A Study of the discrepancy function when modeling a packed bed using Variational Bayes and Gaussian processes

To cite this article: Sang-Ien Sien and A. F. Emery 2018 J. Phys.: Conf. Ser. 1047 012014

View the article online for updates and enhancements.
A Study of the discrepancy function when modeling a packed bed using Variational Bayes and Gaussian processes

Sang-Ien Sien and A. F. Emery
Department of Mechanical Engineering, University of Washington, Seattle, WA 98195
E-mail: Emery@u.washington.edu

Abstract. Estimating properties using maximum likelihood (MLE) gives point values and a rough idea of the uncertainty in the parameters if their distributions are approximately normal. More accurate results require a Monte Carlo approach, often using the Markov Chain Monte Carlo (MCMC) method, which for complex models is unrealistic because of the computational expense. The Variational Bayes approach gives results comparable to MCMC with only a few evaluation of the model. When the model is imprecise or expensive to evaluate, Gaussian processes provide one means of analyzing the noise to determine the uncertainties.

Keywords: Parameter Estimation, Uncertainty, Bayesian Inference, MCMC, Variational Bayes, Gaussian Processes

1. Introduction
It is rare that a physics based model accurately represents a real situation. Firstly because the parameters of the model are usually known only approximately, but more likely because the sub-models are only approximations of the true situation. Probably the most well known deficient sub-model is the turbulence model for which any one of the several different choices may be only a crude approximation to the actual turbulence in the flow.

Let $D$ be the measured response of a system and $M(x, t, \theta)$ be the output of our physics based model where $x, t, \theta$ represent spatial position, time and parameters respectively, then we write

$$D = M + \delta + \epsilon$$

where $\delta$ represents the shortcomings of the model and $\epsilon$ is the measurement noise. $\delta$ is termed the 'discrepancy'. Clearly as a model is improved, $\delta$ should diminish. The improvement can be due to more accurate parameters or properties or by improved sub-models. On the other hand, it is unlikely that the model will improve if correlations are used to represent aspects of the model since most correlations, particularly in heat transfer or fluid flow, are inherently imprecise because they are developed from experimental conditions that often do not accurately relate to the model under question. Using simulations based upon discrepant models to estimate parameters of the model is common, but clearly leads to imprecise estimates. Many papers that treat parameter estimation are often based upon simulated experiments in which the
'experimental data' are either the output of the model corrupted by noise or the output of a model clearly known to be inadequate. For example in heat transfer ignoring known losses, e.g., treating a case with prescribed surface heat fluxes by a model that considers only Dirichlet boundary conditions.

Clearly if we know $\delta$, one can always improve the model. The problem is that all that we learn from the data is $\delta + \epsilon$ and it will be impossible to separate them. Of course when using simulated experiments with known $\epsilon$, $\delta$ is known. The problem becomes more severe when our model is used to estimate parameters. Typically, one scrutinizes the residuals, $res$ defined by

$$res = D - M(\Theta)$$

where $D$ are the data and $\Theta = (\theta_1, \ldots, \theta_n)$ are the parameters to be estimated. If the residuals exhibit desirable traits, e.g., zero mean, constant standard deviation, and near zero autocorrelation we can assign statistical information to our parameter estimates [17]. If the residuals do not have these traits, then much less can be said about our parameter estimates.

A good fitting model $M(x_i, t_i, \theta)$ is one for which the discrepancy is of the same order or less than the measurement error. It is important to recognize that there is no assurance that the estimated parameters, $\hat{\theta}$, of our model, $M(x_i, t_i, \theta)$ are the true values.

Consider an experiment in which $D$ are time series data, for example temperatures measured as a function of time at a specific spatial point. One approach is to assume the form of $\delta(t)$, for example a polynomial, and to increase the order until the residuals exhibit the desired qualities. When the data are time series at several different spatial points, $X$, then one is faced with deriving $\delta(t, X)$. This is rarely effective, particularly when estimating parameters. On the other hand, if we can determine that the residuals, $res(t, X)$ exhibit similar characteristics for all $X$ and all $t$, the problem may be simpler.

In this paper we discuss the estimation of the convective heat transfer coefficient and specific heat of a packed bed using a model that is clearly an approximation. Our interest is in determining if the discrepancies of the model when predicting the behavior of the temperature at different points in the bed share common traits as a function of the Reynold's number by using a Bayesian approach.

