Data-Driven Multi-Objective Optimisation of Coal-Fired Boiler Combustion Systems

Alma A. M. Rahat[a,*], Chunlin Wang[b], Richard M. Everson[a], Jonathan E. Fieldsend[a]

[a] *Department of Computer Science, University of Exeter, UK.*
[b] *Hangzhou Dianzi University, China.*

Abstract

Coal remains an important energy source. Nonetheless, pollutant emissions – in particular Oxides of Nitrogen (NOx) – as a result of the combustion process in a boiler, are subject to strict legislation due to their damaging effects on the environment. Optimising combustion parameters to achieve a lower NOx emission often results in combustion inefficiency measured with the proportion of unburned coal content (UBC). Consequently there is a range of solutions that trade-off efficiency for emissions. Generally, an analytical model for NOx emission or UBC is unavailable, and therefore data-driven models are used to optimise this multi-objective problem. We introduce the use of Gaussian process models to capture the uncertainties in NOx and UBC predictions arising from measurement error and data scarcity. A novel evolutionary multi-objective search algorithm is used to discover the probabilistic trade-off front between NOx and UBC, and we describe a new procedure for selecting parameters yielding the desired performance. We discuss the variation of operating parameters along the trade-off front. We give a novel algorithm for discovering the optimal trade-off for all load demands simultaneously. The methods are demonstrated on data collected from a boiler in Jianbi power plant, China, and we show that a wide range of solutions trading-off NOx and efficiency may be efficiently located.

Keywords: evolutionary multi-objective optimization under uncertainty, coal combustion optimisation, NOx, unburned carbon content in fly ash, Gaussian processes, probabilistic dominance.

Highlights

- Quantification of predictive uncertainty in data-driven models of power plants.
- Multi-objective optimisation under uncertainty to find optimal NOx and UBC trade-off.
- Novel solution selection method based on maximum probability of dominance.

*Corresponding author
Email address: A.A.M.Rahat@exeter.ac.uk (Alma A. M. Rahat)
URL: emps.exeter.ac.uk/computer-science/staff/aamr201 (Alma A. M. Rahat)

- Optimisation for all load demands in a single run: no need for repeated optimisation.
- Methods are demonstrated on the data collected from Jianbi power plant, China.

1. Introduction

Coal is the most abundant fossil fuel on earth [1]. It is therefore heavily used in producing electricity and heat. Nonetheless, coal combustion in power plants is accompanied by the production of pollutants, primarily *Oxides of Nitrogen* (NOx) and Oxides of Sulphur (SOx) [2]. Reducing the emission of such pollutants, in particular NOx, during combustion is a challenging task, and has been the goal of much research in the last few decades [3]. This is due to the harmful effects of the pollutants on the environment and the consequent strict legislation enacted by governments around the globe [4, 5]. These regulations are generally based on the air quality guidelines published by the World Health Organisation (WHO) [6].

To reduce the emission of NOx many de-NOx technologies have been employed, e.g. selective catalytic reduction (SCR), air staging, re-burning etc.; interested readers should refer to [7] for a survey of de-NOx technologies. Despite the success of some of these methods in removing produced NOx (upto 90% with SCR [7]), it remains important to reduce the overall NOx production during the combustion process [8]. Many combustion parameters, such as air flow rate and distribution, oxygen (O_2) percentage, over fire air (OFA), load, etc., may be tuned to achieve lower rates of NOx formation during combustion. However, reducing NOx formation generally requires lower temperatures and lower O_2 levels in the combustion chamber. Consequently, the proportion of the *unburned coal content* (UBC) — a good indicator of boiler inefficiency — increases as NOx production is reduced [9]. Clearly, there is a trade-off between reducing NOx production and increasing efficiency. In this paper, we focus on estimating the optimal trade-off between these conflicting objectives in order to locate a suitable set of operating (input) parameters for optimal combustion in the boiler.

In order to computationally optimise a combustion system, a means of evaluating the boiler's performance, e.g. NOx emission or UBC, for any given operating condition is required so that different sets of operating parameters may be queried by the optimiser to locate good solutions. Unfortunately, coal combustion is a complex process and there are no detailed analytical models available that may be used to predict NOx and UBC performance for specific boilers. In these circumstances we may resort to physical experiments, computational fluid dynamics (CFD) models [10], and data-driven models [11] to relate inputs parameters to performance metrics. Physical experiments involve running the plant with a fixed set of operating parameters and measuring the performance metrics. It is, however, time consuming and financially expensive to configure and run the system for each new condition to be considered. To the best of our knowledge, there has been no attempt to optimise combustion parameters using physical experiments in the literature. This is primarily because of the associated risks and costs of evaluating a solution on a live

system. On the other hand, CFD simulations computationally capture the complex interactions of different fluid flows within the combustion chamber and thus may be used to predict performances; see, for instance, [12] for NOx prediction and [13] for UBC prediction using CFD. It is also time consuming as high quality simulations may take hours or days to converge to a stable solution. Consequently, performing optimisation using CFD may be unreasonable with standard evolutionary methods. For instance, Dal *et al.* minimised the sum of emissions and inefficiency of an industrial 600 MW tangentially-fired pulverized-coal boiler with a genetic algorithm (GA) using CFD to evaluate each tentative solution [14]. The CFD simulation for a solution required 14 hours on an eight-core machine, and the GA required thousands of evaluations to locate a good approximation of the optimal solution. To combat the computational burden, they used a 52-node high-performance computing cluster to perform optimisation of the boiler. It still required several weeks to complete the optimisation process. Clearly, this is impractical for general use at a typical power plant due to time and computational requirements.

We therefore adopt a data-driven modelling approach: we collected data mapping between operating (input) parameters and response (output) performance metrics (NOx and UBC) from physical experiments, and trained independent nonlinear regression models to predict each output from the parameters in order to use these in optimisation. The data-driven models are many orders of magnitude faster to evaluate than either physical experiment or a CFD model, making optimisation feasible.

In recent years a few studies have been conducted on data-driven multi-objective optimisation of coal-fired boilers with a view to reducing NOx emission and increasing efficiency simultaneously. One of the earliest attempt at solving this multi-objective problem is due to Zhou *et al.* [8]. They constructed a multi-layer perceptron (MLP) model for predicting NOx and UBC with 29 input parameters. However, they only used 12 data points for training, which is clearly insufficient to model the relationship accurately. Furthermore, rather than using a multi-objective optimiser, they used a standard GA to reduce UBC below a fixed threshold on NOx emissions: essentially solving a single objective problem. This resulted in a single solution and thus failed to extract the complete trade-off between the objectives. In [15], separate models relating 19 input parameters to predict UBC and NOx were built using support vector regression (SVR) based on 119 data points collected during normal operation of a 300 MW pulverised coal-fired boiler. These models were then used with a multi-objective cellular genetic algorithm (MOCell) to estimate the optimal trade-off front (also known as the Pareto front; see section 4) between UBC and NOx for a particular load. The results achieved superior performance in comparison to hand-tuned configurations from the operators. A further study showed that MOCell also outperformed other contemporary multi-objective evolutionary optimisations methods – Strength Pareto Evolutionary Algorithm (SPEA2) [16], Multi-Objective Particle Swarm Optimiser (OMOPSO) [17] and Archive-Based hYbrid Scatter Search (AbYSS) [18] – in terms of quality of solutions and convergence rates [19]. In [20], experimental data was collected for a 160 MW pulverised coal-fired boiler equipped with a SCR system. The objectives were to reduce NOx and heat rate

(HR), where decreasing HR represents increasing the thermal efficiency, and there were constraints on the O_2 input, UBC and other input parameters. Two independent online support vector regression (OSVR) models were used to predict the NOx and HR objectives from 80 five-dimensional input parameters. The Non-dominated Sorting Genetic Algorithm II (NSGA-II [21]) was used to estimate the optimal Pareto front for the system under full load. Similarly, in [22], a two-output multi-layer neural network (MLNN) model was trained to predict NOx and overall heat loss (as a proxy for overall efficiency) that was trained on a dataset consisting of 20 data points and 17 input parameters. In this case, NSGA-II was used to simultaneously minimise NOx emission and overall heat loss. A recent and comprehensive review on the topic can be found in [23].

A common feature of these previous works is that a point prediction method, such as support vector regression or a neural network, was used to generate predictions of outputs from hypothetical parameter settings (many such modelling approaches in combustion optimisation are presented in [24]). However, these predictions are necessarily inaccurate because of observational errors in the data points themselves and the paucity of training data on which the models are learned. Consequently, straightforward optimisation of these learned models may yield misleading and sub-optimal results. In addition, previous studies have conducted optimisation for a particular, fixed demand loading, whereas in operational situations the demand load fluctuates and different settings are required to achieve optimal performance for a range of loadings.

In this paper, addressing these weaknesses in the literature, we describe a multi-objective evolutionary approach that incorporates Gaussian process ($\mathcal{GP}$) regression models [25] to approximate the optimal trade-off between NOx emission and UBC in fly ash. In outline, our approach is to use Gaussian process models to learn the relationship between plant operating parameters and the performance metrics (UBC and NOx). The great advantage of Gaussian processes is that they quantify the uncertainty in UBC or NOx predictions as predictive posterior distributions. To take advantage of these distributions we adapt a multi-objective evolutionary optimisation algorithm to use probabilistic dominance—the probability that one solution has better NOx and better UBC content that another—to find the probabilistic optimal trade-off. The plant operator must select a single set of parameters and we again use the notion of probabilistic dominance to find the solution that best satisfies the desiderata. The evolutionary optimisation algorithm is first described for the straightforward case of fixed load demand, after which we extend it to simultaneously deal with all loads.

The key contributions of this paper are summarised as:

- For the first time, we show how to quantify uncertainties arising from data scarcity and measurement error in predictive models of power plant emissions. Specifically, we construct Gaussian process models of the NOx and UBC in fly ash from online production data from a coal-fired boiler in the Jianbi power plant, Jiangsu province, China.

- We show how to find the optimal trade-off between UBC and NOx production while accounting for uncertainties in the data-driven model. This is achieved through the use of probabilistic dominance in a multi-objective evolutionary algorithm.
- We give a new method for the selection of boiler operating parameters from the probabilistic trade-off located by the multi-objective algorithm.
- We introduce a novel algorithm to estimate optimal trade-off fronts at all load demands (rather than at a single fixed load).

Although we illustrate these methods on data from the Jianbi power plant, they are generally applicable to engineering problems requiring the optimisation of more than one objective for which no explicit computational model is available.

The rest of the paper is structured as follows. In section 2 we briefly describe the boiler our data relates to, although the methods we present are applicable to a wide range of boilers and other plants. We describe the Gaussian process modelling approach and show relevant results in section 3. The multi-objective evolutionary algorithm and the probabilistic dominance measure are presented in section 4. These are brought together in section 5, which illustrates the evolutionary algorithm and shows how to select an operating point for a probabilistic trade-off front. In section 6 we extend the evolutionary algorithm to locate trade-off fronts for all load demands in a single run. Finally, we draw conclusions in section 7.

