Affordable Bayesian Optimisation

- a question of priors?

Sigrid Passano Hellan

Work with Christopher Lucas and Nigel Goddard

University of Edinburgh

30th June 2023

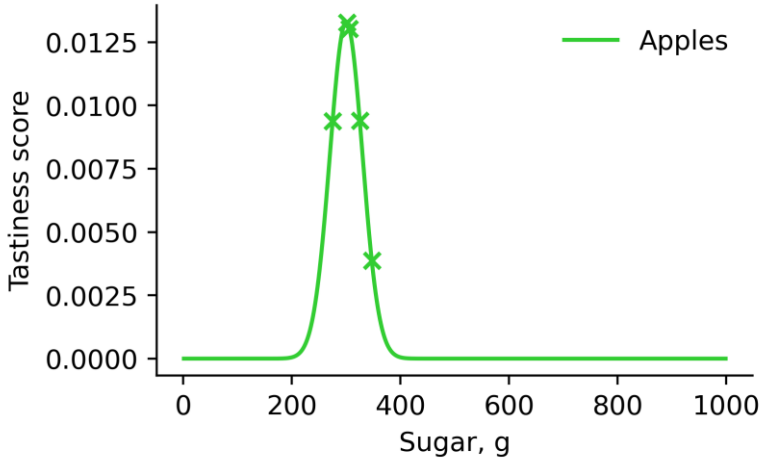
Photo: O. Hellan

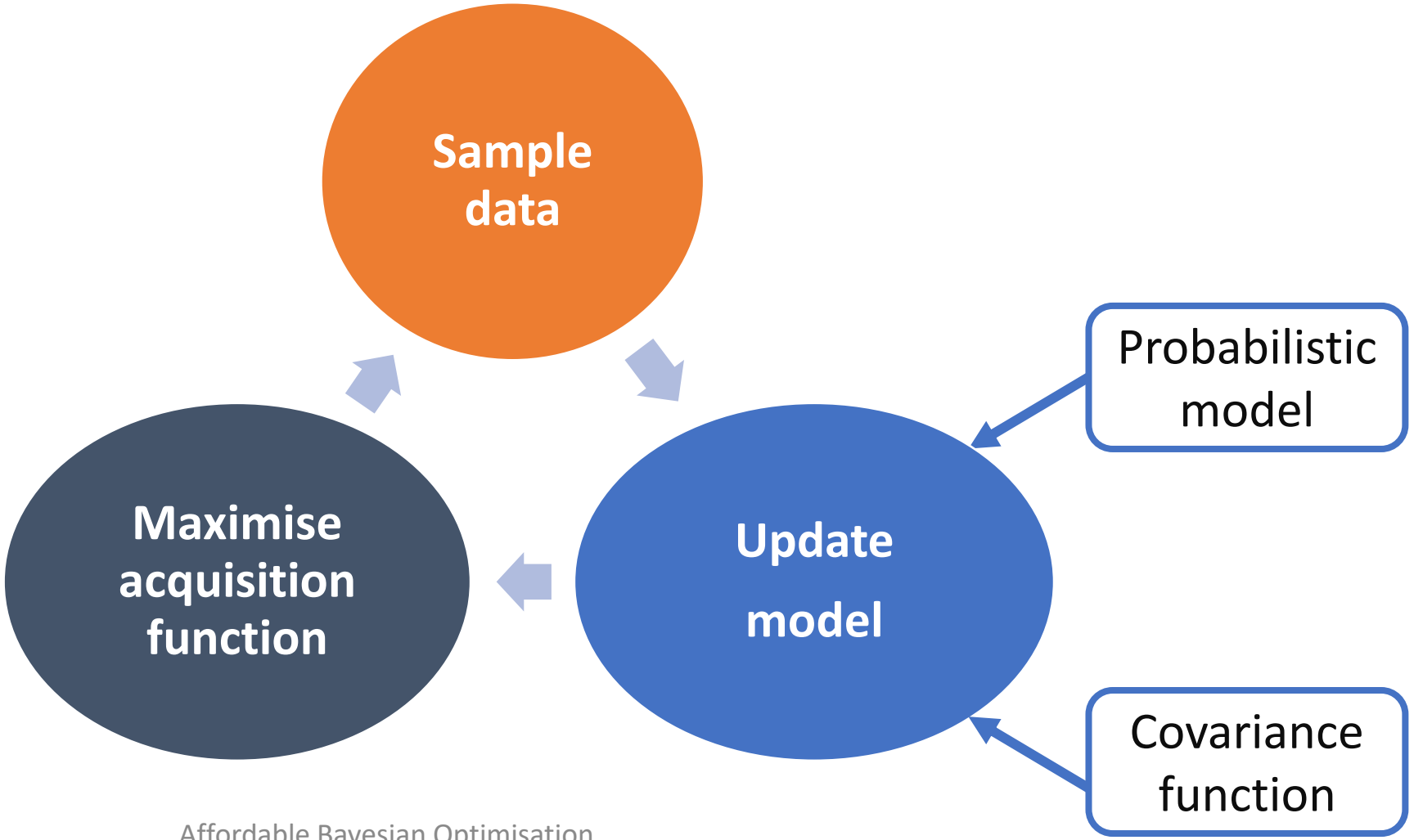
THE UNIVERSITY of EDINBURGH

UK Research and Innovation

DATASCIENCE EPSRC CENTRE FOR DOCTORAL TRAINING

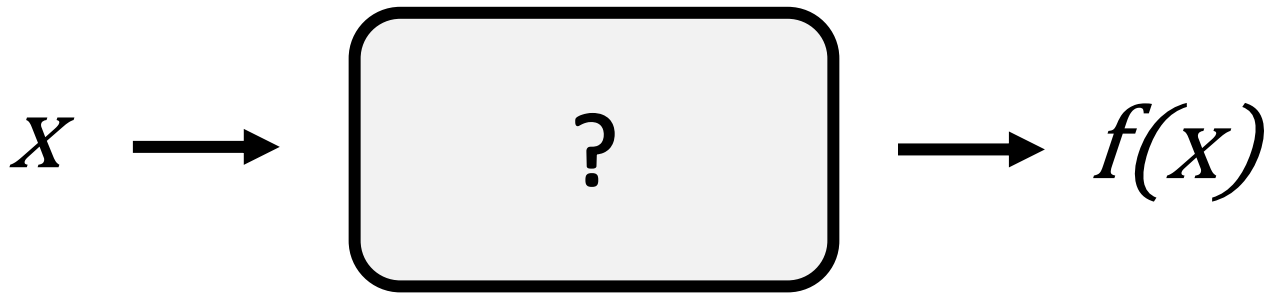