2. Packed Beds

Our packed bed, Figure 1, is a collection of small spheres of solid material contained in a circular tube. Such beds are used extensively in the chemical industry to promote chemical reactions such as absorption, stripping and distillation or as catalysts to promote reactions. Packed beds are also used extensively in thermal regenerators and for energy storage. In this latter case, the beds are often operated in a cyclical manner and a knowledge of the heat transfer coefficient for the interphase heat transfer is important to optimize their design. Because of the complexity of the system due to local variations in the fluid velocity, the transient nature of the solid material temperatures, contact resistance between the particles, and often the existence of strong, but highly irregular radial conduction, and the variation of the void fraction both axially and laterally, it is difficult to analyze the thermal behavior and particularly to optimize their design [32]. We are particularly interested in being able to predict the effects of cyclical operation which requires a knowledge of the heat transfer coefficient and the effective specific heat of the particles [1,34]. Designs are often based upon one dimensional flow and a two temperature model (fluid and packing). In reality, the flow is seldom one-dimensional and a number of experimental
studies have been conducted to determine the radial velocity profile and the effective radial thermal conductivity. Recently, several CFD and multi-physics simulations based upon the Navier Stokes equations have been used to more explicitly study these effects [10, 5, 16, 4, 15]. These studies have used as many as 30 million cells and required up to 11 hours of cpu time.

![Figure 1. Schematic of Packed Bed](image)

### 2.1. One Dimensional Analysis

Most designs are based upon the one dimensional two temperature models, usually referred to as the 'plug flow' method [1]. This approach is similar to the analysis of porous beds and ignores the effect of the radial conductivity and the radial variation of velocity. Reference [27] used this approach to consider rock beds and tubes with phase change and [7] treated a similar problem, but with particles that were very small relative to the bed diameter. The model is clearly deficient and entails a significant discrepancy, $\delta(r, z)$. Adebiyi [2, 23] compared the two equation model with a model that considered two dimensional flow, radial conduction, and losses to the walls for a bed for which the times of heating (the charging time) and cooling (the recovery time) were approximately of the same duration and showed that the two equation model sufficed.

The packed bed used in this experiment consists of a hollow glass cylinder, 0.614 m long, 114.8 mm OD and 102 mm ID packed with 1359 glass marbles each 15 mm in diameter. Air is heated by flowing over an electrical resistance heater to simulate a step change in supply temperature. Data are taken during the transient heating since the sensitivity of the parameters sought is highest then. Typical tests last 15 minutes for the high flow rates to several hours for the low flow rates. Reynolds numbers (see Eq. 4a) ranged from 240 to 7300. The temperatures were measured as a function of time as the bed was heated and then cooled by the flowing air. Along the bed vertical centerline, instrumented marbles were placed approximately every 7 cm from the inlet to the outlet. Each instrumented marble had a small diametral hole with a thermocouple located at its center and another at its surface to measure the air temperature. Figure 2 shows the time history of the marble center temperatures (every 10th increment in time) as a function of the marble position ($z$ in cm) in the bed with the $z=0$ denoting the supply air temperature.

Because the efficiency of the bed in storing and recovering energy is strongly affected by the convective heat transfer between the flowing air and the marbles, we were interested in examining the convective heat transfer coefficient ($h$) during the heating and cooling periods, and the thermal capacitance ($c_s$) of the marbles (of diameter $d$) for a bed whose void fraction was $f=0.50$. The model follows that of Riaz [28] in which the air and bed temperatures are described by
Figure 2. Packed Bed Temperatures as function of Axial Position (z, cm)

\[
\text{Solids} \quad \rho_s c_s (1 - f) \frac{\partial T_s}{\partial t} = h_v (T_s - T_a) \tag{3a}
\]

\[
\text{Air} \quad f \rho_a c_a \left( \frac{\partial T_a}{\partial t} + U_{ave} \frac{\partial T_a}{\partial z} \right) = h_v (T_s - T_a) \tag{3b}
\]

where \( z \) represents the axial distance along the bed measured from the entrance of the bed, \( h_v = h (1 - f) \frac{d}{2} \) is the volumetric heat transfer coefficient, \( U_{ave} \) is the average air velocity, and the subscripts \( a \) and \( s \) refer to air and solids respectively.

2.2. Experimental Results

We analyzed 21 experiments in which the Reynolds number ranged from 240 to 7300 and compared the resulting estimates of the heat transfer coefficient obtained using Least Squares (Sec 3) to the correlations given by Mills [22], Incropera [6] and Wakao [32] assuming that the noise in the measured temperatures was normally distributed. The results presented are based upon an improved model of the bed heat losses and differ from those in [11]. Figure 3 compares the values of \( h \) estimated as described in Section 3.4 with those predicted using Mills’ correlation showing that the agreement is within 20% but with a slightly stronger effect of the Reynolds number. The cause of the stronger effect of \( Re \) is not clear, but we suspect that it is likely due to the transient heating and cooling combined with the effect of heat transfer to the enclosing glass cylinder.