2. Description of the Boiler and Data Collection

The Jianbi power plant is typical of many coal-fired electricity generating facilities in China. It has a capacity of more than 2400 MW with 12 operating boilers. The focus of this paper is on boiler number 7: a 330 MW tangentially fired, dry-bottom boiler, which utilises the air-staged combustion technique. The boiler has eight burner assemblies. As illustrated in Figure 1, each assembly has four levels of primary pulverised coal inlets (denoted AA-DD) interleaved with five levels of secondary air inlets (denoted A-E) and an over-fire air (OFA) inlet. These direct air-coal mixture streams towards the circumference of two circles of approximate radius 0.5 m to achieve good mixing in the burner centres (Figure 1, *right*).

The distributed control system (DCS) of the boiler is capable of reporting the state of various operational parameters, such as the coal feeder frequency of rotation, the flow speeds at various inlets, oxygen concentration, etc. To monitor NOx concentration at the boiler outlet and UBC in fly ash additional monitoring systems were deployed. In particular, an emissions monitoring system (Rosemount and Emerson process management) with two tubes collecting the gasses across the channels following the air heater was used to monitor the NOx concentration.

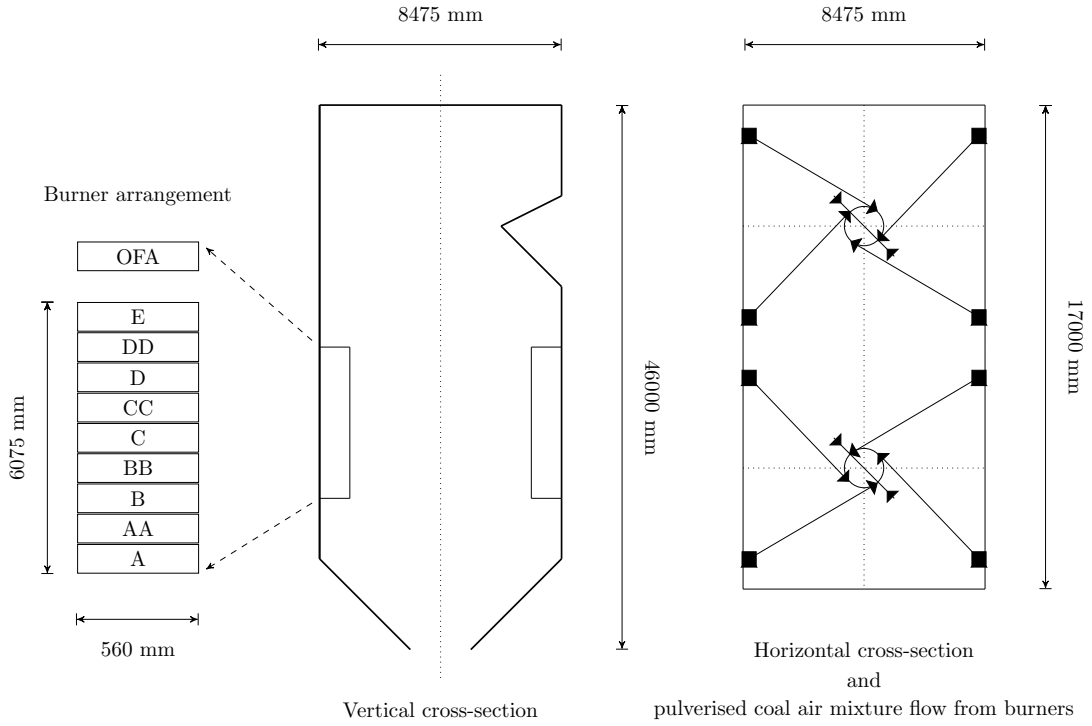


Figure 1: Schematic of the dry-bottom furnace. The arrangement of an individual burner assembly is shown on the *left*: there are four units of primary pulverised coal inlet (AA-DD), five units of secondary air inlet (A-E), and an over-fire air inlet (OFA). The furnace has eight burner assemblies (shown as solid black squares on the *right*). The arrows on the *right* show the approximate direction of the pulverised coal and air mixture entering the furnace from the burner assemblies to form a circular flow in the centre (known as a fireball [26]).

One of the primary goals is to construct a data-driven model that can predict an output performance (e.g UBC or NOx) for a given set of control parameters. We therefore collected 654 sets of training data by measuring the control and the output performance over a period of five days of normal operation. Complete lists of the control parameters and output performance metrics with the associated ranges from the data are presented in Tables 1 and 2. We did not include coal quality in training data as it remained constant over the period. The (bituminous) coal used during the data collection period consisted of 11.06 wt% of ash, 31.96 wt% of volatile gas and 13.60 wt% of moisture content. The heat rating of the coal was 23.49 MJ/kg.

It is well known that there is a strong linear correlation between coal feeder rotational frequency and the output load of the system. However, the collected data only include the coal feeder rotational frequency and the measured external load demand imposed on the system. Usually the operators adjust the feeder frequencies to control the output to match the load demand. The matching process is reliant on the operator's experience, and may be somewhat inaccurate. We therefore represent the relationship between measured load L and coal feeder rotational frequencies as $L = c\sum_k F_k$ where F_k is the k-th coal feeder rotational frequency and c is a constant. We used restricted linear regression to estimate the value of $c \approx 0.155$. The

Table 1: Control parameters.

Parameter description	Symbol	Unit	Lower limit	Upper limit
Primary pulverised coal inlets	AA	m/s	22.94	29.28
	BB		24.32	29.21
	CC		24.12	28.23
	DD		23.23	29.08
Secondary air inlets	A	m/s	24.69	35.83
	B		23.76	31.69
	C		23.44	36.85
	D		24.81	35.34
	E		19.68	39.03
Over fire air inlet	OFA	m/s	2.06	9.87
Oxygen concentration	O_2	%	2.60	4.61
Coal feeder rotational frequency	F_1	rpm	497.72	651.47
	F_2		237.20	505.01
	F_3		0.91	626.33
	F_4		300.90	635.81
Total air consumption	T_A	t/hour	7.91	10.61

Table 2: Output performance metrics.

Description	Symbol	Unit	Lower limit	Upper limit
Oxides of Nitrogen	NOx	ppm	268.84	407.22
Unburned Coal Content	UBC	%	1.43	2.42

standard deviation of the difference between the measured load and the load predicted by $L = c\sum_k F_k$ is $\sigma_L \approx 11$ MW. Using this relation enabled us to exclude the measured load demand from the training data, and use only the feeder rotational frequencies to approximate the output of the boiler system.

Since the UBC output data showed transient behaviour following changes of control parameters, we used straightforward exponential smoothing (e.g., [27]) with a smoothing factor of 0.1 to estimate UBC for model training.

3. Non-linear Regression with Gaussian Processes

In this section we briefly summarise non-linear regression with Gaussian processes ($\mathcal{GP}$s); see [25] for a comprehensive review. The essential benefit for this work conferred by $\mathcal{GP}$s is that they provide not only a

point prediction, but also a posterior probability density indicating how confident the prediction is.

In a typical regression task, we are given a set of N observations $\mathcal{D} = \{(\mathbf{x}_n, f_n) \,|\, n = 1, \ldots, N\}$, where $\mathbf{x}_n \in \mathbb{R}^M$ is the nth input vector with M attributes or features for which f_n is the observed output, which may be corrupted with noise or measurement errors. Here $\mathbf{x}_n$ is the vector of $M = 16$ control parameters listed in Table 1 and f_n is either an UBC or NOx output measured from the boiler operating with control parameters $\mathbf{x}_n$. The regression task is to model the relationship between input features and output, so that predictions about the output $f(\mathbf{x})$ corresponding to a new input $\mathbf{x}$ may be made.

A $\mathcal{GP}$ is defined as a collection of random variables such that any finite number of these have a joint Gaussian distribution [25]. Given some training data $\mathcal{D}$ and an input feature vector $\mathbf{x}$, the $\mathcal{GP}$ predictive probability density of the output $f(\mathbf{x})$ is a Gaussian density:

$$p(f \,|\, \mathbf{x}, \mathcal{D}) = \mathcal{N}(f \,|\, \mu(\mathbf{x}), \sigma^2(\mathbf{x})), \tag{1}$$

where the mean and variance of the prediction are given by

$$\mu(\mathbf{x}) = \mathbf{f}^\top K^{-1} \boldsymbol{\kappa}, \tag{2}$$

$$\sigma^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - \boldsymbol{\kappa}^\top K^{-1} \boldsymbol{\kappa} \tag{3}$$

Here $\mathbf{f}$ is the vector of measured responses, $\mathbf{f} = (f_1, f_2, \ldots, f_N)^\top$. Nonlinearity in the $\mathcal{GP}$ enters through a kernel function $k(\mathbf{x}, \mathbf{x}')$ which models the covariance between two feature vectors. The $N \times N$ covariance matrix K collects these covariances together, $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and $\boldsymbol{\kappa} = \boldsymbol{\kappa}(\mathbf{x})$ is the N-dimensional vector of covariances between the training data and $\mathbf{x}$: $\kappa_n = k(\mathbf{x}, \mathbf{x}_n)$

There is some flexibility of the choice of the covariance function [25]. We examined a range of common covariance functions (squared exponential, Matern52, and multi-layer perceptron) and selected the multi-leayer perceptron (MLP) kernel with automatic relevance determination based on a 10-fold cross-validation of the mean squared error in the prediction. In addition the kernel function depends upon a number of hyperparameters, $\boldsymbol{\theta}$. Training the $\mathcal{GP}$ comprises inferring these hyperparameters by maximising the marginal likelihood of the data $p(\mathcal{D} \,|\, \boldsymbol{\theta})$ given by:

$$\log p(\mathcal{D} \,|\, \boldsymbol{\theta}) = -\frac{1}{2} \log |K| - \frac{1}{2} \mathbf{f}^\top K^{-1} \mathbf{f} - \frac{N}{2} \log(2\pi). \tag{4}$$

Although the log marginal likelihood function landscape may be non-convex and multi-modal, we adopt the standard practice of using a gradient-based optimiser (BFGS) with several random starts to estimate good hyper-parameter values [28].

A further advantage of the Gaussian process model is that uncertainty in the value of the training outputs, perhaps due to process fluctuations or measurement errors, be readily accounted for. If the uncertainties in the output measurement of f_n is quantified by σ_n, then the predictive probability density is given by equations (1)-(3) with $K := K + \mathrm{diag}(\sigma_1^2, \ldots \sigma_N^2)$.

Optimisation requires predictions of both UBC and NOx to be made. While it would be possible to use a multivariate Gaussian process to construct a single model for both UBC and NOx, taking explicit account of the correlations between them [29], we instead model each independently because this reduces the number of additional hyper-parameters that must be learned, which was shown to produce better generalisation performance [30]. We denote the two independent univariate $\mathcal{GP}$s by $\mathcal{G}_{\mathrm{UBC}}(\mathbf{x})$ and $\mathcal{G}_{\mathrm{NOx}}(\mathbf{x})$.

