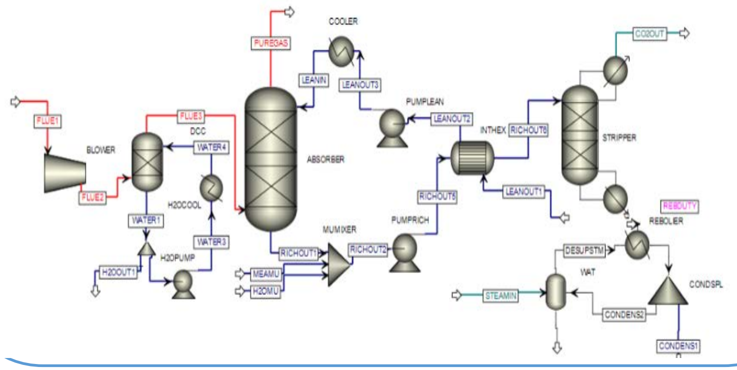
Reduced Order Modelling of Post-combustion CO_2 Capture

Dr Robert Milton, University of Sheffield

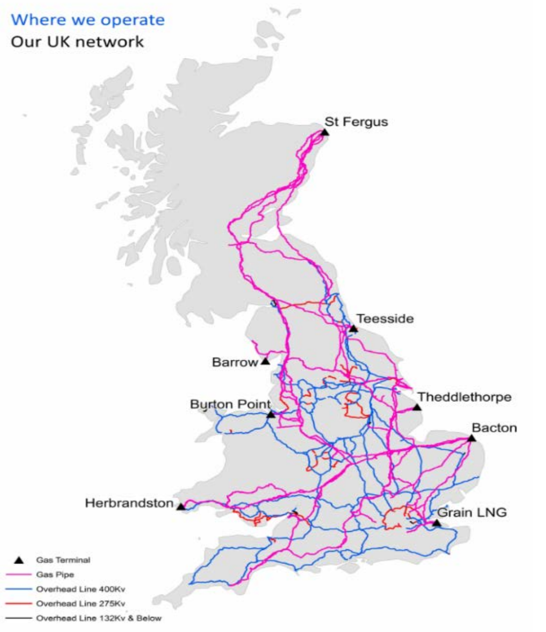
Rationale for a Reduced order model

Process Model (Simulation)



Reduced Order Model
$f_{out}(\boldsymbol{x}_{in})$

UK Energy System Model



Steps to a Reduced Order Model

1. CO_2 capture simulation results

   The curse of dimensionality

2. Gaussian process (GP) surrogate

3. Relevance determination

   Sobol' indices

4. Combined inputs for optimal relevance

CO_2 capture simulation results (Dr Eni Oko)

UKCCS RESEARCH CENTRE

- Solvent (MEA) based PCC for a 453 MWe NGCC power plant.
- 5 input parameters:
  - Flue gas flowrate (kg/s)
  - CO_2 in flue gas (wt%)
  - Solvent flowrate (kg/s)
  - Solvent concentration (wt%)
  - Lean loading (mol CO_2 /mol MEA)
- 8 outputs analysed independently, we shall describe 2:
  - CO_2 capture level (%)
  - Reboiler duty (MWth)
- Statistical dependence between outputs may be considered in future…

CO_2 capture simulation results

UKCCS RESEARCH CENTRE

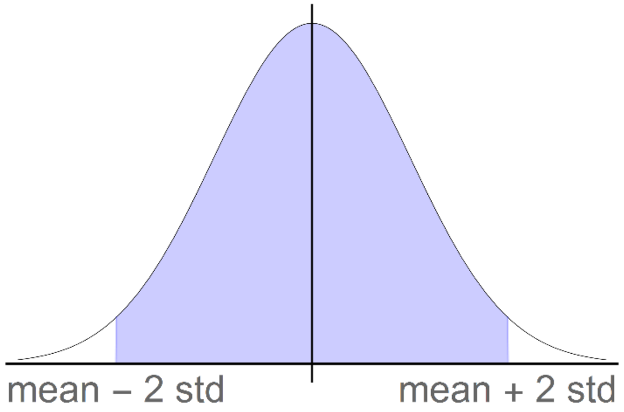
- The curse of dimensionality: with 5 inputs, each output is essentially a 6-dimensional scatter plot.
  - Impossible to usefully visualize.
  - Analysis – such as techno-economic optimization or hazard assessment – is computationally expensive and prone to large errors.
  - High dimensions embody surprising mathematical characteristics, such as the prevalence of extreme conditions. For example, if each input is in “normal” range 80% of the time, all 5 inputs are “normal” less than 33% of the time.

