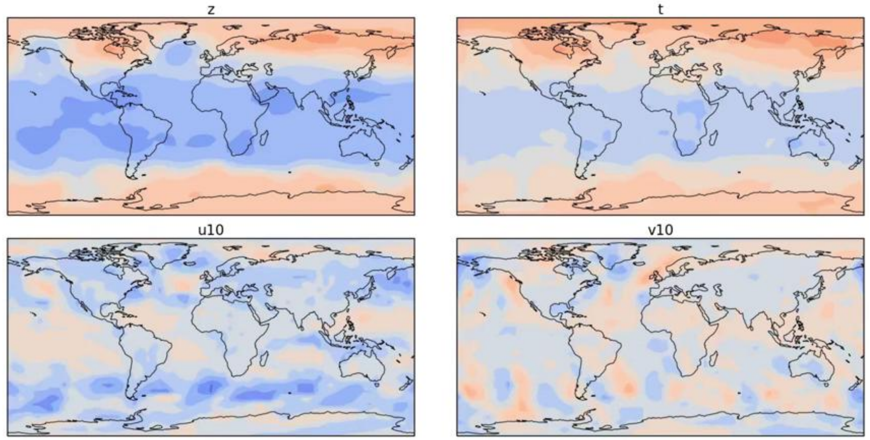


figure from Verma+ 24



video from the project page of Verma+ 24
https://yogeshverma1998.github.io/ClimODE/

Verma+, ClimODE: Climate and weather forecasting with physics-informed neural ODEs, ICLR 2024.

Hybrid modeling in various domains

- Known and practiced for a long time
  - chemical engineering
  - health sciences
  - computational fluid dynamics
  - control
  - etc.

- A review article in chemical engineering with 130+ references [Schweidtmann+ 24]

Digital Chemical Engineering 10 (2024) 100136

Contents lists available at ScienceDirect

Digital Chemical Engineering

journal homepage: www.elsevier.com/locate/dche

ELSEVIER

Review article

A review and perspective on hybrid modeling methodologies

Artur M. Schweidtmann, Dongda Zhang, Moritz von Stosch *

Department of Chemical Engineering, Delft University of Technology, Van der Maasweg 9, Delft 2629 HZ, The Netherlands
Department of Chemical Engineering, the University of Manchester, Engineering Building, Manchester M13 9PL, United Kingdom
DataHow AG, Hagenholzstrasse 111, Zurich, Switzerland

ARTICLE INFO

Keywords:
Hybrid modeling
Hybrid semi parametric modeling
Grey box
Neural networks
Parameter identification

ABSTRACT

The term hybrid modeling refers to the combination of parametric models (typically derived from knowledg about the system) and nonparametric models (typically deduced from data). Despite more than 20 years research, over 150 scientific publications (Agharafeie et al., 2023), and some recent industrial applications o this topic, the capabilities of hybrid models often seem underrated, misunderstood, and disregarded by othe disciplines as "simply combining some models" or maybe it has gone unnoticed at all. In fact, hybrid modelin could become an enabling technology in various areas of research and industry, such as systems and syntheti biology, personalized medicine, material design, or the process industries. Thus, a systematic investigation the hybrid model properties is warranted to scoop the full potential of machine learning, reduce experiment effort, and increase the domain in which models can predict reliably.

1. Introduction

Machine-learning has obtained a lot of attention in recent years performing tasks unthinkable before (Wang et al., 2020). Much hope also relies on machine learning in natural and life science-related fields, dreaming of the description of highly complex systems solely by using data, regressing some inputs (e.g., features, factors, predictors, regressors) to some outputs (e.g., response, target) (Montáns et al., 2019; Jumper et al., 2021). Similarly, seeking to unravel the mechanisms of the system by mechanistic modeling has in the past given rise to fundamental modeling research (Sun et al., 2019; Gernaey et al., 2010; Horstemeyer, 2010). At first sight, machine learning approaches seem to compete with the more traditional fundamental modeling. However, fundamental modeling can be combined with machine-learning approaches as highlighted in literature (Antoniewicz, 2015; Baker et al., 2018; Bikmukhametov and Jäschke, 2020; Hamilton et al., 2017; Zhang et al., 2019, 2020).