Bayesian optimisation





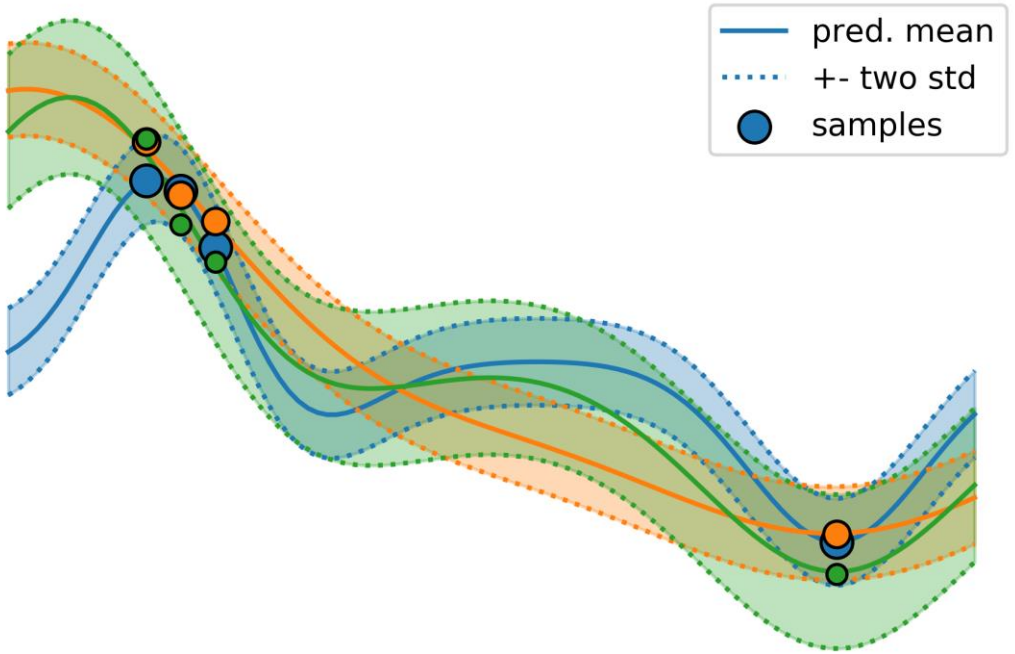
What makes Bayesian optimisation affordable?

- Sample efficient optimisation
- No derivatives needed
- No simulation model needed
- Common application: hyperparameter optimisation (HPO)



What makes Bayesian optimisation **un**affordable?

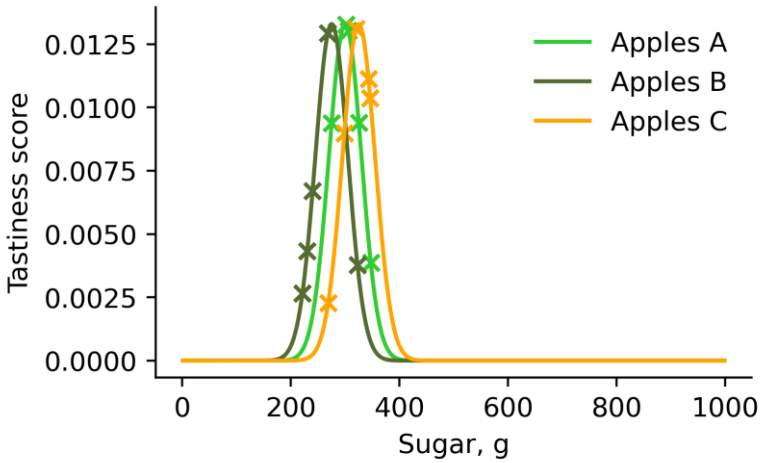
- Scaling in the number of samples
- **Learning model with few samples**