- As more input parameters are included to improve model fidelity, the curse of dimensionality only gets worse…

Gaussian process surrogate

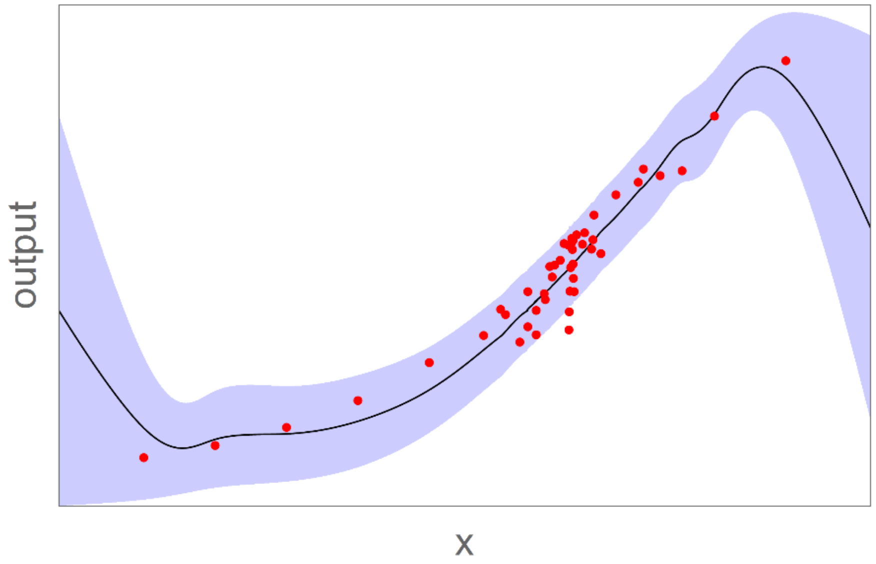
- The GP is a surrogate for full simulation results which:
  - Is fast to compute.
  - Is mathematically convenient, enabling efficient and sophisticated analysis.
  - Automatically measures the uncertainty/error in its predictions.

- The GP gives a Gaussian distribution for the output resulting from an input vector **x**.
  - The mean of this distribution is the best predictor of the output to be seen at **x**.
  - The standard deviation of this distribution measures the predictor's uncertainty.



Gaussian process surrogate

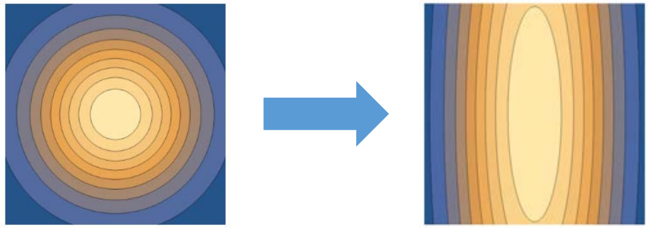
- GP regression learns the output distribution at any input vector **x** from the simulation output (training data).
- Learning presumes that similar inputs probably produce similar output.



- GP automatically smooths (regularizes) the full simulation results, ensuring robust prediction and analysis.
- GPs are only good for interpolation, not extrapolation.

