Physics-Informed Learning via Diffusion Framework for System State Estimation

Tianyi Zeng
Shanghai Jiao Tong University
Shanghai 200240, China
zengtianyi@sjtu.edu.cn

Tianyi Wang*
The University of Texas at Austin
Austin, TX 78712, USA
bonny.wang@utexas.edu

Jiaru Zhang
Shanghai Jiao Tong University
Shanghai 200240, China
jiaruzhang@sjtu.edu.cn

Yajie Zou
Tongji University
Shanghai 201804, China
yajiezou@hotmail.com

Yangyang Wang
Tongji University
Shanghai 201804, China
wyangyang@tongji.edu.cn

Junfeng Jiao
The University of Texas at Austin
Austin, TX 78712, USA
jjiao@austin.utexas.edu

Christian Claudel
The University of Texas at Austin
Austin, TX 78712, USA
christian.claudel@utexas.edu

Xinbo Chen
Tongji University
Shanghai 201804, China
chenxinbo@tongji.edu.cn

Abstract

We propose PILD (Physics-Informed Learning via Diffusion), a novel state estimation framework that combines a diffusion method with physical information. PILD introduces a physics-informed conditional embedding module in diffusion process, ensuring that physical observations stably guide the network's parameter updates. Moreover, we propose a first-principle-based joint loss, achieving an elegant mathematical unification between the diffusion denoising loss and physical loss to ensure the training consistency with physical laws. Experimental results on different systems demonstrate that our PILD achieves superior accuracy, and exhibits strong generalization capability. This framework offers a promising direction for the development of more accurate and more robust engineering systems.

1 Introduction

State estimation for complex engineering systems with noise remains a challenging task mainly because of the complex nonlinear relationship between inputs and outputs [1, 2]. Physics-informed neural networks (PINNs) have demonstrated effectiveness in solving both forward (inference) and inverse (identification) partial differential equation (PDE) problems in nonlinear systems, as well as offering straightforward implementation [3, 4]. Specifically, PINNs can infer unknown parameters in PDEs and reconstruct solutions from partial observations, making them valuable for addressing

*Corresponding Author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: UrbanAI: Harnessing Artificial Intelligence for Smart Cities.

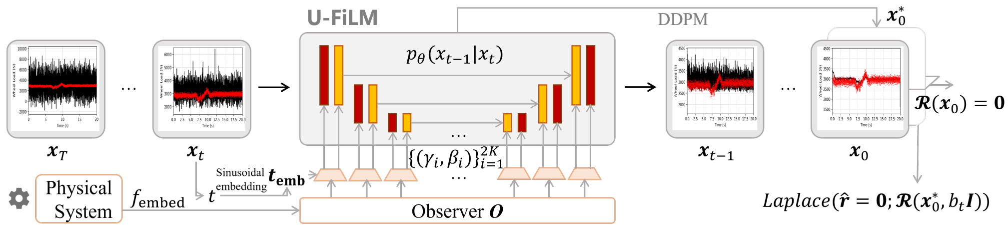

Figure 1: Overview of our PILD (Physics-Informed Learning via Diffusion) framework.

complex state estimation challenges [5]. However, PINNs often lead to non-convergence issues and low accuracy due to system uncertainty and nonlinearity [6].

Diffusion models, where data distributions are learnt by adding noise to data and training neural networks to reverse the process, effectively denoise the data and capture complex patterns [7]. Owing to their rigorous mathematical interpretability grounded in probability theory, diffusion models exhibit a strong capability to learn complex data distributions, leading to state-of-the-art (SOTA) performance across various generative tasks [8]. Even though some physics-informed diffusion models [9, 10] have been proposed to integrate denoising diffusion models with PINNs, such methods still suffer from inconsistency between data distribution and physical laws.

The main contributions of this paper are as follows:

- **PILD (Physics-Informed Learning via Diffusion) Framework** (Figure 1), which incorporates both physical priors and physical posteriors leveraging diffusion models.
- **Physics-Informed Conditional Embedding Module**, which ensures that physical priors can stably guide the update of network parameters.
- **Denoising-PINN Joint Loss**, which probabilistically models complex systems with outliers, unifying physical loss and denoising loss based on the first-principle.