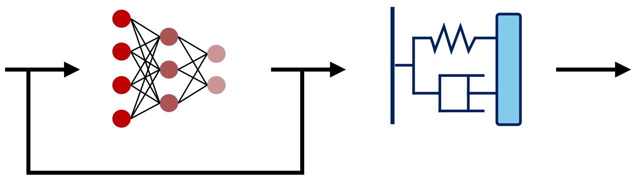
Indeed, the idea of combining mechanistic modeling with data-driven models has been around from the 1990th (Psichogios and Ungar, 1992; Su et al., 1993; Kramer et al., 1992; Johansen and Foss, 1992; Thompson and Kramer, 1994). Since then a significant amount of research has been published using the terms "*hybrid modeling*" in the more process engineering-related research fields and "*grey-box modeling*" in the control and automation field. Though grey-box modeling is understood to include a wider range of models than hybrid modeling, e.g., a system of equations that is derived from first principles and complemented by empirically derived equations or structuring the machine-learning model based on process knowledge (Alhajeri et al 2022; Wu et al., 2020) qualify as grey-box but not as hybrid model Hybrid modeling is understood as the combination of models that a different in their traits, i.e., one part of the model structure is derive from knowledge (hence each parameter has a physical meaning an is normally identifiable, this type of model is named 'parametric' an it is typically represented by white boxes) whereas the other part the structure is derived from data (hence parameters do not have an physical meaning and are normally not identifiable, and this type model is named 'nonparametric' and it is typically represented by blac boxes). As such, and in order to reduce the ambiguity in that the ter hybrid modeling could be understood, the term hybrid semi-parametr modeling has been suggested (Thompson and Kramer, 1994; von Stosc et al., 2014b). In what follows, we use the term hybrid modeling as short version of hybrid semi-parametric modeling.

The current hybrid modeling research and applications have evolve from the area of artificial neural networks, starting with Psichogios an Ungar (Psichogios and Ungar, 1992). They showed that the integratio of fundamental knowledge into neural networks can (1) improve th model's extrapolation performance (the model faithfully predicts th system behavior beyond prior tested conditions), (2) reduce its da requirements, and (3) increase process understanding (e.g., mod interpretability). The origin of these key properties can easily b comprehended considering the examples discussed in Box 1. The key properties were demonstrated and further extended in the lite ature (van Can et al., 1999; Van Can et al., 1998; Schuppert, 200

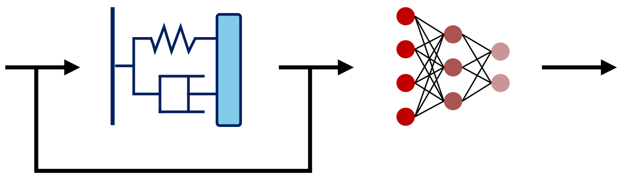
Hybrid model design patterns [e.g., Rudolph+ 24; Schweidtmann+ 24]

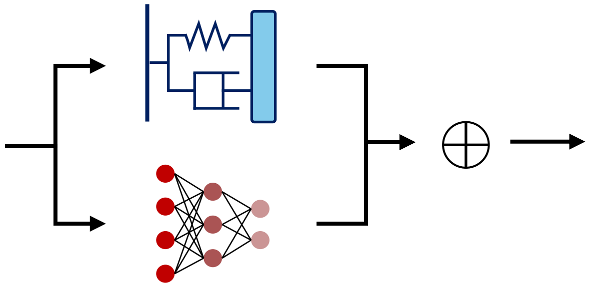
- "ML first"
  - a.k.a. feature extraction / parametrization / submodeling / …

original (non)identifiability of scientific model matters



- "ML last"
  - a.k.a. closures / residual physics / discrepancy modeling / Δ-ML / …





- Scientific models should be (almost everywhere) differentiable
  - world of differentiable simulators

Differentiable simulators

- Scientific models should be (almost everywhere) differentiable
  - must be implemented in a way facilitating differentiation functionality
- Challenges particularly with long rollout, discontinuity, and chaos