3.1. Model Quality

In this paper, we used all the collected data points to train the models for optimisation. Nonetheless, to verify the efficacy of the $\mathcal{GP}$, we show the results from a 10-fold cross-validation in Figure 2. In this figure, the predicted output for each data point is plotted versus the measured output. The 10 fold cross validation means that predictions for each data point were made using a $\mathcal{GP}$ trained on 90% of the data, not including the predicted point[1]. Error bars indicate one standard deviation of the posterior predictive density $p(f \,|\, \mathbf{x}, \mathcal{D})$.

As shown by the proximity of the mean predictions to the actual = predicted diagonal in the figures, the models predicted the outputs well: the measurements are within two standard deviations of the mean predictions. The average root mean squared errors (RMSE) between the mean predictions and the measurements are 0.07% for $\mathcal{G}_{\mathrm{UBC}}$ and 6.67ppm for $\mathcal{G}_{\mathrm{NOx}}$, which is well within the estimated measurement noise variance of 0.2% and 17ppm respectively. However, we draw attention to the fact that the uncertainty in the predictions is a considerable fraction of the range of variation of either UBC or NOx. Consequently, optimising these performance measures without accounting for the uncertainty may lead to unrealistically optimistic and wrong settings of the boiler parameters. We therefore turn our attention to optimisation taking account of this uncertainty.

4. Multi-Objective Optimisation with Uncertainty

Our overall goal is to locate operating parameters that minimise UBC and NOx emission simultaneously and thus approximate the optimal trade-off between these objectives. We denote the UBC and NOx produced by the plant with control parameters $\mathbf{x}$ by $f_U(\mathbf{x})$ and $f_N(\mathbf{x})$. Then this two-objective problem may be expressed as:

$$\begin{aligned} &\underset{\mathbf{x}\in\mathcal{X}}{\text{Minimise}} \quad f_U(\mathbf{x}), \\ &\underset{\mathbf{x}\in\mathcal{X}}{\text{Minimise}} \quad f_N(\mathbf{x}), \end{aligned} \tag{5}$$

[1] In each fold, the Gaussian process model hyperparameters were found by maximising the log-marginal likelihood (4) using a BFGS quasi-Newton algorithm from 10 random restarts. Based on estimated measurement errors, the $\mathcal{GP}$s were trained assuming that there were measurement errors with standard deviations of 0.2% and 17ppm for UBC and NOx respectively.

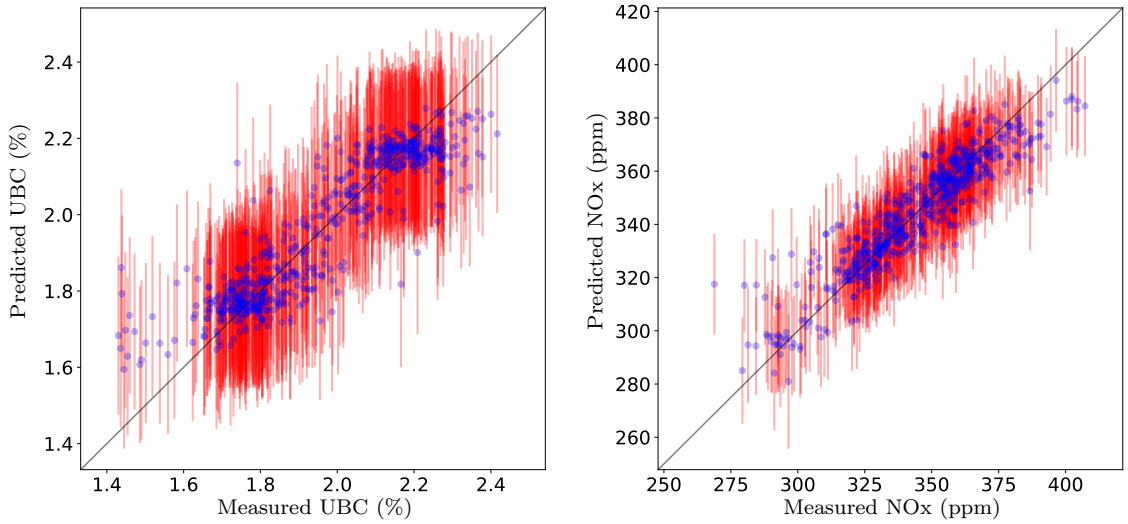


Figure 2: Comparison between predictions and measurements for 10-fold cross-validation of $\mathcal{GP}$s trained for UBC ($\mathcal{G}_{\mathrm{UBC}}(\mathbf{x})$; *left*) and NOx ($\mathcal{G}_{\mathrm{NOx}}(\mathbf{x})$; *right*). Blue dots indicate the mean prediction of the $\mathcal{GP}$, and the red error bars depict uncertainty (one standard deviation) in the prediction. Most of the measurements lie within two standard deviation of the mean prediction, verifying the accuracy of the constructed models.

where the control parameters are constrained to lie within the normal operating ranges $\mathcal{X}$ defined in Table 1.

As observed above, reducing NOx generally results in an increase in UBC and *vice versa*. The goal of solving a multi-objective problem with conflicting objectives like this is to obtain a set of solutions displaying the optimal trade-off between objectives in the sense that if one is reduced then the other is necessarily increased. The operator may then choose an operating point from this set of optimal trade-offs.

The trade-off between solutions is formally characterised by the notion of *dominance* (see, e.g., [31, 32]). In a multi-objective problem for which the objectives can be found without any uncertainty, a solution $\mathbf{x}$ is said to *dominate* another solution $\mathbf{x}'$, if it is not worse than $\mathbf{x}'$ on all objectives and better on at least one. For the two-objectives defined in (5), $\mathbf{x}$ dominates $\mathbf{x}'$, written as $\mathbf{x} \prec \mathbf{x}'$, iff:

$$\begin{aligned} f_N(\mathbf{x}) \leq f_N(\mathbf{x}') \text{ and } f_U(\mathbf{x}) < f_U(\mathbf{x}') \\ \text{or } f_N(\mathbf{x}) < f_N(\mathbf{x}') \text{ and } f_U(\mathbf{x}) \leq f_U(\mathbf{x}'). \end{aligned} \tag{6}$$

Multi-objective optimisation algorithms estimate the set of solutions which are not dominated by any solution in the feasible space: $\mathcal{P} = \{\mathbf{x} \,|\, \mathbf{x}' \not\prec \mathbf{x} \,\forall \mathbf{x}' \in \mathcal{X}\}$. This set is often referred to as the *Pareto optimal set*, while the image of this set in objective space is called the *Pareto front*.

Since accurate deterministic models of $f_N(\mathbf{x})$ and $f_U(\mathbf{x})$ are not available, we use the Gaussian process models, which provide predictive distributions $p(f_N \,|\, \mathbf{x}, \mathcal{D})$ and $p(f_U \,|\, \mathbf{x}, \mathcal{D})$; see equations (1), (2) and (3). The uncertainty inherent in the modelled performance for any set of control parameters means that deter-

ministic dominance cannot be used directly. We therefore instead deal with the probability that one solution dominates another. Modelling the two objectives as independent, the probability of one solution dominating another may be calculated as a product of dominance probabilities due to each model [33]:

$$p(\mathbf{x} \prec \mathbf{x}') = p(f_U(\mathbf{x}) < f_U(\mathbf{x}')) \times p(f_N(\mathbf{x}) < f_N(\mathbf{x}')). \tag{7}$$

Since the predictive distributions are Gaussian, the constituent probabilities may be calculated exactly as [33]:

$$p(f(\mathbf{x}) < f(\mathbf{x}')) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{m}{\sqrt{2}}\right)\right], \tag{8}$$

with

$$m = \frac{\mu(\mathbf{x}') - \mu(\mathbf{x})}{\sqrt{\sigma^2(\mathbf{x}) + \sigma^2(\mathbf{x}')}}, \tag{9}$$

where the mean and variance parameters are given by (2) and (3) for each of the trained UBC and NOx models. Clearly, if $\mathbf{x} = \mathbf{x}'$, then $m = 0$, so that $p(f_s(\mathbf{x}) < f_s(\mathbf{x}')) = \frac{1}{2}$ for $s \in \{U, N\}$ and $p(\mathbf{x} \prec \mathbf{x}') = \frac{1}{4}$ as expected from geometrical considerations.

The definition of dominance may then be extended to account for the probability of dominance: $\mathbf{x}$ is said to α-*dominate* $\mathbf{x}'$, denoted $\mathbf{x} \prec^{\alpha} \mathbf{x}'$, iff $p(\mathbf{x} \prec \mathbf{x}') \geq \alpha$ [34]. We then define the probabilistic Pareto set as:

$$\mathcal{P}^{\alpha} = \{\mathbf{x} \mid \nexists \mathbf{x}' \in \mathcal{X} \text{ such that } \mathbf{x}' \prec^{\alpha} \mathbf{x}\}. \tag{10}$$

That is, $\mathcal{P}^{\alpha}$ is the set of solutions none of which is dominated by a feasible solution with probability greater than α.

Elitist multi-objective evolutionary algorithms (EAs) for deterministic objectives often rely on the notion of dominance as a measure of quality of solutions (see for example [35]). Generally, an initial archive of mutually non-dominated solutions A, an approximation of the optimal Pareto set, is created. As the optimisation progresses, new solutions are stochastically generated through the evolutionary mechanisms of mutation and crossover. A new solution is only added to the archive if it is not dominated by any member in the archive, while any solutions in the archive that is dominated by the new solution are discarded. With uncertainty, elitist algorithms may be adapted by using probabilistic dominance and optimising a probabilistic archive of solutions. As in the deterministic case, candidate solutions are generated through mutation and crossover of existing solutions. However, a new solution is added to the probabilistic archive if no other solution dominates it with probability at least α. If any solution dominated by another with probability at least α, then that solution is discarded from the probabilistic archive. This means that when α is close to 1 solutions are rarely discarded, so when the predictive models are uncertain many solutions are added to the archive and few are deleted, resulting in a "fat" archive. On the other hand, with high certainty

the algorithm can be confident in determining when a solution is dominated, resulting in a relatively "thin" archive as the deterministic limit is approached. An outline of the algorithm is given in Algorithm 1, which we now discuss in more detail.

5. Load-Constrained Optimisation

Combustion plants are usually operated with a specified target load and we therefore describe an evolutionary algorithm to find the optimal trade-off between UBC and NOx production at a specified load L. The algorithm depends on the learned Gaussian process models to predict UBC and NOx for candidate boiler parameterisations which have not actually been observed. However, because the data were collected over long periods of normal operation at various loads, inevitably many of these parameterisations do not correspond to the target load.

