Structured Models with Gaussian Processes

Doctoral defense

Markus Kaiser
www.mrksr.de

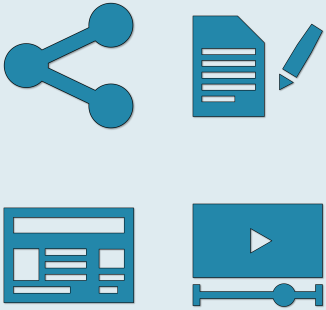
May 18, 2021

Siemens AG
Technical University of Munich

Industrial machine learning

Web ML

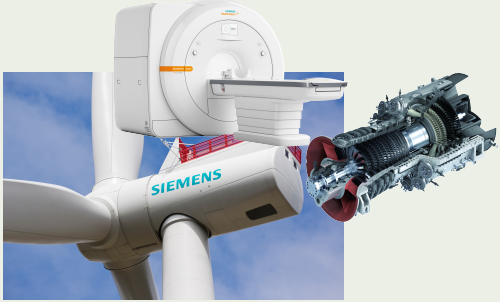
- (Seemingly) mild consequences



Examples: E-commerce, web search, NLP

Industrial ML

- Safety is a main concern

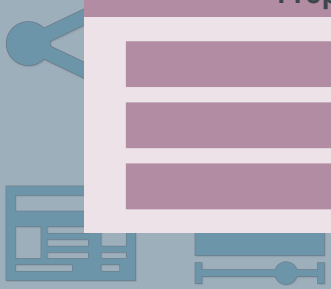


Examples: Machine control, commissioning

Industrial machine learning

Web ML

- (Seemingly) mild consequences



Examples: E-commerce, web search, NLP

Industrial ML

- Safety is a main concern

Examples: Machine control, commissioning

Properties of industrial ML

- Uncertain models
- Domain knowledge
- Need for trust

Research Questions

	Deep learning	Bayesian statistics	
Data-driven insights	Yes	No	RQ1
Strong scalability	Yes	No	
Interpretable results	No	Yes	RQ2
Trustworthy predictions	No	Yes	
Semantic model-selection	No	No	RQ3

How to combine data-driven insights and trustworthy predictions?

Gaussian Processes

Definition

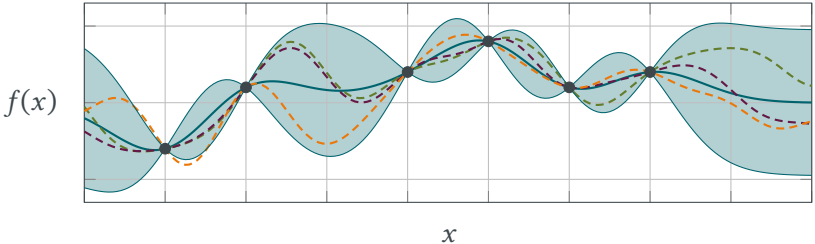
A **Gaussian Process (GP)** is a collection of random variables $\{\mathbf{F}_{\mathbf{x}}\}$, any finite subset of which has a joint Gaussian distribution.

Mean and Kernel Functions

In supervised learning, $\mathbf{F}_{\mathbf{x}}$ models the function value $f(x)$. A GP is completely determined by two functions.

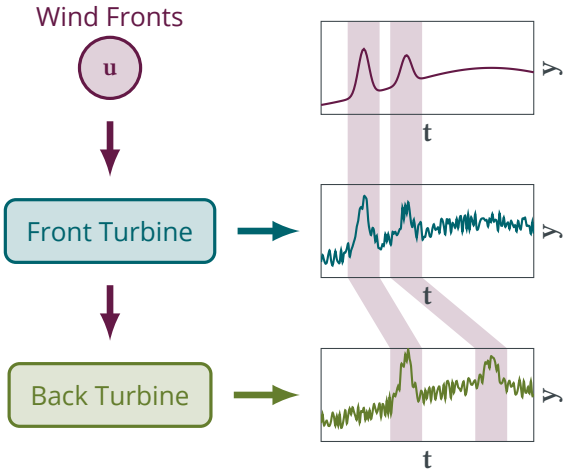
Mean function $\mu_f(x) = \mathbb{E}[f(x)]$