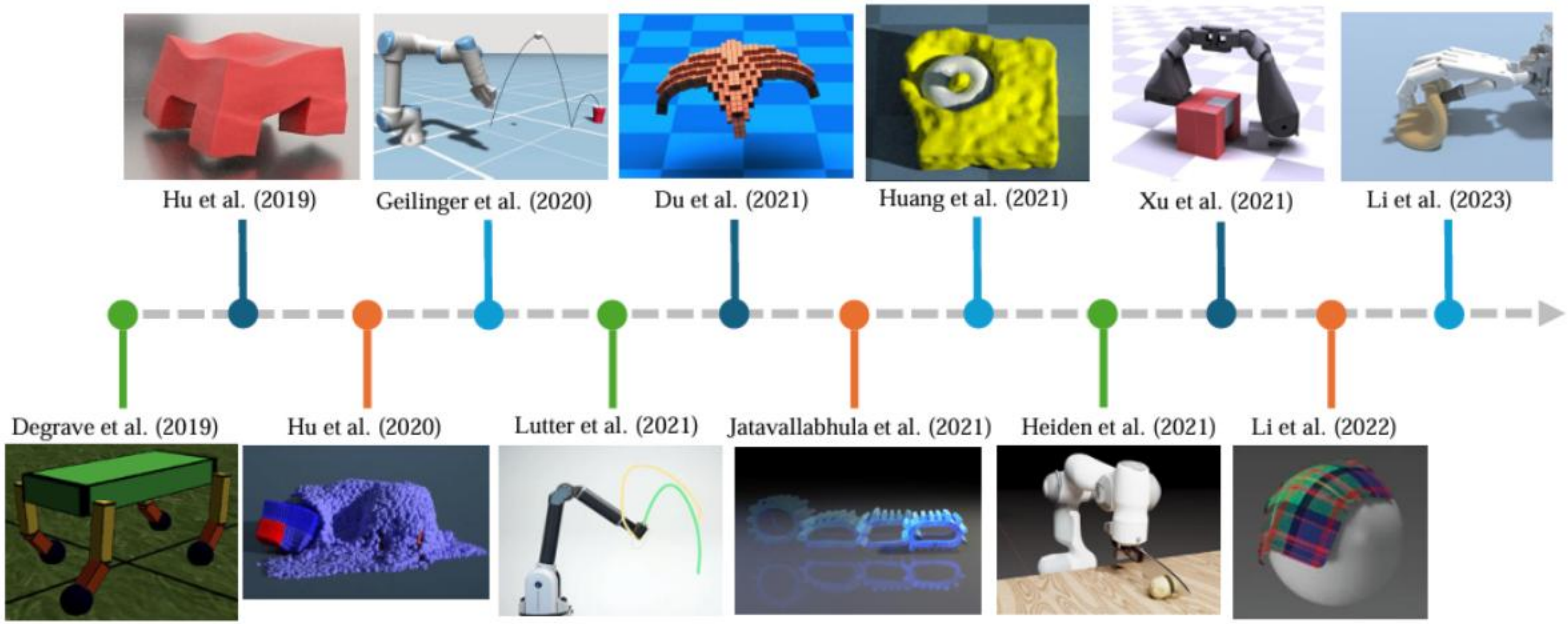
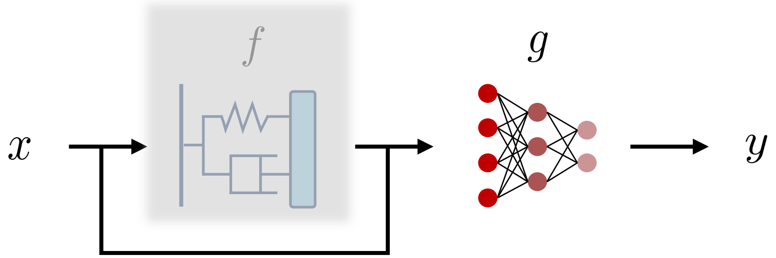


figure from Newbury+24 (and references therein)

(Non)identifiability matters

- Consider an "ML last" hybrid model, especially g being a deep neural net:

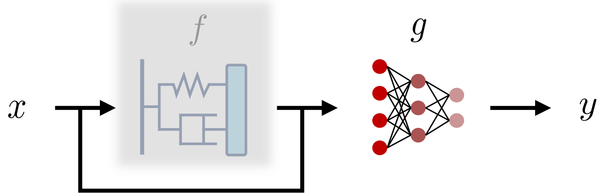
$$y = g\big(x, f(x;\theta);\phi\big)$$



- Regardless of the original identifiability, g makes all of θ nonidentifiable
- Nonidentifiability of ϕ is sometimes fine in practice
- But θ *should be* identified... 🥺
  - they carry the semantics of the scientific model

