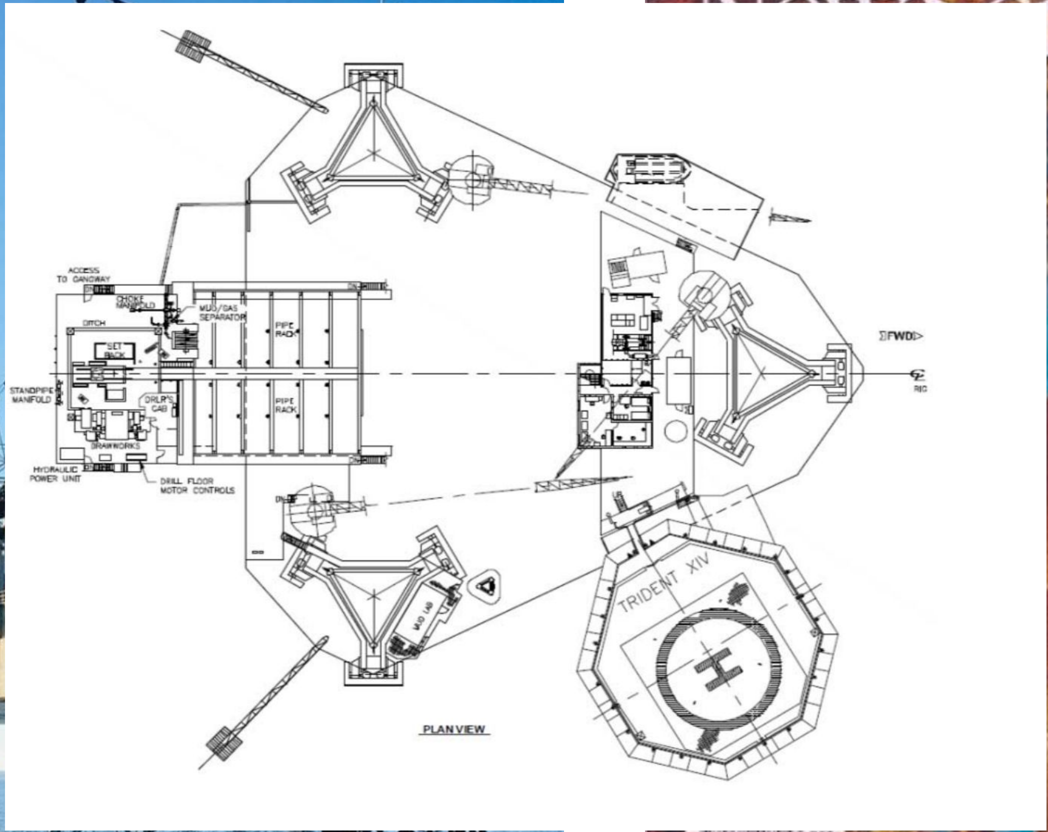
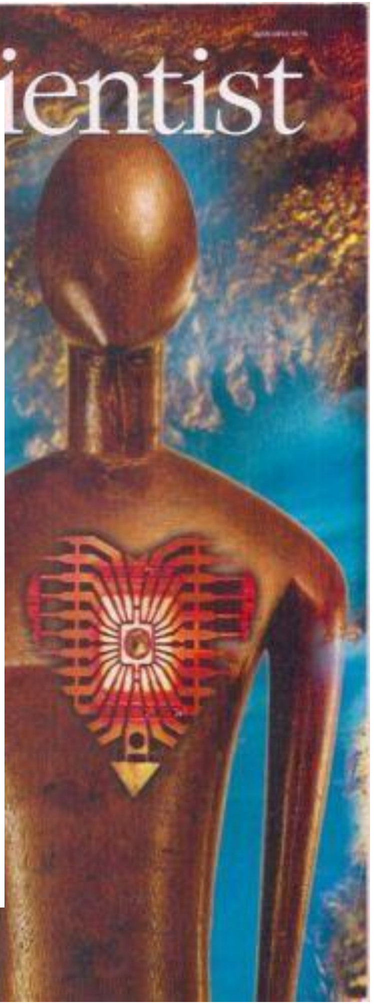
Deep Gaussian processes

Andreas Damianou *(Spotify)*
Neil Lawrence *(University of Cambridge)*

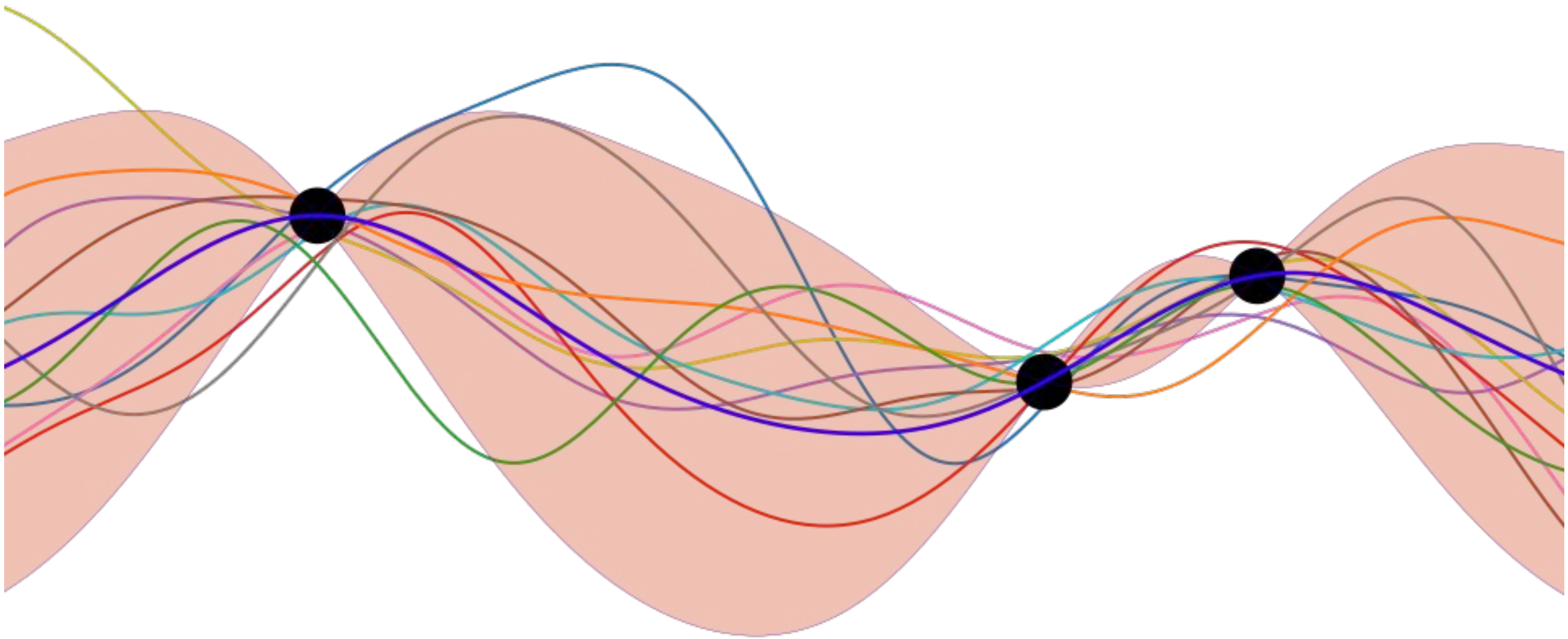
Test of Time Award, AISTATS 2023


© Anon
MarineTraffic.com


ACCESS TO GANGWAY
CHOKE MANIFOLD
MUD/GAS SEPARATOR
DITCH
SET BACK
PIPE RACK
PIPE RACK
STANDPIPE MANIFOLD
DRLR'S CAB
DRAWWORKS
HYDRAULIC POWER UNIT
DRILL FLOOR MOTOR CONTROLS
MUD LAB
TRIDENT XIV
FWD
PLAN VIEW


10 December 1994 No1955 Weekly £1.60

google.co.uk
9 Clarkson Rd
Cambridge, England
Google Street View
Nov 2022
See more dates
Clarkson Road
Grange Rd
Google
Image capture: Nov 2022
© 2023 Google
Spain
Terms
Privacy
Report a problem



direction for further research.

11.1. HAVE WE THROWN THE BABY OUT WITH THE BATH WATER?

According to the hype of 1987, neural networks were meant to be intelligent models which discovered features and patterns in data. Gaussian processes in contrast are simply smoothing devices. How can Gaussian processes possibly replace neural networks? What is going on?

I think what the work of Williams and Rasmussen (1996) shows is that many real–world data modelling problems are perfectly well solved by sensible smoothing methods. The most interesting problems, the task of feature discovery for example, are not ones which Gaussian processes will solve. But maybe multilayer perceptrons can't solve them either.

On the other hand, it may be that the limit of an infinite number of

http://www.inference.org.uk/mackay/gpB.pdf file:///Users/neil/tmp.html

Gaussian Processes in Practice Workshop

Bletchley Park, U.K.
12 - 13 June 2006

Workshop Now Over

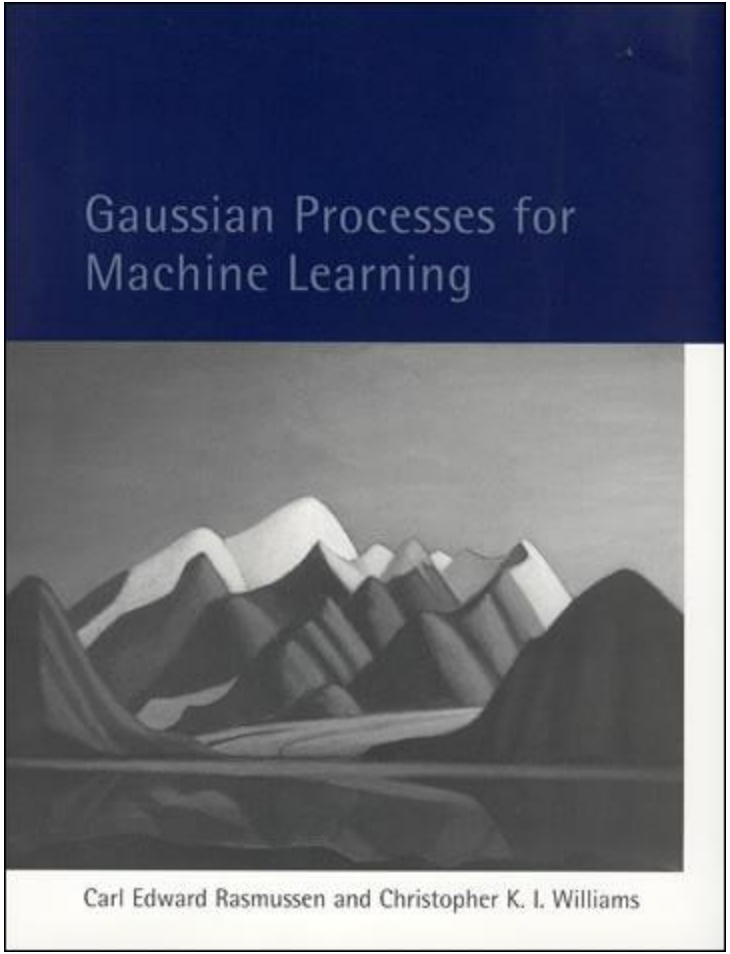
Post-proceedings deadline for submissions is 15th September 2006. See this page for details . The slides presented at the workshop are available from the schedule page .

