On the Identifiability of Hybrid Deep Generative Models: Meta-Learning as a Solution

Yubo Ye, Maryam Toloubidokhti, Sumeet Vadhavkar, Xiajun Jiang, Huafeng Liu and Linwei Wang

NEURAL INFORMATIOI PROCESSING SYSTEM

浙江大学 1897 ZHEJIANG UNIVERSITY

ROCHESTER INSTITUTE OF TECHNOLOGY R·I·T 1829

Computational Biomedicine Lab

Construction of Hybrid-DGMs:

Prior Knowledge: $f_P(\mathbf{x};\mathbf{z}_P)$ Simplified or Inaccurate

Hybrid Models: $F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N]$ Bridge the Potential Gap

Construction of Hybrid-DGMs:

Prior Knowledge: $f_P(\mathbf{x};\mathbf{z}_P)$ Simplified or Inaccurate

Hybrid Models: $F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N]$ Bridge the Potential Gap

Decoder:

$$\mathbf{x} = F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N] + \varepsilon$$

$$p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) = p_\varepsilon(\mathbf{x} - F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N])$$

Prior: $p(\mathbf{z}_P) = N(\mathbf{z}_P \mid \mu_P, \sigma_P^2 \mathbf{I}),\ p(\mathbf{z}_N) = N(\mathbf{z}_N \mid 0, \mathbf{I})$

Hybrid-DGMs: $p_\theta(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) = p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) p(\mathbf{z}_P) p(\mathbf{z}_N)$

Construction of Hybrid-DGMs:

Prior Knowledge: $f_P(\mathbf{x};\mathbf{z}_P)$ Simplified or Inaccurate

Hybrid Models: $F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N]$ Bridge the Potential Gap

Decoder:

$$\mathbf{x} = F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N] + \varepsilon$$

$$p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) = p_\varepsilon(\mathbf{x} - F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N])$$

Prior: $p(\mathbf{z}_P) = N(\mathbf{z}_P \mid \mu_P, \sigma_P^2 \mathbf{I}),\ p(\mathbf{z}_N) = N(\mathbf{z}_N \mid 0, \mathbf{I})$

Hybrid-DGMs: $p_\theta(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) = p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) p(\mathbf{z}_P) p(\mathbf{z}_N)$

Marginal Likelihood: $p_\theta(\mathbf{x}) = \iint p_\theta(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) d\mathbf{z}_N d\mathbf{z}_N$

Unidentifiability of Hybrid-DGMs:

Observed data $D = \{\mathbf{x}^{(i)}\}_{i=1}^{N}$ generated from $p_\theta(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N)$

Goal: learn θ that satisfies $p_{\theta^*}(\mathbf{x}) = p_\theta(\mathbf{x})$

No guarantee $p_{\theta^*}(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) = p_\theta(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N)$

DGMs with unconditional prior are unidentifiable

Unidentifiability of Hybrid-DGMs:

Observed data $D=\{\mathbf{x}^{(i)}\}_{i=1}^{N}$ generated from $p_\theta(\mathbf{x},\mathbf{z}_P,\mathbf{z}_N)$

Goal: learn θ that satisfies $p_{\theta^*}(\mathbf{x})=p_\theta(\mathbf{x})$

No guarantee $p_{\theta^*}(\mathbf{x},\mathbf{z}_P,\mathbf{z}_N)=p_\theta(\mathbf{x},\mathbf{z}_P,\mathbf{z}_N)$

DGMs with unconditional prior are unidentifiable

To recover $\mathbf{z}_P,\mathbf{z}_N$

Hybrid-DGMs need to satisfy:

$$\forall(\theta,\tilde{\theta}): p_\theta(\mathbf{x})=p_{\tilde{\theta}}(\mathbf{x})\Rightarrow\theta=\tilde{\theta}$$

Construction of Identifiable Hybrid-DGMs

The prior of $\boldsymbol{z}_N$ is conditioned on few-shot support samples $p\left(\mathbf{z}_N \mid D^s\right)$

These support samples share the same $\mathbf{z}_P, \mathbf{z}_N$

This conditional prior is assumed to be factorized exponential family distributions with e dimension of sufficient statistic

$$p_{\mathbf{T},\boldsymbol{\lambda}_\varsigma}\left(\mathbf{z}_N \mid D^s\right)=\prod_{i=1}^{n}\frac{Q_i\left(\mathbf{z}_{N,i}\right)}{Z_i\left(D^s\right)}\exp\left[\mathbf{T}_i(\mathbf{z}_{N,i})^T\boldsymbol{\lambda}_{\varsigma,i}\left(D^s\right)\right]$$