Regularization for learning deep hybrid models

$$y = g\big(x, f(x;\theta);\phi\big)$$



- Because $L(\theta, \phi_\theta^*) := \min_\phi L(\theta, \phi) \approx \varepsilon$ for any θ, empirical risk minimization:

$$\min_\theta L(\theta, \phi_\theta^*)$$

  alone does not make much sense

- With $R(\theta, \phi)$ to measure "goodness" of hybridization, we should instead do

$$\min_\theta L(\theta, \phi_\theta^*) + \lambda R(\theta, \phi_\theta^*)$$

- What R? 🤔

Hybrid neural ODEs [Yin+ 2021]

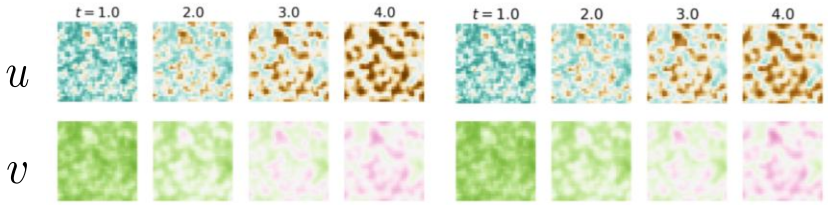
- Physical (misspecified) model of dynamics f + correction by neural net g

$$\boldsymbol{y} = \mathrm{ODEsolve}\left[\frac{d\boldsymbol{s}}{dt} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}; \theta) + \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}; \phi)\right]$$

- Suppress neural net's norm: $R = \|\boldsymbol{g}\|_2^2$
- E.g., learning reaction-diffusion system

$$\frac{\partial u}{\partial t} = a\Delta u + u - u^3 - k - v, \quad \frac{\partial v}{\partial t} = b\Delta u + u - v$$

given as $\boldsymbol{f}$ with $\theta = \{a, b\}$



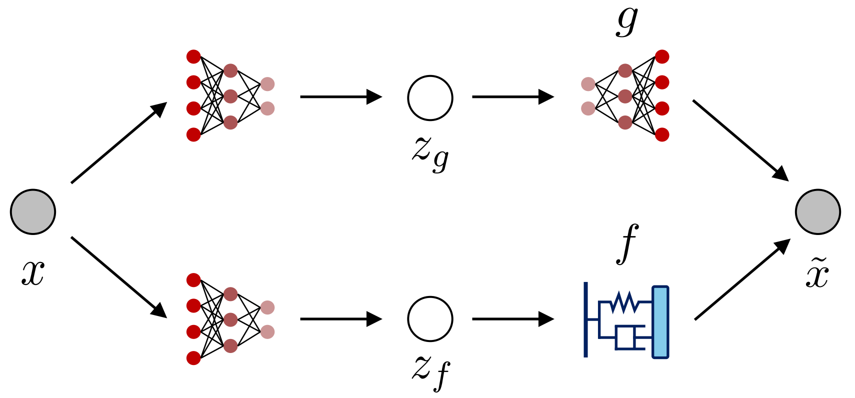
(b) APHYNITY Param PDE (a, b) (c) Ground truth simulation

figure from Yin+ 21

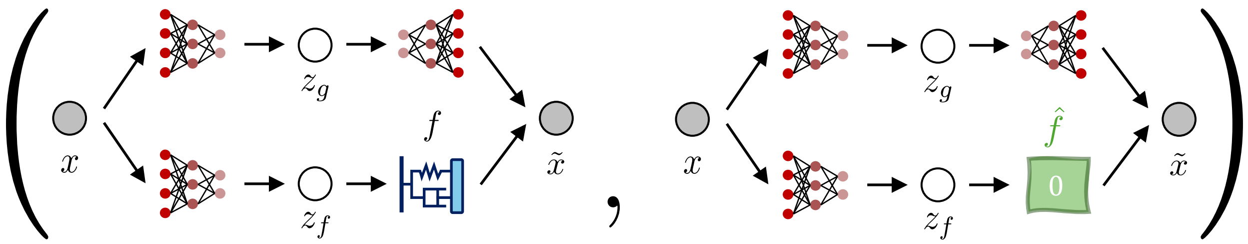
Yin+, Augmenting physical models with deep networks for complex dynamics forecasting, ICLR 2021.