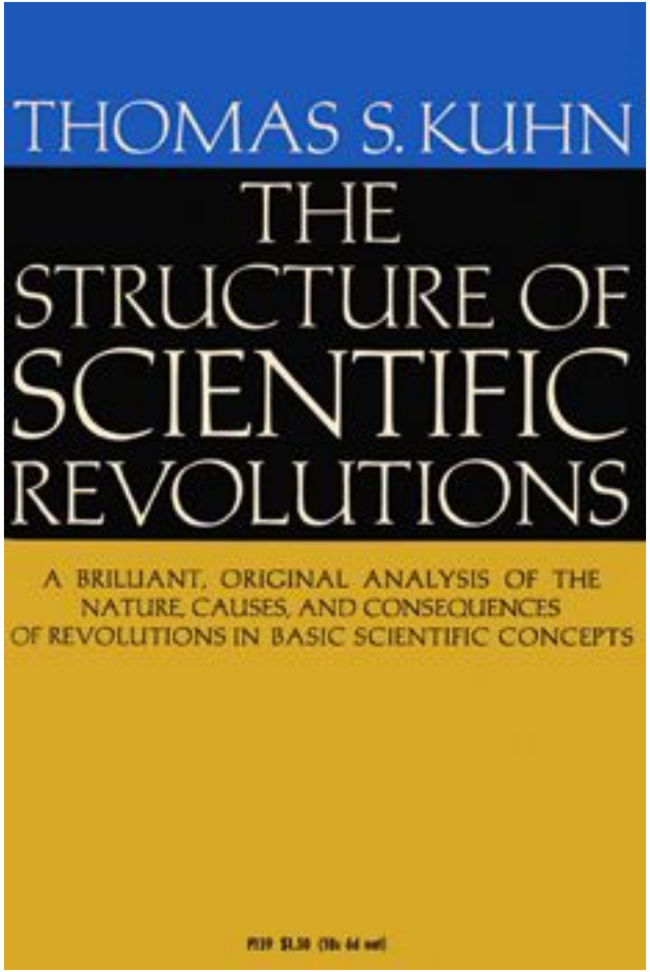
Breaking News

Note: Accommodation in Milton Keynes is filling up fast due to rock concerts and the British Grand Prix.

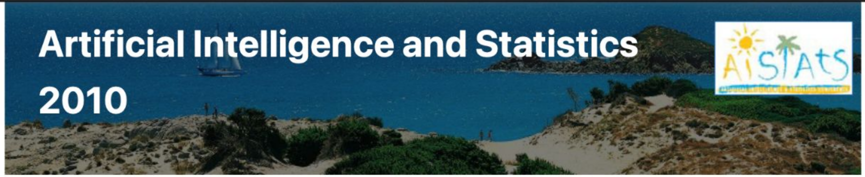
There has been a recent upsurge in interest in Gaussian processes for solving a variety of machine learning problems. Simultaneously there have been recent rapid developments in efficient approximation techniques for Gaussian processes and novel unifying theories of existing approximation techniques.

This workshop will bring experts in Gaussian processes together with individuals who are using Gaussian processes at the forefront of research in their fields. We will start with a set of talks that will relate the latest developments in Gaussian processes (speakers include Chris Williams, Carl Rasmussen, David MacKay and Ed Snelson) while the second portion of talks will discuss application fields where Gaussian processes are being applied as state-of-the-art technologies: Robotics, Graphics and Vision (speakers include Brian Ferris, Aaron Hertzmann and Raquel Urtasun).


Gaussian Processes for
Machine Learning
Carl Edward Rasmussen and Christopher K. I. Williams


THOMAS S. KUHN
THE
STRUCTURE OF
SCIENTIFIC
REVOLUTIONS
A BRILLIANT, ORIGINAL ANALYSIS OF THE
NATURE, CAUSES, AND CONSEQUENCES
OF REVOLUTIONS IN BASIC SCIENTIFIC CONCEPTS

aistats.org



[edit]

- **Home**
- **Accommodation**
- **Call for Abstracts**
- **Call for Papers**
- **Key Dates**
- **Local Information**
- **Organizers**
- **Other Meetings**
- **Proceedings**
- **Registration**
- **Schedule**
- **Invited Speakers**
- **Sponsors**
- **Submission**
- **Proceedings**

The 13th International Conference on Artificial Intelligence and Statistics

[edit]

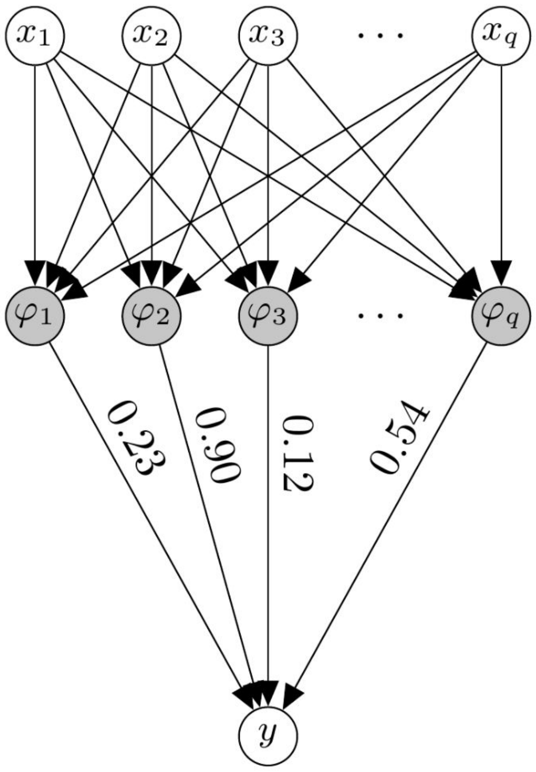
The Proceedings of AISTATS 2010 are now available.

May 13 - 15, 2010
Sardinia, Italy

The 13th International Conference on Artificial Intelligence and Statistics was held in Sardinia, Italy from **Thursday, 13 May 2010 to Saturday, 15 May 2010** at the Chia Laguna Resort.

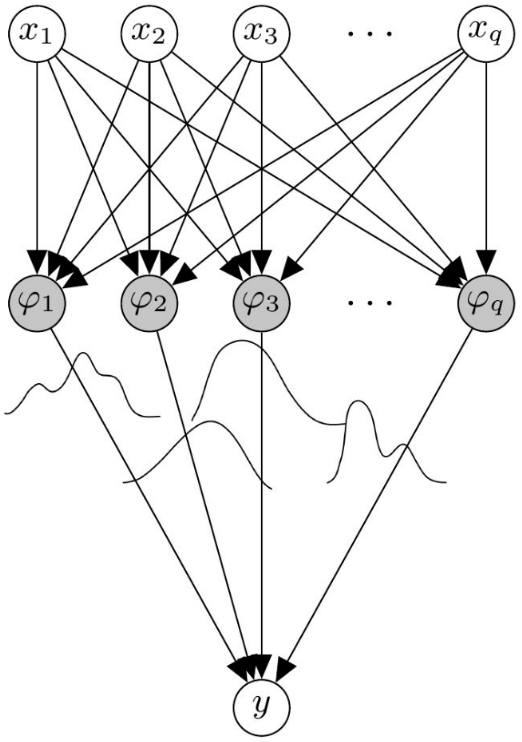
This webpage acts as a record of the program and talks given.

Bayesian Neural Network



Neural Network

$\Rightarrow$



Bayesian Neural Network

From NN to GP

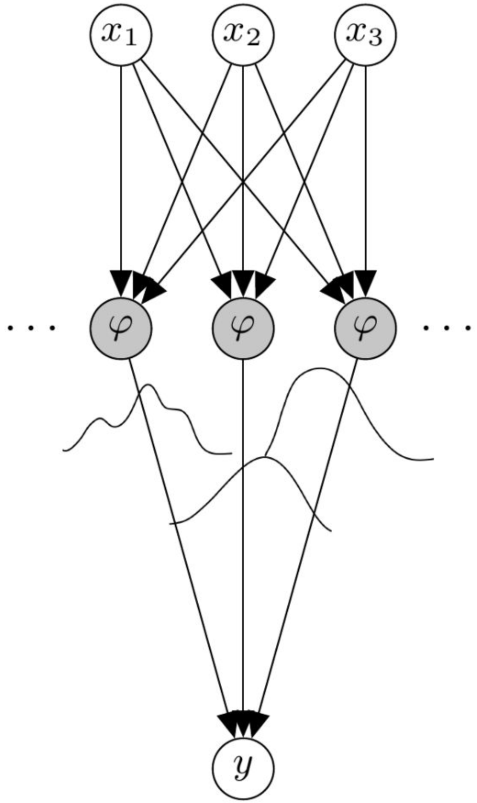
- In the limit of infinite units we obtain a GP [1].

- Think of a function as an infinite dimensional vector.

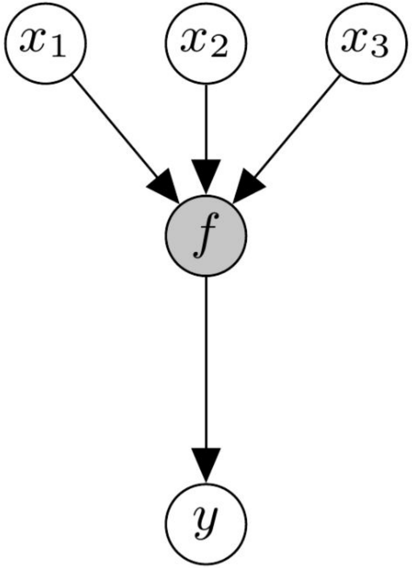
  $f \sim \mathcal{GP}(0, k(x, x'))$. f is stochastic!

- A GP is a distribution over functions (inference directly in function space).

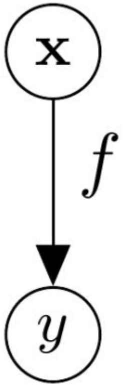
[1] Radford Neal. Bayesian learning for neural networks. 1995



From NN to GP



From NN to GP



Deep GP

- Define a recursive stacked construction

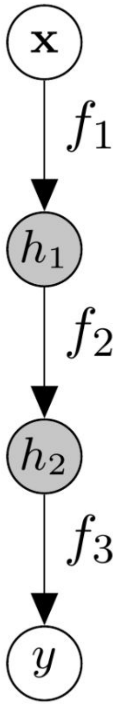
$$f(\mathbf{x}) \rightarrow \text{GP}$$

$$f_L(f_{L-1}(f_{L-2} \cdots f_1(\mathbf{x})))) \rightarrow \text{deep GP}$$

Compare to:

$$\varphi(\mathbf{x})^\top \mathbf{w} \rightarrow \text{NN}$$

$$\varphi(\varphi(\varphi(\mathbf{x})^\top \mathbf{w}_1)^\top \ldots \mathbf{w}_{L-1})^\top \mathbf{w}_L \rightarrow \text{DNN}$$



Damianou & Lawrence, AISTATS 2013

Deep GP

- Define a recursive stacked construction

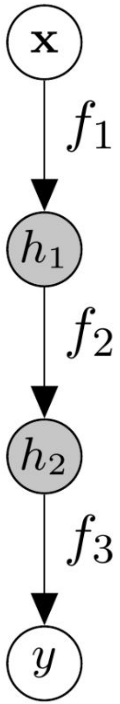
$$f(\mathbf{x}) \rightarrow \text{GP}$$

$$f_L(f_{L-1}(f_{L-2} \cdots f_1(\mathbf{x})))) \rightarrow \text{deep GP}$$

Compare to:

$$\varphi(\mathbf{x})^\top \mathbf{w} \rightarrow \text{NN}$$

$$\varphi(\varphi(\varphi(\mathbf{x})^\top \mathbf{w}_1)^\top \ldots \mathbf{w}_{L-1})^\top \mathbf{w}_L \rightarrow \text{DNN}$$



Damianou & Lawrence, AISTATS 2013

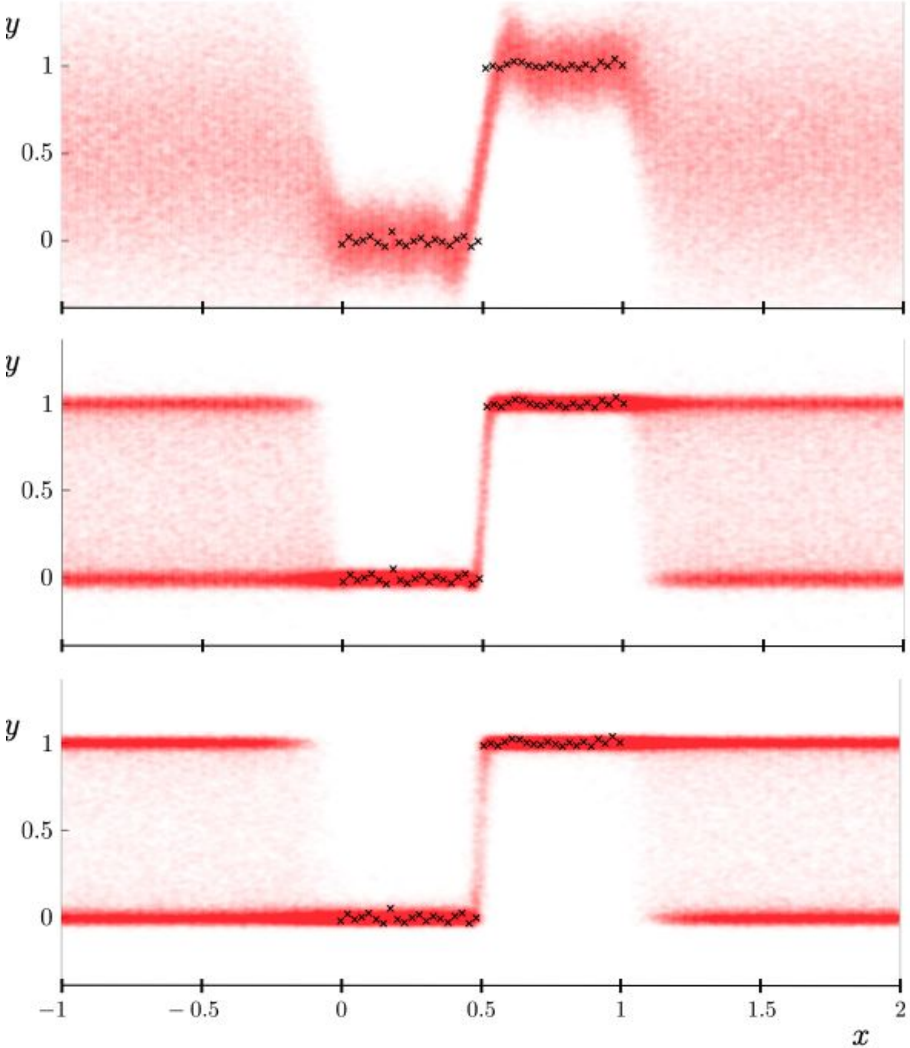
Relax the assumption of global smoothness

Deep GP

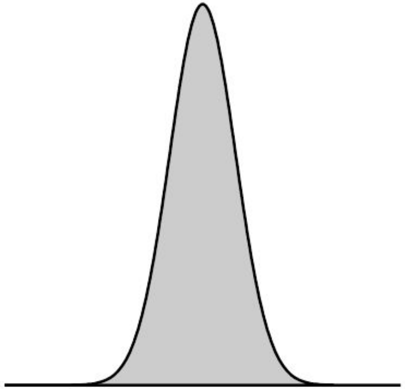
GP

DGP
1 layer

DGP
3 layers

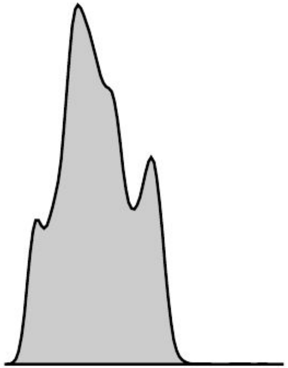


Propagating uncertainty through non-linearities



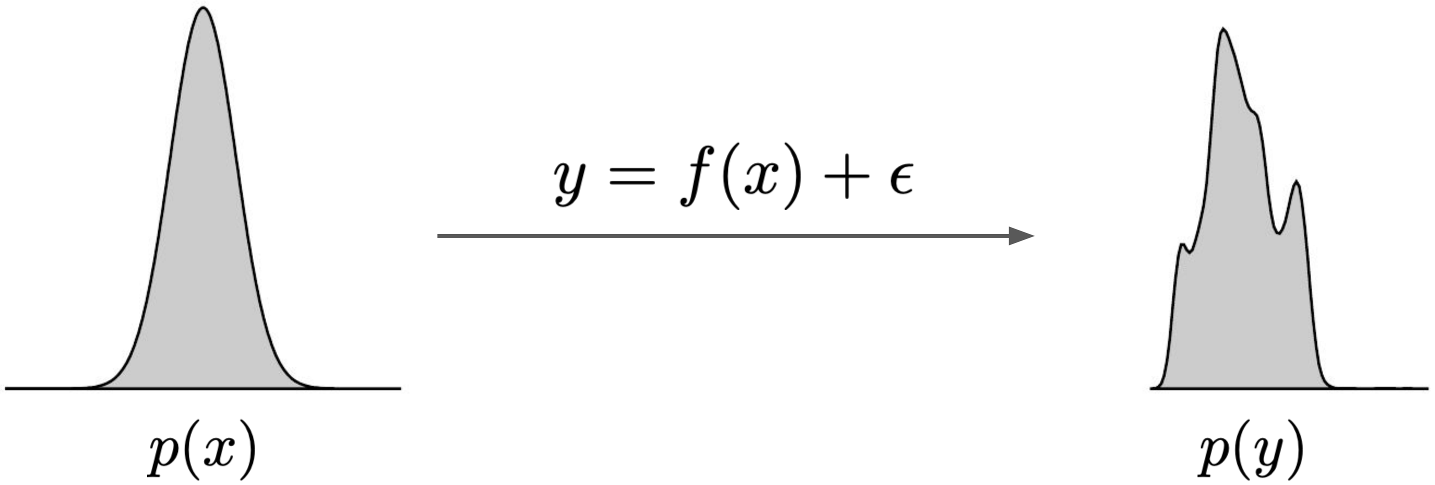
$p(x)$

$$y = f(x) + \epsilon$$



$p(y)$

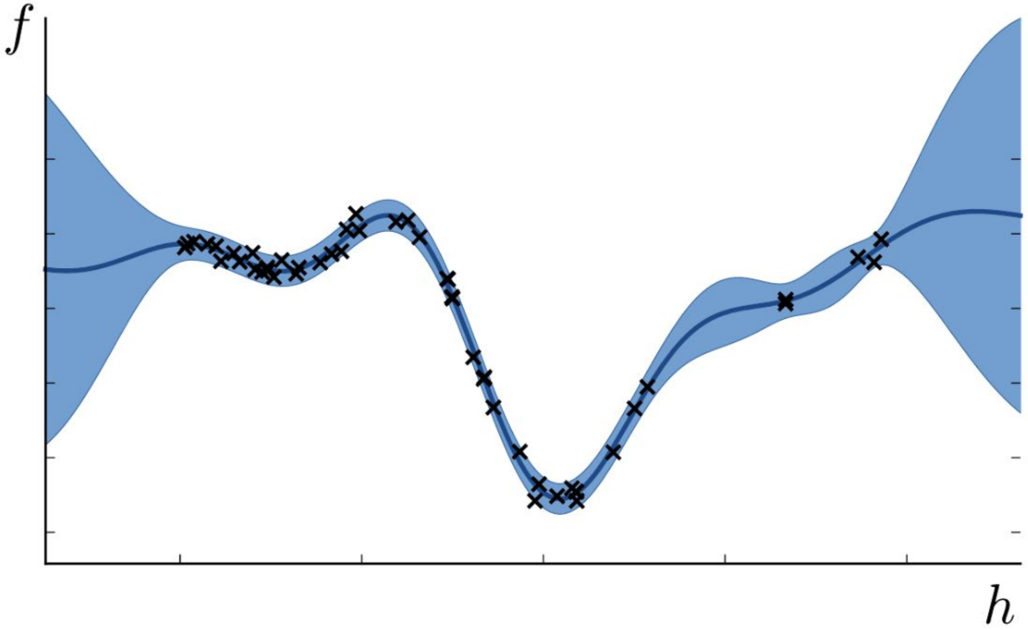
Propagating uncertainty through non-linearities



Without uncertainty propagation I can't have a DGP (I lose the stochasticity).

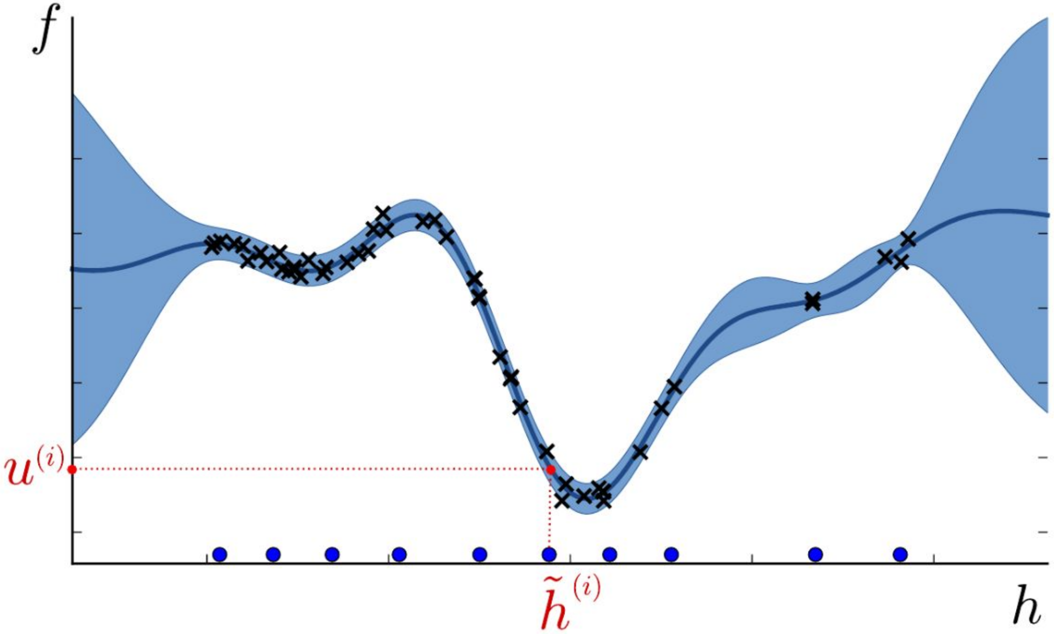
Inducing points in Gaussian processes

$h^{(1)}$	$\mathbf{f}^{(1)}$
$h^{(2)}$	$\mathbf{f}^{(2)}$
$\dots$	$\dots$
$h^{(30)}$	$\mathbf{f}^{(30)}$
$h^{(31)}$	$\mathbf{f}^{(31)}$
$\dots$	$\dots$
$h^{(N)}$	$\mathbf{f}^{(N)}$



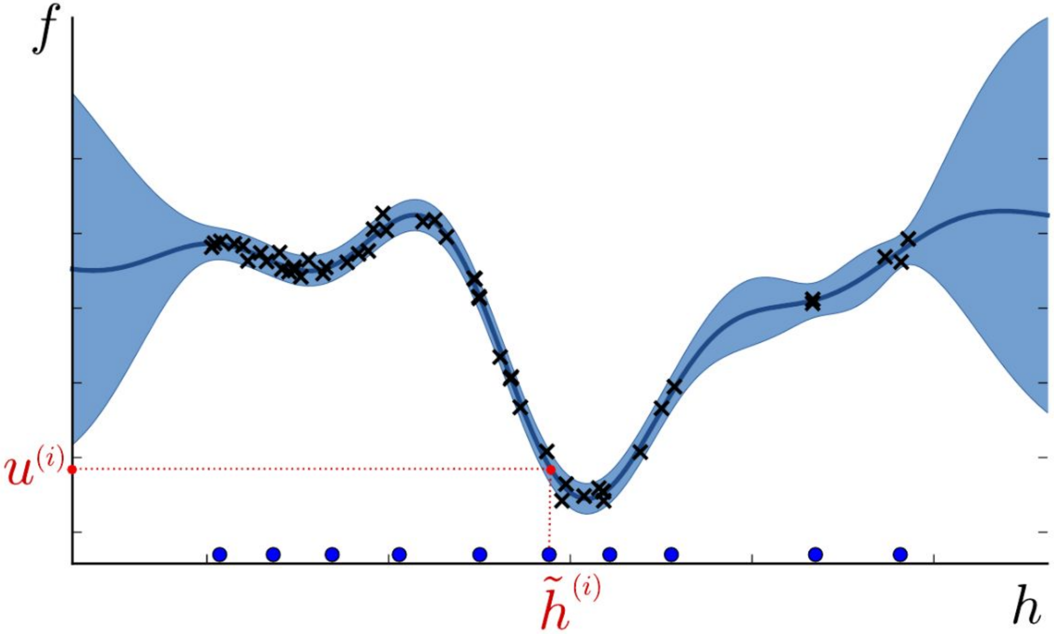
Inducing points in Gaussian processes

$h^{(1)}$	$\mathbf{f}^{(1)}$
$h^{(2)}$	$\mathbf{f}^{(2)}$
$\dots$	$\dots$
$h^{(30)}$	$\mathbf{f}^{(30)}$
$\tilde{h}^{(i)}$	$u^{(i)}$
$h^{(31)}$	$\mathbf{f}^{(31)}$
$\dots$	$\dots$
$h^{(N)}$	$\mathbf{f}^{(N)}$



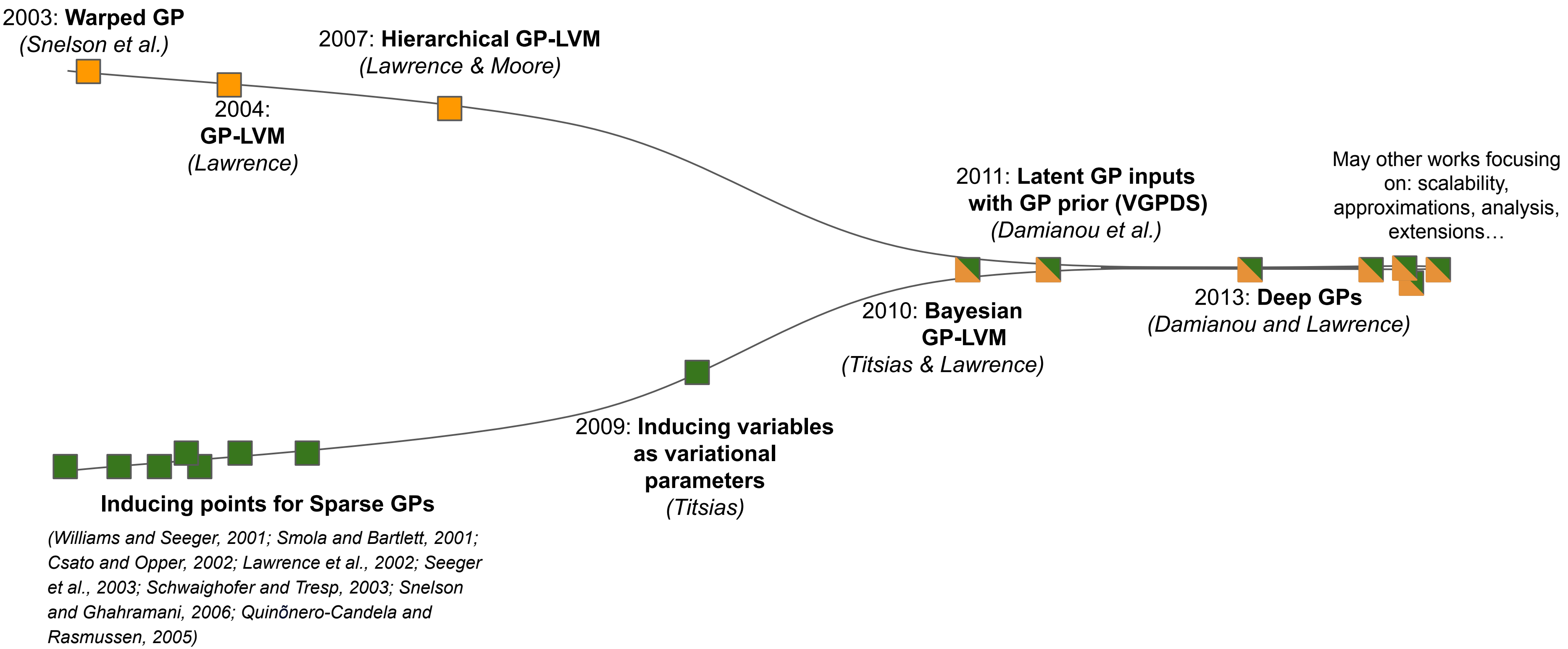
Inducing points in Gaussian processes

$h^{(1)}$	$\mathbf{f}^{(1)}$
$h^{(2)}$	$\mathbf{f}^{(2)}$
$\ldots$	$\ldots$
$h^{(30)}$	$\mathbf{f}^{(30)}$
$\tilde{h}^{(i)}$	$u^{(i)}$
$h^{(31)}$	$\mathbf{f}^{(31)}$
$\ldots$	$\ldots$
$h^{(N)}$	$\mathbf{f}^{(N)}$



- Inducing points originally introduced for faster **(sparse) GPs**
- We can also use them to **induce tractability**: the information in $f = f(h)$ is compressed in u *independently* of h, allowing for the uncertainty of h to be propagated through the nonlinearity *[Titsias & Lawrence 2010, Damianou & Lawrence 2013, Hensman & Lawrence 2014].*

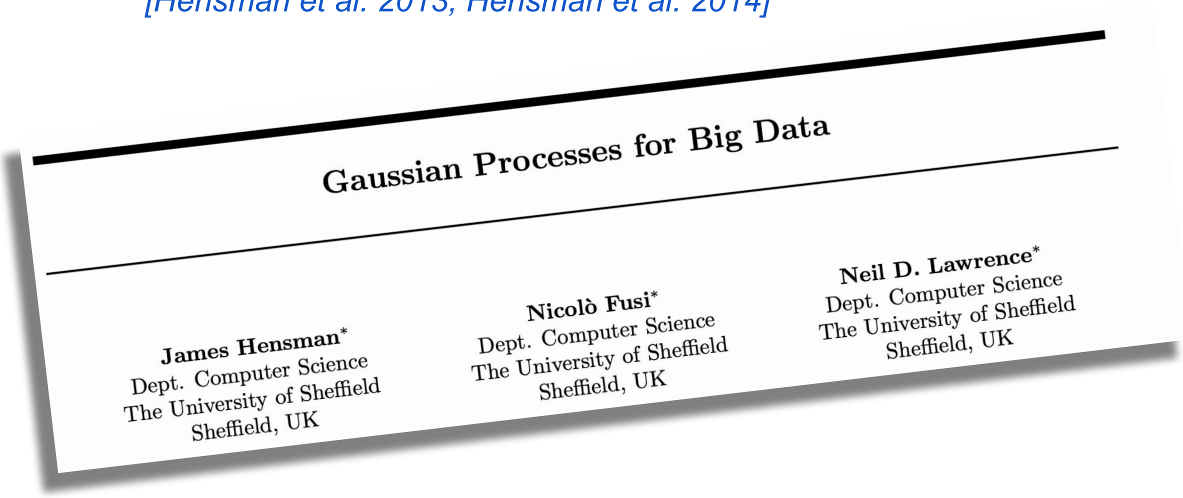
Two lines of work converging: modeling & approximations



Different treatments of the inducings → different properties

Depending on how we treat the variational distribution on u, we can have:

- Distributed computations *[Gal et al. 2014, Dai et al. 2014]* , or
- Fully parallel inference in the style of SVI-GP *[Hensman et al. 2013, Hensman et al. 2014]*

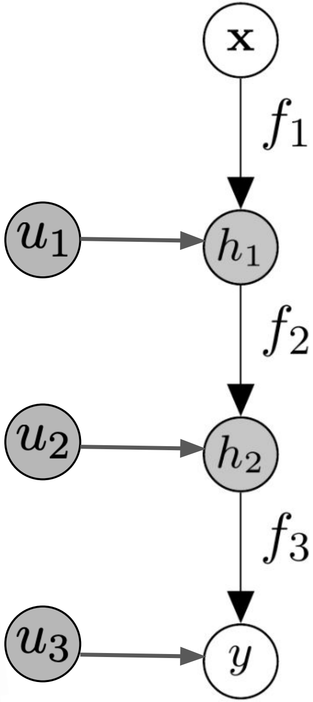


Gaussian Processes for Big Data

James Hensman*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

Nicolò Fusi*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

Neil D. Lawrence*
Dept. Computer Science
The University of Sheffield
Sheffield, UK

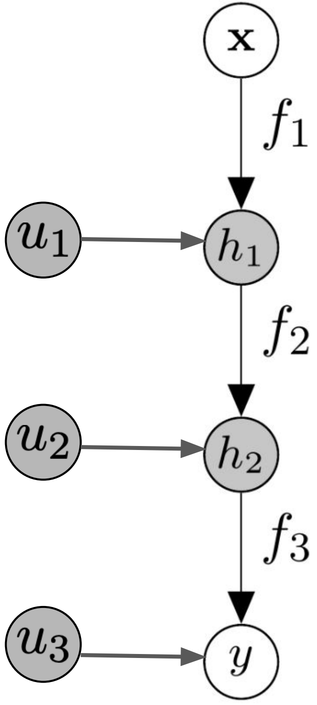


Different treatments of the inducings → different properties

The following distributions are involved in a variational approximation:

$$q(h_1, h_2, \ldots), q(u_1, u_2, \ldots)$$

They're treated differently depending on the particular method.



