Incorporating uncertainty in data driven regression models of fluidized bed gasification: A Bayesian approach

Indranil Pan[a*], Daya Shankar Pandey[b]

[a]*Energy, Environment, Modelling and Minerals (E^2M^2) Research Section, Department of Earth Science and Engineering, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom*

[b]*Carbolea Research Group, Department of Chemical and Environmental Science, University of Limerick, Ireland*

*Corresponding author:
Tel.: +44 7448532764.
E-mail address: *i.pan11@imperial.ac.uk (I. Pan).*

Abstract: In recent years, different non-linear regression techniques using neural networks and genetic programming have been applied for data-driven modelling of fluidized bed gasification processes. However, none of these methods explicitly take into account the uncertainty of the measurements and predictions. In this paper, a Bayesian approach based on Gaussian processes is used to address this issue. This method is used to predict the syngas yield production and the lower heating value (LHV) for municipal solid waste (MSW) gasification in a fluidized bed gasifier. The model parameters are calculated using the maximum a-posteriori (MAP) estimate and compared with the Markov Chain Monte Carlo (MCMC) method. The simulations demonstrate that the Bayesian methodology is a powerful technique for handling the uncertainties in the model and making probabilistic predictions based on experimental data. The method is generic in nature and can be extended to other types of fuels as well.

Keywords: Municipal solid waste; Bayesian statistics; Gaussian processes; Gasification; Fluidized bed gasifier

1. Introduction:

Increasing energy demands and depleting fossil fuel reserves have attracted researchers to explore clean and renewable energy alternatives. On the other hand, the growing world population and rapid urbanization has resulted in an increase of wastes [1] and consequent issues with their disposal. The issues related to waste disposal has led to negative impacts on human health and the environment. Since recent European legislations are trying to enforce a sustainable development strategy, there are good reasons why solid wastes should be addressed by thermal treatment. Thermochemical conversion technologies have been used for converting biomass residues such as wood, rice husk, MSW and dried sewage sludge for energy recovery while addressing disposal issues. Gasification has emerged as an alternative to traditional combustion applications, offering better energy efficiency and lower NO_x, SO_x and particulate emissions. MSW is a heterogeneous fuel containing wide variety of wastes such as wood residues, paper, kitchen garbage, plastic and textiles which make it a better fuel for the gasification process [2].

Recent advances in disposal technology of MSW gasification has been reviewed in [3]. It was proposed that gasification is a viable alternative for waste treatment and subsequent energy recovery (in the form of transportable gas and liquid fuels) [3,4]. An independently verified emissions test indicated that gasification can meet the existing emission standards and this thermal conversion process will reduce the volume of landfill space significantly.

Over the years several contemporary researchers have worked on the experimental and modelling aspects of fluidized bed gasifiers. In the recent past, mathematical modelling and computer simulations have played an important role for conceptual design of the energy system. Often early stage research and

development rely on limited and incomplete information, which inevitably increased the uncertainties in model prediction [5]. The standard way to do uncertainty analysis of the model is to do a simple sensitivity analysis. Here each of the parameters is individually varied keeping the others fixed and the effect on the output is noted. This helps in identifying some of the key parameters which have maximum influence on the uncertainty of the output. However, this approach is mainly suited for a linear model. For non-linear models, which are frequently encountered in the design of energy systems, the effect of joint variation of two or more uncertain parameters might result in a larger deviation of the output. Therefore the simple sensitivity based methods have severe short comings and simulations based solely on these techniques might be misleading, especially, in the comparative analysis of alternative technologies. Therefore it is imperative to have an explicit framework for handling the uncertainties and quantifying their effects on model prediction.

A probabilistic method has been proposed to evaluate the uncertainties involved in an advanced integrated coal gasification combined cycle system for risk assessment, research and planning, process design and cost estimation [6]. It was recommended that the process designer has to consider uncertainties explicitly rather than ignoring it. The probabilistic model can significantly improve the research planning and management while incorporating the implication of uncertainty. Some of the most common representations of uncertainty, in the context of chemical and environmental engineering systems, have been reported by [7], which include interval mathematics, fuzzy theory and the probabilistic approach. Fuzzy set theory has been used to analyse uncertainty associated with process engineering [8,9]. However, the shortcomings of fuzzy set theory in uncertainty analysis have been reported by [10]. Optimization of energy system under uncertainty has been

extensively reviewed in [11]. This work highlighted the inevitability of uncertainty in real systems. Furthermore, it also advised that the use of computationally efficient deterministic equivalent modelling methods to handle such uncertainties in the system. In the experimental work carried out by [12], on a downdraft gasifier, the authors calculated that the total uncertainties involved in the measurement of process parameters such as temperatures, pressure drops, flow rates of wet gases and waste products, air fuel ratio, specific gasification rate and turndown ratio.

Recently, two approaches namely a purely probabilistic Monte Carlo (PMC) and a hybrid probabilistic possibilistic Monte Carlo (HMC) method were proposed to estimate dioxin/furan emissions from the waste gasification plant [13]. Those models were used for analysing the uncertainty propagation in an environmental impact assessment model. This study revealed that a HMC based environmental impact assessment model was more effective in propagating the input uncertainties than a PMC model. A techno-economic assessment and uncertainty analysis with respect to the development of algal bio-refineries were carried out for algae based biodiesel production [14]. A high dimensional model representation method was used for the global sensitivity analysis. A Monte Carlo (MC) method was employed to perform uncertainty analysis. The results of an extensive MC uncertainty analysis provided a better understanding of how uncertainties can influence algae conversion processes cost of algae derived biodiesel production, as well as how it can affect economic viability of the algal biorefineries. A direct Monte Carlo simulation based approach has been presented to quantify uncertainty involved in devolatilization kinetics of coal gasification processes [15]. The study also identified that gasifier temperature has a strong effect on product gas yield and the devolatilization step was a source of high uncertainty.

Increasing interest in biomass gasification has led to the development of different modelling approaches based on thermodynamic equilibrium, kinetic rate, fluid-dynamics and fast meta-models. In the recent past, fast meta-models including artificial neural networks and genetic programming have been presented for process design and optimisation of the gasification system in [16]. An artificial neural network approach was proposed to predict the SO_2 emissions from circulating fluidised bed boilers [17]. A multi-gene genetic programming based model was proposed by [18] for the prediction of LHV and syngas yield production from MSW in a fluidised bed gasifier. However, most of these modelling and experimental studies do not explicitly take the uncertainty into account.

The extensive literature review suggests that it is very important to explore the origin of uncertainties and propagate it through the model for assessing a realistic case. A Bayesian framework offers such flexibility for incorporating the measurement and other model parameter uncertainties and reflecting these in the final prediction. Bayesian methods have been employed in syngas combustion chemistry models for the quantification and propagating uncertainties into simulations as well as evidence based model comparison [19,20]. However, applications of the Bayesian approach have been rarely reported for the modelling of fluidized bed gasifiers. To the best of our knowledge, the optimal design of fluidzed bed gasifier under uncertainty has not been addressed so far. The purpose of this study is to apply Bayesian regression model for incorporating uncertainty in data driven models of fluidized bed gasification.