Gaussian process surrogate

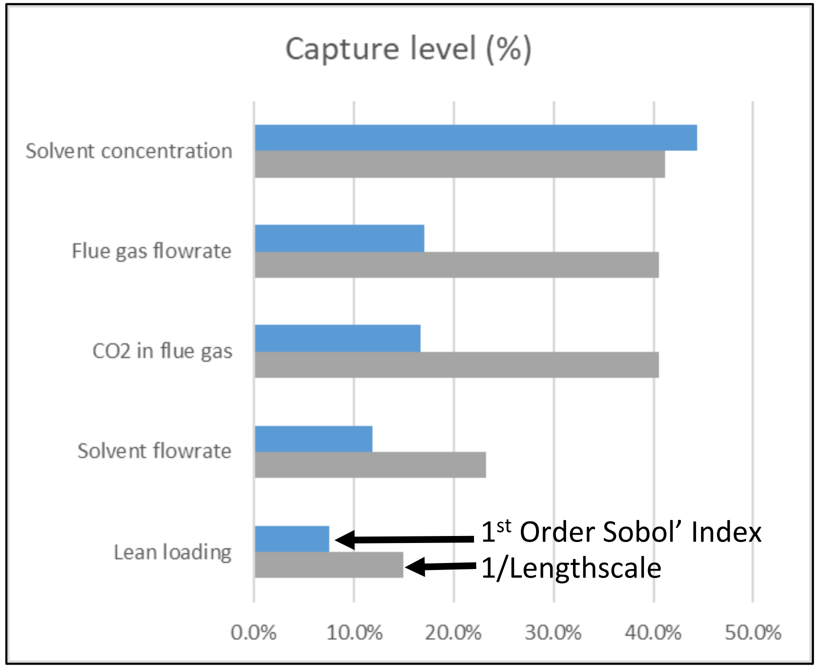
- The GP decomposes the output into signal and noise.
- The signal is determined by similarity between inputs, measured by the kernel.
- The ARD kernel we use employs a different lengthscale along each input dimension to measure similarity or nearness. Longer lengthscales entail smoother output.

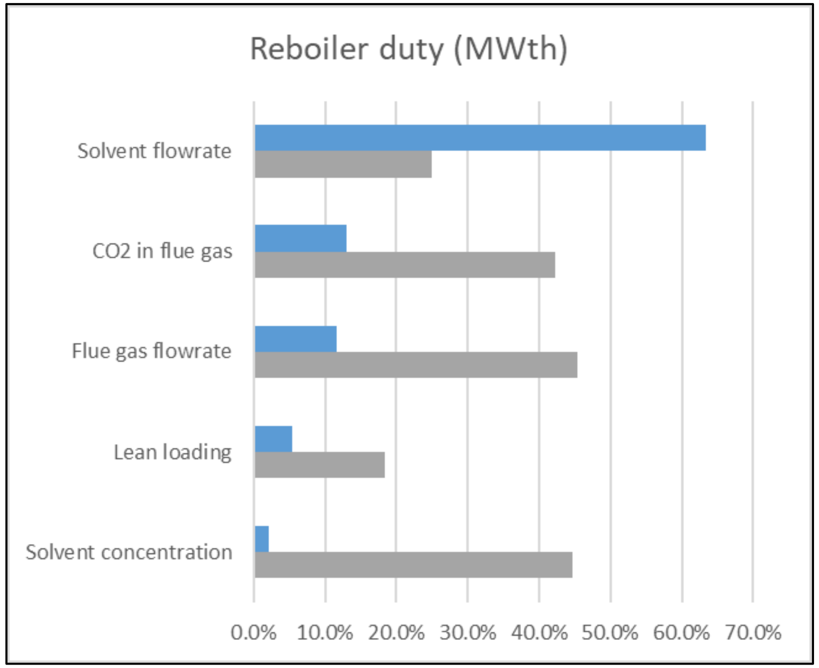


- Kernel lengthscales are chosen to maximize the likelihood of the training data.

Relevance determination

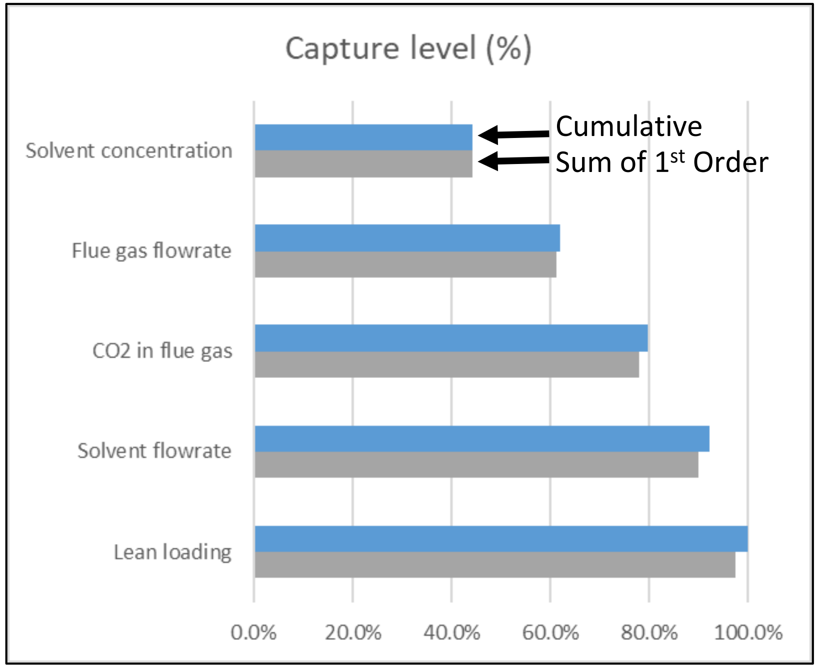
- The 1st order Sobol' index S_i is the fraction of total output variance which is ascribable to the i^{th} input alone.
- Calculating Sobol' indices is hard to do accurately on data, but semi-analytic on a GP surrogate.
- Ordering input dimensions by decreasing S_i orders them by decreasing relevance.