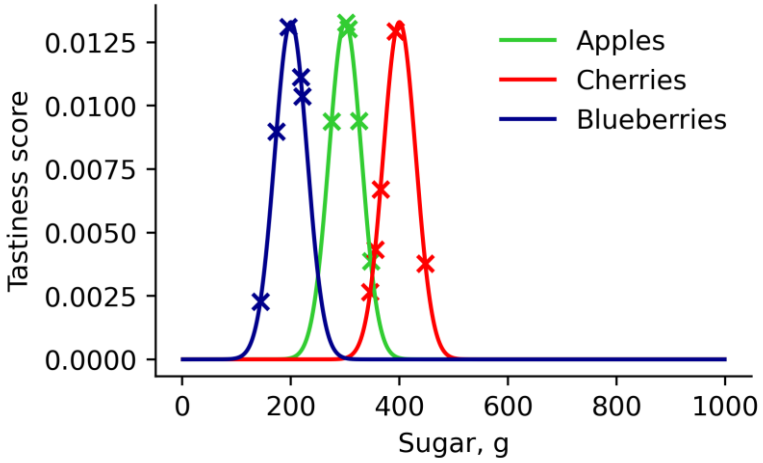


Transfer learning

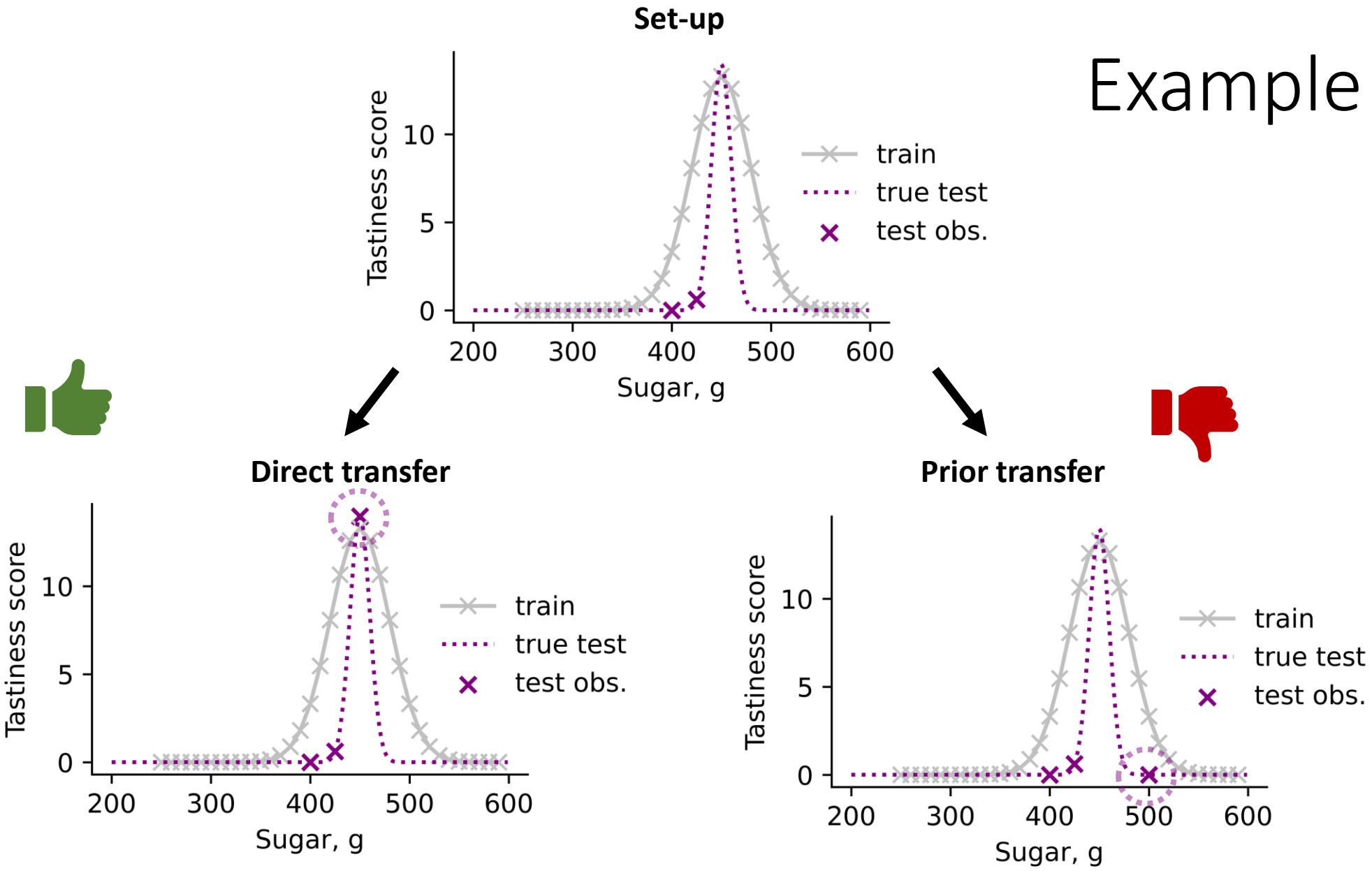
- Assume data from related problems
  - What's the best way to use it?
    - **Direct transfer**: Learn x^* (optimum)
    - **Prior transfer**: Learn shape of $f(x)$