2 Methodology

2.1 Physics-Informed Conditional Embedding

The physics observer $\boldsymbol{O}$ is formally defined as:

$$\boldsymbol{O} = f_{\text{embed}}(\boldsymbol{c}, \boldsymbol{u}_{\text{phys}}) := [\boldsymbol{c}' \,||\, \mathcal{S}(\boldsymbol{u}_{\text{phys}})], \tag{1}$$

where $\boldsymbol{c}'$ is a sub-vector of salient variables from $\boldsymbol{c}$, regarded as key sensor data, $||$ denotes concatenation, and $\mathcal{S}(\cdot)$ is a standardization operator that scales the physics-based prior to have zero mean and unit variance, aligning its distribution with the internal feature space of the diffusion model. The design of this observer enables both the physical model outputs $\boldsymbol{u}_{\text{phys}}$ and key sensor data $\boldsymbol{c}'$ to be fed into the neural network, allowing the network to better learn the physical knowledge.

To ensure the physics observer $\boldsymbol{O}$ exerts a profound and adaptive influence, we integrate it into the network backbone using FiLM (Feature-wise Linear Modulation) [11]. Our denoising network with the residual FiLM blocks $\mathcal{G}_{\text{FiLM}}(\boldsymbol{O}, \boldsymbol{t}_{\text{emb}})$ is equipped with a K-level-depth U-Net structure, denoted as **U-FiLM**, where $\boldsymbol{t}_{\text{emb}}$ is sinusoidal embedding at timestep t. This yields a time-adaptive, scale-aware modulation: coarse physical guidance at large t refines to fine corrections as $t \to 0$.

Each residual block receives its unique pair $(\boldsymbol{\gamma}_i, \boldsymbol{\beta}_i)$ as:

$$\{(\boldsymbol{\gamma}_i, \boldsymbol{\beta}_i)\}_{i=1}^{2K} = \mathcal{G}_{\text{FiLM}}(\boldsymbol{O}, \boldsymbol{t}_{\text{emb}}), \tag{2}$$

and performs the affine transformation on its feature map $\boldsymbol{h}_i$:

$$FiLM(\boldsymbol{h}_i; \boldsymbol{\gamma}_i, \boldsymbol{\beta}_i) = (1 + \boldsymbol{\gamma}_i) \odot \boldsymbol{h}_i + \boldsymbol{\beta}_i, \tag{3}$$

where $\odot$ is element-wise multiplication. Here, we adjust the coefficient to $(1 + \gamma_i)$ instead of γ_i mentioned according to 11.

2.2 Unified Physics-Informed Training Objective

Considering the complex noise present in engineering systems and the Laplace distribution's excellent modeling performance for outliers [12], we model that the virtual residual observable $\hat{\boldsymbol{r}}$ is sampled from a conditional Laplace distribution $q_{\boldsymbol{\mathcal{R}}}(\hat{\boldsymbol{r}}|\boldsymbol{x}_0) = Laplace(\hat{\boldsymbol{r}}; \boldsymbol{\mathcal{R}}(\boldsymbol{x}_0), b_t\boldsymbol{I})$. The probability density function of a multivariate Laplace distribution with zero mean and scale b is proportional to $\exp(-\frac{1}{b}\|\boldsymbol{x}\|_1)$. Minimizing the negative log-likelihood $-\log q_{\boldsymbol{\mathcal{R}}}(\hat{\boldsymbol{r}} = \boldsymbol{0}|\boldsymbol{x}_0)$ would yield the term proportional to $\|\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{x}}_0)\|_1/b_t$. Thus, the overall training objective is interpreted as optimizing a joint log-likelihood over both the data samples and these virtual residual observables:

$$\begin{aligned}\arg\max_{\boldsymbol{\theta}}\{&\mathbb{E}_{\boldsymbol{x}_0\sim q(\boldsymbol{x}_0)}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{O})]+\\ &\mathbb{E}_{\boldsymbol{x}_0\sim p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{O})}[\log q_{\mathcal{R}}(\hat{\boldsymbol{r}}=\boldsymbol{0}|\boldsymbol{x}_0)]\}.\end{aligned} \tag{4}$$

Directly optimizing this joint likelihood involving sampling from $p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{O})$ is computationally challenging [13], our unified loss $\mathcal{L}_{\text{PILD}}$ in Equation 5 then provides a computationally tractable surrogate. The total loss $\mathcal{L}_{\text{PILD}}$ is formulated as a single expectation over the combined objectives, ensuring an elegant integration of data and robust physics:

$$\begin{aligned}&\mathcal{L}_{\text{PILD}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim[1,T],\boldsymbol{x}_0\sim q(\boldsymbol{x}_0),\boldsymbol{O},\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\\ &\left[\lambda_t\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta\|^2 + \frac{1}{b_t}G(t)\|\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{x}}_0)\|_1\right],\end{aligned} \tag{5}$$

where $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, \boldsymbol{O}, t)$ and $\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{x}}_0(\boldsymbol{x}_t, \boldsymbol{O}, t))$ have been abbreviated as $\boldsymbol{\epsilon}_\theta$ and $\boldsymbol{\mathcal{R}}(\hat{\boldsymbol{x}}_0)$ due to typesetting constraints, λ_t is a time-dependent Min-SNR weighting [14] that naturally balances the denoising objective across different noise levels, fundamentally arising from the score matching equivalence for maximizing data likelihood [13], $b_t = B_t/c$ is a rescaled scale parameter, $c > 0$ is a single hyperparameter effectively dictating the strength of the physical penalty, B_t is the fixed variance of denoising process which has been mentioned before, $G(t) = \log(1 + \frac{T}{t})$ weakens physical penalties when t is larger, and T is total step.

2.3 Diffusion Inference and Sampling

Given the set of conditional variables $\boldsymbol{c}$, and the corresponding Physics Observer $\boldsymbol{O}$, the sampling process then proceeds as outlined in Algorithm 1.

Algorithm 1 PILD Model Sampling

Input: Trained noise prediction network $\boldsymbol{\epsilon}_\theta$, conditional variables $\boldsymbol{c}$, number of diffusion steps T, variance schedule $\overline{\alpha}_t$, β_t.
Output: Generated sample $\boldsymbol{x}_0$.

Construct Physics Observer $\boldsymbol{O}$ from $\boldsymbol{c}$.
Draw initial sample $\boldsymbol{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
for $t = T$ **down to** 1 **do**
  Predict noise using the trained network: $\hat{\boldsymbol{\epsilon}} = \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, \boldsymbol{O}, t)$
  Compute the mean of the reverse process posterior:
  $\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\hat{\boldsymbol{\epsilon}}\right)$
  Calculate the variance for the current step: $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$
  Sample the next state: $\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \sigma_t\boldsymbol{z}$
  where $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\boldsymbol{z} = \mathbf{0}$ for $t = 1$.
end for
return $\boldsymbol{x}_0$ =0

The network training is conducted on a 4090/24G GPU. The dimensions of each layer in U-FiLM are 128, 256, 512, and 1024, respectively. Experimental verification shows that it is recommended that the time embedding dimension of the encoding network should be at least larger than the dimension of the input vector, which is set to 96 in this paper. The diffusion step count T is 200, and β_T is 0.02. In the loss function, hyperparameter c is set to 0.1.

3 Experiments

To verify the role of PILD in both providing stable physical guidance and improving physical consistency of training, we perform experiments on dynamic wheel load estimation.

Table 1: Quantitative results.

Method	Wheel Load Estimation			Trajectory Tracking					
	Aggressive	Smooth	Sporty	Downtown Driving			Rural Driving		
	$e_F\downarrow$	$e_F\downarrow$	$e_F\downarrow$	$e_{x\&y}\downarrow$	$e_{\psi}\downarrow$	$e_{v_x}\downarrow$	$e_{x\&y}\downarrow$	$e_{\psi}\downarrow$	$e_{v_x}\downarrow$
Basic									
FCN [15]	1257.633	601.559	794.780	9.567	7.977	14.284	9.623	6.606	10.972
ResNet [16]	1121.367	585.561	647.976	9.236	6.899	9.094	5.517	10.361	9.315
PINN									
PINN-FCN [17]	1012.849	551.508	654.644	8.824	14.488	17.287	8.279	10.349	15.251
PINN-ResNet [17]	994.7839	642.110	653.944	5.923	9.753	11.286	3.729	9.134	8.248
B-PINN [18]	1002.531	598.617	661.279	5.867	6.597	10.563	6.034	6.511	9.656
LSTM									
LSTM-RNN [19]	1302.589	720.567	788.163	8.953	10.199	11.298	9.643	8.831	8.168
LSTM-GA [1]	1150.434	582.840	642.638	9.019	6.579	8.442	9.005	5.994	10.837
Diffusion									
DDPM [13]	980.328	560.796	660.897	9.653	6.109	4.317	4.697	7.138	4.936
PIDM [20]	988.642	579.546	630.543	4.845	**4.733**	5.671	4.293	5.874	4.061
PILD (Ours)	**958.578**	**520.631**	**607.274**	**4.796**	5.391	**3.048**	**3.562**	**5.837**	**3.033**

3.1 Wheel Load Estimation

Since tire parameters change with temperature, we treat the tire temperature data collected by sensors as an independent variable, and thus the system is expressed as PDEs [1]. The wheel load is:

$$\boldsymbol{F_z} = \boldsymbol{\mathcal{P}}(D_{\mathrm{v}}; \boldsymbol{c}), \tag{6}$$

where D_{v} is the set of vehicle parameters, and $\boldsymbol{c}$ is the set of sensors' data. We evaluate our approach on a racing car chassis dynamic dataset [1] with three sets of working conditions, including aggressive, sporty, and smooth driving.

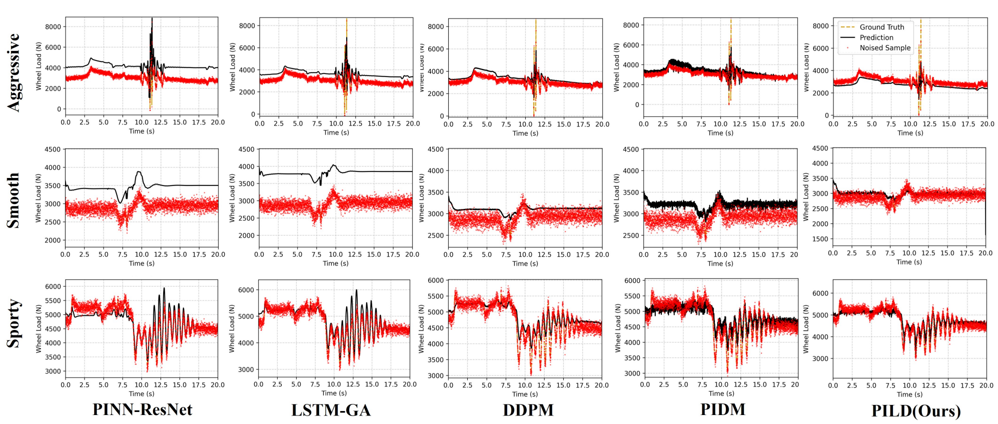

Figure 2: Experiment results on wheel load estimation tasks.

The quantitative results are presented in Table 1. The bold data is the best one and the underlined one has the second best performance. e_F is the RMSE error of wheel load (N). We select five methods for demonstration in Figure 2. Since all sensors for load estimation are installed in the vehicle chassis, the collected data contain significant noise. Different type of vehicles have totally different chassis dynamics, posing significant challenges to physical consistency.

Basic methods perform poorly in this task, as they are greatly affected by both noise and unknown model parameters. The PINN methods are also misguided by idealized physical models. The LSTM methods fail to effectively extract patterns from unknown models with noisy data. Among these methods, only the diffusion methods achieve better performance. Furthermore, we observe that the results of the PIDM method contain noise. This is because the modeling of the Gaussian distribution for physical residuals in PIDM is less effective in handling system outliers compared to the Laplace distribution used in the PILD. The proposed **PILD** method, which ensures stable physical guidance and training consistency, demonstrates the best generalizability and results.

3.2 Trajectory Tracking

Vehicle tracking requires a vehicle dynamic model to predict the future state. The Ackerman steering model is provided in the supplementary materials, including longitudinal, lateral, and yaw motions. The system is treated as an ODE system.

Using the Ackerman model, we compute the position, yaw angle, and speed of the vehicle at next time step via the following concise equations:

$$[x_{t+1}, y_{t+1}, \psi_{t+1}, v_{x_{t+1}}]^\top = \boldsymbol{\mathcal{P}}(D_\mathrm{v}; [x_t, y_t, \psi_t, v_{x_t}]^\top, \boldsymbol{c}), \tag{7}$$

where $[\cdot]\top$ is the vehicle state vector (x coordinate, y coordinate, yaw angle, forward velocity), the system transfer physical model is denoted as $\boldsymbol{\mathcal{P}}$, D_v is vehicle's parameter set, and $\boldsymbol{c}$ is the set of all required sensor's data.

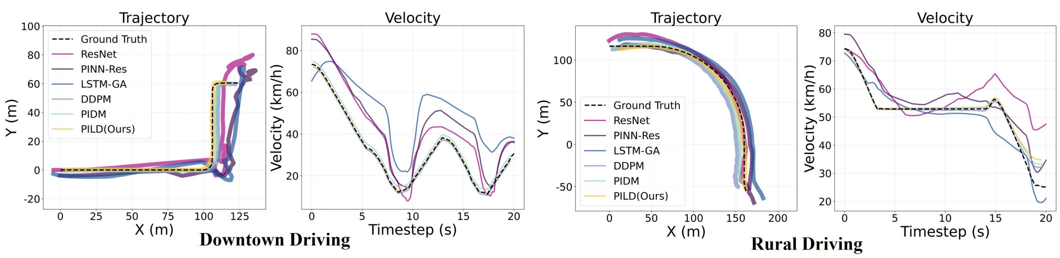

Figure 3: Experiment results on trajectory tracking tasks.

The quantitative results are presented in Table 1. $e_{\{x\&y,\psi,v_x\}}$ in the table is RMSE error of position (m), yaw (deg) and velocity (km/h). We take two scenarios for demonstration in Figure 3. Results show that our method performs well in the tracking task. When dealing with vehicle parameters that have never been encountered before, the Basic methods produce large errors due to the lack of physical guidance. Meanwhile, the PINN methods may be misled by fixed physical models, preventing the network from balancing the differences between physical models and real-world data. The LSTM methods also fail to achieve optimal results due to the randomness of trajectory sequences and the uncertainty of the model itself. Compared with other diffusion methods, the **PILD** incorporates stable guidance from a physical observer and uses a Laplace distribution to model system outliers, demonstrating good results.

4 Conclusion

This paper propose a physics informed learning framework via diffusion (**PILD**), aiming to solve the state estimation problem for complex engineering systems. In our **PILD** framework, the conditional physical observer is deeply embedded, and the diffusion method is introduced to generate the system state. The proposed **PILD** loss function leverages Laplace distribution in handling outliers of complex engineering systems and unifies the physical loss and diffusion denoising loss from the first principles. The experimental results show that **PILD** can achieve favorable performance under different conditions. However, the proposed method currently has limitations: due to the complexity of its network structure, it is more computationally intensive than non-diffusion methods. Future research will focus on reducing the computational cost of our method while maintaining superior performance.

References

[1] T. Zeng, T. Wang, L. Yu, Z. Liu, H. Chen, and X. Chen, "Wheel load estimation and anti-roll bar control using suspension analysis with neural network," in *IFToMM World Congress on Mechanism and Machine Science*. Springer, 2023, pp. 332–342.