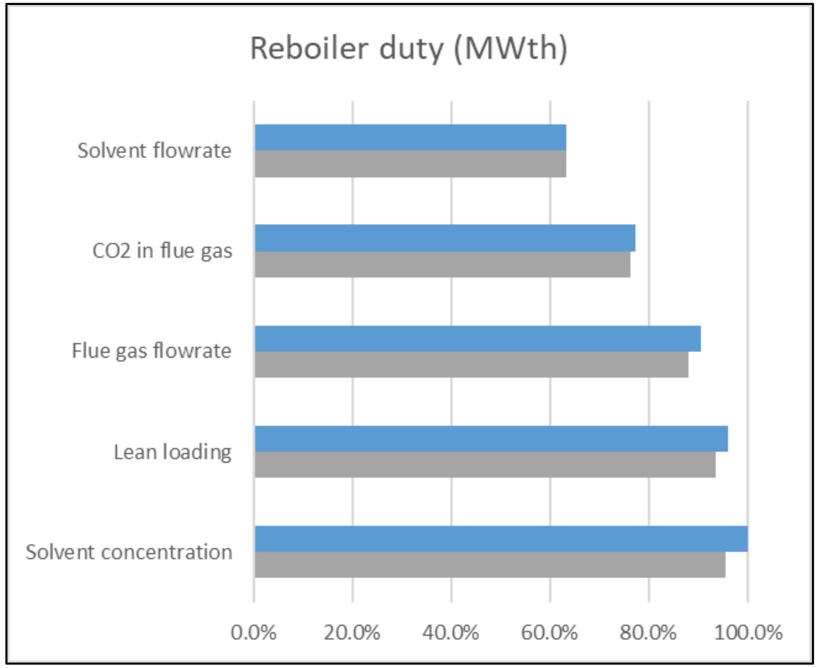




Relevance determination

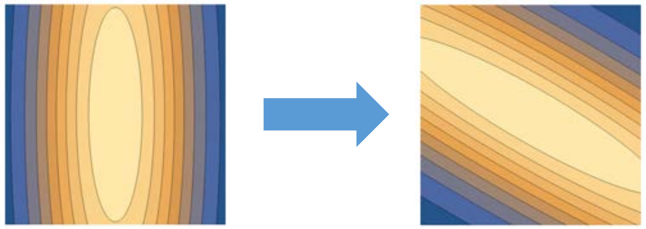
- The cumulative Sobol' index $S_{1,..,j}$ captures interactions between inputs 1,…,*j*.
- $\sum_{i=1}^{j} S_i \leq S_{1,\ldots,j} \leq S_{1,\ldots,5} = 100\%$
- Ordering input dimensions to maximize $S_{1,..,j}$ for *j*=1 to *j*=5 successively maximizes the relevance of the first *j* input dimensions taken together.