Kernel function $\mathscr{K}(x, x') = \mathrm{cov}[f(x), f(x')]$



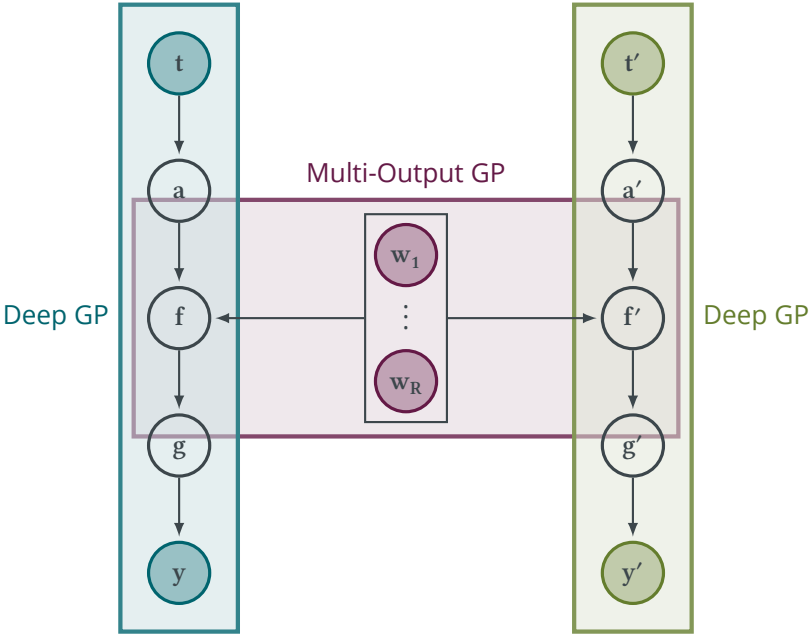
Deep GP model for wind propagation

Lillgrund Wind Farm



- Interpretable factorized uncertainties
- Strong priors enforce physical plausibility
- Experts can validate their expectations

A Bayesian graphical model



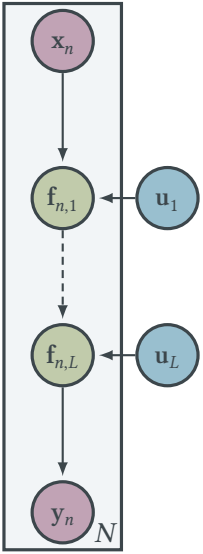
A primer for Deep GPs

A stack of L functions modeld with GPs

$$\mathbf{y} = f_L(f_{L-1}(\cdots(f_1(\mathbf{X})))) + \epsilon$$

with a joint likelihood

$$\mathrm{p}(\mathbf{y}|\mathbf{X}) = \int \mathrm{p}(\mathbf{y}, \mathbf{f}_1, \ldots, \mathbf{f}_L|\mathbf{X})\, \mathrm{d}\mathbf{f}_1 \ldots \mathrm{d}\mathbf{f}_L$$



A primer for Deep GPs

A stack of L functions modeld with GPs

$$\mathbf{y} = f_L(f_{L-1}(\cdots(f_1(\mathbf{X}))))+\epsilon$$

with a joint likelihood

$$\mathrm{p}(\mathbf{y}|\mathbf{X}) = \int \mathrm{p}(\mathbf{y}, \mathbf{f}_1, \dots, \mathbf{f}_L|\mathbf{X})\,\mathrm{d}\mathbf{f}_1 \dots \mathrm{d}\mathbf{f}_L$$

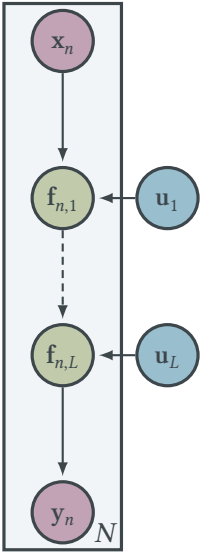
Variational approximation yields the bound

Likelihood Prior

$$\mathscr{L}_{\mathrm{NVC}} \geq \boxed{\log \mathcal{N}\left(\mathbf{y}\,|\,\mathbf{\Psi}_L \mathbf{K}_{\mathbf{u}_L\mathbf{u}_L}^{-1}\mathbf{m}_L, \sigma_n^2 \mathrm{I}\right)} - \boxed{\sum_{l=1}^{L} \mathrm{KL}(\mathrm{q}(\mathbf{u}_l)\,\|\,\mathrm{p}(\mathbf{u}_l))}$$

$$\boxed{\begin{aligned} &- \frac{1}{2\sigma_1^2}\,\mathrm{tr}(\mathbf{K}_{11} - \mathbf{Q}_{11}) - \sum_{l=2}^{L} \frac{1}{2\sigma_l^2}\left(\psi_l - \mathrm{tr}\left(\mathbf{\Psi}_l \mathbf{K}_{\mathbf{u}_l\mathbf{u}_l}^{-1}\right)\right) \\ &- \sum_{l=2}^{L} \frac{1}{2\sigma_l^2}\,\mathrm{tr}\left(\left(\mathbf{\Phi}_l - \mathbf{\Psi}_l^{\mathsf{T}}\mathbf{\Psi}_l\right)\mathbf{K}_{\mathbf{u}_l\mathbf{u}_l}^{-1}\left(\mathbf{m}_l\mathbf{m}_l^{\mathsf{T}} + \mathbf{S}_l\right)\mathbf{K}_{\mathbf{u}_l\mathbf{u}_l}^{-1}\right) \end{aligned}}$$