The corresponding parameters conditioned on D^s

$$\boldsymbol{\lambda}_{\varsigma,i}\left(D^s\right)=\frac{1}{\left|D^s\right|}\sum_{\mathbf{x}^s\in D^s}h_{\varsigma,i}\left(\mathbf{x}^s\right)$$

Construction of Identifiable Hybrid-DGMs

The prior of $\mathbf{z}_N$ is conditioned on few-shot support samples $p(\mathbf{z}_N \mid D^s)$

These support samples share the same $\mathbf{z}_P, \mathbf{z}_N$

This conditional prior is assumed to be factorized exponential family distributions with e dimension of sufficient statistic

$$p_{\mathbf{T},\boldsymbol{\lambda}_\varsigma}(\mathbf{z}_N \mid D^s) = \prod_{i=1}^{n} \frac{Q_i(\mathbf{z}_{N,i})}{Z_i(D^s)} \exp\left[\mathbf{T}_i(\mathbf{z}_{N,i})^T \boldsymbol{\lambda}_{\varsigma,i}(D^s)\right]$$

The corresponding parameters conditioned on D^s

$$\boldsymbol{\lambda}_{\varsigma,i}(D^s) = \frac{1}{|D^s|} \sum_{\mathbf{x}^s \in D^s} h_{\varsigma,i}(\mathbf{x}^s)$$

Conditional Hybrid-DGMs:

$$p_\phi(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) = p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) p(\mathbf{z}_P) p_{\mathbf{T},\boldsymbol{\lambda}_\varsigma}(\mathbf{z}_N \mid D^s)$$

Learn-to-Identify Hybrid-DGMs via Meta-learning

Dataset $D = \{D_m\}_{m=1}^{M}$ with M different latent generated variables $[\mathbf{z}_P, \mathbf{z}_N]$

$D_m = \{D_m^s, D_m^q\}$, support samples $D_m^s = \{\mathbf{x}^{s,i}\}_{i=1}^{k}$, query samples $D_m^q = \{\mathbf{x}^{q,i}\}_{i=1}^{l}$, $k << l$

Learn-to-Identify Hybrid-DGMs via Meta-learning

Dataset $D = \{D_m\}_{m=1}^{M}$ with M different latent generated variables $[\mathbf{z}_P, \mathbf{z}_N]$

$$D_m = \{D_m^s, D_m^q\}, \text{support samples } D_m^s = \{\mathbf{x}^{s,i}\}_{i=1}^{k}, \text{ query samples } D_m^q = \{\mathbf{x}^{q,i}\}_{i=1}^{l}, k << l$$

Amortized Variational Inference:

$$q_\eta(\mathbf{z}_P \mid \mathbf{x}) \approx p(\mathbf{z}_P \mid \mathbf{x}) \qquad q_\varsigma(\mathbf{z}_N \mid \mathbf{x} \cup D^s) \approx p_\varsigma(\mathbf{z}_N \mid \mathbf{x}, D^s)$$

ELBO:

$$\sum_{m=1}^{M} \sum_{\mathbf{x}^q \in D_m^q} \log p(\mathbf{x}^q \mid D_m^s) \geq \sum_{m=1}^{M} \sum_{\mathbf{x}^q \in D_m^q} \{E_{q_\eta(\mathbf{z}_P|\mathbf{x}^q), q_\varsigma(\mathbf{z}_N|\mathbf{x}^q \cup D_m^s)}[\log p_\theta(\mathbf{x}^q \mid \mathbf{z}_P, \mathbf{z}_N)]$$
$$- \mathrm{KL}(q_\eta(\mathbf{z}_P \mid \mathbf{x}^q) \| p(\mathbf{z}_P)) - \mathrm{KL}(q_\varsigma(\mathbf{z}_N \mid \mathbf{x}^q \cup D_m^s) \| p_\varsigma(\mathbf{z}_N \mid D_m^s))$$

Identifiability Theory:

Conditional Hybrid-DGMs with parameter $(\theta, \mathbf{T}, \boldsymbol{\lambda}_\varsigma)$

$$p_\phi(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N) = p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) p(\mathbf{z}_P) p_{\mathbf{T}, \boldsymbol{\lambda}_\varsigma}(\mathbf{z}_N \mid D^s)$$

$$p_\theta(\mathbf{x} \mid \mathbf{z}_P, \mathbf{z}_N) = p_\varepsilon(\mathbf{x} - F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N])$$

$$p_{\mathbf{T}, \boldsymbol{\lambda}_\varsigma}(\mathbf{z}_N \mid D^s) = \prod_{i=1}^{n} \frac{Q_i(\mathbf{z}_{N,i})}{Z_i(D^s)} \exp\left[\mathbf{T}_i(\mathbf{z}_{N,i})^T \boldsymbol{\lambda}_{\varsigma,i}(D^s)\right]$$

Definition 1: DGMs with parameter ϕ is ~-identifiable if

$$\forall(\phi, \tilde{\phi}): p_\phi(\mathbf{x}) = p_{\tilde{\phi}}(\mathbf{x}) \Rightarrow \phi \sim \tilde{\phi}$$

Definition 2: The ~ equivalence relation on Hybrid-DGMs defined as follows:

$$(\theta, \mathbf{T}, \boldsymbol{\lambda}_\varsigma) \sim (\tilde{\theta}, \tilde{\mathbf{T}}, \tilde{\boldsymbol{\lambda}}_\varsigma) \Leftrightarrow \exists \mathbf{A}, \mathbf{c} : \mathbf{T}(F_\theta^{-1}(\mathbf{x})) = \mathbf{A}\tilde{\mathbf{T}}(F_{\tilde{\theta}}^{-1}(\mathbf{x})) + \mathbf{c}, \text{ for all } \mathbf{x}$$

$$\Leftrightarrow \exists \mathbf{A}, \mathbf{c} : [\mathbf{z}_P, \mathbf{z}_N] = \mathbf{A}[\tilde{\mathbf{z}}_P, \tilde{\mathbf{z}}_N] + \mathbf{c}, \text{ for all } \mathbf{x}$$

If $\mathbf{A}$ is invertible, we denote this relation by $\sim_A$-identifiable

Theorem:

Assume we observe data sampled from $p_{\phi}(\mathbf{x}, \mathbf{z}_P, \mathbf{z}_N)$ with parameter $\phi = (\theta, \mathbf{T}, \boldsymbol{\lambda}_{\varsigma})$

Also assume the following holds:

(i) The set $\{\mathbf{x} \in X \mid \varphi_{\varepsilon}(\mathbf{x}) = 0\}$ has measure zero, where φ_{ε} is the characteristic function of the density p_{ε} defined in $p_{\varepsilon}(\mathbf{x} - F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N])$

(ii) The hybrid mixing function $F_{\theta} = F[f_P, f_{N_\theta}; \mathbf{z}_P, \mathbf{z}_N]$ is injective

(iii) The sufficient function $T_{i,j}$ are differentiable almost everywhere, and linearly independent on any subset of X of measure greater than zero

(iv) There exist $ne+1$ distinct support samples $D^{s,0}, D^{s,1}, \ldots, D^{s,ne}$ such that $ne \times ne$ matrix $\mathbf{L}$ defined as follows is invertible

$$\mathbf{L} = \left(\boldsymbol{\lambda}_{\varsigma}\left(D^{s,1}\right) - \boldsymbol{\lambda}_{\varsigma}\left(D^{s,0}\right), \ldots, \boldsymbol{\lambda}_{\varsigma}\left(D^{s,ne}\right) - \boldsymbol{\lambda}_{\varsigma}\left(D^{s,0}\right)\right)$$

Then after learning (maximum the ELBO),

the learned parameter $\widetilde{\phi}$ and the true parameter ϕ are $\sim_A$-identifiable

Experiment: Synthetic Datasets

(1) Forced damped pendulum system governed by:

$$\frac{d^2\varphi(t)}{dt^2}+\omega^2\sin\varphi(t)+\xi\frac{d\varphi(t)}{dt}-A\cos(2\pi\phi t)=0$$

Prior knowledge: $\frac{d^2\varphi(t)}{dt^2}+\omega^2\sin\varphi(t)=0$

(2) Advection-diffusion system governed by:

$$\frac{\partial T}{\partial t}-a\frac{\partial^2 T}{\partial t^2}+b\frac{\partial T}{\partial s}=0$$

Prior knowledge: $\frac{\partial T}{\partial t}-a\frac{\partial^2 T}{\partial t^2}=0$