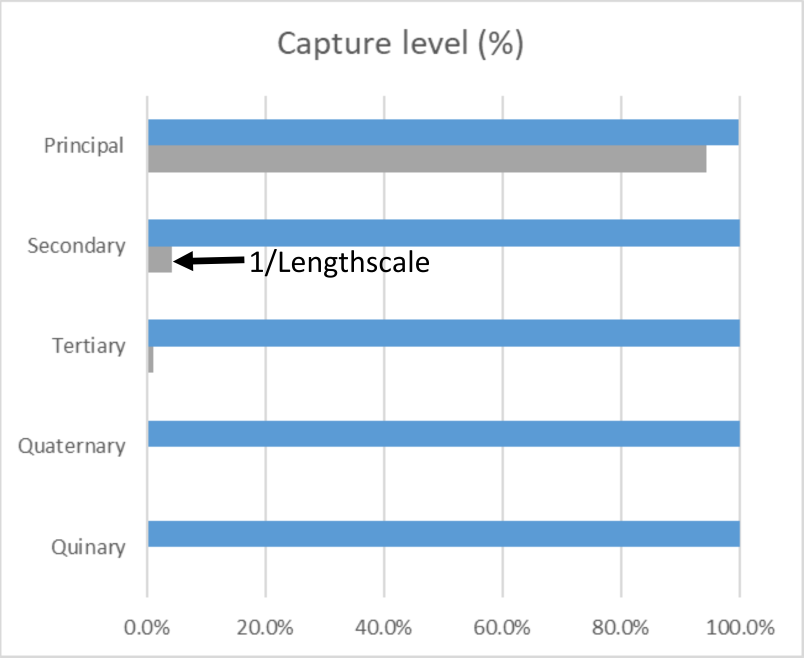


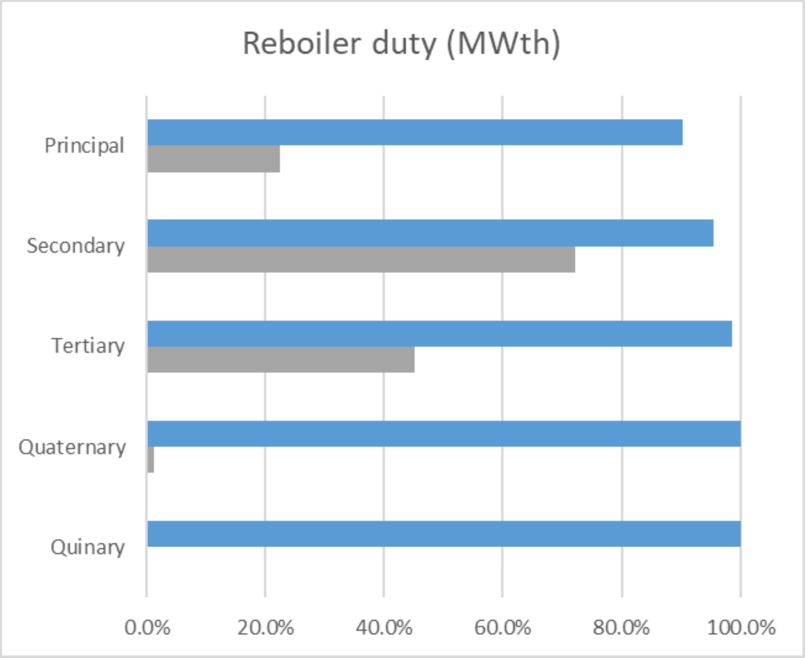


Combined inputs for optimal relevance

Combined input directions are found by maximizing $S_{1,..,j}$ for j=1 to j=5 successively. The calculation is onerous, but effective.