Various Deep GP approximations

NON-EXHAUSTIVE LIST

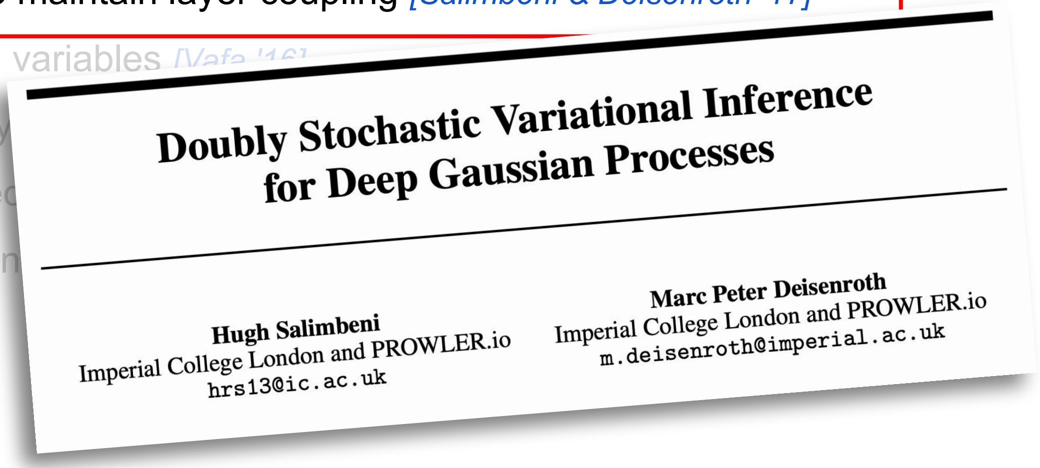
- Mean-field, re-parameterized *[Damianou & Lawrence '13, Damianou '15]*
- Amortized with NNs *[Dai et al. '14]*
- Approximate scalable EP *[Bui et al. '16]*
- Projected $q(h)$ distribution in nested variational inference. *[Hensman & Lawrence '14]*
- Sample through the $q(f_{1:L})$ chain to maintain layer coupling *[Salimbeni & Deisenroth '17]*
- Sampling + FITC + MAP for inducing variables *[Vafa '16]*
- Approximate kernel's spectral density + VI *[Cutajar et al. '17]*
- DeepGPs & NN regularization connections *[Gal & Ghahramani '15; Louizos & Welling '16]*
- Variational distribution with correlation across layers *[Ustyuzhaninov et al. '20]*

See also: Keynote @ NeurIPS workshop on Advances in Approximate Bayesian Inference, A. Damianou, 2017

Various Deep GP approximations

NON-EXHAUSTIVE LIST

- Mean-field, re-parameterized *[Damianou & Lawrence '13, Damianou '15]*
- Amortized with NNs *[Dai et al. '14]*
- Approximate scalable EP *[Bui et al. '16]*
- Projected $q(h)$ distribution in nested variational inference. *[Hensman & Lawrence '14]*
- Sample through the $q(f_{1:L})$ chain to maintain layer coupling *[Salimbeni & Deisenroth '17]*
- Sampling + FITC + MAP for inducing variables *[Vafa '16]*
- Approximate kernel's spectral density
- DeepGPs & NN regularization connec
- Variational distribution with correlation



Doubly Stochastic Variational Inference for Deep Gaussian Processes

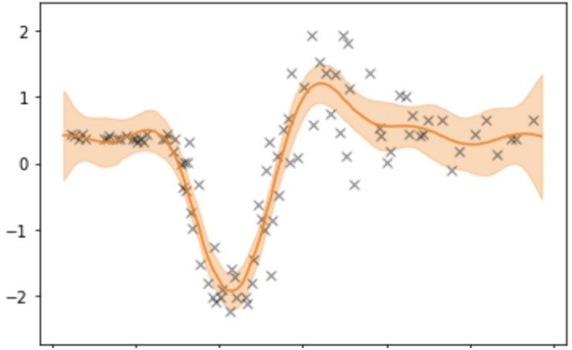
Hugh Salimbeni
Imperial College London and PROWLER.io
hrs13@ic.ac.uk

Marc Peter Deisenroth
Imperial College London and PROWLER.io
m.deisenroth@imperial.ac.uk

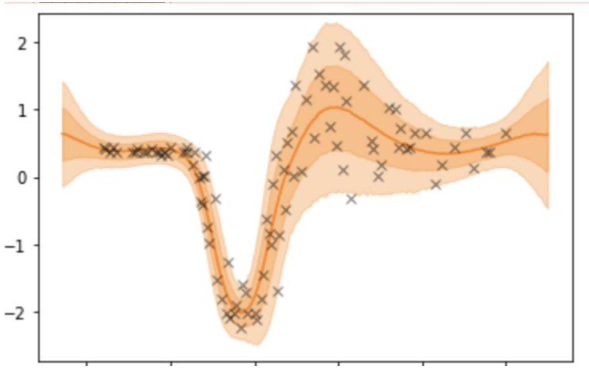
Doubly stochastic VI for DGPs *[Salimbeni & Deisenroth 2017]*

$$q(\{f_l, u_l\}_{l=1^L}) = \prod_{l=1}^{L} p(f_l|u_l; f_{l-1})q(u_l)$$

No “bottleneck” layers (noise corrupted intermediate outputs)



Doubly-stochastic VI



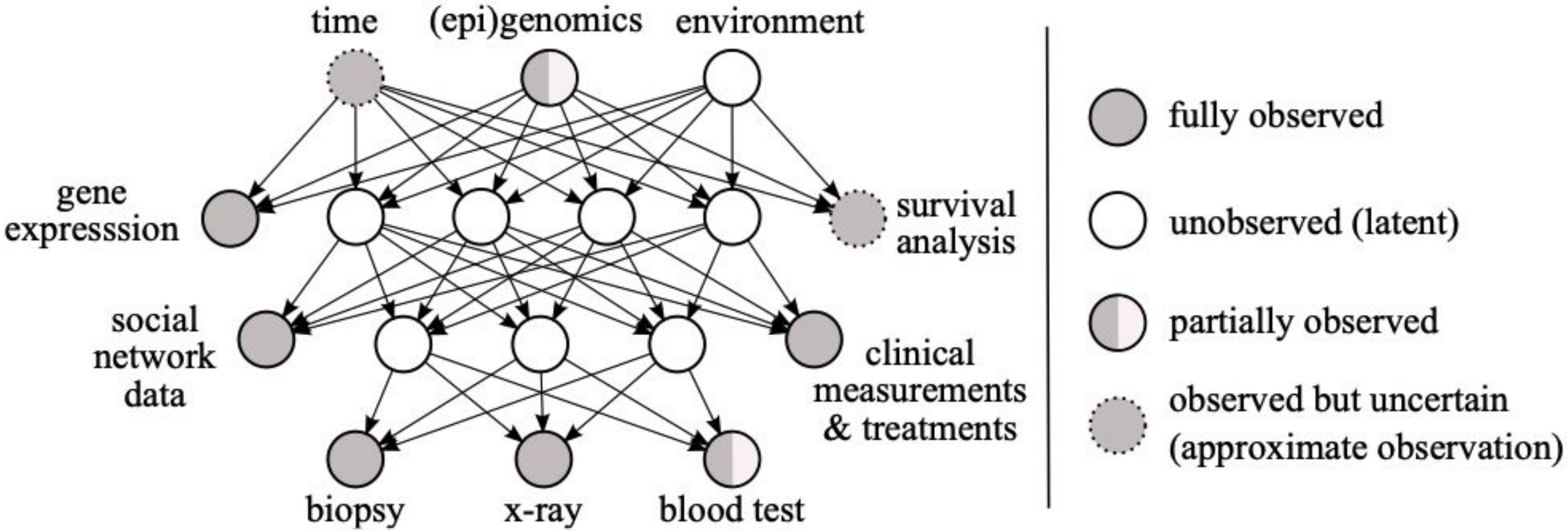
Mean-field VI

** Img from GPflux.*
Thanks to Vincent Dutordoir for the insights.

This can *sometimes* have trouble with modeling heteroscedastic noise.

But! It makes the model more practical: No “bottleneck” layers h, no need to worry about initializing them!
Very popular approximation choice in DGP software.

Emphasis on latent variable or function learning?



A deep, non-linear, non-parametric multi-view extension of factor analysis.

See *Manifold Relevance Determination* [Damianou et al. 2012], [Damianou et al. 2016]

Software for DGPs

- **GPflux**: A library for DGPs. *[Dutordoir et al. 2021]* Based on GPflow and Tensorflow.
- Bayesian Layers (based on **Edward** prob. programming); *[Tran et al. 2019]*
- **MXFusion** based on MXnet prob. programming *[Dai et al. 2018]*

- **GPyTorch** *[Gardner et al. 2018]* (focus on conj. gradients for scalability)
- **GPy** (PyDeepGP)
- **MATLAB deepGP** *[Damianou et al. 2013]*

- Others (see Appendix A from GPflux library)

Google Scholar "deep gaussian processes for"

Deep Gaussian process regression for lithium-ion battery health prognosis and degradation mode diagnosis

Piyush Tagade [a], Krishnan S. Hariharan [a], Sanoop Ramachandran [a], Ashish Khandelwal [a], Arunava Naha [a], Subramanya Mayya Kolake [a], Seong Ho Han [b]

Deep Gaussian Processes for the Analysis and Optimization of Complex Systems
-
Application to Aerospace System Design

A dissertation submitted by

Ali Hebbal

MEMES: Machine learning framework for Enhanced MolEcular Screening†‡

Sarvesh Mehta [a], Siddhartha Laghuvarapu§ [a], Yashaswi Pathak§ [a], Aaftaab Sethi [b], Mallika Alvala [c] and U. Deva Priyakumar *[a]

Empirical Assessment of Deep Gaussian Process Surrogate Models for Engineering Problems

Dushhyanth Rajaram, Tejas G. Puranik, S. Ashwin Renganathan, WoongJe Sung, Olivia Pinon Fischer, Dimitri N. Mavris and Arun Ramamurthy

Published Online: 18 Sep 2020 • https://doi.org/10.2514/1.C036026

Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data

A Deep Gaussian Process-Based Flight Trajectory Prediction Approach and Its Application on Conflict Detection

by Zhengmao Chen, Dongyue Guo and Yi Lin *

Gait Prediction and Variable Admittance Control for Lower Lim Exoskeleton With Measurement Delay and Extended-State-Observer

Publisher: IEEE Cite This PDF

Zhenlei Chen ; Qing Guo ; Tieshan Li ; Yao Yan ; Dan Jiang All Authors

Forecast the Plausible Paths in Crowd Scenes *

Hang Su, Jun Zhu, Yinpeng Dong, Bo Zhang

Indoor Radio Map Construction and Localization With Deep Gaussian Processes

Publisher: IEEE Cite This PDF

Xiangyu Wang ; Xuyu Wang ; Shiwen Mao ; Jian Zhang ; Senthilkumar C. G. Periaswamy ; Justin Patton All Authors

Automatically Designed Deep Gaussian Process for Turbomachinery Application

Yuan Jin, Jin Chai, Olivier Jung

Google Scholar "deep gaussian processes for"

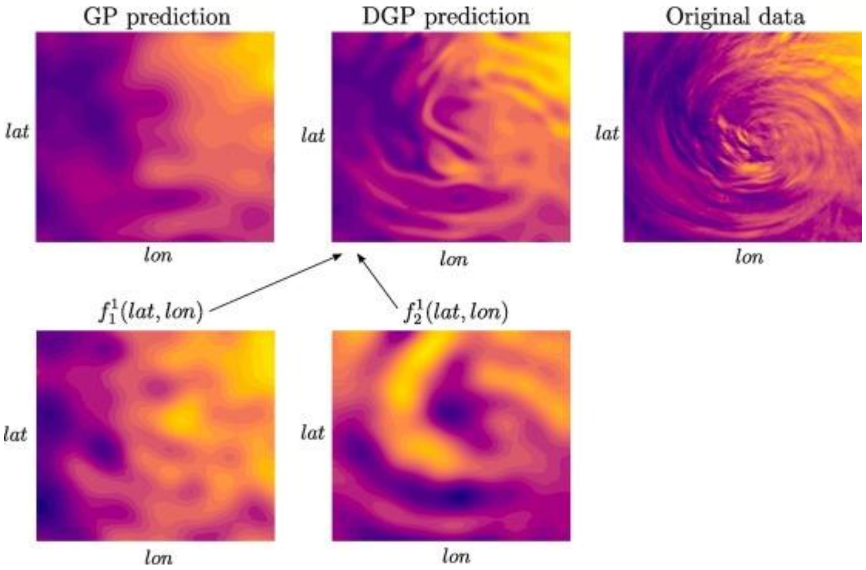
NON-EXHAUSTIVE

- **Object detection, image classification, DGP + convolutions** *[Damianou, 2015; Kumar et al., 2018; Blomqvist et al., 2019]*
- **Speech synthesis** *[Koriyama and Kobayashi, 2019; Mitsui et al. 2021]*
- **Novelty detection** *[Domingues et al., 2018]*
- **(Inverse) RL** *[Jin et al., 2015; Gadd et al. 2020]*
- **Transfer learning** *[Kandemir, 2015]*
- **Disease identification & diagnosis.** Tumor detection *[Kandemir, 2015]*; Survival analysis *[Alaa and van der Schaar, 2017]*; Classification of fetal heart rate tracings *[Feng et al., 2018]*, Molecular screening *[Mehta et al. 2021]*; EEG signal analysis *[Román et al. 2022]*
- **Physical sciences.** Atmospheric data modeling for assessment modeling for nuclear plants *[Jančič et al., 2018]*, Crop yield prediction *[You et al. 2017]*, Remote Sensing & Earth observation *[Svendsen et al. 2020]*
- **Engineering & simulation.** Crowd motion modeling *[Sou et al. 2017]*, Metamodeling response surfaces *[Dutordoir et al. 2017];* Comp. fluid dynamics *[Park et al. 2018]*, Nuclear steam turbine generator simulation *[Zhao et al., 2019]*, Nuclear reactor simulation *[Radaideh and Kozlowski, 2020]*, Flight trajectory prediction *[Chen et al. 2020]*, Multi-fidelity modeling *[Perdikaris et al. 2017, Cutajar et al. 2019]*, Turbomachinery *[Jin et al. 2021]*, Aerospace system design *[Hebbal, 2021]*, Control for lower limb exoskeleton *[Chen et al. 2022],* Map Construction & Localization *[Wang et al. 2020],* Battery health prognosis *[Tagade et al. 2020]*, Antenna Optimization *[Zhang et al. 2020],* Bayesian optimization in Engineering *[Rajaram et al. 2020],* Emulation *[Ming et al. 2022]*

*(List bootstrapped by **Ali Hebbal**'s nice summary in his 2021 thesis)*

Hurricane structure modeling with DGPs

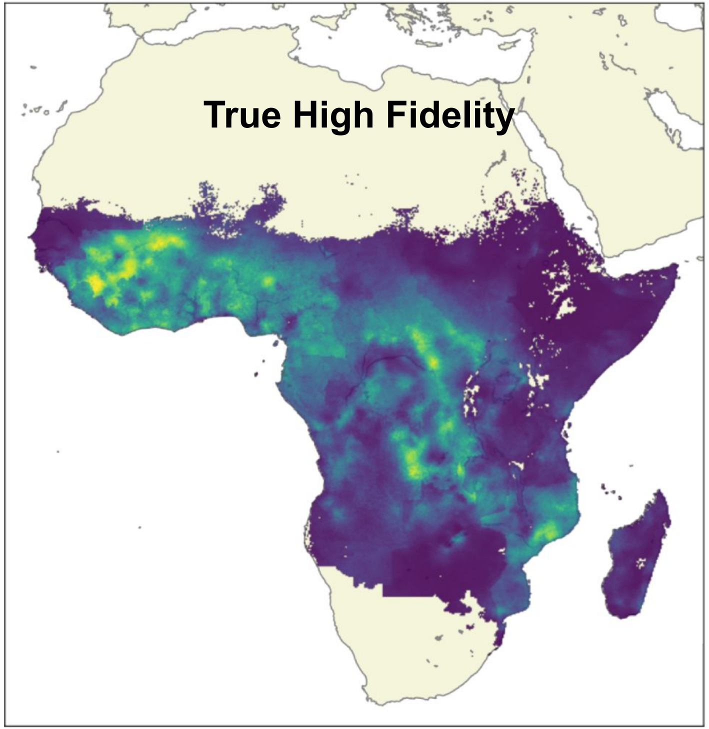
The models to be inverted are highly hierarchical and modular, calling for a layered approach.

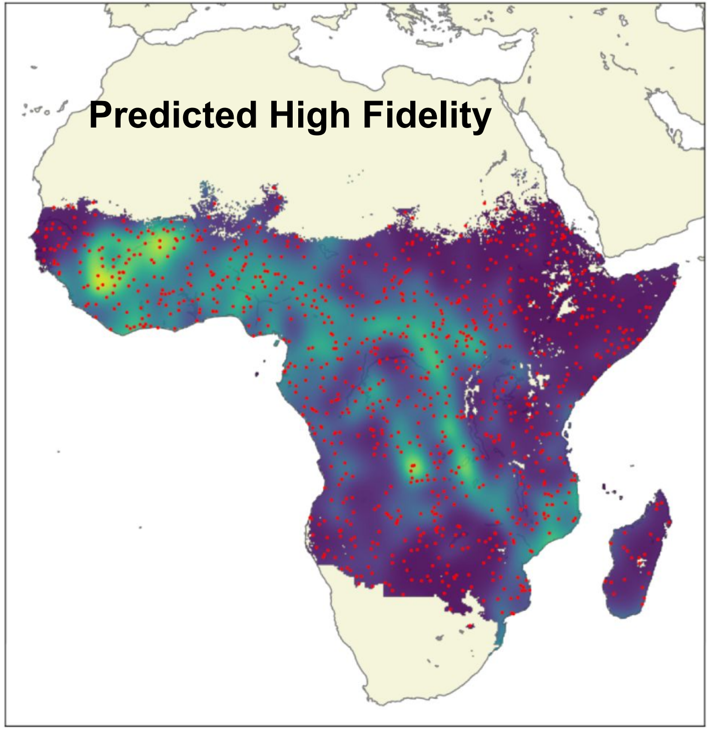


Layer 1 captures low frequencies; layer 2 focuses on the hurricane structure. A single GP layer leads to a too blurry prediction, unable to capture the whirl structure.

[Svendsen et al. 2020]

DGP Multi-fidelity modeling *for infection rates of plasmodium falciparum*





- **High fidelity:** Few samples from left figure (2015 data).
- **Low fidelity:** Many samples from 2005 data.

[Cutajar et al. 2018, Perdikaris et al. 2017]

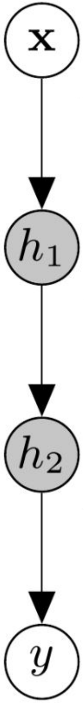
Derivative and related models

NON-EXHAUSTIVE

- **Deep kernel GP** *[Wilson et al. 2015]*
- **SDEs transforming GP inputs** *[Hedge et al. 2019]*

- **Recurrent DGP** *[Mattos et al. 2015, Föll et al. 2017]*
- **Convolutional DGP** *[Blomqvist 2018, Kumar et al. 2018, Singh 2018, Dutordoir et al. 2019]*
- **State-space DGP** *[Zhao et al. 2021]*
- **DGPs over Graphs** *[Li et al. 2020, Opolka and Lió 2022, Jiang et al. 2022]*
- **Connections of DGPs to Transformers** *[Chen et al. 2023]*

Limit properties

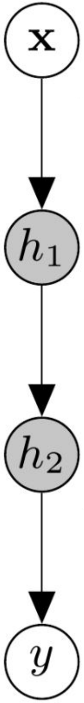


$dim(h_1) \rightarrow \infty \Rightarrow$ GP

h_1 <u>fixed</u> and $dim(h_2) \rightarrow \infty \Rightarrow$ GP

$dim(h_1)$ and $dim(h_2)$ $\rightarrow \infty \Rightarrow$ DeepGP ?