(3) Double Pendlum system governed by:

$$\frac{d}{dt}\begin{bmatrix}\dot{\varphi}_1\\ \dot{\varphi}_2\end{bmatrix}=\begin{bmatrix}\dfrac{G\sin\varphi_2\phi_1-\phi_2(L_1\dot{\varphi}_1^2\phi_1+L_2\dot{\varphi}_2^2)-(\widetilde{m}+1)G\sin\varphi_1}{L_1\left(\widetilde{m}+\phi_2^2\right)}\\ \dfrac{(\widetilde{m}+1)(L_1\dot{\varphi}_1^2\phi_2-G\sin\varphi_2+G\sin\varphi_1\phi_1)+L_2\dot{\varphi}_2^2\phi_1\phi_2}{L_2\left(\widetilde{m}+\phi_2^2\right)}\end{bmatrix}$$

Prior knowledge:

$\widetilde{m}=1$

Baselines:

1. APHYNITY; 2.Hybrid-VAE

Proposed Method: Meta-Hybrid-VAE

Identifiable Results:

MCC evaluate $\sim_A$-identifiable relation
between true parameters and learned parameter $[\boldsymbol{z}_P, \boldsymbol{z}_N]$

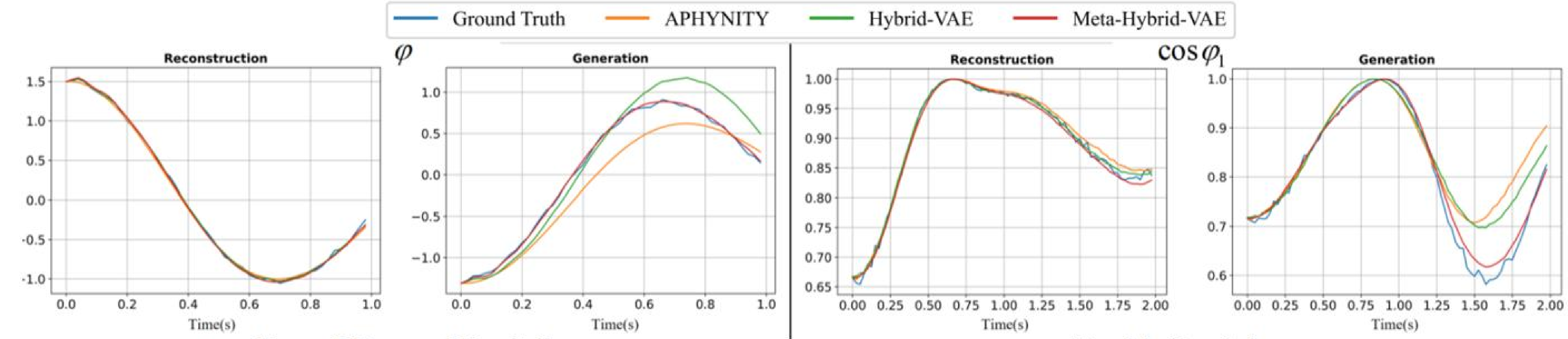
	Forced Damped Pendulum			Advection-Diffusion System		
	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE
MSE of $\mathbf{z}_P$ ↓	6.96(0.01)e-2	4.14(0.29)e-2	**1.59(0.07)e-2**	1.9(0.3)e-2	7.12(5.32)e-4	**1.34(0.75)e-4**
MCC ↑	0.79(0.02)	0.59(0.03)	**0.99(0.00)**	0.98(0.01)	0.94(0.00)	**0.99(0.00)**

	Double Pendulum		
	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE
MSE of $\mathbf{z}_P$ ↓	4.58(0.04)e-1	3.97(0.06)e-1	**3.85(0.19)e-1**
MCC ↑	0.50(0.00)	0.51(0.00)	**0.98(0.00)**

Performance on reconstruction and prediction (generation) task :