As discussed in section 2, plant operators adjust the coal feeder rotational frequencies F_k to achieve a particular target load using the relationship

$$c \sum_k F_k = L \tag{11}$$

where we determined c by restricted linear regression. We exploited this relationship to map solutions that do not match the load constraint to the target load by replacing the vector (F_1, F_2, F_3, F_4) of rotational speeds with its orthogonal projection onto the plane defined by (11). As changing the feeder speeds may change the system outputs, the load-constrained solutions were then evaluated with the $\mathcal{GP}$ models to determine the appropriate objective values. This procedure for enforcing this equality constraint was used to generate the initial archive of solutions which comprises those load-mapped measured control parameters $\{\mathbf{x}_n\}_{n=1}^N$ which are not dominated at the α level. This is shown in steps 3 to 8 of Algorithm 1, where the function NonDominated$(A, \alpha, \mathcal{G}_{\mathrm{UBC}}, \mathcal{G}_{\mathrm{NOx}})$ uses the trained $\mathcal{GP}$ models to return the set of solutions in A which are not dominated by any other solution with probability α:

$$\text{NonDominated}(A, \alpha, \mathcal{G}_{\mathrm{UBC}}, \mathcal{G}_{\mathrm{NOx}}) = \{\mathbf{x} \mid \nexists \mathbf{x}' \in A \text{ such that } \mathbf{x}' \prec^{\alpha} \mathbf{x}\}. \tag{12}$$

The archive, which is the current approximation to the Pareto set, is unconstrained and thus there is no limit on the number of mutually non-dominated solutions that may be retained (see, for instance, [35]).

At each generation of the evolution, two unique parent solutions from the archive, $\mathbf{x}_1, \mathbf{x}_2 \in A$, are selected (in step 10) using Partitioned Quasi Random Selection (PQRS) [35]. PQRS randomly selects solutions from the current archive, but attempts to avoid any bias that may arise from spatial crowding in particular regions of the estimated Pareto front. The selected parents $\mathbf{x}_1$ and $\mathbf{x}_2$ are then subjected to evolutionary combination and mutation in the next stage.

Algorithm 1 Multi-objective combustion optimisation.

Inputs

$\mathcal{D}$: Measured data, $\mathcal{D} = \{(\mathbf{x}_n, f_{U,n}, f_{N,n})\}_{n=1}^{N}$

L : Target load demand

α : Confidence level for probabilistic dominance

T : Number of epochs

Steps

1: $\mathcal{G}_{\text{UBC}} \leftarrow \text{TrainModel}(\mathcal{D})$ ▷ Train UBC model

2: $\mathcal{G}_{\text{NOx}} \leftarrow \text{TrainModel}(\mathcal{D})$ ▷ Train NOx model

3: $A \leftarrow \emptyset$

4: **for** $n = 1 \rightarrow N$ **do**

5: $\quad \mathbf{x}'_n \leftarrow \text{ConstrainLoad}(\mathbf{x}_n, L)$

6: $\quad A \leftarrow A \cup \{\mathbf{x}'_n\}$

7: **end for**

8: $A \leftarrow \text{NonDominated}(A, \alpha, \mathcal{G}_{\text{UBC}}, \mathcal{G}_{\text{NOx}})$ ▷ Initialise archive

9: **for** $i = 1 \rightarrow T$ **do**

10: $\quad \{\mathbf{x}_1, \mathbf{x}_2\} \leftarrow \text{PQRS}(A)$ ▷ Select two parent solutions

11: $\quad \{\mathbf{c}_1, \mathbf{c}_2\} \leftarrow \text{Crossover}(\mathbf{x}_1, \mathbf{x}_2)$

12: $\quad$ **for** $k = 1, 2$ **do** ▷ For each child

13: $\quad\quad \mathbf{c}_k \leftarrow \text{Mutate}(\mathbf{c}_k)$

14: $\quad\quad \mathbf{c}_k \leftarrow \text{ConstrainLoad}(\mathbf{c}_k, L)$

15: $\quad\quad A \leftarrow \text{NonDominated}(A \cup \{\mathbf{c}_k\}, \alpha, \mathcal{G}_{\text{UBC}}, \mathcal{G}_{\text{NOx}})$ ▷ Retain if probabilistically non-dominated

16: $\quad$ **end for**

17: **end for**

18: **return** A ▷ Approximation of the Pareto set

The evolutionary mechanisms used in this paper are *crossover* and *mutation*. The crossover operator takes some parts of one parent and combines them with parts of the other parent to generate two complementary children; well constructed crossover operators construct children which are convex combinations of the parents [36]. This enables the algorithm to generate child solutions that are not in the vicinity of current members of the archive and thus promotes exploration. On the other hand, the mutation strategy randomly perturbs a single solution, usually with the intention of locating a better solution in the vicinity of the current solution.

In Algorithm 1, we always apply crossover between the selected parents $\mathbf{x}_1$ and $\mathbf{x}_2$ (step 11). A child $\mathbf{c}_1$ is constructed by selecting attributes either from $\mathbf{x}_1$ or from $\mathbf{x}_2$ with a probability of q and $1-q$ respectively.

The other child $\mathbf{c}_2$ is constructed as the complement of $\mathbf{c}_1$. Each of the children is mutated with probability 0.8[2]. If a child is mutated (step 13), each parameter forming the child is altered with probability 0.2. Parameters are perturbed by replacing them with a random draw from a Gaussian distribution centred on the old parameter value and with standard deviation equal to 10% of the range of the parameter (Table 1); draws outside the range are redrawn. Following crossover and mutation the mutated child is constrained to the required load (step 14). We then apply the repair mechanism on the mutated child $\mathbf{c}_k$ in step 14 to ensure that the mutant satisfies the load demand. The child solutions are then added to the archive A is the probability of them being dominated by an existing solution is less than α, and any solutions in the archive that are dominated with probability at least α are discarded (step 15). This process continues either for a predetermined number of iterations or until the archive is judged to have converged.

The performance of a multi-objective optimiser may be monitored and evaluated by the hypervolume measure [37], which is the volume/area of objective space that is dominated by solutions in the estimated Pareto front. If there is no uncertainty associated with the model, then the hypervolume dominated by the archive A with respect to a reference vector $\mathbf{r} = (r_U, r_N)$ in the objective space may be defined as:

$$H(A, \mathbf{r}) = \text{vol}_{\mathbf{z}}\{(f_U(\mathbf{x}), f_N(\mathbf{x})) \prec \mathbf{z} \prec \mathbf{r} \wedge \mathbf{x} \in A\}, \tag{13}$$

where $\mathbf{z} = (z_U, z_N)$ is an arbitrary vector in the objective space. In this paper, we consider the hypervolume dominated by the mean prediction $(\mu_U(\mathbf{x}), \mu_N(\mathbf{x}))$ as suggested in [38] to monitor progress of the optimiser.

Figure 3 shows the estimated Pareto front after $20,000$ generations and the hypervolume dominated by the mean predictions durning the optimisation, which used $\alpha = 0.5$. The figure shows the mean predicted performance of the $22,068$ solutions in the archive, together with error bars indicating the uncertainty (one standard deviation) in the model prediction for some of the solutions. For clarity error bars are not shown for all solutions, but the magnitude of the predictive uncertainty is comparable for all solutions (c.f., Figure 2).

As can be readily seen from the figure, the estimated probabilistic front is thicker than a traditional estimated Pareto front. This is because the mutually non-dominated solutions do not dominate each other with a probability of at least $\alpha = 0.5$, and hence are retained in the archive. In contrast, fronts where there is no model uncertainty are thinner as dominated solutions can be precisely determined and removed from the archive.

The progression of the dominated hypervolume indicates that the front is well converged after $20,000$ generations and there is little to gain from additional iterations. To check we ran the algorithm for 10^5 generations, but only achieved a small increase in dominated hypervolume as indicated by the h in the right panel of Figure 3. We point out that each generation requires predictions to be made for two children so the

[2]Mutation probabilities were chosen empirically after a few short runs of the optimiser.

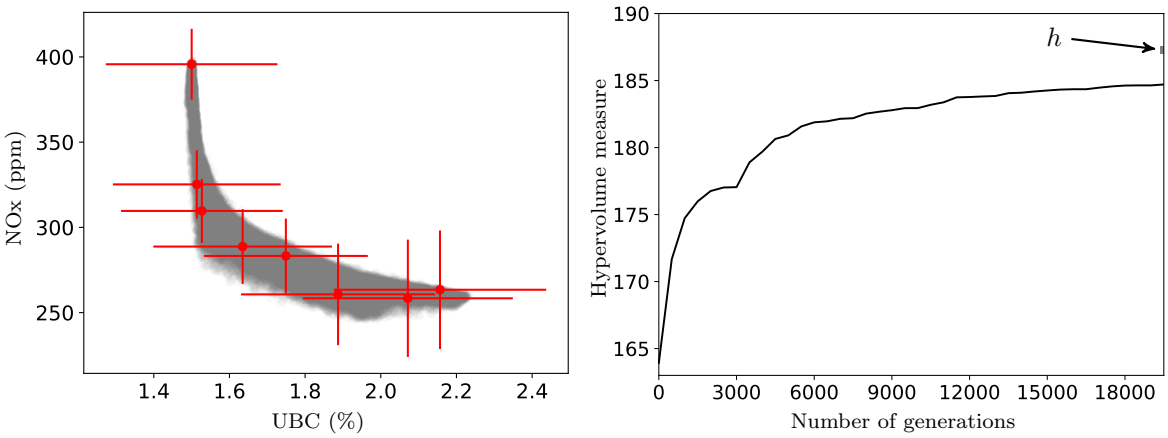


Figure 3: Estimated probabilistic Pareto front (*left*) and the hypervolume progression of the non-dominated mean predictions (*right*) after 20,000 generations for $L = 300$ MW. Grey dots on the *left* panel, depict the mean predictions from the solutions in the front. The red errorbars show the uncertainty (one standard deviation) in the mean prediction for ten selected members of the front. We chose to show the errorbars on only few for clarity. Other neighbouring solutions' predictions have similar level of uncertainty associated with them. The hypervolume (h) after 10^5 generations is depicted with a solid square (*right*).

number of $\mathcal{GP}$ function evaluations was 40,000. However, although the *training* of the $\mathcal{GP}$s, which has to be done only once, is relatively slow (but taking less than 10 s for each of these models) the evaluation for a new parameter vector is very fast, requiring only a dot product to compute the mean (2) and the solution of N linear equations to find the predictive variance (3).

As a result of the optimisation, a wide range of trade-off solutions are found. Considering only the mean predictions, the best solution in terms of UBC produces 1.49% unburned carbon content while producing 381.7ppm NOx. At the other extreme for UBC, the carbon content emission is 2.23%, a worse performance by 49% with an improvement of 47% in NOx production performance. Considering both the extremes in terms of NOx, the solutions show a change of 61% in NOx production while the associated UBC production changes by 29%. However, the large uncertainties associated with all the points in the Pareto archive makes the selection of an operating point rather challenging: it is not obvious how to discriminate between the mean predictions or how the uncertainties should be incorporated in order to make an informed decision. We address this issue in the next section.