Given that several mathematical models have been evolved for the prediction of syngas production from the gasifier, it is important to incorporate sources of uncertainty involved in these models. Uncertainties are broadly categories in two

types – aleatoric and epistemic. Aleatoric (statistical) uncertainties are due to unknown variables that differ in multiple runs of the same experiment. Epistemic (systematic) uncertainties arise due to neglecting the effects of all possible parameters in the model. Generally uncertainty quantification techniques are geared towards reducing epistemic uncertainty. The present methodology, is generic and can incorporate both these effects depending on how the magnitude of the noise variance is defined. For example, the observations can themselves be noisy due to measurement variability in the experiments. In such a case, the model would reflect the aleatoric uncertainties. However, since only 9 input parameters are included in the model and other extraneous variables are neglected, it is possible to add some more noise variance to incorporate this (over and above that due to aleatoric uncertainty). The model would then reflect both these types of uncertainties. Ultimately, from an engineering perspective, it is the uncertainty and accuracy of the prediction that matters most. This is appropriately handled by the present method.

The rest of the paper is organised as follows. Section 2 provides a brief overview of the Bayesian approach and discusses the relative merits and de-merits of the same with its frequentist counterpart. It also introduces a Bayesian modelling approach, namely Gaussian processes, and demonstrates its applicability for regression problems involving uncertainty in the measurements and the predictions. Section 3 presents the simulation results and the discussions. The paper ends in Section 4 with the conclusions followed by the references.

2. Overview of the Bayesian modelling paradigm:

In a Bayesian context, the uncertainty or ‘degree of belief’ is quantified by probability. The Bayesian inference method allows for the quantification of prior knowledge that can be incorporated in the current model. The observations or experimental measurements can then be used to update the priors in a consistent fashion (using Bayes theorem) to give the posterior distribution of the beliefs. The Bayesian modelling paradigm considers the data to be fixed and the unknown model parameters to have a distribution. This is in contrast to the frequentist approach where the unknown parameters have a fixed point estimate. This inclusion of uncertainty in the model parameters results in more realistic predictions for the Bayesian methodology as compared to the frequentist approaches. Other advantages of the Bayesian method over the frequentist ones include less problems with over-fitting (due to the integration over the model parameters), avoiding problems with model identification and inclusion of informative priors.

Bayesian methods are often criticized due to the subjective nature of the priors and also due to the choice of the likelihood function. Previously, researchers mainly used conjugate priors for modelling (i.e. using a likelihood-prior pair which results in a posterior with the same functional form). This was done to enable analytical computation of the posterior density without having to bother about the denominator (marginal likelihood) in the Bayes’ rule. This simplification of-course led to inferior quality models as they did not reflect the actual distribution of the observed data. However, with the advent of powerful algorithms like MCMC, this problem can now be averted and hitherto computationally intractable distributions can be used with ease.

Popular methods for data-driven regression modelling include neural networks and genetic programming. From a Bayesian standpoint, the choice of a specific neural network model for regression can be thought of as defining a prior probability distribution over non-linear functions [21]. The learning process of the neural network can be interpreted as the posterior probability distribution over the unknown functions [21].

There are multiple toolboxes in the R language which allow for Gaussian Process regression like GPfit [22], GP toolkit [23] and Gaussian Process function data analysis package[24,25].There are also similar toolboxes in Matlab like GPML [24] GPstuff [26] etc. The GPstuff toolbox is used for the analysis conducted in this paper. The readers can use any of the above toolboxes to extend the present work or apply the methodology to their own dataset.

In this paper, Gaussian processes are used for the non-linear regression problem of mapping the inputs parameters of the gasification process to the outputs of interest (viz. LHV and yield). In the limit of very large neural networks, the prior distribution over non-linear functions as represented by the Bayesian interpretation of the neural network is a part of a larger family of probability distributions given by Gaussian processes [27].

Consider the regression problem of estimating a non-linear function $\zeta(\gamma)$ parameterised by w, for a set of input vectors $\Gamma_N = \gamma_i \ \forall\, i \in \{1,2,\cdots,N\}$ and the corresponding set of target values $\Theta_N = \chi_i \ \forall\, i \in \{1,2,\cdots,N\}$. The inference of the function $\zeta(\gamma)$ can be expressed in the form of the posterior probability distribution [21]

$$P\left(\zeta(\gamma)\middle|\Theta_N,\Gamma_N\right)=\frac{P\left(\Theta_N\middle|\zeta(\gamma),\Gamma_N\right)P\left(\zeta(\gamma)\right)}{P\left(\Theta_N\middle|\Gamma_N\right)} \quad (1)$$

where, the term $P\left(\Theta_N\middle|\zeta(\gamma),\Gamma_N\right)$ indicates the probability of the target values given the function $\zeta(\gamma)$ and the term $P\left(\zeta(\gamma)\right)$ is the prior distributions of the functions which are assumed by the model. Any choice of a parametric model implicitly specifies this prior in terms of smoothness and continuity. For any prediction problem involving unknown values of the target variable χ, only the prior distribution $P\left(\zeta(\gamma)\right)$ and the assumed noise model is important. The explicit parameterization of the function $\zeta(\gamma;w)$ is not required [21]. The key concept in modelling with Gaussian processes is to place a prior $P\left(\zeta(\gamma)\right)$ on the function space without explicitly parameterizing the function $\zeta(\gamma;w)$. The next sub-sections give a very brief introduction to Gaussian processes for regression problems along with a simple one dimensional example.

2.1. Gaussian processes and their application for regression

A Gaussian process is a random function $f(\psi)$, where ψ is the d dimensional input vector, which has the property that the joint distribution of any finite set of realizations $f(\psi_1),\cdots f(\psi_n)$ is a multi-dimensional Gaussian [24]. The mathematical definitions of Gaussian processes are given in Appendix A. Gaussian processes offer a flexible way to do regression. In contrast to assuming a fixed structure for the model (say linear, quadratic), the GP places a prior on the function space without explicitly parameterizing the function, as discussed in Section 2. The GP approach is not fully non-parametric though. It expresses the nature of the latent function (e.g. smoothness, variability) through the mean and covariance functions. This helps in

giving importance to the data itself and the final model is adaptable to the given dataset. This is in contrast to specifying a fixed model structure at the very onset and trying to fit the model parameters assuming that the data can be fitted by such a structure (which is not always the case).

To ensure computational tractability, the observations (y) are assumed to be conditionally independent given a latent function ($f(\psi)$), so that the likelihood ($p(y|f)$) factorises over the cases. In general, modelling with Gaussian processes can be defined using the following generic definition [29]

$$\text{Observation model:} \qquad y|f,\phi \sim \prod_{i=1}^{n} p(y_i|f_i,\phi) \tag{2}$$

$$\text{GP prior:} \qquad f(x)|\theta \sim \mathcal{GP}\left(m(\psi), k(\psi,\psi'|\theta)\right) \tag{3}$$

$$\text{Hyper-prior:} \qquad \theta,\phi \sim p(\theta)\,p(\phi) \tag{4}$$

where, $\{\theta,\phi\}$ are the parameters for the covariance function and the observation model respectively. The next section demonstrates a simple one dimensional regression problem with Gaussian processes, to illustrate the concept in detail.

