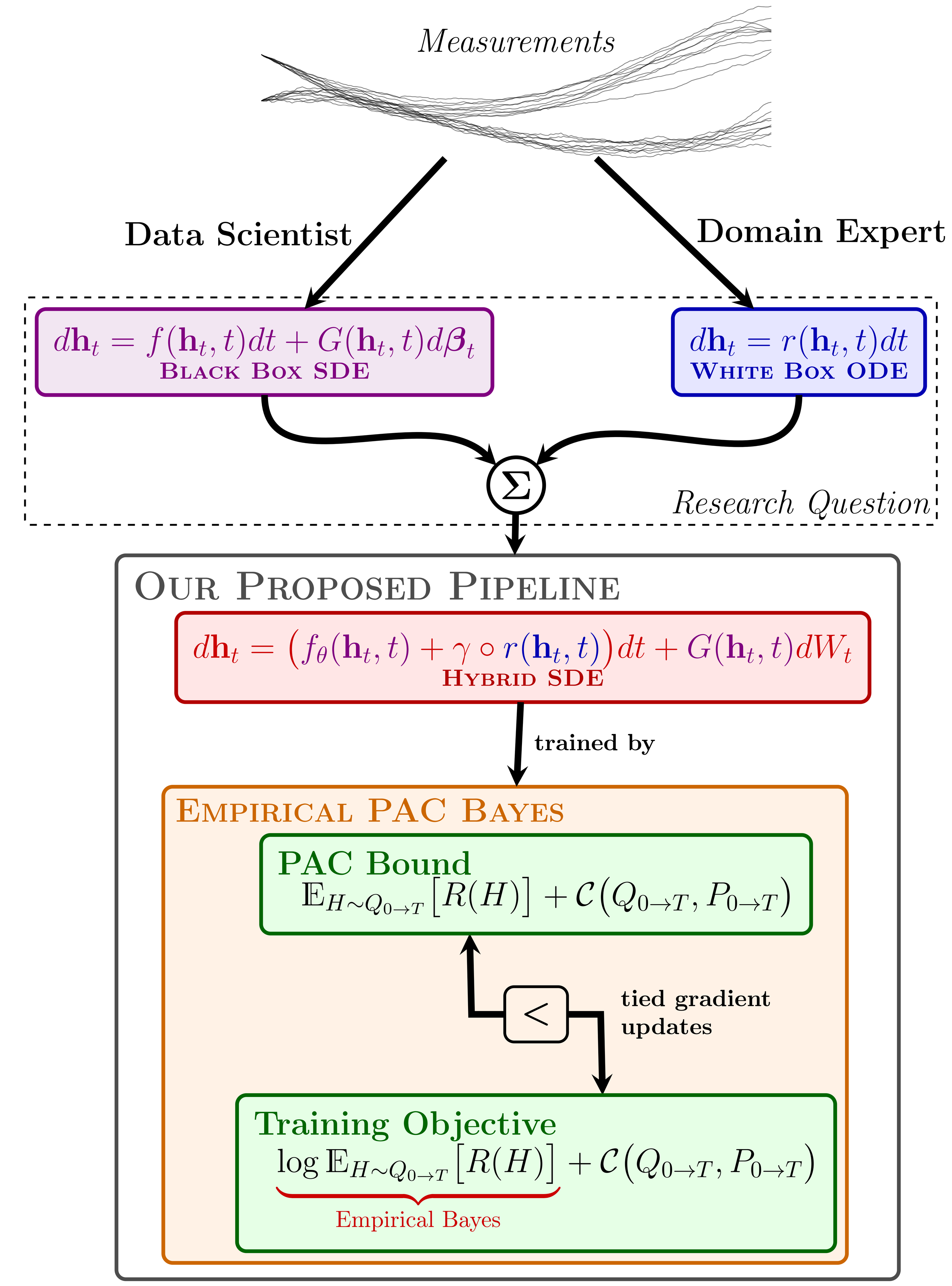
Learning Partially Known Stochastic Dynamics with Empirical PAC Bayes

Manuel Haußmann[1,*] Sebastian Gerwinn[2,*] Andreas Look[2] Barbara Rakitsch[2] Melih Kandemir[2]
[1]HCI/IWR, Heidelberg University, Germany [2]Bosch Center for Artificial Intelligence, Renningen, Germany [*]equal contribution



Summary

Neural Stochastic Differential Equations model dynamical environments with neural nets assigned to their drift and diffusion terms. The high expressive power of their nonlinearity comes at the expense of instability in the identification of the large set of free parameters.