5.1. Operating Point Selection

A straightforward way of selecting a solution at which to operate from the probabilistic fat front is, after the optimisation is complete, assume that the uncertainty around the mean prediction is zero and thus determine the set of non-dominated mean predictions. The operator will then typically select a solution close to the "knee" in the estimated front where the gradient changes rapidly. This is analogous to the procedure used above to monitor the hypervolume during optimisation. This method, however, fails to account for the

uncertainty in the predicted performance during the decision making process.

To take account of the prediction uncertainties we first characterise each point in the objective space by the maximum probability that it is dominated by a solution in the archive. If $\mathbf{z} = (z_U, z_N)$ is a location in the UBC-NOx objective space, then the probability a solution $\mathbf{x} \in A$ has a better performance than $\mathbf{z}$ is the probability that $\mathbf{z}$ is dominated[3] by $\mathbf{x}$, namely:

$$p(\mathbf{x} \prec \mathbf{z}) = p(f_U(\mathbf{x}) < z_U) \times p(f_N(\mathbf{x}) < z_N) \tag{14}$$

$$= \frac{1}{4}\left[1 + \operatorname{erf}\left(\frac{z_U - \mu_U(\mathbf{x})}{\sqrt{2}\sigma_U(\mathbf{x})}\right)\right]\left[1 + \operatorname{erf}\left(\frac{z_N - \mu_N(\mathbf{x})}{\sqrt{2}\sigma_N(\mathbf{x})}\right)\right]. \tag{15}$$

Figure 4a shows

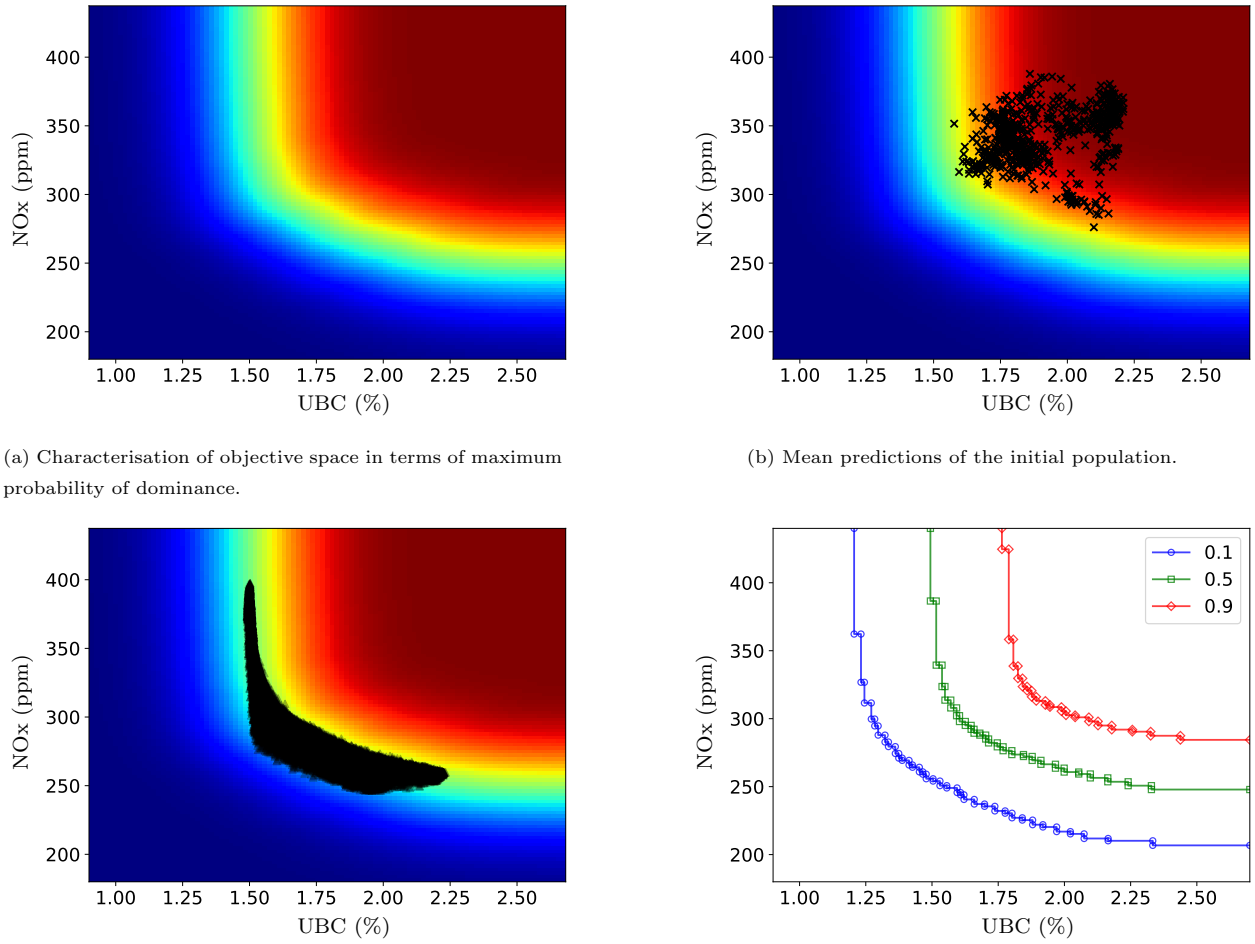
$$\rho(\mathbf{z}) = \max_{\mathbf{x} \in A} p(\mathbf{x} \prec \mathbf{z}) \tag{16}$$

for the archive shown in Figure 3. The spatial characterisation of the objective space using probability of dominance clearly depicts the performance that may be achieved using any solutions from the archive A. The plots in Figure 4 show the probability that any given performance can be achieved using a solution from the archive; clearly there is a low probability of obtaining very good performance on both objectives. To select a particular performance, a point on the objective space with desired performance level may be chosen, and then the solution that has the maximum probability of dominating the desired point should be selected as the operating point.

The maximum probability of dominance $\rho(\mathbf{z})$ plot also allows assessment of the algorithm performance. Figure 4b shows the location of the initial population of solutions, while the mean predictions from the optimised archive are shown in Figure 4c. As the figure shows, all the initial solutions are dominated by a member of the final archive with probability at least 50%. The repeatability of the algorithm may be assessed by displaying the summary attainment surfaces (introduced in [39] and a method for efficient computation is described in [40]). The summary attainment surfaces shown in Figure 4d show the median position of the $\rho(\mathbf{z}) = 0.1$, 0.5 and 0.9 contours, where the median (50th percentile) was taken over 10 independent runs of the optimiser. These contours clearly display the UBC-NOx trade-offs. Attainment surfaces corresponding to the 10th and 90th percentile attainment are very close to the median attainment; they are not plotted because they would be hidden by the symbols used for plotting the median attainment. The very small variation between 10th and 90th percentile curves indicates the desirable repeatability and convergence properties of the algorithm.

With the probabilistic archive on hand, the operator of the boiler is able to select an operating point for the boiler with a knowledge of the range of trade-off possibilities available. In making this decision

[3]Note that $\mathbf{x}$ is a vector of control parameters, while $\mathbf{z}$ is a performance, so the notation $\mathbf{x} \prec \mathbf{z}$ extends the usual dominance notation.



(a) Characterisation of objective space in terms of maximum probability of dominance.

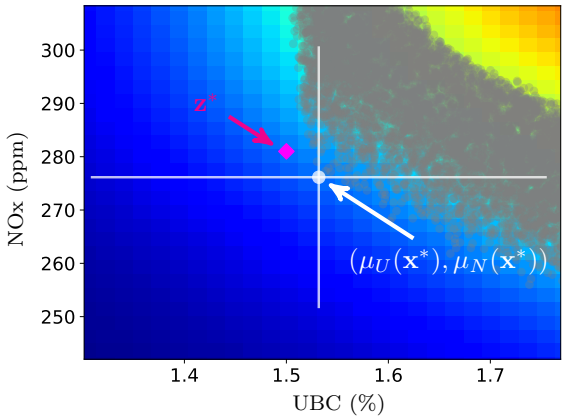
(b) Mean predictions of the initial population.

(c) Mean predictions of the solutions from the estimated Pareto front.

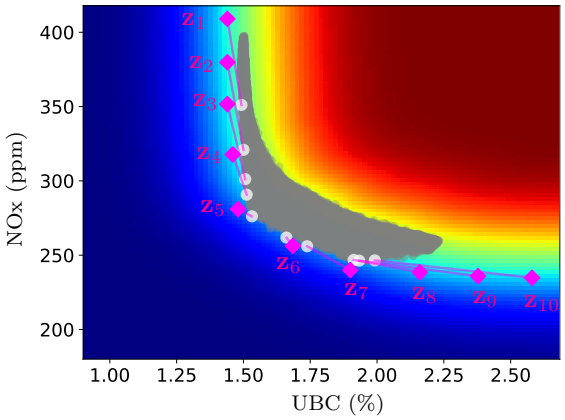
(d) Summary attainment surfaces.

Figure 4: Characterisation of the objective space by the maximum probability that a solution in the archive dominates it (a). Black crosses in (b) show the mean performance of the initial solutions, and black triangles in (c) depict the mean of the predicted performance for the solutions in the probabilistic front A. (d) Summary attainment fronts for maximum probability of dominance at 0.1 (blue circles), 0.5 (green squares) and 0.9 (red diamonds) computed with the estimated Pareto fronts from 10 independent simulation runs.

the uncertainty associated with solutions in the archive must be taken into account. In Figure 5 we show how a solution that best meets a desired performance may be located. Denote the desired performance by $\mathbf{z}^*$; for example in Figure 5a this performance is indicated by the magenta diamond. To best achieve this



(a) Selection of an operating point with desired performance $\mathbf{z}^*$ (magenta diamond): $\mathbf{x}^*$ is selected so that it has the maximum probability of dominating $\mathbf{z}^*$.



(b) Selection of ten solutions with performances as close to the given performances $\mathbf{z}_1, \ldots, \mathbf{z}_{10}$. The corresponding mean predictions are connected to the desired operating points with magenta lines.

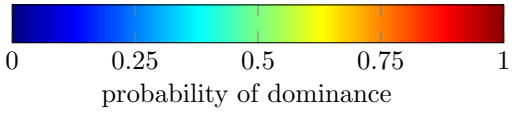


Figure 5: Selection an operating point based on probability of dominance in the objective space. The magenta diamonds $\mathbf{z}_i$ in (a) and (b) illustrate desired performances in the objective space. The grey dots portray the mean predictions from the solutions in the estimated probabilistic Pareto archive A, and the white error bars with crosses indicate the posterior predictive distributions of the selected solutions from A with its uncertainty (one standard deviation) that have the maximum probability of dominating the desired points.

performance we select the solution which has the greatest probability of dominating $\mathbf{z}^*$; that is

$$\mathbf{x}^* = \underset{\mathbf{x} \in A}{\operatorname{argmax}}\, p(\mathbf{x} \prec \mathbf{z}). \tag{17}$$

The performance of this solution is indicated by the one-standard deviation errorbars in Figure 5a.

