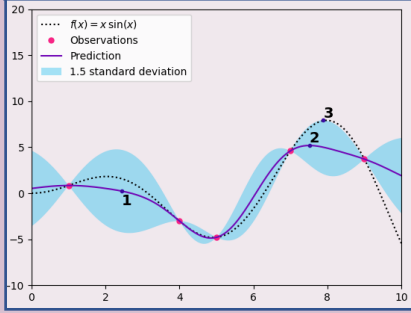
Bayesian optimisation in Chemistry

Rubaiyat Khondaker[1], Stephen Gow[2], Mahesan Niranjan[3], Jeremy Frey[2]
[1]Department of Mathematics, University of Cambridge; [2]Department of Chemistry, University of Southampton; [3]Department of Electronics and Computer Science, University of Southampton
rmk47@cam.ac.uk

Skills 4 Scientists

Introduction to Bayesian Optimisation



From initial observations, construct **statistical model**. Then choose new observation based on **acquisition function**:

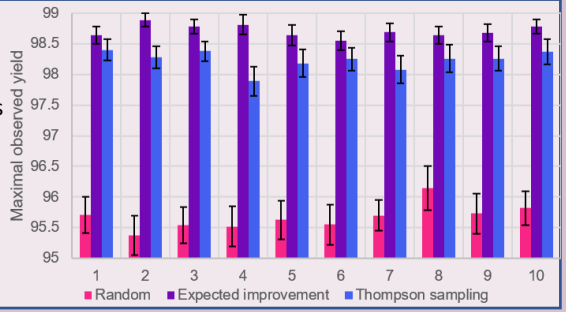
1) Max uncertainty
2) Max mean
3) Max mean + uncertainty

In practice a balanced approach like **3** is preferred.

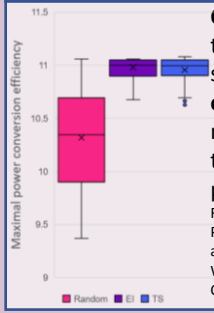
Suzuki reaction yield dataset

From **EDBO optimiser** paper (Experimental Design in Bayesian Optimisation). Tested with **50 random initialisations** per configuration. Aim was to determine dependence of maximal observed yield on **batch size**. Results indicate **stable performance**. Expected improvement (EI) consistently outperforms Thompson Sampling (TS).

Reference: Bayesian reaction optimization as a tool for chemical synthesis, Nature, 2021, *590* (7844), 89–96.



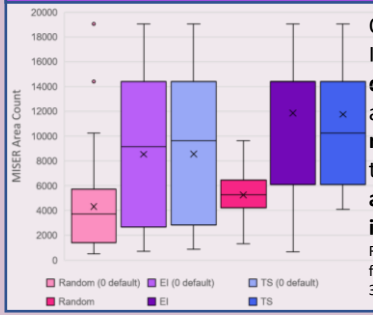
Harvard Clean Energy



Organic photovoltaic data tabulating molecular structures and **power conversion efficiency**. Global max c. 11.1%. EI took ~ 4 times as long as TS, but produced similar results.

Reference: The Harvard Clean Energy Project: Large-Scale Computational Screening and Design of Organic Photovoltaics on the World Community Grid, Journal of Physical Chemistry Letters, 2011, 2, 2241–2251.

Nanomole-scale high-throughput screening



Optimising Area Count (LC-MS). Initial domain had "holes" due to **disallowed combinations**, so coded 0 as default for these. Produced **poor results** (pale colours). Subsequently these combinations were excluded – **average performance markedly improved**.

Reference: Nanomole-scale high-throughput chemistry for the synthesis of complex molecules, Science, 2015, 347 (6217), 49-53

Conclusions

Bayesian optimisation is a promising technique with the potential to be used across a **wide variety of problems**. Only **small modifications** were required to transfer an algorithm built for reaction yield optimisation into a very different domain. Future work could explore **noisy** objective functions, **'generative'** optimisation, **time-dependent** objective functions, or **other problem domains.**

AI³ Science Network+

PSDS PHYSICAL SCIENCES DATA-SCIENCE SERVICE

RSC INTEREST GROUP CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP