

# Thomas Pinder

📞 +44 7939 668 405 • ✉ [t.pinder2@lancaster.ac.uk](mailto:t.pinder2@lancaster.ac.uk) • 🌐 [tpin.uk](http://tpin.uk) • [GitHub](#)

I am a PhD student in Statistics at Lancaster University, focussing on Gaussian process models and scalable inference with applications to air quality data. I am interested in developing more scalable techniques to enable tractable computations on large, complex datasets. Previously I studied for an MSc in Data Science where my thesis concerned Bayesian neural networks. Before the MSc, I completed a BSc in Mathematics and Statistics with a year in industry.

## Education

- **Lancaster University** **Lancaster**  
2018–Present  
*Statistics PhD (Supervised by [David Leslie](#), [Christopher Nemeth](#) and [Paul Young](#))*  
Developing scalable inferential techniques for Gaussian Processes in order to better model air quality data.
- **Lancaster University** **Lancaster**  
2017–2018  
*Data Science MSc, **Distinction***  
**Awards:** Best Student and Data Science Institute Award for Outstanding Performance.  
**Dissertation:** Adversarial Detection in Computer Vision & Reinforcement Learning Through Bayesian Neural Networks. [\[Thesis\]](#)
- **University of Reading** **Reading**  
2013–2017  
*Mathematics & Statistics With A Year In Industry, BSc, **2:1***  
**Dissertation:** Survival Analysis of Individuals Suffering from the Herpes Zoster Virus.

## Publications and Current Projects

1. **Thomas Pinder**, Kathryn Turnbull, Christopher Nemeth, David Leslie, 2020. **Gaussian processes on hypergraphs.**, Submitted [\[ArXiv\]](#).
2. **Thomas Pinder**, Christopher Nemeth, David Leslie, 2020. **Stein variational Gaussian processes.**, Submitted [\[ArXiv\]](#) [\[Package Repository\]](#).
3. **Thomas Pinder**, Michael Hollaway, Christopher Nemeth, Paul Young, David Leslie, 2020. **A Probabilistic Assessment of the COVID-19 Lockdown on Air Quality in the UK.**, Submitted [\[pdf\]](#).
4. **Thomas Pinder**, 2019. **Barycentric data fusion in Gaussian process models.**, *In progress.*
5. Jamie Fairbrother, Christopher Nemeth, Maxime Rischard, Johanni Brea and **Thomas Pinder**, 2019. **GaussianProcesses.jl: A Nonparametric Bayes package for the Julia Language.**, Submitted [\[ArXiv\]](#) [\[Package Repository\]](#).
6. Shafipour Yourdshahi, E., **Pinder, T.**, Dhawan, G., Soriano Marcolino, L. and Angelov, P.P., 2018. **Towards Large-Scale Ad-hoc Teamwork.**, *IEEE International Conference on Agents.*

## Previous Employment

- **Amazon** **Cambridge**  
June 2021–Present  
*Applied Science Intern*  
I spent time in the supply chain optimisation technologies team lead by [James Hensman](#) working on emulating doubly intractable problems using Gaussian processes and control variates. Within the work, we use the uncertainty in our emulator to optimally select the set of observations that yield the highest utility under some user-specified constraints.
- **Mind Foundry** **Oxford**  
November 2020–February 2021  
*Applied Machine Learning Intern*  
I spent 3 months working with [Alessandra Tosi](#) on multi-fidelity Bayesian optimisation over conditional parameter spaces. This methodological developments were applied to model selection problems where the resultant model should be the most efficient model and not necessarily the most accurate.
- **Relative Insight** **Lancaster**  
December 2017–September 2018  
*NLP Data Scientist*  
I was responsible for researching, prototyping and implementing an unsupervised, abstractive text summarisation algorithm. The final summarisation tool utilised tf-idf, word2vec and LSTM neural networks, all implemented in Python TensorFlow. Originally for internal uses, my work is currently being integrated into Relative Insight's customer dashboard.
- **novi.digital** **Lancaster**  
September 2017–December 2017  
*Data Science Intern*

Using a combination of traditional machine learning methods, I developed a Flask web application to rank a company's web pages based upon their SEO and Google Analytics metrics. Complimentary to this, pages that are underperforming were also considered, with an emphasis on interpretability so we could see why they were under/overperforming. This work was done using Python and predominantly random forest algorithms.

**Institute for Environmental Analytics**

**Reading**

- *Data Analytics Intern*

*August 2016–January 2017*

Working alongside Telefonica I developed a large-scale directed network to facilitate inference into the movement and behaviours of Telefonica phone users. An emphasis for this project was visualisation, and consequently, the Python-based analysis was visualised in a D3.js web application.

**SAP**

**London**

- *Data Analytics Intern*

*July 2015–July 2016*

My primary role focussed on developing forecasting methods to aid the UKI's weekly and quarterly forecasting tool. Further to this, I provided data analysis for sales consultants on prospective sales opportunities.

## Notable Projects

---

- **Masters Project:** *'Adversarial Detection in Computer Vision & Reinforcement Learning Through Bayesian Neural Networks.'* [\[Paper\]](#) [\[Code\]](#)

This project explored the capability of Bayesian neural networks to express reasonable uncertainty estimates in the presence of adversarial examples. This notion was explored in the domain of computer vision and deep reinforcement learning.

- **Distributed AI Project:** *'A Comparison of Multi-Armed Bandit Heuristics in Monte-Carlo Tree Search Based Chess Engine.'* [\[Code\]](#)

This work explored the use of Monte-Carlo tree search to solve *end game* problems in chess. I empirically explored the varying performance of different online learning heuristics, such as Thompson sampling and epsilon-greedy.

- **BSc Thesis:** *'Survival Analysis Approaches of the Herpes Zoster Virus and the Effects Caused from Reactivation within the Trigeminal Nerve.'* [\[Paper\]](#)

A detailed review of existing method in survival analysis was first presented, both from a theoretical and a practical perspective. Building upon this literature review, an accelerated failure time model was devised to model the time-to-recovery for patients suffering with shingles.

## Teaching

---

I have acted as a teaching assistant for the following modules: Probability theory (MATH220), Multivariate statistics for machine learning (MATH336), Programming for data scientists (MATH550), Statistical inference (CFAS406), Statistical models (MATH333) and Computational mathematics (MATH245).

## Technical and Personal skills

---

- **Programming Languages:** Proficient in Python, Julia, R, and JavaScript.  
Basic ability with SAS and MATLAB.
- **Machine Learning Frameworks:** Proficient in TensorFlow, Jax, GFlow and PyTorch  
Basic ability: Stan, Pyro and PyMC3.
- **Supplementary Technical Skills:** AWS (EC2, RedShift, S3 and SageMaker), Bash, SQL, Tableau, Git, Docker.

## Interests and extra-curricular activity

---

- I'm a keen endurance athlete, specifically focusing on ultra-marathons and long-distance cycling. Over the years I've raised over £5,000 through self-supported efforts such as cycling home to the UK from Pisa, Italy and running the full 300km of the Thames Footpath in 3 days.
- In my spare time I enjoy woodwork and playing chess.

## References

---

- References are available on request.