This is further illustrated in Figure 5b for desired operating points $\mathbf{z}_1$ to $\mathbf{z}_{10}$ chosen to cover the ranges of available performances[4]. The length of the lines connecting the desired performance (magenta diamonds) with the mean performance of the solution with greatest probability of dominating the desired performance demonstrates that the solution in the archive "nearest" to a desired operating point is not necessarily the solution that dominates the operating point with greatest probability. This is particularly apparent near to the extremes of the front.

These results demonstrate the need to consider the uncertainties predicted performance during optimisation and in decision making.

[4]The $\mathbf{z}_i$ are located on equally spaced rays emanating from the top-right hand corner of the plot.

	UBC (%)	NOx (ppm)
$\mathbf{x}_1$	1.49	351.12
$\mathbf{x}_2$	1.50	320.92
$\mathbf{x}_3$	1.51	301.18
$\mathbf{x}_4$	1.51	290.74
$\mathbf{x}_5$	1.53	276.15
$\mathbf{x}_6$	1.66	261.83
$\mathbf{x}_7$	1.73	256.00
$\mathbf{x}_8$	1.91	246.80
$\mathbf{x}_9$	1.93	246.33
$\mathbf{x}_{10}$	1.99	246.49

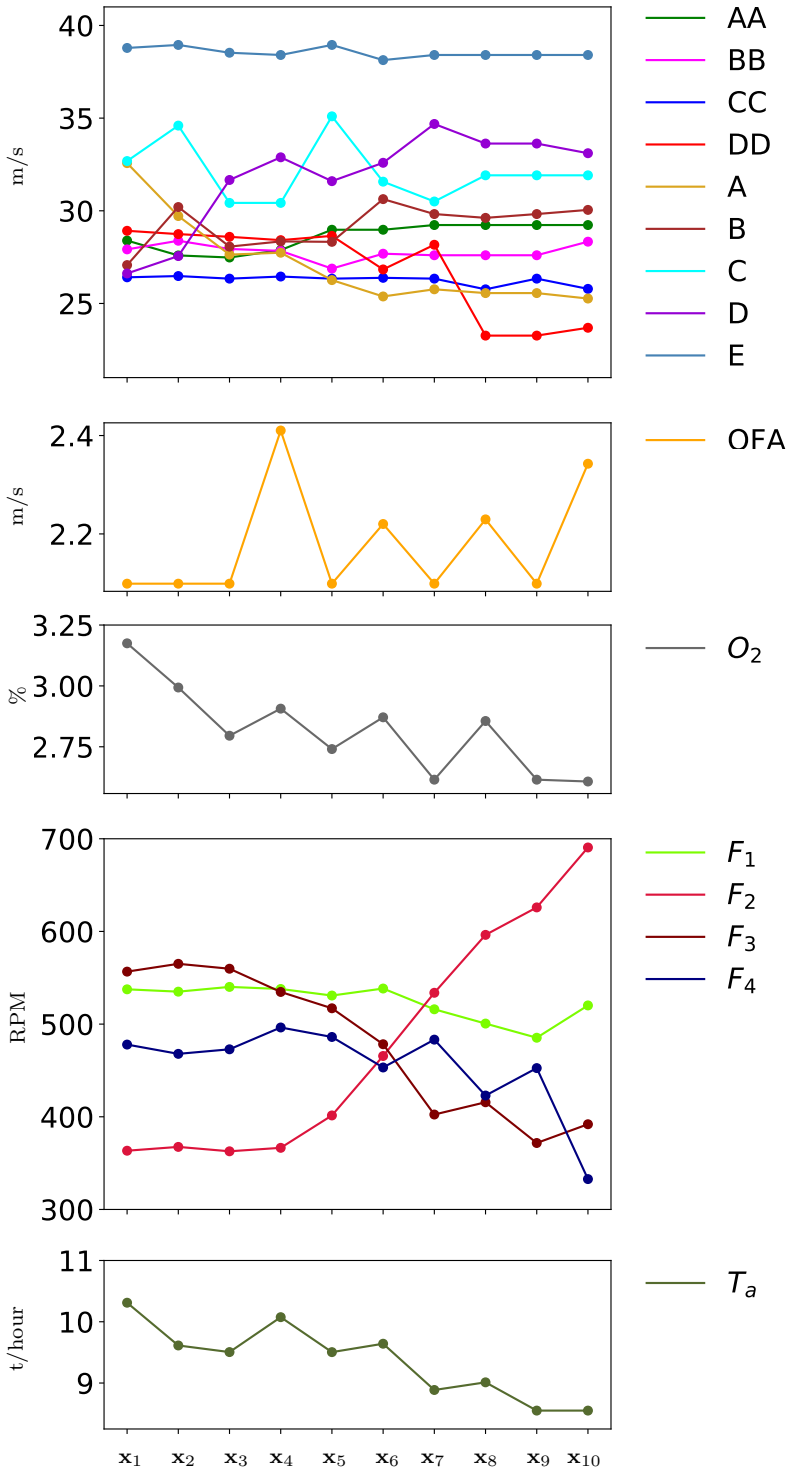


Figure 6: Variation in boiler operating parameter values for selected solutions $\mathbf{x}_1$ to $\mathbf{x}_{10}$ covering the range of available trade-offs shown in Figure 5. Generally, UBC increases from $\mathbf{x}_1$ to $\mathbf{x}_{10}$, while NOx is reduced, as shown in the table on the *left*, which gives the mean predicted UBC and NOx objectives, $\mu_U(\mathbf{x})$ and $\mu_N(\mathbf{x})$. The constraint on load demand was 300 MW. See Figure 1 and Table 1 for a schematic diagram of the boiler and the key to parameter names.

5.2. Variation of parameters across the Pareto front

In Figure 6, we investigate the variation in the boiler parameter values $\mathbf{x}_1$ to $\mathbf{x}_{10}$ the performances corresponding to the desired operating points $\mathbf{z}_1$ to $\mathbf{z}_{10}$ shown in Figure 5b that cover the range of trade-offs between UBC and NOx. As the front is traversed from $\mathbf{x}_1$ to $\mathbf{x}_{10}$, UBC increases and the NOx decreases, as shown in Figure 5. Corresponding to this change along the front, it may be expected that the total air (T_A) and Oxygen (O_2) should decrease, while OFA should increase. This is because lower T_A and O_2 result in a higher proportion of the coal remaining unburned and lower NOx formation in the primary combustion zone. In the burnout zone, above the primary combustion zone, over-fire air (OFA) is introduced to complete the combustion process; greater OFA should ensure greater use of unburned coal and drive NOx formation down. Indeed, T_A and O_2 show downward trends with increasing UBC and decreasing NOx. However, the change in OFA is not so obvious. We hypothesise that the change in air distribution through primary and secondary inlets, especially at the inlets near the top of the boiler (see Figure 1), sometimes compensates for the relative insensitivity to OFA. For instance, although secondary air flow at E remains stable, flow at D increases to support OFA, and therefore helps NOx reduction. We note that there is little variation in the primary air flows (AA-DD) across the front. The coal feeder speed F_2 increases gradually as the front is traversed, while the other feeder speeds F_1, F_3 and F_4 reduce accordingly to satisfy the constraint of 300 MW load.

It should be noted that the models are approximations of the actual physical process. Nonetheless, the changes in parameter values with respect to the change in UBC and NOx generally agree with operators' experience and expectations. Thus using these models is an effective way of determining a good operating parameter set.

6. Optimising for all load demands

We have presented an evolutionary algorithm that uses measured data, rather than a computational model, to optimise the boiler performance for a single load. A limitation of this approach is that changing load demands would require re-optimisation for each new load. Although it would be possible to build up a library of solutions for a range of loads, we here propose a new algorithm to generate trade-off solutions for all loads in a single optimisation run. In outline this algorithm is similar to augmenting the two-objective problem with an additional objective, namely the load demand. However, we do not seek to minimise the load, but to generate solutions (which are Pareto optimal in the UBC-NOx plane) across a wide range of load demands. We thus generate an archive of solutions which are probabilistically non-dominated with respect to UBC and NOx given that they correspond to similar loads. Modifications to the load-constrained algorithm to achieve this are as follows.

Exclusion of load constraint. To perform constrained optimisation, recall that we mapped the coal feeder speeds $\bar{F} = (F_1, F_2, F_3, F_4)$ to the load hyperplane defined by (11) by orthogonally projecting them onto that plane. Since all load demands are simultaneously considered, this constraint is no longer required and we omit steps 5 and 14 of Algorithm 1.

Comparing solutions for probabilistic dominance. Since the performance of a solution (in terms of UBC and NOx) is associated with a particular load, solutions with differing loads should not be compared directly for dominance between them. However, as discussed in section 2, the linear load model indicated that there is uncertainty in the load calculation of $\sigma_L \approx 11$ MW and we therefore regard solutions as having equivalent load if their loads are within σ_L. We compare solutions with a maximum difference in load of σ_L for probabilistic dominance. Hence, rather than the probabilistic Pareto set $\mathcal{P}^\alpha$ (10) we seek the set: $\mathcal{P}^\alpha_L = \{\mathbf{x} \mid \mathbf{x} \prec^\alpha \mathbf{x}' \wedge |L(\mathbf{x}) - L(\mathbf{x}')| \leq \sigma_L \wedge \mathbf{x} \neq \mathbf{x}'; \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}\}$, where $L(\mathbf{x})$ is the load associated with $\mathbf{x}$. The archive A in Algorithm 1 is thus constructed replacing the NonDominated function defined in (12) with

$$\text{CloseLoadNonDominated}(A, \alpha, \sigma_L, \mathcal{G}_{\text{UBC}}, \mathcal{G}_{\text{NOx}}) = \{\mathbf{x} \mid \nexists \mathbf{x}' \in A \text{ such that } \mathbf{x}' \prec^\alpha \mathbf{x} \wedge |L(\mathbf{x}) - L(\mathbf{x}')| < \sigma_L\}. \tag{18}$$

Selection strategy. For selecting $\mathbf{x}_1, \mathbf{x}_2 \in A$ (step 10), we extend the PQRS selection to incorporate the calculated load of the solutions. In PQRS, one of the objectives is chosen with uniform probability. Then the associated objective values across all the members of the archive are sorted into bins of equal width spanning the range of objective values for that objective. A solution is then selected by randomly choosing a bin and then a solution from that bin. The coarse-grained sorting into bins ensures that selection does not favour parts of the front where solutions are dense [35]. For unconstrained load optimisation we treat the load as if it were additional objective in PQRS, but biasing selection towards it by selecting on the basis of load 50% of the time, and on the basis of UBC and NOx for 25% each. This ensures that solutions selected for crossover and mutation come from a wide range of loads, enhancing exploration of the load space.