The packed bed correlations, [22, Eq. 4.131], [6, Eq. 7.72], [32] are respectively

\[
N_u = 0.5 \left( Re_M^{0.5} + 0.2 Re_M^{2/3} \right) Pr^{1/3} \left( \frac{\mu_s}{\mu_b} \right) -0.14 \text{ with } Re_M = \frac{\rho_a U_{ave} d}{(1 - f) \mu_a} \tag{4a}
\]

\[
f \, jh = 2.06 Re^{-0.573} \text{ with } Re = \frac{\rho_a U_{ave} d}{\mu_a} \tag{4b}
\]

\[
N_u = 2.0 + 1.1 Pr^{1/3} Re^{0.6} \text{ with } Re = \frac{\rho_a U_{ave} d}{\mu_a} \tag{4c}
\]

\[
N_u = 1.2 Re_M^{0.728} \text{ fitted values} \tag{4d}
\]
in which the correlation, Eq. 4a, is reported to be accurate to ±25% [34]. Over the range of Reynolds numbers used, Wakao’s correlation, Eq.4c is uniformly about 5% higher than that of Mills, Eq.4a. That of Incropera, Eq.4b, ranges from approximately twice at low Reynolds numbers to just slightly less at high values with an average value being approximately 20% higher. The coefficients of the fit, Eq.4d, have standard deviations of 0.225 and 0.020 respectively which are comparable with those reported by [30]. The correlations are for steady state. Kuwahara [19] suggested that Wakao’s correlation needed some adjustment for the transient cases.

![Figure 3. Estimated values of h (left) and compared to Mills’ correlation (right) (5a)](image)

\[ \hat{\Theta} = (S^T S)^{-1} S^T D \]  
\[ \sigma^2 = \frac{1}{n - N_\theta} (D - M(\Theta))^T (D - M(\Theta)) \]  
\[ \text{cov}(\Theta) = \sigma(\epsilon) (S^T S)^{-1} \]  

where \( S \) is the sensitivity of the model to the parameters, \( dM(\Theta)/d\Theta \) and \( \sigma(\epsilon) \) is the measurement noise and \( N_\theta \) is the number of parameters.
What we really want to know is the probability distribution of the parameters that we are searching for, not just point values, i.e., the most probable value. To do this we make use of Bayes’ relation [24]

\[
p(\theta|D) = \frac{p(D|\theta)\pi(\theta)}{p(D)} = \frac{p(D|\theta)\pi(\theta)}{\int p(D|\theta)\pi(\theta)d\theta}
\]  

where \( D \) represents data (experimental results), \( p(D|\theta) \) is the likelihood of obtaining the data given values of the parameters \( \theta \) and \( \pi(\theta) \) represents our prior knowledge (belief) about the parameters. The difficulty with Eq. 6 is the evaluation of the denominator (often termed the ‘evidence’) which must be evaluated if \( p(\theta|D) \) is to be a proper probability distribution, i.e, the area under the curve sums to one. When several parameters are to be evaluated, the evaluation of the denominator is numerically difficult and often very computationally expensive.

For linear problems with the assumption of no prior information, i.e., \( \pi(\theta) \) is non-informative, the posterior distribution, \( p(\theta|D) \) will be normal. For other priors or for non-linear models, the posterior can be non-normal and estimating the uncertainty in \( \theta \) is difficult. It is not uncommon to make use of Laplace’s approximation [29] that relates the degree of uncertainty to the Hessian of the model.

3.1. MCMC

One approach to resolving the problem of evaluating the denominator and the need to know good initial values when using least squares is to use the Markov Chain Monte Carlo (MCMC) method. In this approach, one starts with some values of the parameters, \( \theta_0 \), evaluates \( p(D|\theta_0)\pi(\theta_0) \), then chooses a new set of parameters, \( \theta_1 \) using what is called a ‘proposal’ distribution and checks to see if the probability has increased. If it has, the new values are accepted, if not, they are accepted with some random probability. After continuing this process, usually for a great number of times, the probability \( p(\theta|D) \) is found to converge to a stable distribution. How to choose the sequence of parameters, i.e., the proposal distribution, is not trivial. Typically it is either a random walk or a Gaussian step. The sequence of choices is called a ‘chain’, see [3] for a useful discussion of MCMC.