2.1.1. Sample 1D example

Figure 1 shows a sample case of 1D regression with Gaussian processes. The actual measurements/observations are indicated by the red points. The mean of the predicted latent function is shown by a blue line and the filled area represents the 95% confidence interval for the predictions. The details of the model building procedure along with the priors, covariance functions etc. are outlined in Appendix B.

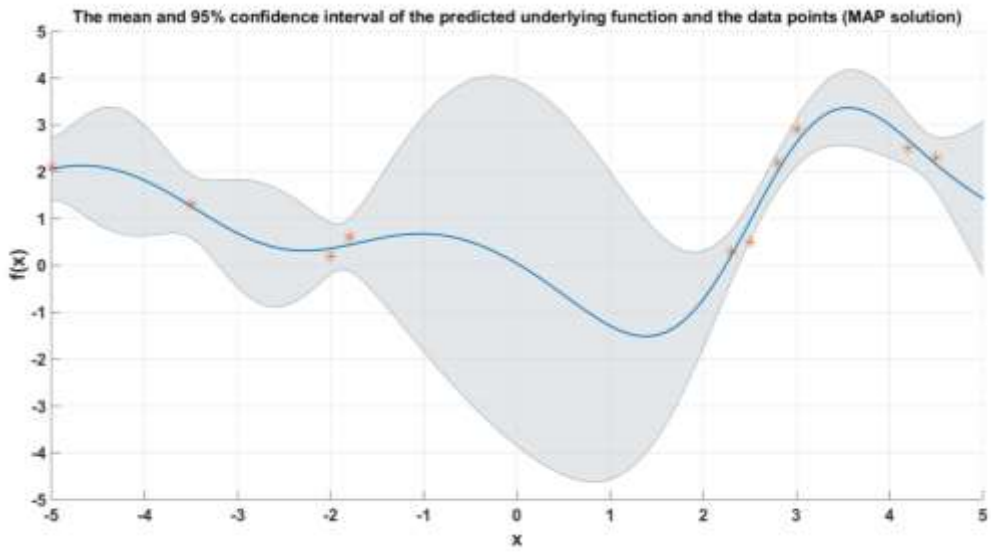

Figure 1: 1D regression of a function with Gaussian processes considering measurement uncertainty

A Gaussian likelihood function and a squared exponential covariance function are used for the demonstration. The actual measurements are considered to have some uncertainty associated with them. This is taken care by the likelihood function which specifies a distribution for the uncertainty (in terms of the noise variance σ^2). This is reflected in the prediction results as well, since the predicted mean of the latent functions do not pass through all the measurement points exactly. Also the corresponding confidence intervals are much smaller at the measurement points, but not exactly zero (as would have been the case with no measurement uncertainty). Another thing to note is that the uncertainty is much higher for the values of x near the centre (between -2 and 2). This is due to the fact that there are no observations in that region and also the property of the squared exponential covariance function which makes the correlations decay at points which are further away from the actual measurements. Figure 2 shows the same GP based regression (with the same data-points) without any consideration of measurement uncertainty. Clearly, this is reflected in the values of the confidence intervals at the measurement points which are zero, due to the consideration of precise measurements.

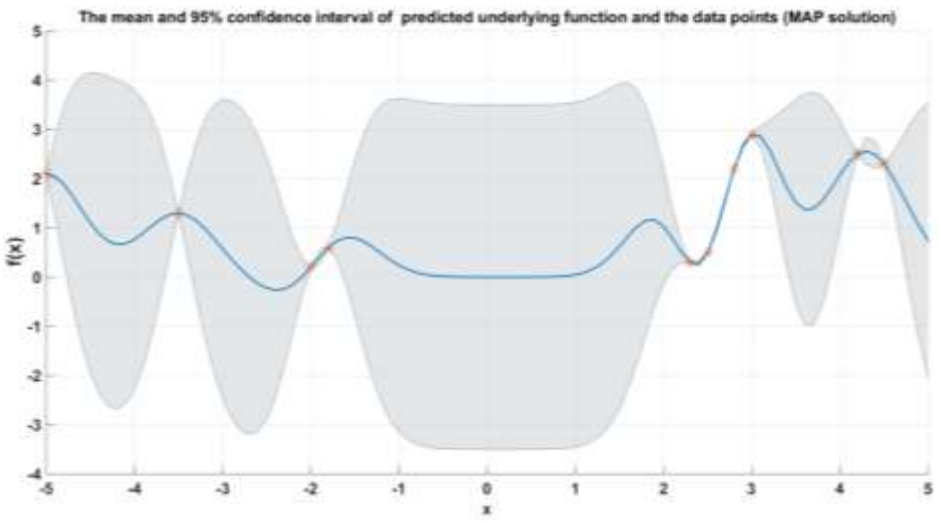

Figure 2: 1D regression of a function with Gaussian processes without any measurement uncertainty

The maximum a-posterior (MAP) estimate, as done in the present example, is preferred since it is computationally fast and easy to calculate. This is due to the fact that the objective function involving the log-marginal likelihood or its approximations are differentiable with respect to the parameters and therefore can be maximised with a gradient based optimisation algorithm in a few iterations. However, one of the fundamental shortcomings of the MAP estimate is that it underestimates the uncertainty in the model parameters $\{\theta, \phi\}$. This is because, the MAP algorithm essentially assign point values at the mode of the posterior distribution of the parameters.

Improvisation over the MAP entails approximating the marginal of the latent function by integrating over the parameters $\{\theta, \phi\}$ in some way. One of the ways is to use grid integration which entails a weighted summation over a grid of points in the space of $\{\theta, \phi\}$. However, this method quickly becomes computationally intractable as the number of parameters increases. The MCMC method can be useful in such a case to deal with the scaling issue. One of the issues with MCMC is that there is a

significant autocorrelation among the samples (even after rejecting the initial samples to account for burn-in and ensuring that the chain has converged to the stationary distribution) which implies that it is difficult to make independent draws of the samples. A chain thinning procedure can be done which rejects some of the samples at regular intervals and alleviates this issue to some extent.

2.2. Predictive information criteria for comparing different Bayesian models

There are many ways to assess different Bayesian models, for e.g., posterior predictive checks, prior predictive checks, Bayes factors and continuous model expansion. [28]. However, since the objective of such non-linear regression models is to accurately predict samples outside the training dataset, it is prudent to compare them based on their predictive accuracy. K-fold cross-validation is a useful way to estimate the out-of-sample prediction errors. In most cases, K is chosen as 5 or 10 if the dataset is large. In the extreme case when K is equal to the number of samples in the dataset, then the method is known as leave-one-out cross validation (LOO-CV).