Objective and constraint space characterisation. The resulting archive spans three dimensions: UBC, NOx and load. As illustrated in Figure 7 this space is a combined objective and constraint space. Any slice at constant load through this space results in a probabilistic trade-off front between UBC and NOx at that load. In order to select a solution at which to operate with a desired load L, we first select all solutions close to L, $A' = \{\mathbf{x} | \mathbf{x} \in A \wedge |L(\mathbf{x}) - L| \leq \sigma_L\}$ and then select the solution in A' that dominates the desired operating point $\mathbf{z}$ with maximum probability: $\mathbf{x}^\star = \text{argmax}_{\mathbf{x} \in A'} p(\mathbf{x} \prec \mathbf{z})$.

With these modifications to Algorithm 1 and using the same parameters for evolution (i.e. crossover rate, mutation rate, etc.) as before, we ran the optimiser for 150×10^3 and 250×10^3 generations. Convergence of

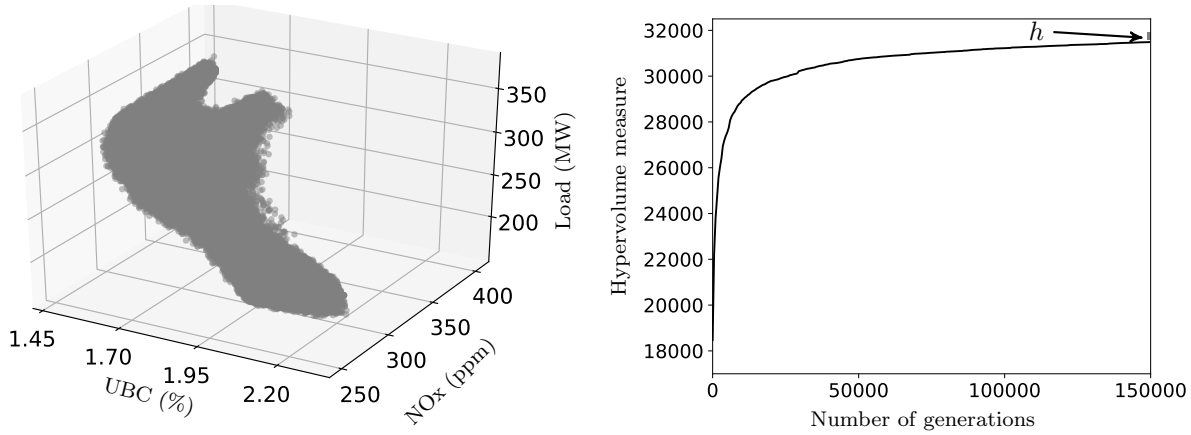


Figure 7: UBC, NOx, and load archive resulting from unconstrained load optimisation (*left*) and the hypervolume progression of the non-dominated mean predictions (*right*) after $150,000$ generations for all loads. Grey dots in the *left* panel depict the mean predictions from the solutions in the archive. The hypervolume (h) after $250,000$ generations is depicted with a grey solid square (*right*), showing that there is little to gain with 10^5 additional generations.

this algorithm may be assessed by the dominated hypervolume in each UBC-NOx "slice" summed over all slices. As shown in the right panel of Figure 7, the hypervolume progression suggests that the optimisation is well converged after 50×10^3 generations.

In Figure 7 (*left* panel) we present the mean predictions from the estimated Pareto front A. A wide range of solutions are located with UBC from 1.48% to 2.3% and NOx from 248.68 ppm to 404.63 ppm for load demands ranging between 160.7 MW and 374.8 MW. The figure also clearly shows that the UBC-NOx trade-off front varies considerably with load.

To better represent and understand the trade-off between solutions, Figure 8 presents the UBC-NOx-load space with isosurfaces drawn at various levels of probability of dominance. It is clear that for a fixed load $L = 300$ MW, we achieve a similar characterisation of the objective space (c.f. Figures 5 and 8). The variation in the dominated region can be seen by fixing UBC or NOx to a particular value as shown in the lower two panels of Figure 8. Interestingly, at any single fixed load larger areas are dominated when the load is close to 330 MW, indicating that the best performances in terms of both UBC and NOx may be obtained by operating with approximately this load. Increasing load from 330 MW results in a drastic deterioration in performance, while decreasing load also diminishes performance but at a lower rate. This concurs with the fact that the boiler is designed to be efficient at a 330 MW load, so poorer performance is to be expected away from this load.

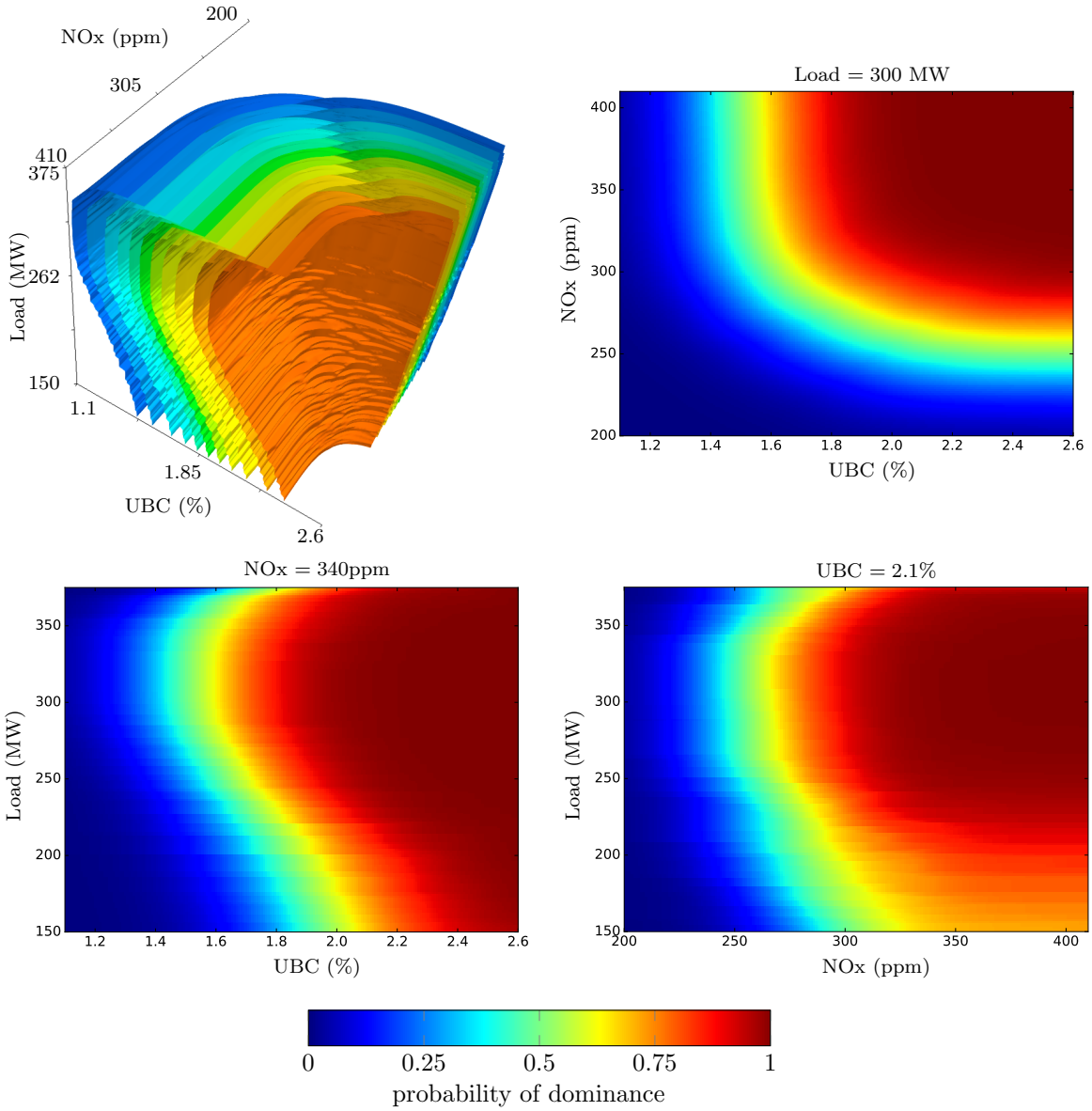


Figure 8: UBC-NOx-load performance for all loads in a single optimisation run, depicted with isosurfaces in 3-dimensions (*top-left*). Considering a slice at $L = 300$ MW (*top-right*), similar results are obtained in comparison to load-constrained optimisation, c.f. Figure 5. Slices with NOx = 340 ppm (*bottom-left*) and UBC = 2.1% (*bottom-right*). When load demand is between 250 and 350 MW larger areas of objective space are dominated, indicating that better operating points may be located.

7. Conclusion

Coal remains an important energy source for power generation, and it is likely to be around for many years to come. However, coal combustion generates pollutants, such as NOx. Optimising to reduce the emission of NOx, especially through the manipulation of input parameters during combustion, is known

to be detrimental to the coal burning efficiency [9]. Therefore, this may be considered as a multi-objective optimisation problem with two conflicting objectives: reducing NOx emission and increasing efficiency, which is equivalent to minimising unburned coal content in the fly ash.

There is no analytical model for the coal-fired boiler combustion system leading to data-driven approaches. In recent studies, these models have usually been built with a point prediction method, such as support vector regression, disregarding the uncertainties in the predictions of the model. Also, the observational errors and noise inherent in the measurements was ignored. In this paper, we addressed these issues by using Gaussian process regression models to model UBC and NOx. The benefit of using $\mathcal{GP}$s is that they provide a natural framework to account for measurement uncertainty and importantly they produce a complete posterior predictive distribution allowing confidence in predicted operation to be assessed. Compared to other studies, we also used a large number of data points from the Jianbi power station to train the model.

Generally, evolutionary algorithms are used to locate good trade-off solutions for this multi-objective optimisation problem assuming a fixed load demand. As traditional approaches utilise models without uncertainty in prediction, typical evolutionary algorithms may be used. Unlike these approaches, here we presented an evolutionary algorithm based on probabilistic dominance to estimate the optimal trade-off between UBC and NOx considering the uncertainties in the predictions. We showed that a wide range of solutions may be discovered in the form of probabilistic front using our approach for a particular load demand. We showed how an operating point may be selected using the maximum probability of dominance by any solutions from the probabilistic front for a desired performance level in the objective space. We also observed sensible changes in input parameter values as the UBC and NOx changed between selected solutions, indicating the effectiveness of a data-driven optimisation approach.

With the change in load demand, the system should be re-optimised to locate good operating points as previously located solutions may not perform as well under the new load. To find good operating parameters at all loads we extended our algorithm to estimate optimal solutions for all loads in a single run. We showed that our approach can locate many solutions showing trade-off between the objectives for all load demands. The results also indicated that the most effective solutions may be discovered when the load demand is close to 330 MW as would be natural for a system designed for that particular load.

The methodology developed here is general and may be applied to the optimisation of any system for which a computational or analytic model is not available, but whose operation can be characterised through measurement. The innovation of optimisation while taking account of the uncertainties arising from measurement error and the relative paucity of data allows confidence in the results of the optimisation and in the choice of operating parameters.