- *Imagine you're making jam…*

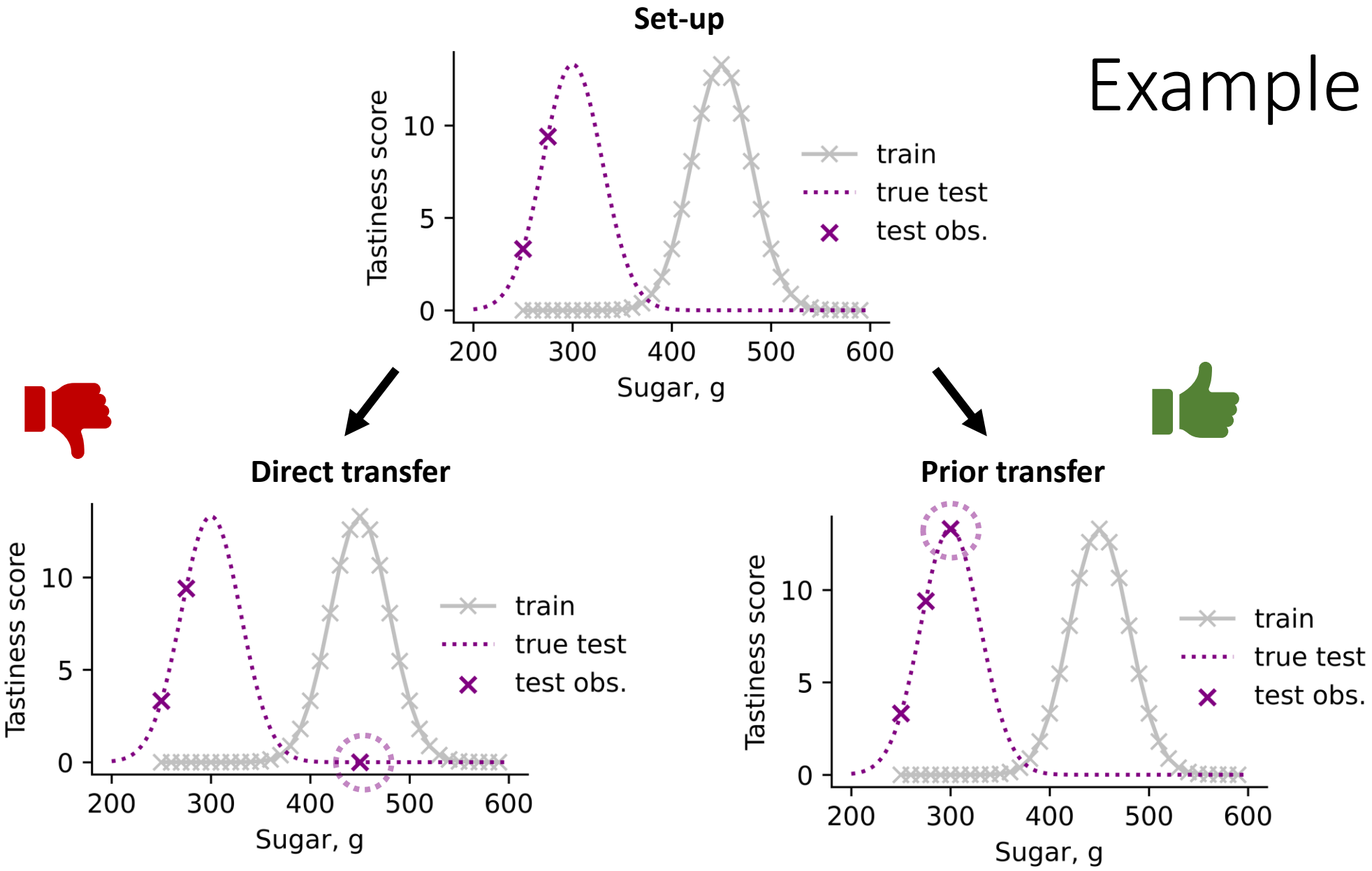




Example 1


Set-up
Tastiness score
Sugar, g
train
true test
test obs.
Direct transfer
Prior transfer