If the LOO-CV method is employed, then the model's fit to the new data can be quantified using the root mean squared error (RMSE) value. Since, for the Bayesian case, the output is predicted in the form of a range of values, the expected value of the output is taken as the model prediction, for the calculation of this metric. The lower the value of RMSE, the better is the model fit. The RMSE is easy to calculate but is not apt for models which do not follow a normal distribution [29].

The log of the predictive density (log-likelihood) gives a more generalized summary of the predictive fit. This quantity is related to the Kullback-Leibler (KL) information measure. For models which are constructed from a large number of

samples, the one with the highest value of the mean of the log predictive density (MLPD) is the best model with the highest posterior probability. This also implies that this best model has the lowest value of KL information [29].

However, the LOO-CV is computationally intensive and might have issues with sparse data. To circumvent this issue, other measures like Akaike's Information Criterion (AIC), Deviance Information Criterion (DIC) and Watanabe Information Criterion (WAIC) can be used. These are essentially some sort of approximation to different versions of cross-validation. These criteria also incorporate some penalty term for the effective number of model parameters. This is important when comparing different types of models and selecting the appropriate one. Models with larger number of parameters would inherently be more flexible than one with fewer parameters. Therefore, they would be able to fit the data-set more accurately (with lower value of RMSE) than the simpler model. However, the added complexity of the model due to many parameters makes it less transparent. These information criteria can be used to assess whether it is worthwhile to use a model that better fits the data, at the cost of higher complexity.

The AIC [30] uses the maximum likelihood estimate (MLE) of the model parameters to compute the log-likelihood and adjusts it with a bias correction factor depending on the number of parameters of the model. This measure is useful for linear models with flat priors, but is not adequate for informative priors and hierarchical model structures [29]. The DIC [31] is to some extent a Bayesian version of AIC which uses the posterior mean instead of the MLE estimate as in AIC. It also uses a data based bias correction factor unlike the one based on model parameters as in the case of AIC. The WAIC [32] is a further improvement over the AIC and the DIC and is a more fully Bayesian approach to calculating the out-of-sample expected

log predictive density for a whole data set. The advantage of the WAIC over DIC and AIC is that it averages over the posterior distribution instead of conditioning on a point estimate like the MLE. This is consequently a better measure of predictive accuracy in the Bayesian context. There are two slightly different versions of the Watanabe Information Criterion, viz. $WAIC_V$ and $WAIC_G$. The $WAIC_V$ uses the functional variance along with the number of training inputs and the Bayesian training utility. The $WAIC_G$ is computed using the Bayesian training utility along with the Gibbs training utility. A lower value of WAIC implies a better model.

3. Results and discussions:

The input and validated datasets were obtained from the lab-scale fluidised bed gasifier. Experiments were performed in a lab-scale fluidised bed gasifier (560 mm high and an internal diameter of 31 mm) operating at an atmospheric pressure. The gasifier temperature was maintained externally by an electric heater. Silica sand was used as a bed material (particle size 0.250–0.355 mm). The gasifier consists with electric heater, screw feeder to supply the feed, filter for collecting elutriated char and ash and gas-bag for off-line sampling of produced gas. The details of the gasifier can be found elsewhere [33]. Hongkong MSW has been gasified in a small scale gasifier to assess the feasibility of installing MSW gasifier in Hong Kong University of Science and Technology [34]. Experiments were performed at different temperatures ($400 \leq$ Temperature $< 800\ ^0$C) and equivalence ratios ($0.2 \leq ER \leq 0.6$).

The total number of data-points used in this study is 67. Out of this, 57 experimental data points were obtained from [33] and another 10 data points from the Hongkong MSW [34]. The difference in the two types of feed stocks and gasification procedures are reflected in the 9 input variables. The data-set of the

experimental measurements consists of nine input variables viz. % by weight of carbon, hydrogen, nitrogen, sulphur, oxygen, moisture content and ash, along with equivalence ratio (ER) and temperature of gasifier (in °C). The definition of ER as considered in this paper is the ratio of actual air to fuel ratio to that of the stoichiometric air to fuel ratio required for complete combustion. The data-set has two output variables viz., LHV (kJ/Nm^3) and gas yield (Nm^3/kg). Each of the inputs and outputs are normalized within a range of [0, 1]. The mean value of the input and output variables are $[43.81\ \ 5.11\ \ 0.68\ \ 0.17\ \ 36.53\ \ 4.21\ \ 9.55\ \ 0.4\ \ 581]$ and $[3153\ \ 2.86]$ respectively. Similarly, their corresponding standard deviations are $[0.12\ \ 0.69\ \ 0.59\ \ 0.18\ \ 6.26\ \ 5.95\ \ 10.67\ \ 0.28\ \ 154.14]$ and $[835.80\ \ 2.62]$.

The Gaussian Process approach fits a generic model, leveraging on both the datasets, so that it can predict LHV and yield for other types of gasification processes as well, where these 9 input parameters are known.

The LOO-CV scheme is used for the present study. The data is partitioned into two sets, with the training set having all the data points except one and the validation set having only one data point. The model is trained on the training dataset (66 data points in this case) and the MSE of the prediction is tested on the remaining data point. This is repeated by partitioning the data in all similar possible combinations resulting in 67 such repetitions. The average of these 67 runs is taken as the MSE.

The LOO-CV is computationally demanding but is better than a hold-out cross validation strategy (with some % of data for training and the remaining for validation), since it reduces variability due to averaging and gives a better estimate of the model's predictive performance.

For fitting the data-set with the Gaussian processes, two different cases are considered – Case A, where the experimental observations are assumed to be perfect with negligible measurement uncertainty, Case B, where the experimental observations have an appreciable level of uncertainty associated with them.

For the simulations, the observation model in Eqn. (2) is considered to be Gaussian with a standard deviation of 0.01 and 0.1 in Cases A and B respectively. The mean function ($m(\psi)$), in Eqn. (3) is considered to be zero and a squared exponential covariance function is constructed for $k(\psi,\psi')$ in Eqn. (3), with length-scale and variance parameters of 0.01 and 0.01^2 respectively for Case A and 0.1 and 0.1^2 respectively for Case B. The observation model in Eqn. (2) is given a log uniform hyper-prior (i.e. the parameter ϕ) and the hyper-priors for the length scale and the variance parameters (θ) are set to uniform distributions (i.e. non-informative). Table 1 shows the LOO cross-validation results for the two cases (A and B) with the different inference methods.

Table 1: Fitness measures of the LOO-CV results with different inference methods for different cases

	Fitness Measures	LHV prediction		Yield prediction	
		Methods		**Methods**	
		MAP	MCMC	MAP	MCMC
Case A	$RMSE_{LOO}$	0.1389	0.1108	0.018	0.019
	$MLPD_{LOO}$	0.3325	0.7893	2.8006	2.6893
	$WAIC_G$	5.3115	0.9661	3.0584	3.0253
	$WAIC_V$	5.1171	0.8723	2.9567	2.9109
Case B	$RMSE_{LOO}$	0.1026	0.1668	0.018	0.0191
	$MLPD_{LOO}$	0.9091	0.2544	2.8002	2.6905
	$WAIC_G$	1.068	0.9117	3.0599	3.029
	$WAIC_V$	0.9837	0.603	2.9576	2.9142

From Table 1, it can be observed that the different fitness measures for the yield prediction is almost same for both the MAP and the MCMC based approach.