We noted that the outputs sometimes show transient behaviour as the operating parameters are changed. Therefore significant step changes in operating parameters with changing demands may elicit unexpected

and undesirable behaviour from the system. Future work will investigate data-driven models to achieve time-dependent control to ensure smooth changes in the system response as operating parameters are changed to meet new loadings.

Nomenclature

Acronyms

AbYSS Archive-Based hYbrid Scatter Search

CFD Computational fluid dynamics

DCS Distributed control system

EA Evolutionary Algorithm

$\mathcal{GP}$ Gaussian process

GA Genetic algorithm

HR Heat rate

MLNN Multi-layer neural network

MOCell Multi-objective cellular genetic algorithm

NOx Oxides of Nitrogen [ppm]

OFA Over fire air

OMOPSO Multi-Objective Particle Swarm Optimiser

OSVR Online support vector regression

PQRS Partioned quasi random selection

SCR Selective catalytic reduction

SOx Oxides of Sulpher

SPEA Strength Pareto Evolutionary Algorithm

SVR Support vector regression

UBC Unburned coal content [%]

NSGA Non-dominated sorting genetic algorithms

Symbols

α	Confidence level in probabilistic dominance	
A	Archive: current approximation of the probabilistic Pareto set $\mathcal{P}^\alpha$ at any stage of evolution	
c	Constant factor relating load demand and coal feeder frequencies	[MW/rpm]
$\mathcal{D}$	Data set consisting of observed control parameters $\mathbf{x} \in X$ and associated response $\mathbf{y}$	
$\rho(\mathbf{z})$	Maximum probability that $\mathbf{z}$ is dominated by a member of the archive A	
f	Predicted function response for a control parameter $\mathbf{x}$	
F_k	kth coal feeder frequency	[rpm]
f_n	Measured function response $f(\mathbf{x}_n)$ for control parameters $\mathbf{x}_n$	
$\mathcal{G}_{\mathrm{NOx}}(\cdot)$	Trained $\mathcal{GP}$ model for NOx	
$\mathcal{G}_{\mathrm{UBC}}(\cdot)$	Trained $\mathcal{GP}$ model for UBC	
$H(A, \mathbf{r})$	The hypervolume of the non-dominated set A with respect to a reference vector $\mathbf{r}$	
K	Covariance matrix constructed from $\mathcal{GP}$ training data	
L	Load demand	[MW]
M	Number of control parameters	
N	Number of observations	
$\mathcal{P}^\alpha$	Probabilistic Pareto set	
$\mathcal{P}$	Optimal Pareto set	
$\mathcal{P}_L^\alpha$	Probabilistic Pareto set in the combined objective and constraint space Γ	
$\mathbf{r}$	A predefined reference vector in the objective space for hypervolume calculation	
σ_L	Standard deviation in load demand measurements	[MW]
σ_n^2	Variance quantifying uncertainty in measurement of f_n	
$\boldsymbol{\theta}$	Vector of covariance function hyperparameters	
$\mathbf{x}_n$	nth M-dimensional vector of control parameters	
$\mathcal{X}$	Feasible decision space	
$\mathbf{z}$	An arbitrary vector in the objective space	

Acknowledgement

This work was supported by the Engineering and Physical Sciences Research Council, United Kingdom [grant number: EP/M017915/1], the State Nature Science Foundation of China [grant numbers: 61375078 and 61304211], and the China Scholarship Council.

References

[1] S. Shafiee, E. Topal, When will fossil fuel reserves be diminished?, Energy Policy 37 (1) (2009) 181 – 189.

[2] A. L. Moretti, C. S. Jones, P.-G. Asia, Advanced emissions control technologies for coal-fired power plants, Babcock & Wilcox Power Generation Group. (BR-1886).

[3] Y. Zhao, S. Wang, L. Duan, Y. Lei, P. Cao, J. Hao, Primary air pollutant emissions of coal-fired power plants in china: Current status and future prediction, Atmospheric Environment 42 (36) (2008) 8442 – 8452.

[4] C. T. Bowman, Control of combustion-generated nitrogen oxide emissions: Technology driven by regulation, Symposium (International) on Combustion 24 (1) (1992) 859 – 878.

[5] J. Smrekar, P. Potočnik, A. Senegačnik, Multi-step-ahead prediction of NOx emissions for a coal-based boiler, Applied Energy 106 (2013) 89 – 99.

[6] WHO, Air Quality Guidelines: Global Update 2005. Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide, World Health Organization (WHO), 2006.

[7] R. K. Srivastava, R. E. Hall, S. Khan, K. Culligan, B. W. Lani, Nitrogen oxides emission control options for coal-fired electric utility boilers, Journal of the Air & Waste Management Association 55 (9) (2005) 1367–1388.

[8] H. Zhou, K. Cen, J. Fan, Multi-objective optimization of the coal combustion performance with artificial neural networks and genetic algorithms, International journal of energy research 29 (6) (2005) 499–510.

[9] S. Li, T. Xu, S. Hui, X. Wei, NOx emission and thermal efficiency of a 300MWe utility boiler retrofitted by air staging, Applied Energy 86 (9) (2009) 1797 – 1803.

[10] C. R. Choi, C. N. Kim, Numerical investigation on the flow, combustion and nox emission characteristics in a 500mwe tangentially fired pulverized-coal boiler, Fuel 88 (9) (2009) 1720 – 1731.

[11] S. R. Dindarloo, J. C. Hower, Prediction of the unburned carbon content of fly ash in coal-fired power plants, Coal Combustion and Gasification Products 7 (2015) 19–29.

[12] E. H. Chui, H. Gao, Estimation of NOx emissions from coal-fired utility boilers, Fuel 89 (10) (2010) 2977 – 2984.

[13] R. Weber, N. Schaffel-Mancini, M. Mancini, T. Kupka, Fly ash deposition modelling: Requirements for accurate predictions of particle impaction on tubes using rans-based computational fluid dynamics, Fuel 108 (2013) 586 – 596.

[14] S. Dal Secco, O. Juan, M. Louis-Louisy, J.-Y. Lucas, P. Plion, L. Porcheron, Using a genetic algorithm and cfd to identify low nox configurations in an industrial boiler, Fuel 158 (2015) 672–683.

[15] F. Wu, H. Zhou, T. Ren, L. Zheng, K. Cen, Combining support vector regression and cellular genetic algorithm for multi-objective optimization of coal-fired utility boilers, Fuel 88 (10) (2009) 1864 – 1870.

[16] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the strength pareto evolutionary algorithm, TIK-report 103.

[17] M. R. Sierra, C. A. C. Coello, Improving pso-based multi-objective optimization using crowding, mutation and ϵ-dominance, in: International Conference on Evolutionary Multi-Criterion Optimization, Springer, 2005, pp. 505–519.

[18] A. J. Nebro, F. Luna, E. Alba, B. Dorronsoro, J. J. Durillo, A. Beham, Abyss: Adapting scatter search to multiobjective optimization, IEEE Transactions on Evolutionary Computation 12 (4) (2008) 439–457.

[19] F. Wu, H. Zhou, J.-P. Zhao, K.-F. Cen, A comparative study of the multi-objective optimization algorithms for coal-fired boilers, Expert Systems with Applications 38 (6) (2011) 7179 – 7185.

[20] F. Si, C. E. Romero, Z. Yao, E. Schuster, Z. Xu, R. L. Morey, B. N. Liebowitz, Optimization of coal-fired boiler {SCRs} based on modified support vector machine models and genetic algorithms, Fuel 88 (5) (2009) 806 – 816.

[21] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, in: M. S. et al. (Ed.), Proceedings of Parallel Problem Solving from Nature - PPSN VI, Springer, 2000, pp. 849–858.

[22] T. Yu, H. Zhu, C. Peng, Multi-objective optimization of coal-fired boiler combustion based on NSGA-II, JNW 8 (2013) 1300–1306.

[23] H. Zhou, K. Cen, Combustion Optimization Based on Computational Intelligence, Advanced Topics in Science and Technology in China, Springer Singapore, 2018.

[24] S. A. Kalogirou, Artificial intelligence for the modeling and control of combustion processes: a review, Progress in Energy and Combustion Science 29 (6) (2003) 515 – 566.

[25] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[26] S. H. Baek, H. Y. Park, S. H. Ko, The effect of the coal blending method in a coal fired boiler on carbon in ash and NOx emission, Fuel 128 (2014) 62 – 70.

[27] E. S. Gardner, Exponential Smoothing: The State of the Art, Journal of Forecasting 4 (1) (1985) 1–28.

[28] GPy, GPy: A gaussian process framework in python, `http://github.com/SheffieldML/GPy` (since 2012).

[29] M. A. Alvarez, N. D. Lawrence, Computationally efficient convolved multiple output gaussian processes, Journal of Machine Learning Research 12 (May) (2011) 1459–1500.

[30] J. P. Kleijnen, E. Mehdad, Multivariate versus univariate kriging metamodels for multi-response simulation models, European Journal of Operational Research 236 (2) (2014) 573 – 582.

[31] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms, Wiley, Chichester, 2001.

[32] C. A. Coello Coello, G. B. Lamont, D. A. V. Veldhuizen, Evolutionary Algorithms for Solving Multi-Objective Problems, Springer, 2007.

[33] E. J. Hughes, Evolutionary Multi-objective Ranking with Uncertainty and Noise, in: E. Zitzler, L. Thiele, K. Deb, C. A. Coello Coello, D. Corne (Eds.), Proceedings of First International Conference on Evolutionary Multi-Criterion Optimization, Springer, Berlin, Heidelberg, 2001, pp. 329–343.

[34] J. E. Fieldsend, R. M. Everson, Multi-objective optimisation in the presence of uncertainty, in: IEEE Congress on Evolutionary Computation, Vol. 1 of CEC '05, 2005, pp. 243–250.

[35] J. E. Fieldsend, R. M. Everson, S. Singh, Using Unconstrained Elite Archives for Multiobjective Optimization, IEEE Transactions on Evolutionary Computation 7 (3) (2003) 305–323.

[36] A. Moraglio, R. Poli, Topological interpretation of crossover, in: K. Deb (Ed.), Genetic and Evolutionary Computation – GECCO 2004, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 1377–1388.

[37] E. Zitzler, Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications, Ph.D. thesis, ETH Zurich, Switzerland (1999).

[38] N. Beume, B. Naujoks, M. Emmerich, Sms-emoa: Multiobjective selection based on dominated hypervolume, European Journal of Operational Research 181 (3) (2007) 1653 – 1669.

[39] C. Fonseca, P. Fleming, On the Performance Assessment and Comparison of Stochastic Multiobjective Optimizers, in: H.-M. Voigt, W. Ebeling, I. Rechenberg, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature - PPSN IV, Vol. 1141 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 1996, pp. 584–593.

[40] J. Knowles, A Summary-Attainment-Surface Plotting Method for Visualizing the Performance of Stochastic Multiobjective Optimizers, in: Proceedings of the International Conference on Intelligent Systems Design and Applications, ISDA '05, 2005, pp. 552–557.