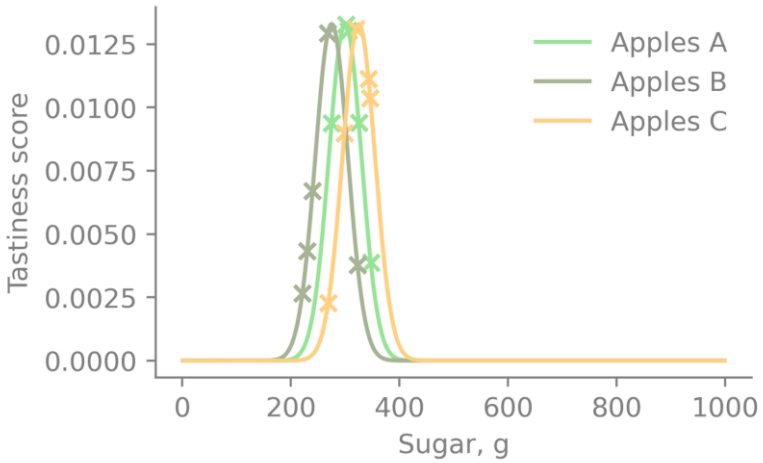
Example 2

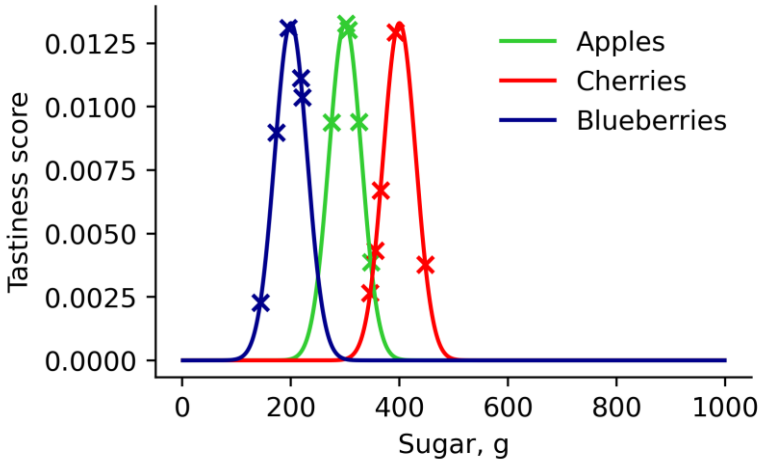

Set-up
Tastiness score
Sugar, g
train
true test
test obs.
Direct transfer
Prior transfer

Transfer approaches

- Direct transfer
- Few-Shot
- Prior learning

Prior transfer





Transfer approaches

- **Direct transfer**
  - Learn feature vectors x s.t. $f(x) \approx f^*$
  - Examples:
    - Joint model (Yogatama 2014)
    - Warm-starting (Feurer 2015)
    - Reduce search space (Perrone 2019)
- Few-Shot
- Prior learning

Yogatama, D. & Mann, G. *Efficient Transfer Learning Method for Automatic Hyperparameter Tuning*. AISTATS, 2014

Feurer, M., Springenberg, J. T. & Hutter, F. *Initializing Bayesian Hyperparameter Optimization via Meta-Learning*. AAAI, 2015

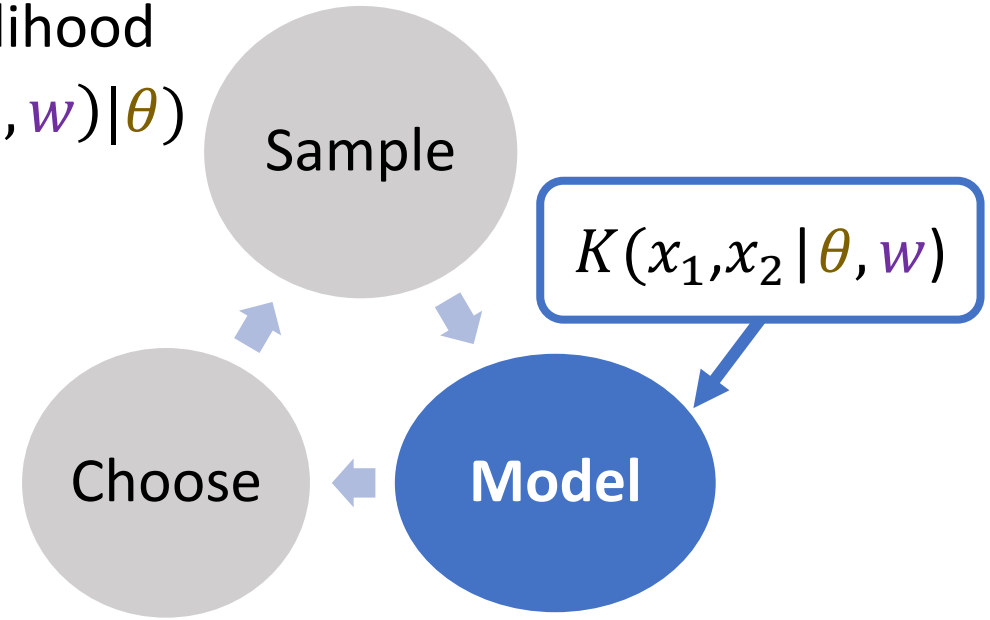
Perrone, V., Shen, H., Seeger, M.W., Archambeau, C. & Jenatton, R. *Learning search spaces for Bayesian optimization: Another view of hyperparameter transfer learning*. NeurIPS 2019

Transfer approaches

- Direct transfer
- **Few-Shot** (Wistuba & Grabocka 2021)
  - Deep kernels
  - φ is a two-layer neural network 128 -> 128, with weights w
  - Learn θ, w by gradient descent on marginal likelihood

$$K_{deep}(x_1, x_2|\theta, w) = K_{RBF}(\varphi(x_1, w), \varphi(x_2, w)|\theta)$$