Limit properties



$dim(h_1) \rightarrow \infty \Rightarrow$ GP

h_1 fixed and $dim(h_2) \rightarrow \infty \Rightarrow$ GP

$dim(h_1)$ and $dim(h_2)$ $\rightarrow \infty \Rightarrow$ ~~Deep~~GP

[Lee et al. 2018, Matthews et al. 2018]

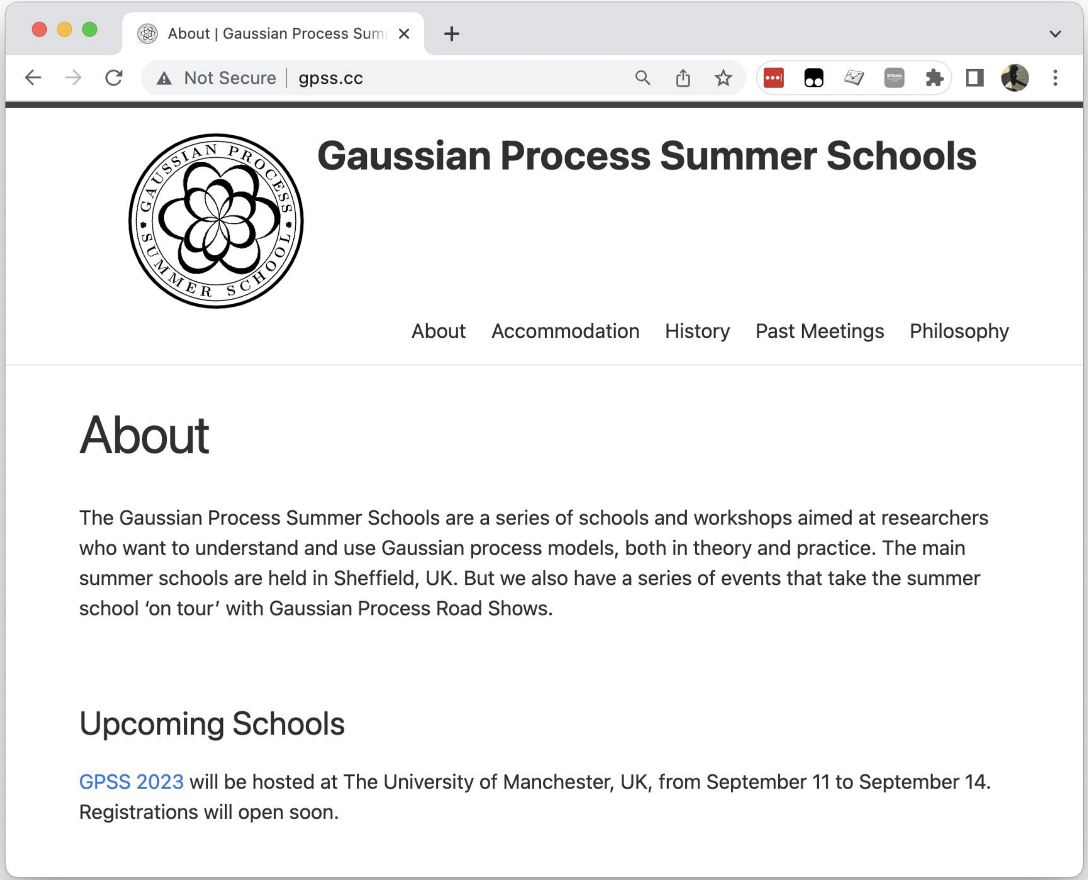
Theoretical analysis of DGPs

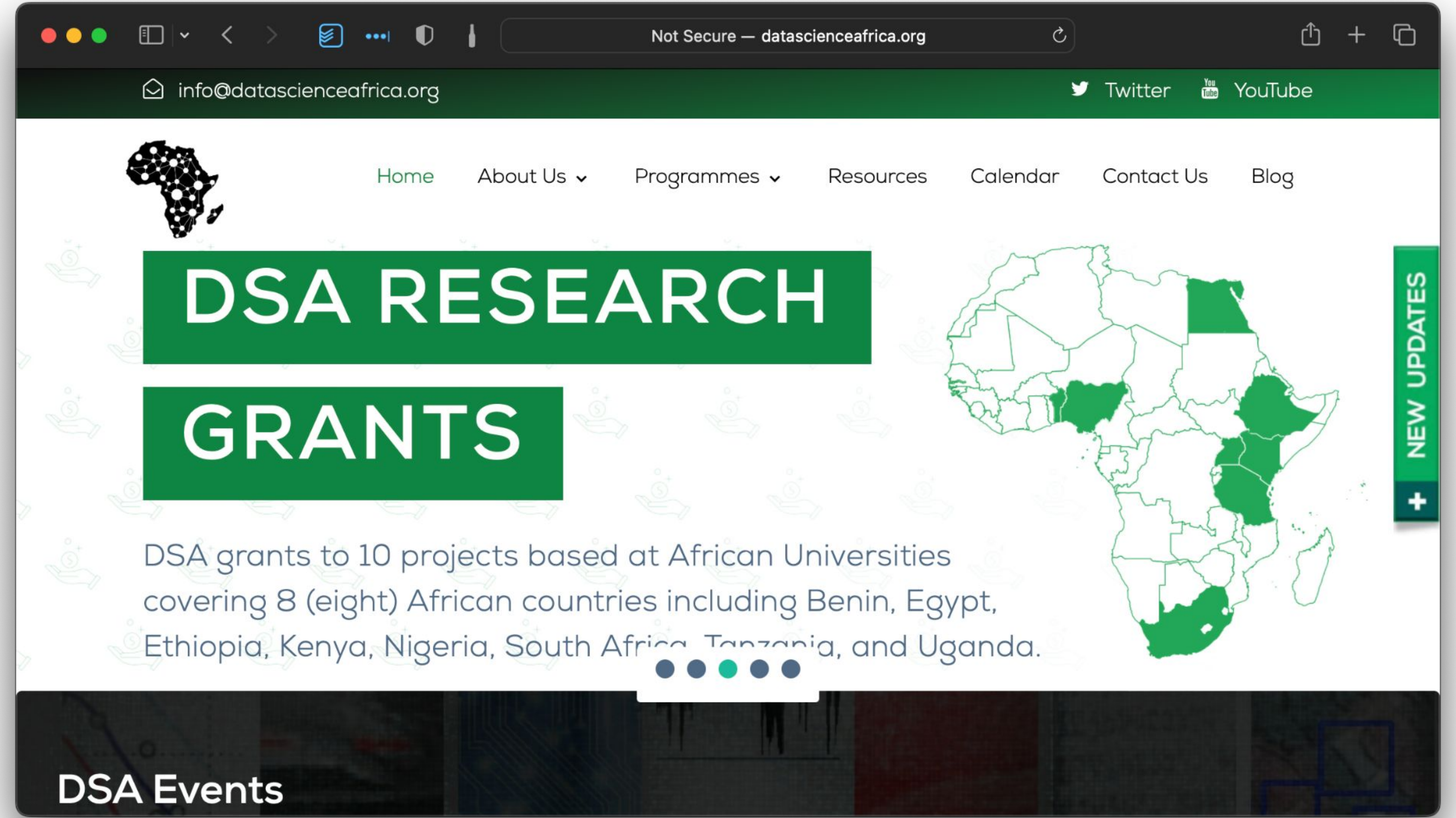
- **Degrees of freedom** reduces across layers for *very deep networks* *[Duvenaud et al.,2014, Dunlop et al. 2018]*. But this can be fixed in the DGP case:
  - Connect inputs to every layer *[Duvenaud et al. 2014]*.
  - Identity mean function *[Salimbeni & Deisenroth 2017]*.
  - Increase width of bottleneck layers *[Dunlop et al. 2018].*

- **Ergodicity analysis** (DGP samples form a Markov chain across layers) *[Dunlop et al. 2018]*
  - Convergence analysis and effect of numbers of layers

- **PAC-Bayesian bounds** for (some) DGP models *[Föll et al. 2019]*

- **DGPs** $\longleftrightarrow$ **DNNs** *[Dutordoir et al. 2021]*: Build a DGP layer with mean that resembles a NN layer


About | Gaussian Process Sum
Not Secure | gpss.cc
Gaussian Process Summer Schools
About Accommodation History Past Meetings Philosophy
About
The Gaussian Process Summer Schools are a series of schools and workshops aimed at researchers who want to understand and use Gaussian process models, both in theory and practice. The main summer schools are held in Sheffield, UK. But we also have a series of events that take the summer school 'on tour' with Gaussian Process Road Shows.
Upcoming Schools
GPSS 2023 will be hosted at The University of Manchester, UK, from September 11 to September 14. Registrations will open soon.

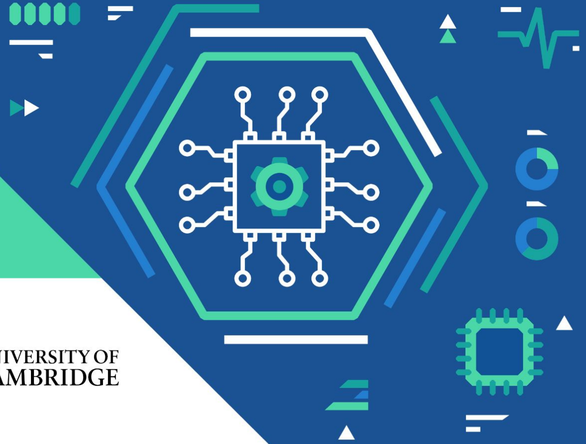

Not Secure — datascienceafrica.org
info@datascienceafrica.org
Twitter
YouTube
Home
About Us
Programmes
Resources
Calendar
Contact Us
Blog
DSA RESEARCH
GRANTS
DSA grants to 10 projects based at African Universities covering 8 (eight) African countries including Benin, Egypt, Ethiopia, Kenya, Nigeria, South Afr
ia, and Uganda.
NEW UPDATES
DSA Events

Amazon’s new drone delivery system:

https://www.youtube.com/watch?v=3HJtmx5f1Fc&t=1s

Machine learning from innovation to deployment

A strategic research agenda for AutoAI



UNIVERSITY OF CAMBRIDGE

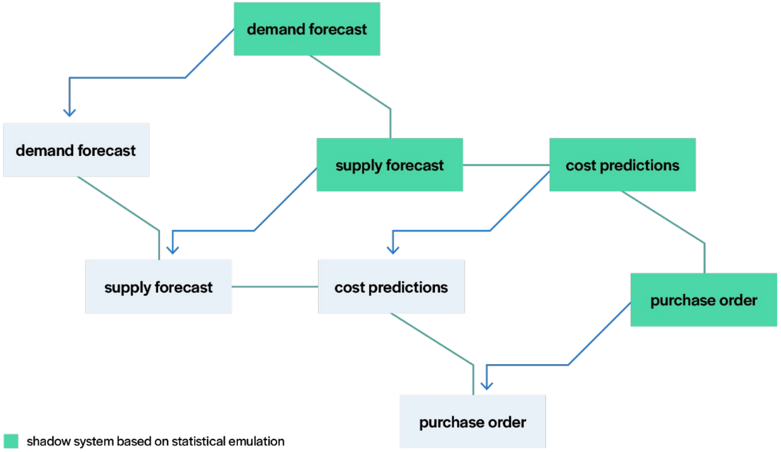
3

Putting systems at the heart of AI deployment

The three D's of AI systems design

Figure 4

A potential path of models in a machine learning system



Careful design of emulators to answer a given question leads to efficient diagnostics and understanding of the system, but in a complex interacting system an exponentially increasing number of questions can be asked. This calls for a system of automated construction of emulators which selects the right structure and redeploys the emulator as necessary. Automatically deploying these families of emulators for full system understanding requires advances in engineering infrastructure, emulation and Bayesian optimisation.

Jupyter Notebook Viewer

nbviewer.org/github/emukit/emukit/blob/main/notebooks/index.ipynb

jupyter nbviewer

JUPYTER FAQ

emukit / notebooks

Emukit tutorials

Emukit tutorials can be added and used through the links below. The goal of each of these tutorials is to explain a particular functionality of the Emukit project. These tutorials are stand-alone notebooks that don't require any extra files and fully sit on Emukit components (apart from the creation of the model).