However, for the LHV predictions there is a significant difference between the MAP and the MCMC methods. In case A, the MCMC method clearly outperforms the MAP method as it shows a lower value of the $RMSE_{LOO}$, and WAIC and higher value of $MLPD_{LOO}$ as compared to the MAP method. This implies that the predictive accuracy for out-of-sample cases is better with the MCMC method. For Case B, a lower value of $RMSE_{LOO}$ and a higher value of $MLPD_{LOO}$ for the MAP case than the MCMC case indicates that the former method is better. However, as discussed in Section 2.2, these alone might not be a good indicator of better models as they do not use a bias correction factor and penalize the model complexity. The WAIC values which look at both the predictive fits and the model complexity are lower for the MCMC case as compared to the MAP method. Therefore the MCMC method turns out to be better using this criterion.

Figure 3 shows the output of the Gaussian process based Bayesian model with respect to the actual measurements. The model parameters are obtained by the maximum a-posterori (MAP) inference method. It can be observed in Case A, that there is almost no uncertainty in the measurements and the corresponding model predictions are very accurate (i.e. they lie on the blue line) and precise (the error bars on the predictions are very small). For case B, the measurement uncertainty is indicated by the blue horizontal bars, for each of the data points marked in red. This is fed into the model and the corresponding model output can be read off the ordinate. The associated uncertainty in the model predictions is given by the vertical orange bars. It can be observed that all of the model predictions in Case B do not fall on the blue line, unlike that in Case A. This is expected since there is significant uncertainty associated with the measurements themselves.

Figure 3 also shows the simulation results for the yield prediction. In Figure 3, it can be observed that, unlike the LHV case, the MAP estimate gives lower values of uncertainty in the yield prediction for case B (even though there is appreciable uncertainty in the measurements). The yield prediction outputs for Case A have a slightly higher level of uncertainty than the corresponding MAP predictions for LHV in Case A.

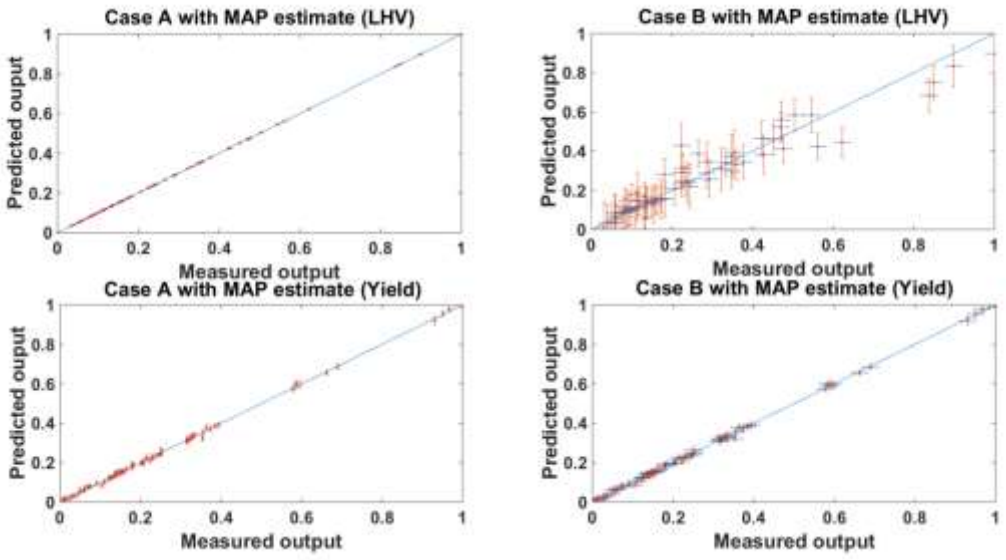

Figure 3: Plot of model predictions versus experimental measurements for LHV and yield prediction in both cases (A & B) with MAP estimate of the model parameters

For the MCMC method, the sample chains of the model parameters are generated for 10,000 iterations. The first 500 samples are rejected to take into account the effects due to burn-in. This helps to give the algorithm some iterations to reach the equilibrium distribution and overcome the issue of starting from a point with a very low probability under the equilibrium distribution. The sample chains of length scale and magnitude of the GP covariance function (i.e. the model parameters) are shown in the upper row in Figure 4. Also shown in Figure 4 is the predictions made by the MCMC method vis-à-vis the measured output. It can be observed that this is similar to the case with the MAP estimate of LHV for Case B in Figure 3.

The lower row of Figure 4 shows the sample chains from the MCMC algorithm, along with the corresponding MCMC yield predictions for case B. The sample chains do not deviate significantly from their mean values, unlike the case for LHV prediction. This can also be observed in the lower row of Figure 5, where the distributions do not have fat tails. Figure 5 is essentially a histogram plot of the sample chains of the model parameters as obtained by the MCMC method in Figure 4. It also shows the superimposed MAP estimate on the histograms.

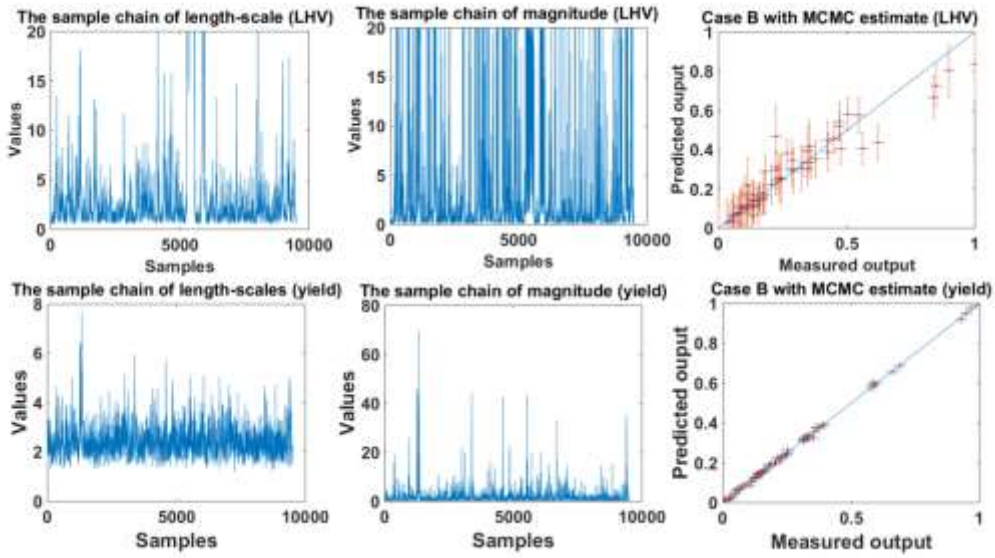

Figure 4: Sample chains from the MCMC algorithm for the length-scale and the magnitude parameter in the Gaussian process model along with the LHV and yield model predictions for Case B

The MAP estimate works almost as good as the MCMC estimate for the case of yield predictions while it does not perform that well for the LHV predictions. The reason for this can be deduced from Figure 5. The model parameters for the LHV predictions in Figure 5 has a long tailed distribution and therefore approximating the distribution by its mode (as done in the MAP case) is not very representative of the overall distribution. However, in the case of yield predictions as in the lower row of Figure 5, the distributions tend to be closer to Gaussians and therefore the MAP method works better.