- Prior learning

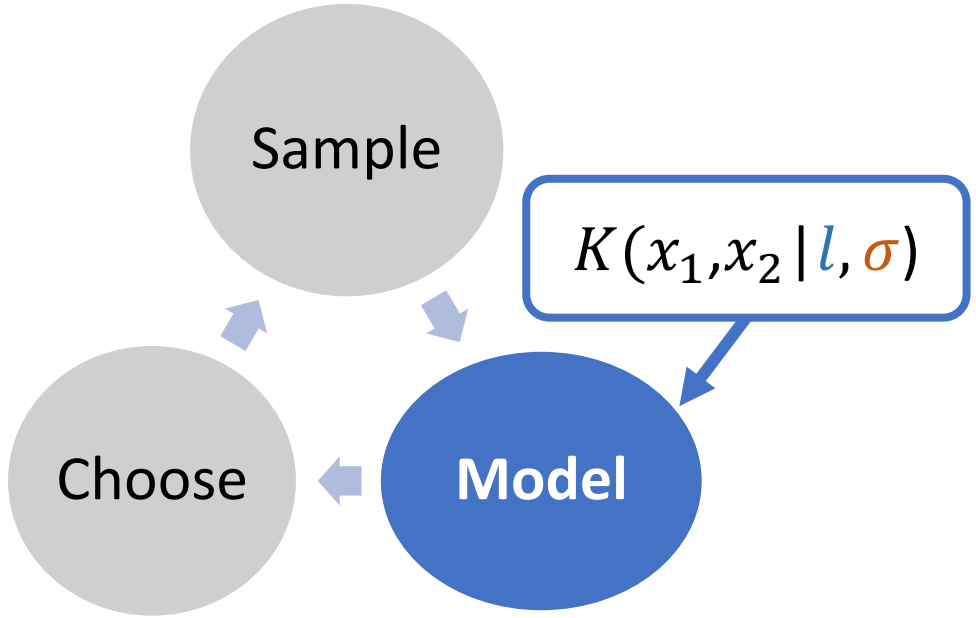


Wistuba, M. & Grabocka, J. *Few-Shot Bayesian Optimization with Deep Kernel Surrogates*. ICLR 2021

Transfer approaches

- Direct transfer
- Few-Shot
- **Prior learning (Ours)**
  - Learn distributions for hyperparameters l and $\sigma \in \theta$

$$K_{RBF}(x_1, x_2|l, \sigma) = \sigma^2 \exp^{-\frac{1}{2}\frac{\|x_1 - x_2\|_2^2}{l^2}}$$



Hellan, S. P., Lucas, C. G., & Goddard, N. H. *Bayesian Optimisation for Active Monitoring of Air Pollution*. AAAI, 2022

Our prior learning approach

Preprocessing:

- Learn priors

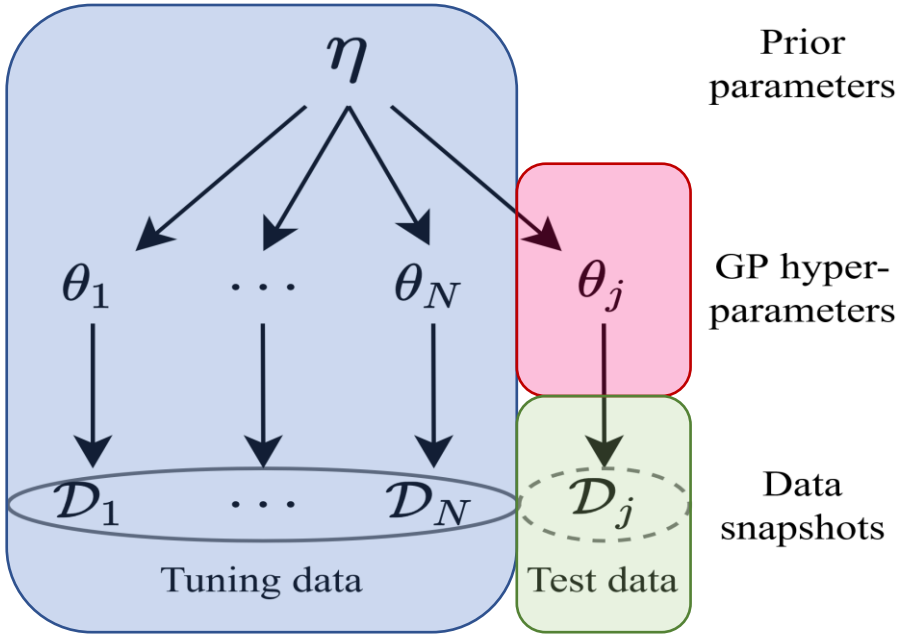
MCMC

Sample candidates

Within BO loop:

- Weighted acquisition function
  - Fit to the observed data

Importance weighting



Key differences

	Prior learning (ours)	**Few-Shot**
• Approach	• Learn l and σ	• Learn w, l and σ
• Output	• Distribution	• Point estimates
• How?	• MCMC	• Train neural network
• BO loop	• Importance weighting	• Fine-tune neural network

• Application	• Air pollution
	  • ~ 200 tasks, ~ 40 data points
		• Hyperparameter optimisation
		  • ~ 50 tasks, ~ 100 data points
		  • ~ 30 tasks, ~ 25 000 data points

Air pollution monitoring

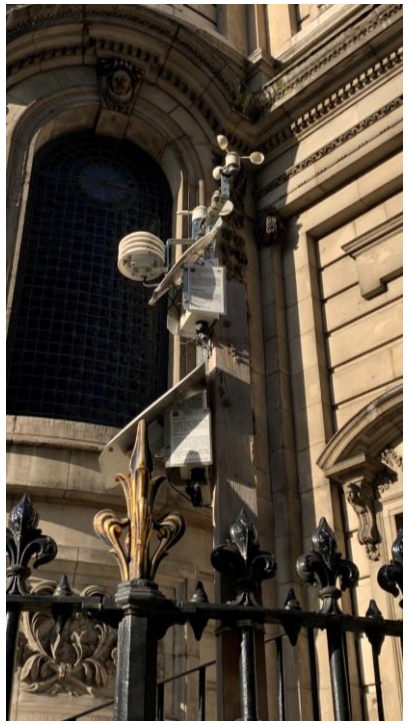
Want to find pollution maximum in a city as quickly as possible by iteratively placing sensors