Hybrid VAEs [Takeishi & Kalousis 21]

- VAE with hybrid decoder



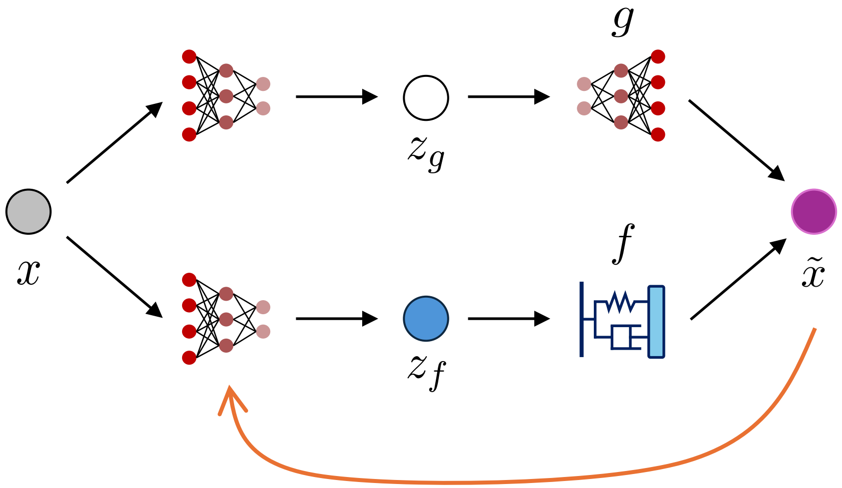
- Minimize the difference between full model (above) and "reduced" model
  - to make a reduced model, replace f with "null augmentation": $\hat{f} = 0$, $\hat{f} = \mathrm{Id}$, etc.

$R = D\Big($  $\Big)$

T & Kalousis, Physics-integrated variational autoencoders for robust and interpretable generative modeling, NeurIPS 2021.

Hybrid VAEs [Takeishi & Kalousis 21]

- VAE with hybrid decoder



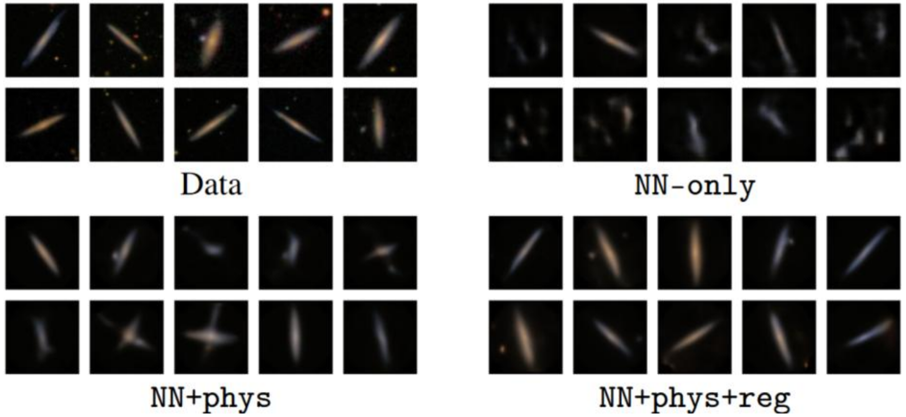
- Data augmentation
  - manipulate z_f (scientific model's latent variable),
  - generate new x, and
  - use it for supervision of the encoder

Hybrid VAEs [Takeishi & Kalousis 21]

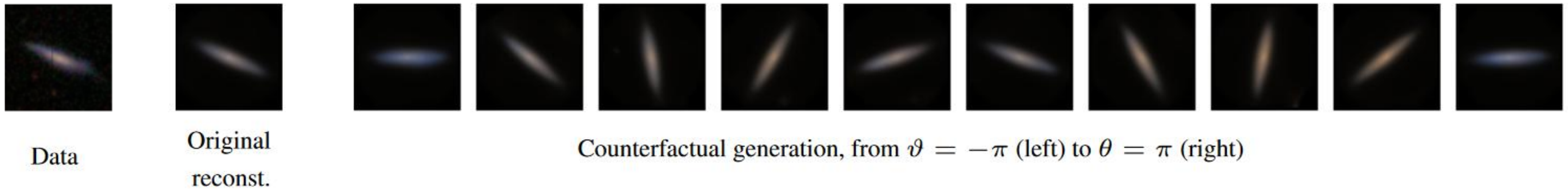
- E.g., controlled generation of galaxy images

  f = exponential profile of light distribution; $\theta = \{\mathrm{intensity}, \mathrm{size}, \mathrm{angle}\}$

  g = anything else (colors, background, etc.)



might seem trivial,
but pure ML cannot do so
easily; cf. disentanglement



What regularizers to use?