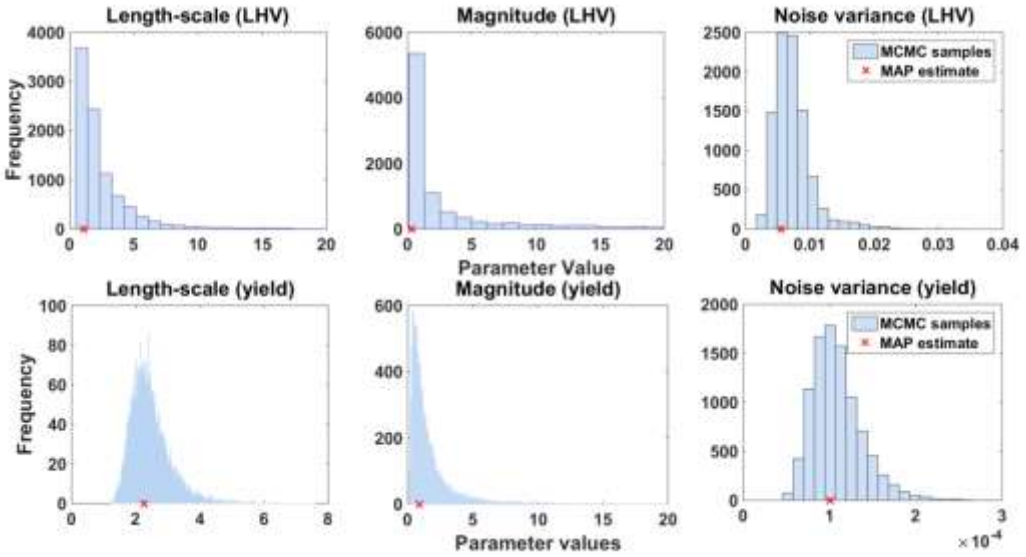

Figure 5: A comparison of the MAP estimate versus the MCMC samples for the different LHV and yield model parameters in Case B

Figures 6 and 7 show the conditional predictions for each of the nine individual co-variates, when all the others are fixed to their mean values. Comparing Figures 6 and 7, it is clear that the uncertainty in the predictions for LHV is much greater than that of the yield. This is also supported by the quantitative error metrics in Table 1.

It is possible to improve the predictions of the model by understanding how each of the covariates behave and identifying what are the ranges of the important covariates for which the experiments should be designed and the data needs to be collected.

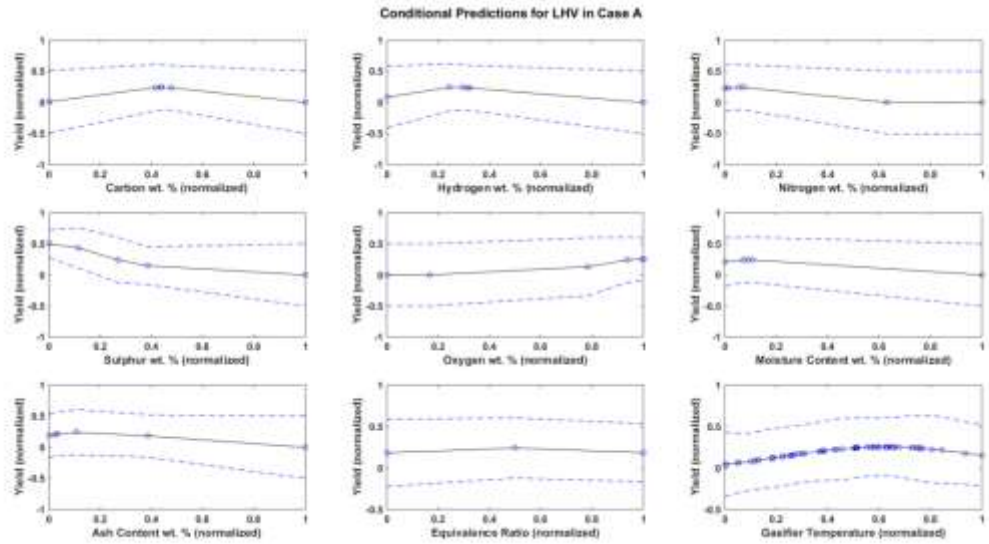

Figure 6: Conditional predictions for LHV for Case A

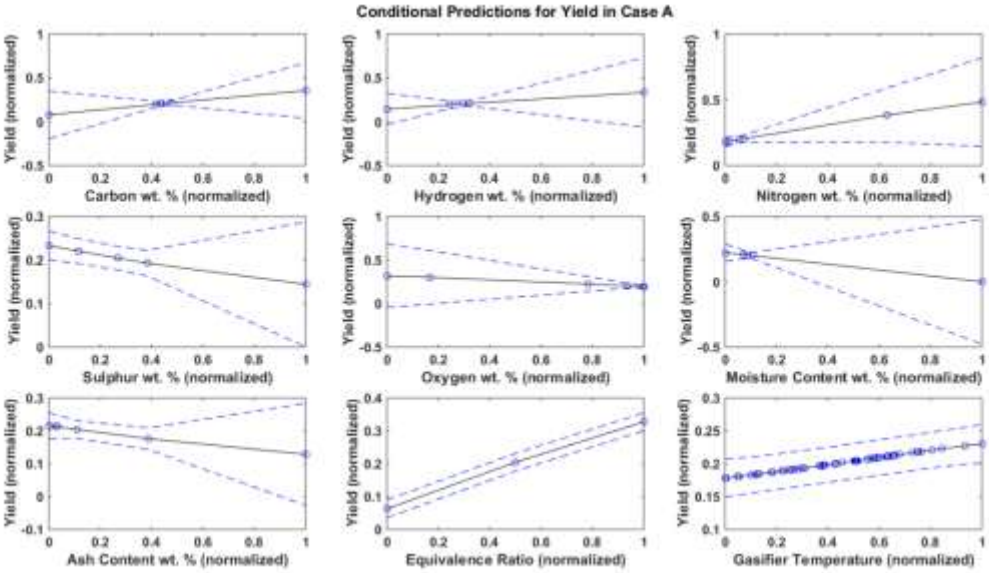

Figure 7: Conditional predictions for Yield for Case A

As shown in Figure 7, most of the data-points for moisture content or ash content are very closely clustered, i.e., they do not uniformly cover the whole range of the abscissa. Therefore the uncertainties in the conditional predictions are smaller near the data-points, but have much higher variance at points where there is no data available. This is not the case for gasifier temperature, for example. It covers the whole range of the abscissa uniformly and therefore has similar levels of uncertainty

predictions throughout. So from Figure 7, it is clear that to improve the predictions, it is important to conduct more experiments with fuels having higher values of moisture and ash content. Even though the equivalence ratio does not have a very uniform distribution of the data points over the whole range of the abscissa, the uncertainties in the conditional predictions are lesser. This implies that in the experimental design, the equivalence ratio does not need to be varied much in order to get more confident predictions. Therefore, it is more apt to identify the different experimental designs for which the gasification experiments can be conducted and the data points can be used to significantly improve the predictive power of the models.