Air pollution monitors

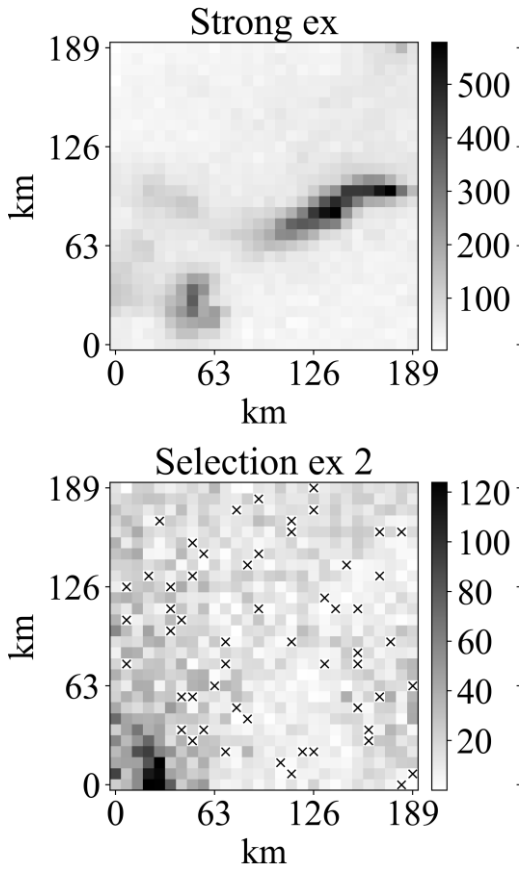
Edinburgh

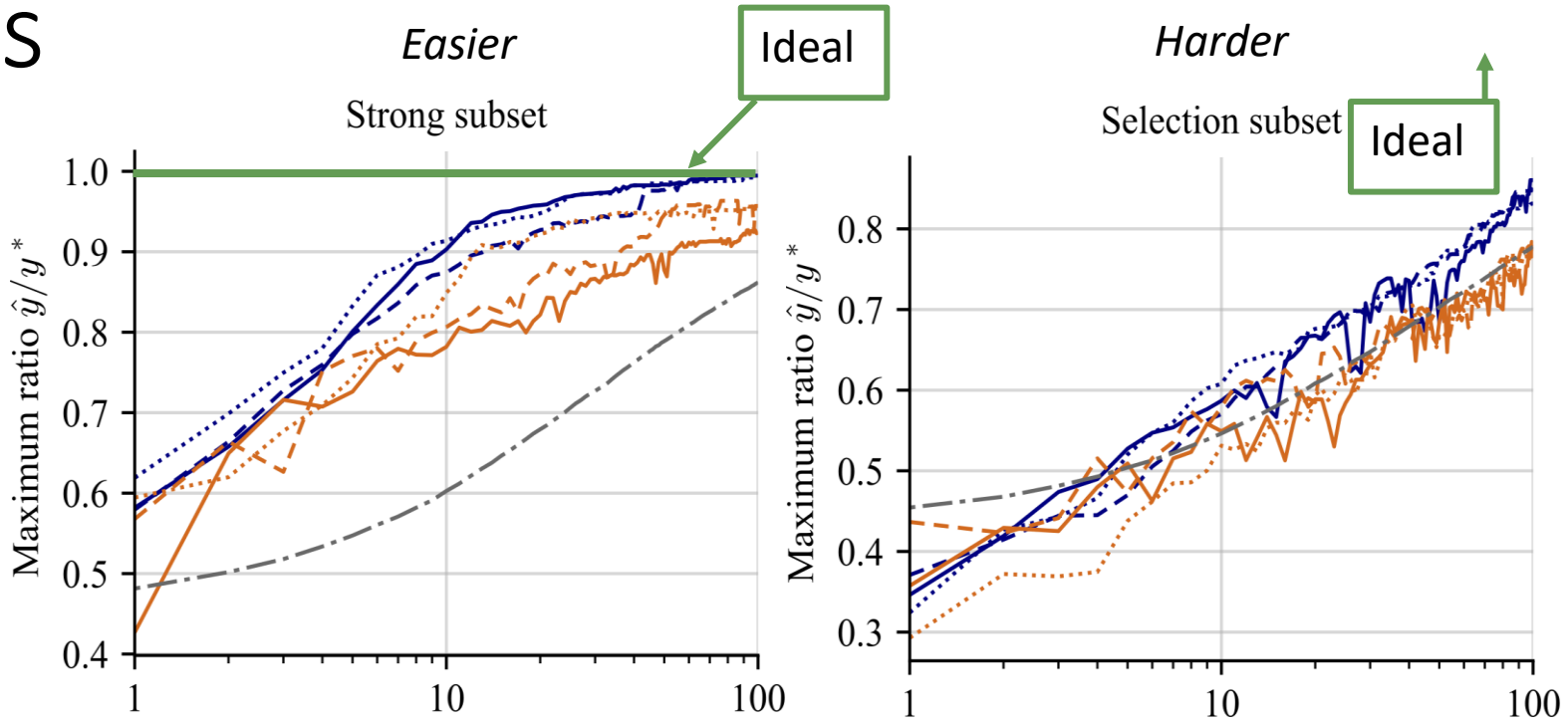
Berlin

London, low-cost



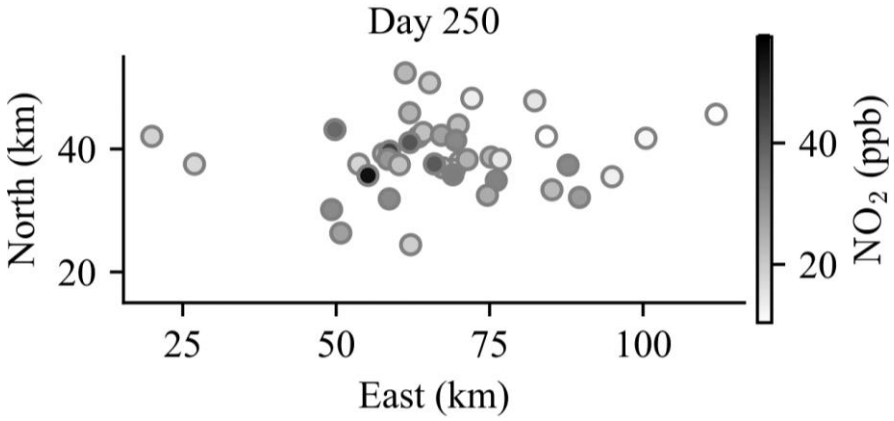
Satellite results

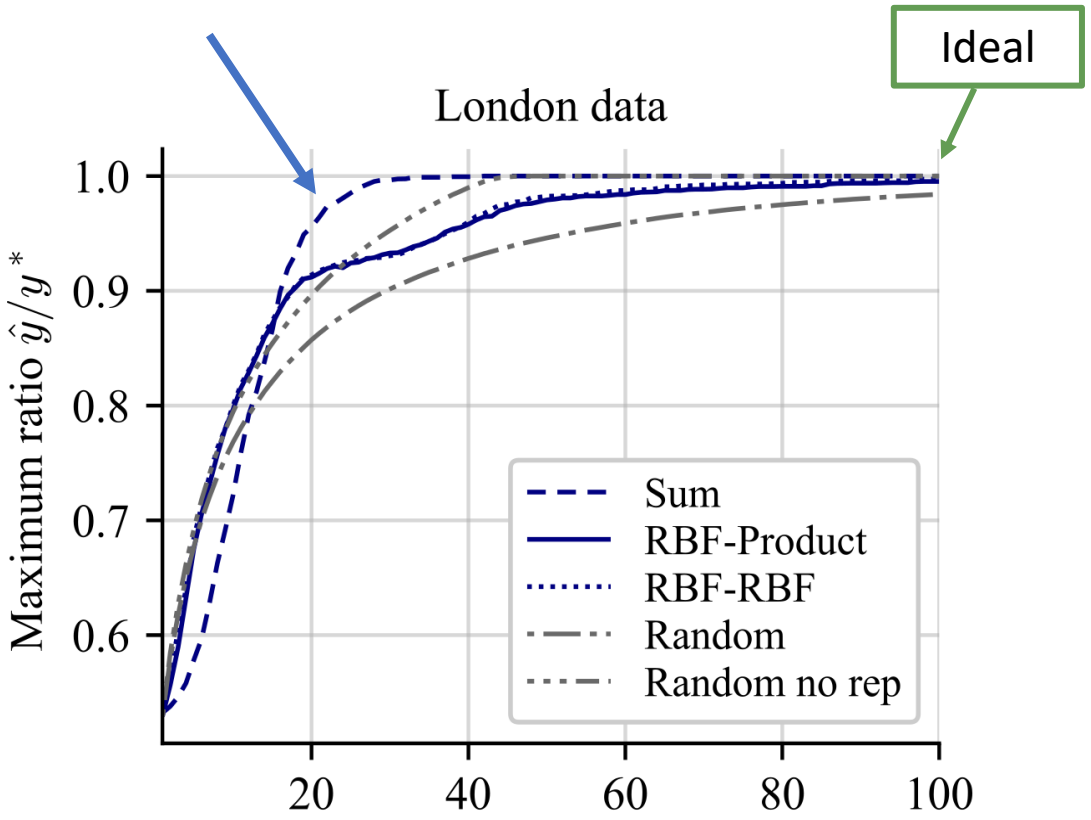




Grey: Random sampling with replacement
Orange: Simpler priors
Blue: Prior learning (ours)

London results





Grey: Random sampling, with and without replacement
Blue: Prior learning (ours)

https://www.londonair.org.uk/

Future work

- Comparison to Few-Shot
  - Show that ours needs less training data?
- Results on synthetic data
- Ablations
- Other applications? Come talk to me

Thank you!

Questions?

All figures by authors unless otherwise stated.

Yogatama, D. & Mann, G. *Efficient Transfer Learning Method for Automatic Hyperparameter Tuning*. AISTATS, 2014

Feurer, M., Springenberg, J. T. & Hutter, F. *Initializing Bayesian Hyperparameter Optimization via Meta-Learning*. AAAI, 2015

Perrone, V., Shen, H., Seeger, M.W., Archambeau, C. & Jenatton, R. *Learning search spaces for Bayesian optimization: Another view of hyperparameter transfer learning*. NeurIPS 2019

Wistuba, M. & Grabocka, J. *Few-Shot Bayesian Optimization with Deep Kernel Surrogates*. ICLR 2021

Hellan, S. P., Lucas, C. G., & Goddard, N. H. *Bayesian Optimisation for Active Monitoring of Air Pollution*. AAAI, 2022