Nested Variational Compression



Deep Multi-Output GPs in the wind model

Convolutional MO-GP models covariances between turbines.

$$\mathrm{cov}[f_d(\mathbf{x}), f_{d'}(\mathbf{x}')] = \sum_{r=1}^{R} \int k_{d,r}(\mathbf{x}-\mathbf{z}) k_{d',r}(\mathbf{x}'-\mathbf{z})\, \mathrm{d}\mathbf{z}$$

$$= \sum_{r=1}^{R} \frac{(2\pi)^{\frac{K}{2}} \sigma_{d,r}\sigma_{d',r}}{\prod_{k=1}^{K} \hat{\ell}_{d,d',r,k}^{-1}} \exp\left(-\frac{1}{2}\sum_{k=1}^{K} \frac{(x_k - x_k')^2}{\hat{\ell}_{d,d',r,k}^2}\right)$$

Need to derive joint Psi-statistics

$$\psi_f = \mathbb{E}_{q(\mathbf{a})}[\mathrm{tr}(\mathbf{K}_{\mathbf{ff}}) \,|\, \mathbf{w}_1, \dots, \mathbf{w}_R]$$

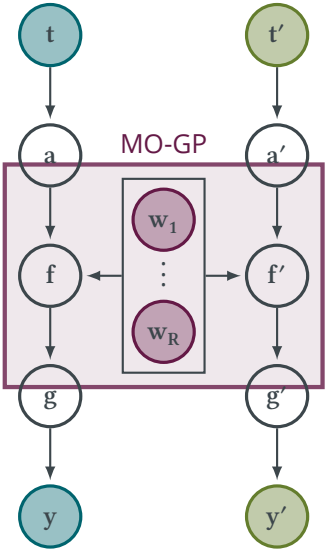
$$\mathbf{\Phi}_{\mathbf{f}} = \mathbb{E}_{q(\mathbf{a})}[\mathbf{K}_{\mathbf{uf}}\mathbf{K}_{\mathbf{fu}} \,|\, \mathbf{w}_1, \dots, \mathbf{w}_R]$$

$$\mathbf{\Psi}_{\mathbf{f}} = \mathbb{E}_{q(\mathbf{a})}[\mathbf{K}_{\mathbf{fu}} \,|\, \mathbf{w}_1, \dots, \mathbf{w}_R]$$

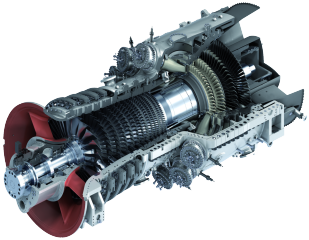
for a variational bound, for example

$$(\mathbf{\Psi}_{\mathbf{f}})_{ni} = \int \mathrm{cov}\left[\hat{f}(\mathbf{a}_n), \hat{f}(\mathbf{Z}_i)\right] q(\mathbf{a}_n)\, \mathrm{d}\mathbf{a}_n$$

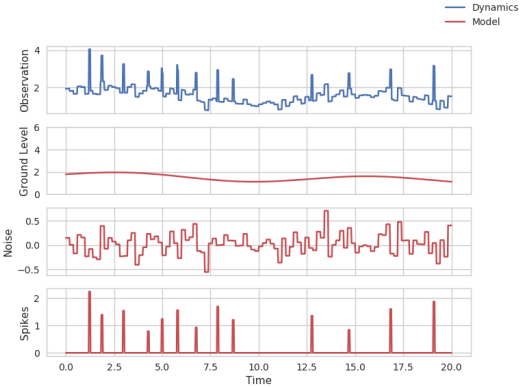
$$= \hat{\sigma}_{ni}^2 \sqrt{\frac{(\Sigma_{\mathbf{a}})_{nn}^{-1}}{\hat{\ell}_{ni} + (\Sigma_{\mathbf{a}})_{nn}^{-1}}} \exp\left(-\frac{1}{2}\frac{(\Sigma_{\mathbf{a}})_{nn}^{-1}\hat{\ell}_{ni}}{(\Sigma_{\mathbf{a}})_{nn}^{-1} + \hat{\ell}_{ni}} ((\boldsymbol{\mu}_{\mathbf{a}})_n - \mathbf{Z}_{\mathbf{i}})^2\right).$$