One issue with using conditional predictions is that it keeps all the other covariates fixed at their mean values. This might not show the true effect of the covariate. To investigate the effect of each parameter on the model, ideally the parameter must be excluded and all possible model combinations must be done with the remaining eight covariates and then compared with similar models using all the nine covariates. However, this would result in a huge number of model combinations which is realistically infeasible to evaluate.

Traditional regression analyses have been additionally done using the same dataset. Table 2 showing the comparison of these methods with the Gaussian Process approach and with other research works employing ANN. The comparison can only be done for the case without uncertainty as none of the previous literatures have explicitly considered uncertainty in their data modelling techniques.

The results from Table 2 indicate that the traditional regression methods have higher relative errors of prediction. This might be due to outliers, in which case

robust variants of regressions might be more effective. The accuracy is also dependent on the structure of the regression model which is assumed a-priori.

Table 2 Comparison of regression models and corresponding relative errors

Modelling technique	LHV Relative Error (%)		Yield Relative Error (%)	
ANN [21]	20.5	-20.5	12.9	-12.9
GP (no uncertainty, MAP solution)	0.04	-0.03	17.53	-18.34
Traditional regression models				
Linear	476.00	-228.12	97.00	1009.60
Interactions	181.84	-302.98	135.31	-32.45
Purequadratic	406.94	-233.96	93.81	-1110.40
Quadratic	184.36	-299.16	107.62	-28.04

where,

$$relative\,error = \left(\frac{predicting\,value - \exp erimental\,value}{\exp erimental\,value} \right) \times 100\% \tag{5}$$

The mathematical structures for the traditional regression models are shown below:

$$y_{linear} \sim 1 + \sum_{i=1}^{9} x_i \tag{6}$$

$$y_{\text{interactions}} \sim 1 + \sum_{i,j=1}^{9} x_i x_j \tag{7}$$

$$y_{\text{purequadratic}} \sim 1 + \sum_{i=1}^{9} x_i + \sum_{i=1}^{9} x_i^2 \tag{8}$$

$$y_{\text{quadratic}} \sim 1 + \sum_{i,j=1}^{9} x_i x_j + \sum_{i=1}^{9} x_i^2 \tag{9}$$

It is imperative to stress that even though the GP gives good predictive accuracy for this case, the whole purpose of the Bayesian philosophy is to make predictions

incorporating uncertainty. Even though this is important for comparison purposes with other contemporary works, judging the utility of the approach just based on results presented in Table **2** would miss out the whole purpose of using GP in the first place.

4. Conclusions:

This paper applied a Bayesian methodology for non-linear regression, to predict the LHV and the syngas yield production, for the MSW gasification in a fluidized bed gasifier. The methodology is demonstrated to incorporate the uncertainties in the experimental measurements and reflect it in the probabilistic nature of the predictions. A cross-validation approach is used to estimate the model parameters which ensures that the model does not over-fit the training data-set and can give good predictions on untrained data-sets as well. Comparison among different models can also be done using this technique as shown in the simulation results. Having confidence intervals instead of point estimates for the model predictions is important, especially in cases where the output of the gasifier model is a part of a more complex system level model. This method also helps to improve the predictions of the model by understanding how each of the covariates behave and identifying what are the ranges of the important covariates for which the experiments should be designed and the data needs to be collected. This uncertainty quantification method can also be useful for estimating the feasibility of different alternate energy pathways and making policy decisions based on the probabilistic outputs. Future work can look at application of these data-driven Bayesian methods to other areas in bio-resource technologies.

Appendix A: Mathematical definition of Gaussian processes

A Gaussian process can be mathematically defined in the form of the mean function ($m(\psi)$) and the covariance function ($k(\psi,\psi')$) as follows:

$$m(\psi) = \mathrm{E}\left[f(\psi) \right] \tag{1}$$

$$k(\psi,\psi') = \mathrm{E}\left[\left(f(\psi) - m(\psi) \right)\left(f(\psi') - m(\psi') \right) \right] \tag{2}$$

The joint distribution of the finite set of realizations $f(\psi_1),\cdots f(\psi_n)$ can then be expressed in terms of the mean and covariance functions as follows:

$$\begin{pmatrix} f(\psi_1) \\ \vdots \\ f(\psi_n) \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} m(\psi_1) \\ \vdots \\ m(\psi_n) \end{pmatrix}, \begin{pmatrix} k(\psi_1,\psi_1) & \cdots & k(\psi_1,\psi_n) \\ \vdots & \ddots & \\ k(\psi_n,\psi_1) & & k(\psi_n,\psi_n) \end{pmatrix} \right) \tag{3}$$

Appendix B: Mathematical underpinnings of Gaussian process regression

For building the model in Figure 1, the measurements are considered to satisfy

$$y = f(\psi) + \varepsilon \tag{4}$$

where, $f(\cdot)$ is the latent function and ε is the error term which is distributed with a zero mean and standard deviation σ, i.e., $\varepsilon \sim N\left(0,\sigma^2\right)$. A zero mean Gaussian process prior is placed on $f(\cdot)$. In other words, at the measured points ψ_{meas}, the latent variables have the prior

$$p\left(f \middle| \psi_{meas} \right) \sim N\left(0, K_{n,n}\right) \tag{5}$$

where, n is the number of data points, $K_{n,n}$ is the covariance matrix and its elements ($\left[K_{n,n}\right]_{ij}$) are expressed in terms of its co-variance function (k) and its corresponding hyper-parameters θ as follows:

$$K_{ij} = k\left(\psi_i, \psi_j \middle| \theta\right) \tag{6}$$

A commonly used covariance function is the stationary squared exponential given by

$$k_{se}\left(\psi_i, \psi_j\right) = \sigma_{se}^2 e^{\left(-\sum_{d=1}^{D}\left(\psi_{i,d}-\psi_{j,d}\right)^2 \middle/ l_d^2\right)} \tag{7}$$

where, σ_{se} is the scaling parameter, D is the total number of dimensions and l_d is the length scale which dictates the decay in correlation along the direction d. Therefore the hyper-parameters denoted by θ in Eqn. (consists of $\left\{\sigma_{se}^2, l_1, \cdots, l_D, \sigma^2\right\}$. As both the likelihood and the prior are Gaussian, the marginal likelihood is also Gaussian and can be expressed as

$$p\left(y \middle| \psi_{meas}, \theta\right) = N\left(0, K_{n,n} + \sigma^2 I\right) \tag{8}$$