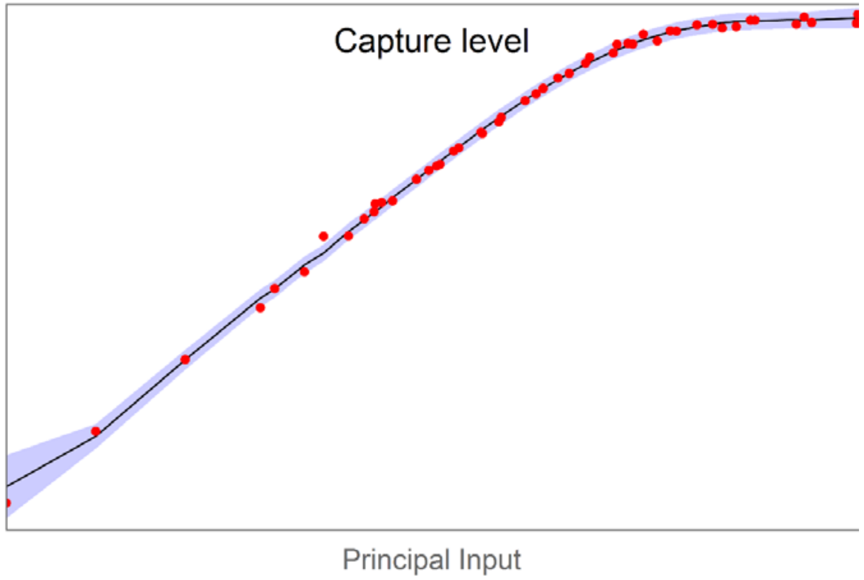




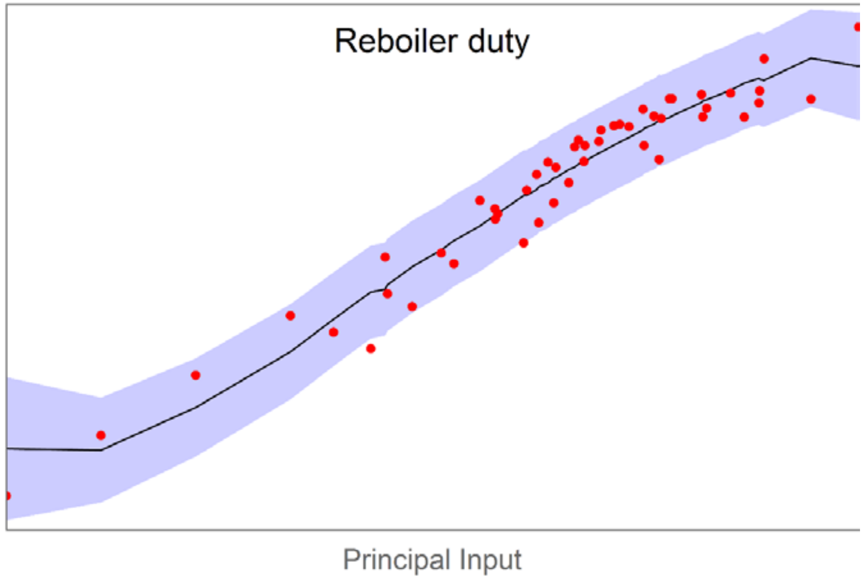


Combined inputs for optimal relevance

Rejecting all but the principal combined input, GP regression provides quite accurate predictions.



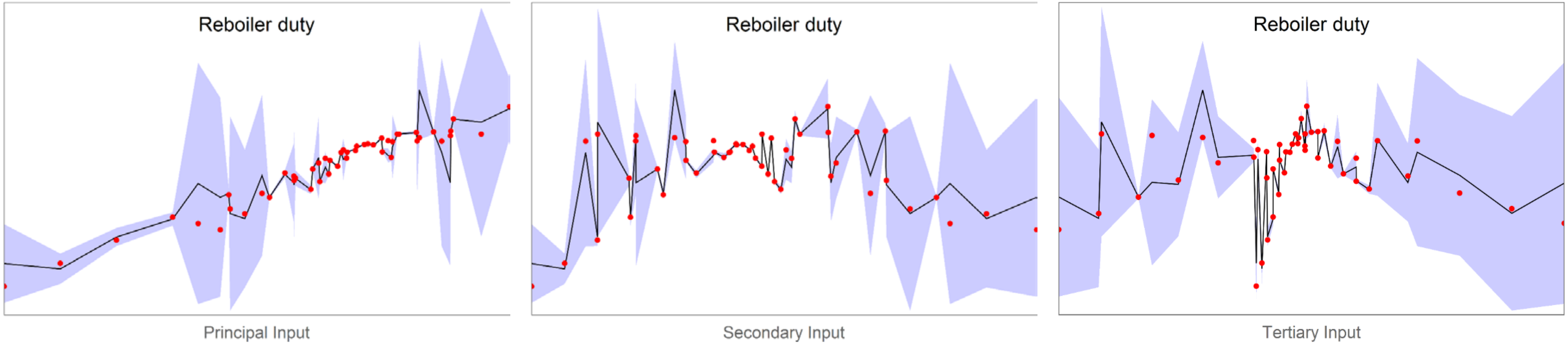
Principal Input ≈ 0.7 Solvent concentration – 0.4 Flue gas flowrate – 0.4 CO_2 in flue gas



Principal Input ≈ 0.8 Solvent flowrate + 0.4 Flue gas flowrate + 0.4 CO_2 in flue gas