- Principle of "least action" of correction term g
  - minimize $\|g\|$
  - minimize $\|g - \mathrm{Id}\|$
  - ☹ needs definition of "least action"
- Scientific model f alone should predict as well as possible
  - minimize $\|y - f(x)\|$
  - maximize dependency between y and $f(x)$
  - ☹ needs dissimilarity or dependency measure; not obvious when y and $f(x)$ live in different spaces
- Two models, f and g, should work independently
  - minimize (nonlinear) correlation of $f(x)$ and $g(x)$
  - ☹ needs dissimilarity measure; again not obvious necessarily
- Total output should be sensitive enough to $f(x)$
  - make $\partial(g \circ f)(x)/\partial f(x)$ large to some extent
  - ☹ to what extent?

...

...... 🤯

Choosing regularizer is hard [Takeishi & Kalousis 23]

- Choice of regularizer solely depends on user's belief: what's a good model?
  - exploratory analysis might help

- Revisit: learning reaction-diffusion system

data: $\dfrac{\partial u}{\partial t} = a\Delta u + u - u^3 - k - v, \quad \dfrac{\partial v}{\partial t} = b\Delta u + u - v$

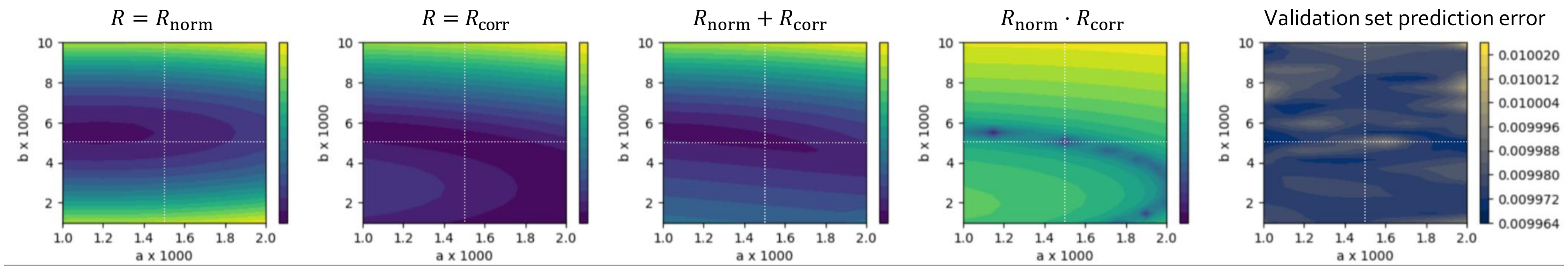
model: $\boldsymbol{y} = \mathrm{ODEsolve}\left[\dfrac{d\boldsymbol{s}}{dt} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}; \theta) + \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}; \phi)\right]$

data-generating value:
$a = 5.0 \times 10^{-3}, b = 1.5 \times 10^{-3}$

- Regularizer landscapes

$R_{\mathrm{norm}} = \|g\|_2, \;\; R_{\mathrm{corr}} = |\langle f, g\rangle|,$ and their combinations

this only hits
the data-generating value



T & Kalousis, Deep grey-box modeling with adaptive data-driven models toward trustworthy estimation of theory-driven models, AISTATS 2023.

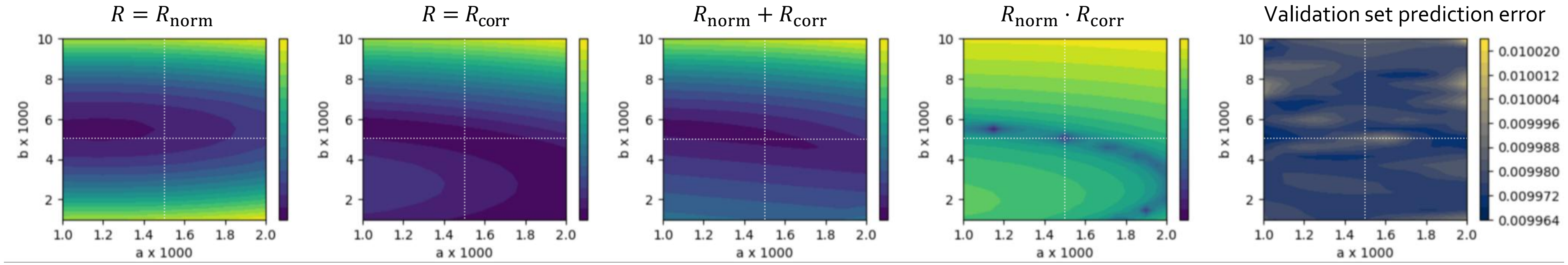
Choosing regularizer is hard [Takeishi & Kalousis 23]