[2] T. Zeng, Z. Liu, C. He, Z. Zeng, H. Chen, F. Zhang, K. Fu, and X. Chen, "Analysis and design of suspension state observer for wheel load estimation," in *SAE Technical Paper 2024-01-2285*, SAE, Ed. Warrendale, PA: SAE International, 2024, p. 09 Apr 2024.

[3] X. Guan, X. Wang, H. Wu, Z. Yang, and P. Yu, "Efficient bayesian inference using physics-informed invertible neural networks for inverse problems," *Machine Learning: Science and Technology*, vol. 5, no. 035026, 2024.

[4] Y. Sun, D. Mukherjee, and Y. Atchadé, "On the estimation rate of bayesian pinn for inverse problems," *arXiv preprint arXiv:2406.14808v1*, 2024.

[5] J. Long, A. Q. M. Khaliq, and K. M. Furati, "Identification and prediction of time-varying parameters of covid-19 model: a data-driven deep learning approach," *arXiv preprint arXiv:2103.09949v1*, 2021.

[6] P. Rathore, W. Lei, Z. Frangella, L. Lu, and M. Udell, "Challenges in training pinns: A loss landscape perspective," in *Proceedings of the 41st International Conference on Machine Learning*. Vienna, Austria: PMLR, 2024, pp. 1–13.

[7] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M.-H. Yang, "Diffusion models: A comprehensive survey of methods and applications," *ACM computing surveys*, vol. 56, no. 4, pp. 1–39, 2023.

[8] M. Rixner and P.-S. Koutsourelakis, "A probabilistic generative model for semi-supervised training of coarse-grained surrogates and enforcing physical constraints through virtual observables," *Journal of Computational Physics*, vol. 434, p. 110218, 2021.

[9] J. Qiu, J. Huang, X. Zhang, Z. Lin, M. Pan, Z. Liu, and F. Miao, "Pi-fusion: Physics-informed diffusion model for learning fluid dynamics," *arXiv preprint arXiv:2406.03711*, 2024.

[10] Z. Shi, L. Yu, T. Xie, and C. Zhang, "Diffusion-pinn sampler," *arXiv preprint arXiv:2410.15336*, 2024.

[11] E. Perez, F. Strub, H. De Vries, V. Dumoulin, and A. Courville, "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[12] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, 2012.

[13] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[14] T. Hang, S. Gu, C. Li, J. Bao, D. Chen, H. Hu, X. Geng, and B. Guo, "Efficient diffusion training via min-snr weighting strategy," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 7441–7451.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[17] K. Long, X. Shi, and X. Li, "Physics-informed neural network for cross-dynamics vehicle trajectory stitching," *Transportation Research Part B: Methodological*, 2022.

[18] A. Pensoneault and X. Zhu, "Efficient bayesian physics informed neural networks for inverse problems via ensemble kalman inversion," *Journal of Computational Physics*, vol. 508, p. 113006, 2024.

[19] A. Ip, L. Irio, and R. Oliveira, "Vehicle trajectory prediction based on lstm recurrent neural networks," in *2021 IEEE 93rd vehicular technology conference (VTC2021-Spring)*. IEEE, 2021, pp. 1–5.

[20] J.-H. Bastek, W. Sun, and D. M. Kochmann, "Physics-informed diffusion models," *arXiv preprint arXiv:2403.14404*, 2024.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We presented our contribution and scope of this paper clearly in the abstract and introduction.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitation and future work are claimed in the conclusion section.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: This paper doesn't include any theoretical derivation and proof.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.

- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The results can be easily reproduced based on the technical details of this paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The authors don't release the data and code, but the results can be reproduced based on the technical details of this paper.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.

- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The experimental setup is clearly presented in this paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The statistical significance of the experiments is introduced in the paper.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Compute resources are provided in experimental setup.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.

- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conform with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This paper discusses the potential societal impacts in conclusion section.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper doesn't have such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper doesn't use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The authors don't release the data and code.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper doesn't involve human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper doesn't involve human subjects.

    Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: The core method in this research doesn't involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.