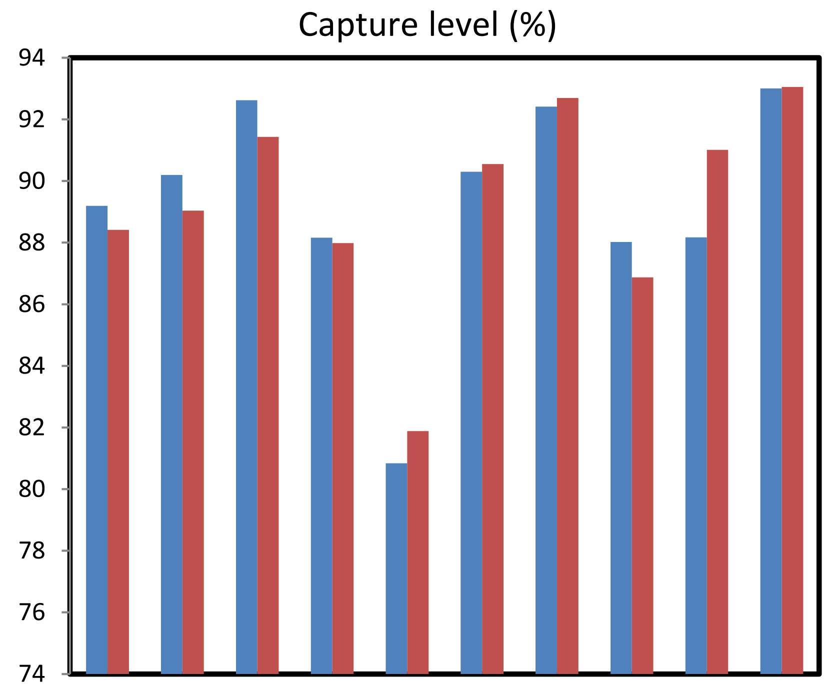
Combined inputs for optimal relevance

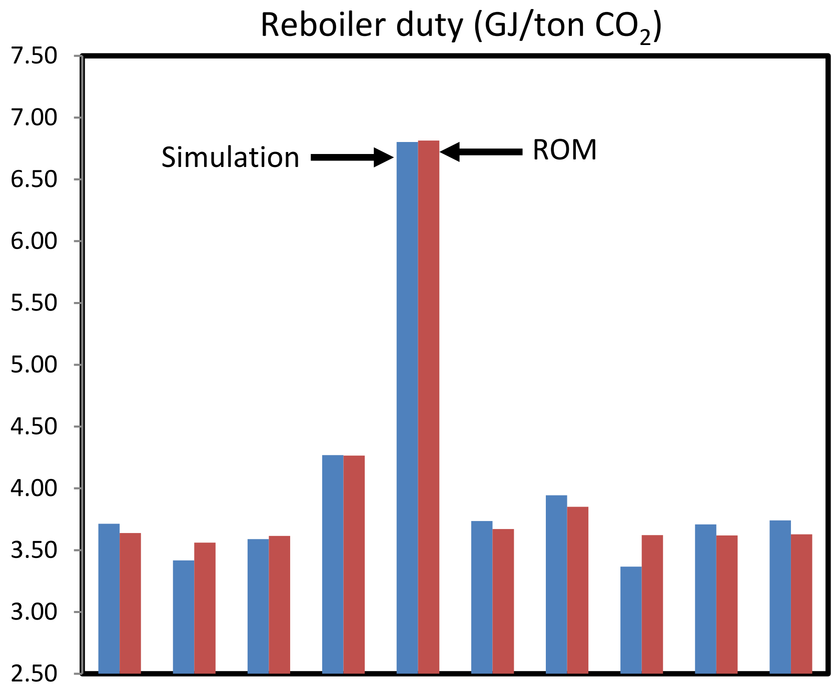
The curse of dimensionality returns when a GP is trained on the 3 most relevant combined inputs for Reboiler duty.



Ongoing data iteration (Dr Eni Oko)

As more simulation data becomes available the ROM is being tested and re-trained. Prediction accuracy remains impressive.





Conclusions

- GP surrogates enable robust and efficient analysis of complex data, with built-in uncertainty quantification.
- Training and analysing GP surrogates can be somewhat onerous. A trained GP for CO_2 capture can be provided as a simple Python function. Prediction using a trained GP is almost instant.
- Relevance determination and optimization using GPs discovers the most significant combinations of inputs. These can be provided as a small matrix.
- Python software developed for Reduced Order Modelling can be applied to many other problems besides CO_2 capture, and is freely available at https://github.com/C-O-M-M-A/rom-comma