- For drawing the landscapes efficiently, we used an adaptive hybrid model:

$$\boldsymbol{y} = \mathrm{ODEsolve}\left[\frac{d\boldsymbol{s}}{dt} = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}; \theta) + \boldsymbol{g}(\boldsymbol{s}, \boldsymbol{x}, \mathrm{stopgrad}[\boldsymbol{f}(\boldsymbol{s}, \boldsymbol{x}; \theta)]; \phi)\right]$$

  to predict with any θ without re-training

- During training, "marginalize out" θ : minimize $\mathbb{E}_{\theta\sim p(\theta)}[L(\theta, \phi)]$



Next challenge: Hybrid model architecture search

- Common assumptions
  - we know how we should combine two models
  - scientific model is (in a sense) not wrong; e.g., it does not contain irrelevant terms
- Dropping these assumptions, automatic architecture search would be beneficial
- Related work: hybrid model architecture search with LLMs [Holt+ 24]
  - as code generation & refinement through interactions with human experts

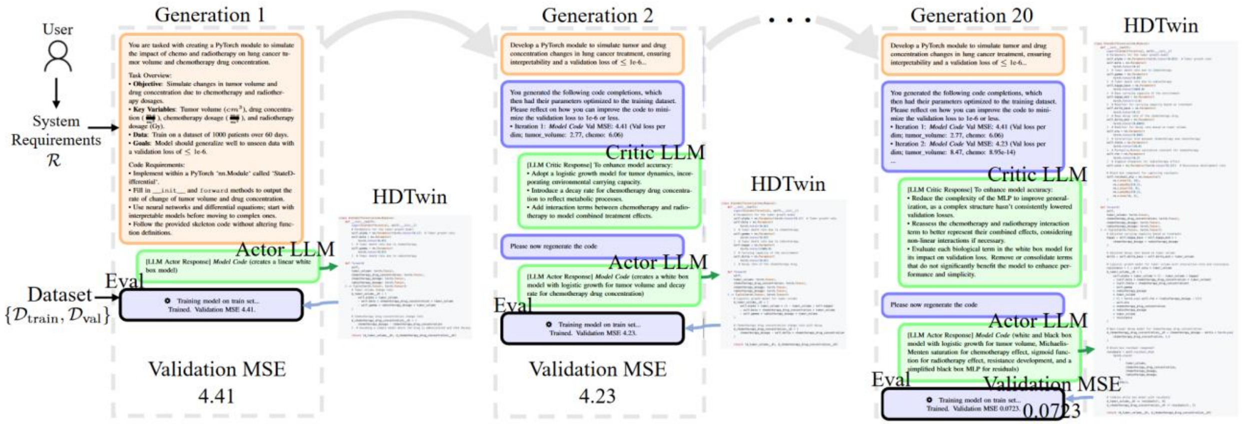
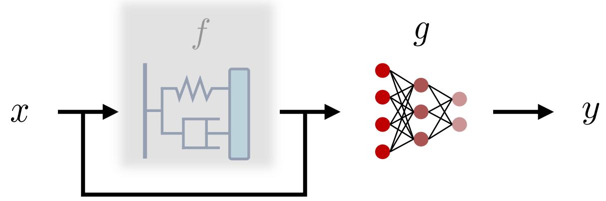


figure from Holt+ 24

Holt+, Automatically learning hybrid digital twins of dynamical systems, NeurIPS 2024.

Summary

- Hybrid modeling
  - to take the best of both worlds: scientific models and machine learning
- Regularization
  - minimize $\|g\|$
  - minimize $\|y - f(x)\|$ ......
  - designing a generic regularizer is difficult



- Open questions
  - architecture search
  - analysis of when to hybridize
  - more real applications
  - bridge to related topics
    - semiparametric models
    - double machine learning