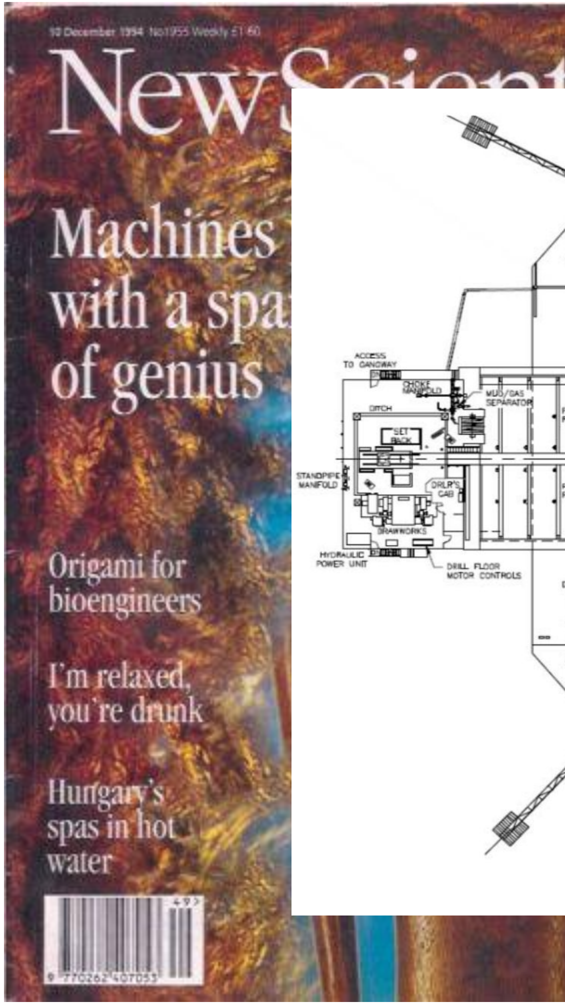
Some tutorials have been written with the purpose of explaining some scientific concepts and can be used for learning about different topics in emulation and uncertainty quantification. Other tutorials are a small guide to describe some feature of the library.

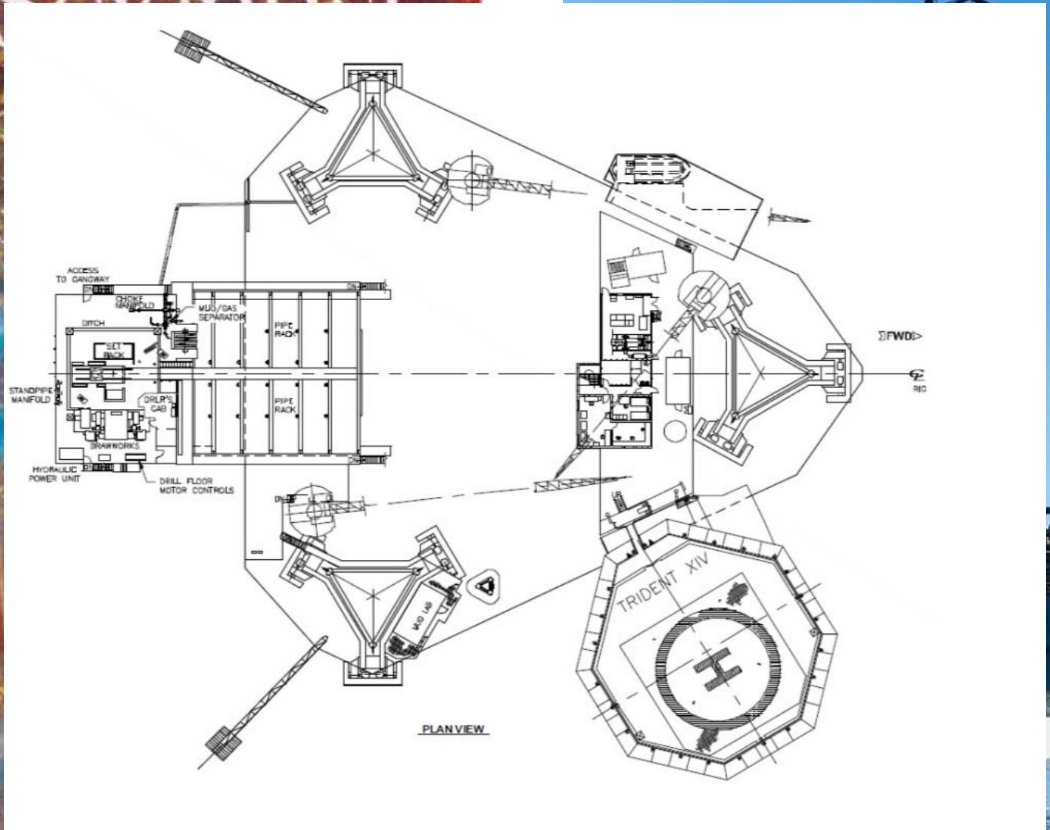
Another great resource to learn Emukit are the examples which are more elaborated modules focused either on the implementation of a new method with Emukit components or on the analysis and solution of some specific problem.

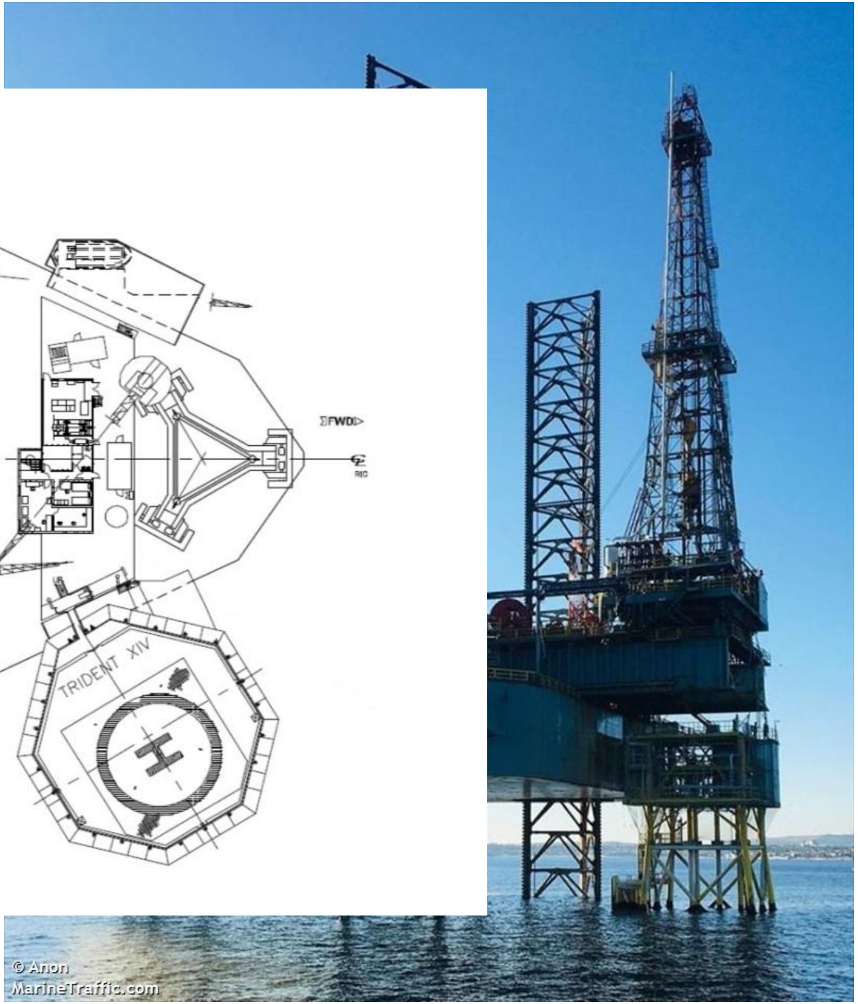
Getting Started

Tutorials in this section will get you up and running with Emukit as quickly as possible.

- 5 minutes introduction to Emukit
- Philosophy and Basic use of the library


10 December 1994 No1955 Weekly £1.60
New
Machines
with a spa
of genius
Origami for bioengineers
I'm relaxed, you're drunk
Hungary's spas in hot water

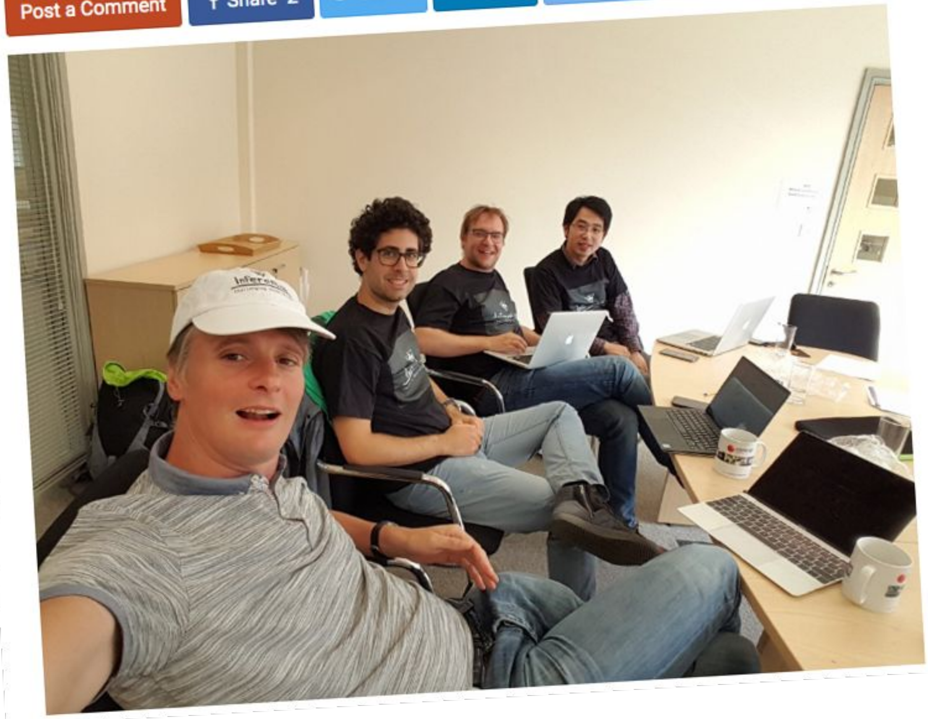

ACCESS TO GANGWAY
CHOKE MANIFOLD
MUD/GAS SEPARATOR
DITCH
SET BACK
PIPE RACK
PIPE RACK
STANDPIPE MANIFOLD
DRLR'S CAB
DRAWWORKS
HYDRAULIC POWER UNIT
DRILL FLOOR MOTOR CONTROLS
MUD LAB
TRIDENT XIV
FWD
RIG
PLAN VIEW


© Anon
MarineTraffic.com

Amazon beefs up machine learning presence in UK with new team of researchers

BY MONICA NICKELSBURG on September 2, 2016 at 8:59 am

Post a Comment | f Share 2 | Tweet | in Share | Reddit | Email



Thanks!