We present a recipe to improve the prediction accuracy of such models by

- accounting for epistemic uncertainty by assuming probabilistic weights,
- incorporating partial knowledege on the state dynamics,
- training the combined model via a PAC-Bayesian based generalization bound objective

$\Rightarrow$ Improved model fit and better predictive uncertainty

Pipeline

1. Step: Replace deterministic drift with BNN

   Place a prior $p_\phi(\theta)$ over the weights θ of the drift

2. Step: Marginalize weights via Empirical Bayes

   Compute the marginal likelihood to be optimized

$$\arg\max_{\phi} \int p(\mathcal{D}|\mathbf{H})p(\mathbf{H}|\theta)p_\phi(\theta)d(\mathbf{H}, \theta)$$

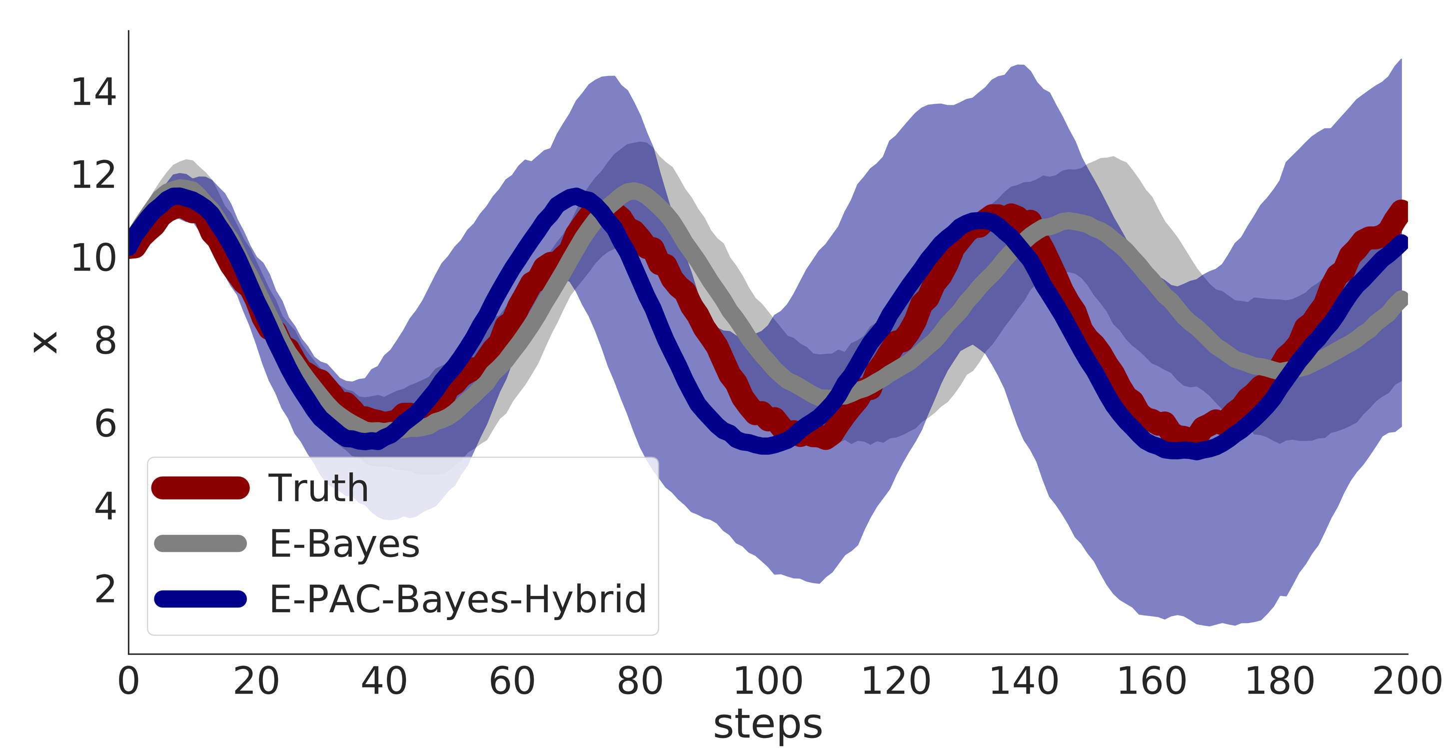
   approximating the integral via samples