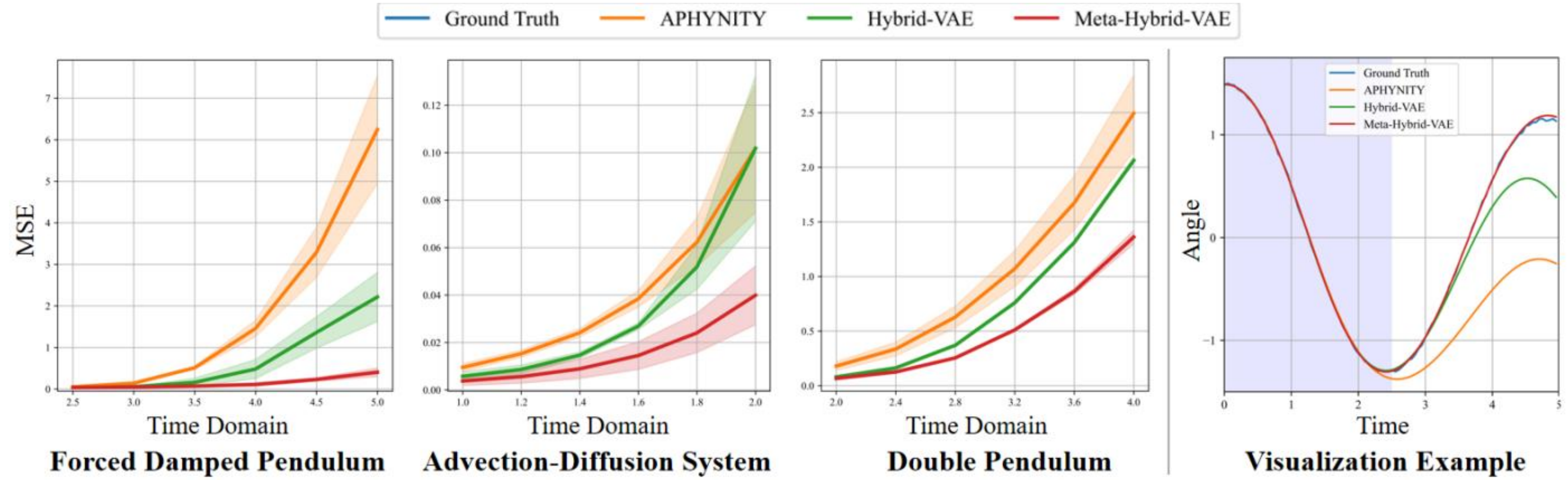
	Forced Damped Pendulum			Advection-Diffusion System		
	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE
MSE of **x** (Rec) ↓	5.2(0.01)e-2	**2.66(0.03)e-2**	2.91(0.06)e-2	6.22(0.41)e-3	4.90(0.75)e-3	**2.93(0.22)e-3**
MSE of **x** (Pre) ↓	2.37(0.67)	1.74(0.20)	**6.85(1.11)e-2**	8.25(0.58)e1	8.53(5.72)e-2	**5.63(1.34)e-3**

	Double Pendulum		
	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE
MSE of **x** (Rec) ↓	4.88(0.00)e-2	4.05(0.25)e-2	**3.83(0.06)e-2**
MSE of **x** (Pre) ↓	1.29(0.23)e1	5.52(0.06)	**2.80(0.48)e-1**