Data-Association model for gas turbines



Siemens gas turbine



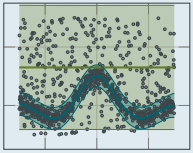
Combustion Dynamics

- Data from different operational regimes
- Robust inference for faulty sensors
- Trustworthy Reinforcement Learning

Contributions

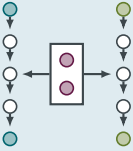
RQ1: Reduce model-bias of white-box models

- Bayesian interpretation of data association
- Application to dynamics of a gas turbine



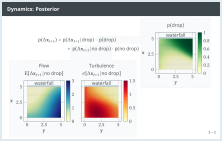
RQ2: Reduce data-bias of black-box models

- Bayesian deep nonlinear time-series alignment
- Enforces physical plausibility with strong priors



RQ3: Semantic model-selection

- Semantic decomposition of RL system dynamics
- Surrogate models for informative structure in BO



Patents

- Depeweg, Stefan, Markus Michael Geipel, Markus Kaiser, and Steffen Udluft (Mar. 18, 2020). “Computerimplementiertes Verfahren zum Abschätzen eines technischen Verhaltens einer Vorrichtung”. European pat. 3623881A1. Siemens AG.
- Egedal, Per, Peder Bay Enevoldsen, Alexander Hentschel, et al. (July 31, 2019). “Verfahren und Vorrichtung zur kooperativen Steuerung von Windturbinen eines Windparks”. European pat. 3517774A1. Siemens Gamesa Renewable Energy AS.
- Geipel, Markus Michael, Thomas Hubauer, Markus Kaiser, and Anja von Beuningen (Mar. 11, 2020). “Transferlernen von Modellen des maschinellen Lernens unter Verwendung einer Wissensgraphdatenbank”. European pat. 3620997A1. Siemens AG.
- Geipel, Markus Michael and Markus Kaiser (Dec. 2, 2020). “Verfahren zur Bestimmung einer Vielzahl von trainierten Modellen des Maschinellen Lernens”. European pat. 3745326A1. Siemens AG.
- Heesche, Kai and Markus Kaiser (Oct. 8, 2020). “Verfahren zum Steuern einer Gasturbine”. European pat. 2020057424W. Siemens AG.
- Heesche, Kai, Markus Kaiser, and Volkmar Sterzing (Nov. 18, 2020). “Informierte unbemerkbare Untersuchung auf Basis von Kontrollrichtlinien”. European pat. 3739525A1. Siemens AG.
- Kaiser, Markus and Marc Christian Weber (Aug. 21, 2019). “Verfahren und Vorrichtungen zur automatischen Ermittlung und/oder Kompensation des Einflusses einer Wirbelschleppe auf eine Windkraftanlage”. European pat. 3527817A1. Siemens AG.

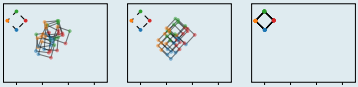
Own Publications

- Bodin, Erik, Markus Kaiser, Ieva Kazlauskaite, et al. (Feb. 24, 2020). “Modulating Surrogates for Bayesian Optimization”. In: *Proceedings of the International Conference on Machine Learning (ICML) 119*. arXiv: 1906.11152.
- Kaiser, Markus, Clemens Otte, Thomas A. Runkler, and Carl Henrik Ek (2018). “Bayesian Alignments of Warped Multi-Output Gaussian Processes”. In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, et al. Curran Associates, Inc., pp. 6995–7004. arXiv: 1710.02766.
- — (Sept. 2019a). “Data Association with Gaussian Processes”. In: *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD) 2019*. arXiv: 1810.07158.
- — (2019b). “Interpretable Dynamics Models for Data-Efficient Reinforcement Learning”. In: *Computational Intelligence and Machine Learning* ESANN 2019 proceedings, p. 6.
- — (Apr. 10, 2020). “Bayesian Decomposition of Multi-Modal Dynamical Systems for Reinforcement Learning”. In: *Neurocomputing*. ISSN: 0925-2312. DOI: 10.1016/j.neucom.2019.12.132.
- Ustyuzhaninov, Ivan, Ieva Kazlauskaite, Markus Kaiser, et al. (Feb. 25, 2020). “Compositional Uncertainty in Deep Gaussian Processes”. In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. arXiv: 1909.07698.

Future Research

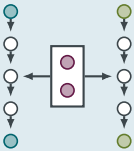
1. How to make effective use of uncertainties?

- Hierarchical posteriors are complex
- Which uncertainties should we care about?



2. What do hierarchical priors mean?

- Systems combine well-understood priors
- How do they interact?



3. What makes a model or system trustworthy?

- Performance metrics are not enough
- Can we reason about system-context?