3. Step: Use PAC-Bayes for proper regularization

   The final objective is given via the following theorem:

   Theorem: *The expected true risk is bounded above with* $\mathbb{P} \geq 1-\delta$ *by* $\mathbb{E}_{H\sim Q_{0\to T}}[R(H)] \leq \mathbb{E}_{H\sim Q_{0\to T}}[R_{\mathcal{D}}(H)] + \mathcal{C}_\delta(Q_{0\to T}, P_{0\to T})$, *with empirical risk* $R_{\mathcal{D}}(H)$ *and tractable complexity* $\mathcal{C}_\delta(\cdot,\cdot)$.

Lorenz Attractor



Setup: prior knowledge on z dimension of the chaotic Lorenz attractor
Result:

- Better fit than the pure BNN in MSE
- Increasing predictive uncertainty with increasing time

CMU Walking Data Experiment

Method	Reference	Bayesian	Hybrid	+KL	Test MSE	Test NLL
DTSBN-S	(Gan et al., 2015)	No	No	No	34.86 ± 0.02	Not Applicable
npODE	(Heinonen et al., 2018)	No	No	No	22.96	Not Applicable
Neural-ODE	(Chen et al., 2018)	No	No	No	22.49 ± 0.88	Not Applicable
ODE^{2}VAE	(Yildiz et al., 2019)	Yes	Yes	No	10.06 ± 1.40	Not Reported
ODE^{2}VAE-KL	(Yildiz et al., 2019)	Yes	Yes	Yes	8.09 ± 1.95	Not Reported
D-BNN (SGLD)	(Look and Kandemir, 2019)	Yes	No	No	13.89 ± 2.56	747.92 ± 58.49
D-BNN (VI)	(Hegde et al., 2019)	Yes	No	Yes	9.05 ± 2.05	452.47 ± 102.59
E-Bayes	Baseline	Yes	No	No	8.68 ± 1.56	433.76 ± 77.78
E-PAC-Bayes	Ablation	Yes	No	Yes	9.17 ± 1.20	489.82 ± 67.06
E-Bayes-Hybrid	Ablation	Yes	Yes	No	9.25 ± 1.99	462.82 ± 99.61
E-PAC-Bayes-Hybrid	Proposed	Yes	Yes	Yes	$\mathbf{7.84 \pm 1.41}$	$\mathbf{415.38 \pm 80.37}$

Notation

$\mathcal{D}$ is the observed data; $\mathbf{H}$ the latent dynamics; $Q_{0\to T}$ the hybrid process; $P_{0\to T}$ the prior process; $r(\cdot,\cdot)$ the prior knowledge; $\gamma \in [0,1]^D$ an importance regularizer; $G(\cdot,\cdot)$ the diffusion;
See `https://arxiv.org/abs/2006.09914` **for the full paper.**

References

R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural Ordinary Differential Equations. In *NeurIPS*, 2018.

Z. Gan, C. Li, R. Henao, D.E. Carlson, and L. Carin. Deep Temporal Sigmoid Belief Networks for Sequence Modeling. In *NIPS*, 2015.

P. Hegde, M. Heinonen, H. Lähdesmäki, and S. Kaski. Deep Learning with Differential Gaussian Process Flows. In *AISTATS*, 2019.

M. Heinonen, C. Yildiz, H. Mannerström, J. Intosalmi, and H. Lähdesmäki. Learning Unknown ODE Models with Gaussian Processes. In *ICML*, 2018.

A. Look and M. Kandemir. Differential Bayesian Neural Nets. In *4th NeurIPS Workshop on Bayesian Deep Learning*, 2019.

C. Yildiz, M. Heinonen, and H. Lahdesmaki. ODE2VAE: Deep Generative Second Order ODEs with Bayesian Neural Networks. In *NeurIPS*. 2019.