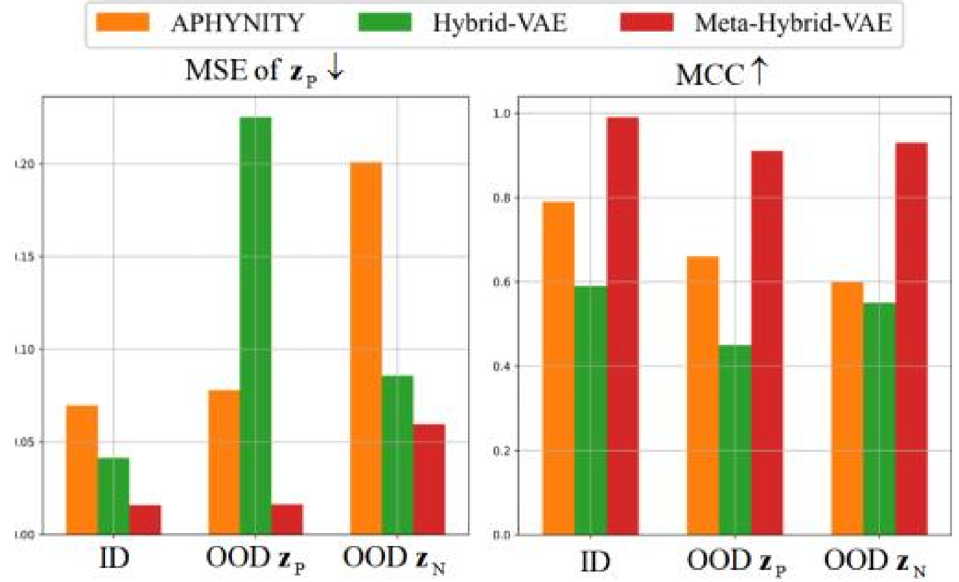


Forced Damped Pendulum **Double Pendulum**

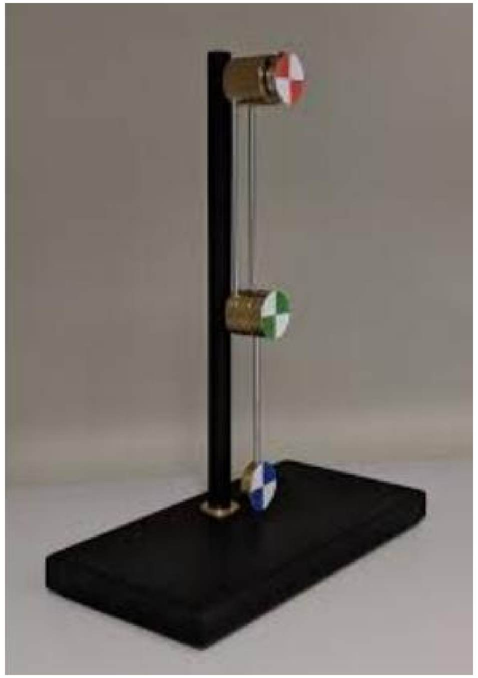
Performance over longer time domains :



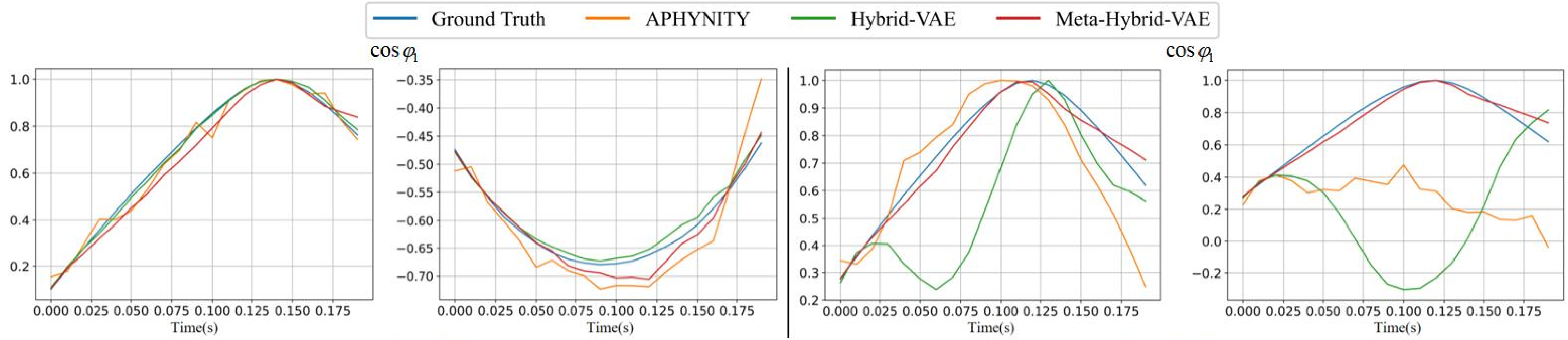
Out-of-distribution Performance :



Results of real world dataset-double pendulum:



	Real Double Pendulum		
	APHYNITY	Hybrid-VAE	Meta-Hybrid-VAE
MSE of $\mathbf{z}_P$ ↓	/	/	/
MCC ↑	/	/	/
MSE of $\mathbf{x}$ (Rec) ↓	3.78(0.52)e-2	**2.20(0.14)e-3**	2.67(0.34)e-2
MSE of $\mathbf{x}$ (Pre) ↓	5.16(0.69)	1.87(0.13)	**2.88(0.34)e-2**



Reconstruction **Prediction**

Conclusion:

1. In this work, we propose a meta-learning based method to make hybrid-DGMs achieve A identifiable and provide the the corresponding theoretical and empirical evidence is provided

2. Our experiments show that, with identifiability, hybrid-DGMs can perform better on many downstream tasks (like generation, long time domian predction and OOD).

Conclusion:

1. In this work, we propose a meta-learning based method to make hybrid-DGMs achieve A identifiable and provide the the corresponding theoretical and empirical evidence is provided

2. Our experiments show that, with identifiability, hybrid-DGMs can perform better on many downstream tasks (like generation, long time domian predction and OOD).

Limitation:

While the meta-formulation foregoes explicitly observed auxiliary variable, it does assume the ability to pair support and query samples from the same data generation process, which is not always available..