A prior $p(\theta)$ is placed on the hyper-parameters θ and the maximum a-posterior (MAP) estimate $\hat{\theta}$ is obtained by maximizing the following objective function

$$\hat{\theta} = \arg\max_{\theta} \quad \left[\log\left(p\left(\theta\right)\right) - \frac{1}{2}\log\left|K_{y,y}\right| - \frac{1}{2} y^T K_{y,y}^{-1} y\right] \tag{9}$$

where $K_{y,y} = K_{n,n} + \sigma^2 I$. Instead of the above MAP method of hyper-parameter estimations, approximations of the posterior of the parameters can also be obtained using computationally suitable methods like MCMC or other integration approximations like grid integration. The hyper-parameter values obtained by any of

the other methods can then be used with the data to make predictions $f(\psi_*)$ at any new value ψ_* using the following equation

$$p\left(f_* \middle| y, \theta, \psi_*\right) = N\left(K_{*,n}K_{y,y}^{-1}y, K_{*,*} - K_{*,n}K_{y,y}^{-1}K_{n,*}\right) \quad (10)$$

References:

[1] Alavi Moghadam M, Mokhtarani N, Mokhtarani B. Municipal solid waste management in Rasht City, Iran. Waste Management 2009;29:485–9.

[2] Fabry F, Rehmet C, Rohani V, Fulcheri L. Waste gasification by thermal plasma: a review. Waste and Biomass Valorization 2013;4:421–39.

[3] Arena U. Process and technological aspects of municipal solid waste gasification. A review. Waste Management 2012;32:625–39.

[4] Gómez-Barea A, Vilches L, Leiva C, Campoy M, Fernández-Pereira C. Plant optimisation and ash recycling in fluidised bed waste gasification. Chemical Engineering Journal 2009;146:227–36.

[5] Gong B. Methodology for technology evaluation under uncertainty and its application in advanced coal gasification processes. Doctoral Thesis. Massachusetts Institute of Technology.Department of Chemical Engineering, 2011.

[6] Frey HC, Rubin ES. Evaluation of advanced coal gasification combined-cycle systems under uncertainty. Industrial & Engineering Chemistry Research 1992;31:1299–307.

[7] Tatang MA. Direct incorporation of uncertainty in chemical and environmental engineering systems. Doctoral Thesis. Massachusetts Institute of Technology. Department of Chemical Engineering, 1995.

[8] Kraslawski A, Koiranen T, Nystrom L. Concurrent engineering: robust design in fuzzy environment. Computers & Chemical Engineering 1993;17:S447–S452.

[9] Kraslawski A. Review of applications of various types of uncertainty in chemical engineering. Chemical Engineering and Processing: Process Intensification 1989;26:185–91.

[10] Spiegelhalter DJ. A statistical view of uncertainty in expert systems. Artificial Intelligence & Statistics Edited By Gale, WA, Addison Wesley 1986.

[11] Sahinidis NV. Optimization under uncertainty: state-of-the-art and opportunities. Computers & Chemical Engineering 2004;28:971–83.

[12] Dogru M, Midilli A, Howarth CR. Gasification of sewage sludge using a throated downdraft gasifier and uncertainty analysis. Fuel Processing Technology 2002;75:55–82.

[13] Ripamonti G, Lonati G, Baraldi P, Cadini F, Zio E. Uncertainty propagation in a model for the estimation of the ground level concentration of dioxin/furans emitted from a waste gasification plant. Reliability Engineering & System Safety 2013;120:98–105.

[14] Brownbridge G, Azadi P, Smallbone A, Bhave A, Taylor B, Kraft M. The future viability of algae-derived biodiesel under economic and technical uncertainties. Bioresource Technology 2014;151:166–73.

[15] Gel A, Chaudhari K, Turton R, Nicoletti P. Application of uncertainty quantification methods for coal devolatilization kinetics in gasifier modeling. Powder Technology 2014;265:66–75.

[16] Puig-Arnavat M, Bruno JC, Coronas A. Review and analysis of biomass gasification models. Renewable and Sustainable Energy Reviews 2010;14:2841–51.

[17] Krzywanski J, Czakiert T, Blaszczuk A, Rajczyk R, Muskala W, Nowak W. A generalized model of SO_2 emissions from large-and small-scale CFB boilers by artificial neural network approach: Part 1. The mathematical model of SO_2 emissions in air-firing, oxygen-enriched and oxycombustion CFB conditions. Fuel Processing Technology 2015;137:66–74.

[18] Pandey DS, Pan I, Das S, Leahy JJ, Kwapinski W. Multi-gene genetic programming based predictive models for municipal solid waste gasification in a fluidized bed gasifier. Bioresource Technology 2015;179:524–33.

[19] Braman K, Oliver TA, Raman V. Bayesian analysis of syngas chemistry models. Combustion Theory and Modelling 2013;17:858–87.

[20] Braman K, Oliver T, Raman V. Bayesian methods for the quantification of uncertainties in syngas chemistry models. AIAA 2012-0496 2012.

[21] MacKay DJ. Information theory, inference, and learning algorithms. Cambridge University Press; 2003.

[22] MacDonald B, Ranjan P, Chipman H. GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs. Journal of Statistical Software 2015;64:1–23.

[23] Kalaitzis AA, Honkela A, Gao P, Lawrence ND. Gaussian Processes Tool-Kit-An R package for GP regression. University of Sheffield, UK 2015.

[24] Rasmussen CE, Williams C. Gaussian processes for machine learning. The MIT Press 2006.

[25] Febrero-Bande M, Oviedo de la Fuente M. Statistical computing in functional data analysis: the R package fda. usc. Journal of Statistical Software 2012;51:1–28.

[26] Vanhatalo J, Riihimäki J, Hartikainen J, Jylänki P, Tolvanen V, Vehtari A. GPstuff: Bayesian modeling with Gaussian processes. The Journal of Machine Learning Research 2013;14:1175–9.

[27] Neal RM. Bayesian learning for neural networks. Doctoral dissertation. University of Toronto; 1995.

[28] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Taylor & Francis; 2014.

[29] Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Statistics and Computing 2014;24:997–1016.

[30] Akaike H. Information theory and an extension of the maximum likelihood principle. Selected Papers of Hirotugu Akaike, Springer; 1998, p. 199–213.

[31] Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2002;64:583–639.

[32] Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. The Journal of Machine Learning Research 2010;11:3571–94.

[33] Xiao G, Ni M, Chi Y, Jin B, Xiao R, Zhong Z, et al. Gasification characteristics of MSW and an ANN prediction model. Waste Management 2009;29:240–4.

[34] Choy KK, Porter JF, Hui C-W, McKay G. Process design and feasibility study for small scale MSW gasification. Chemical Engineering Journal 2004;105:31–41.