We used the MCMC program DREAM [31] which employs an adaptive parallel multi-chain method and assumed that the data are normally distributed to estimate the heat transfer coefficient and the specific heat. Typical chain lengths after the burn-in, i.e., number of posterior samples, is NL, where N is the number of parallel chains and L the length of each chain. For low Reynolds numbers the burn in involved about 5000 evaluations of the model while for high Reynolds numbers it was of the order of 2500. Acceptance rates were of the order of 15% and 20% respectively. In both cases it took about 20000 model evaluations to get well defined probability distributions of the parameters. The number of evaluations depends upon how accurately one wants to estimate the posterior distributions. It took only about 20% of these values if one were satisfied with an estimate of the parameters and their standard deviation. A reduction of about 50% can be achieved if one starts with initial values taken from the MLE solutions.

3.2. Variational Bayesian Inference (VB)

Variational Bayesian methods are a family of techniques that are popular in Machine Learning. They allow one to re-write the statistical inference problems as an optimization problem, thus solving parameter estimation problems using modern optimization methods. The goal of variational Bayes is to represent \( p(D) \) and \( p(\theta|D) \) in simpler forms.
The most commonly used variational algorithm is based upon the mean field approach in which the posterior distribution of the parameters, $\theta$, is represented by

$$q(\theta) = \prod_i q_i(\theta_i)$$

(7)

Despite the computational convenience of the mean-field approximation it is not appropriate when there are strong dependencies between the different $\theta$. The alternate choice is the weighted Gaussian distribution

$$q(\theta) = \sum_{i=1}^{N_0} w_i N(\theta | \mu_i, \Sigma_i) = \sum_{i=1}^{N_0} w_i N(\theta | \mu_i, \sigma_i I)$$

(8)

where $w_i$ represent weights and the latter form assumes isotropic covariances (see [12]). Such mixtures can model reasonably complex, even multi-modal, distributions.

The evidence, $p(D)$ is represented as

$$\log(p(D)) = F(q(\theta)) + KL(q(\theta), p(\theta|D))$$

(9a)

where

$$F(q(\theta)) = E_q[\log \frac{p(D, \theta)}{q(\theta)}]$$

(9b)

and

$$KL(q(\theta), p(\theta|D)) = E_q[\log \frac{q(\theta)}{p(\theta|D)}]$$

(9c)

where $E_q$ is the expectation over the assumed distribution, $q(\theta)$. The KL term measures the dissimilarity, i.e., the divergence, between $q(\theta)$ and $p(\theta|D)$. The more similar they are the closer KL approaches zero. The KL-divergence is always positive and equals zero only when the two distributions are equal. $F(q(\theta))$ is termed the free energy and due to the non-negativity of the KL-divergence, the free energy is always less than or equal to $p(D)$ and thus is a lower bound to it. In other words, the VB method aims to find a variational distribution that is as similar as possible to the posterior distribution, which is equivalent to maximizing the free energy. Thus the VB method has converted the estimation of parameters into a maximization problem which may be simpler than the numerical evaluation of the denominator in Eq.6.

The Variational Bayesian approach can yield results equivalent to MCMC with a computational expense that is several orders of magnitude less. Particularly good explanations of VB can be found in References [14, 26, 12].

Maximizing $F(q)$ with the constraints posed by the prior probabilities is often not a trivial numerical problem. Typically some form of Newton’s method based upon the Hessian or its approximation is used, e.g., BFGS or L-BFGS [21].

3.3. Comparison of Methods
The most common method for parameter estimation is least squares. Conclusions about the uncertainty in the estimated parameters and the degree of correlation are valid only if
the residuals are normally distributed. The Variational Bayes method can utilize different distributions, Eq.8. On the other hand, the MCMC approach does not require one to make assumptions about the behavior of the residuals.

To compare the different parameter estimation methods consider the temperature distribution in a slab of thickness $L$ with generation rate $Q$, surface heat transfer coefficient $h$ and conductivity $k$ that satisfies

$$T - T_\infty = \frac{Q}{h} L + \frac{QL}{2k}(L^2 - x^2)$$

(10)

Data were simulated by adding normally distributed noise with a standard deviation equal to 10% of $T(0) - T(L)$. We chose an error set for which the residuals were close to being normally distributed to give an idea of the best performance and most comparable performance of the methods. The results presented in Table 1 are in nondimensional form, e.g., $h^* = h_{est}/h_{actual}$. The VB results were obtained using the mean field approximation using the VBA toolbox [9] and following the algorithm suggested by Gershman [12]. Since Eq.10 is non-linear in $k$ and $h$, but linear in $1/k$, $1/h$, the computations were done for both representations as indicated in the table. The VB approach showed a strong effect on the estimated uncertainties of $k$, but the estimated values of $k$ and $h$ are only slightly changed while the number of model evaluations was halved because of the linearity of the model as also seen in the MLE case.

Table 1. Fitting $T - T_\infty = \frac{Q}{h} L + \frac{QL}{2k}(L^2 - x^2)$

|       | $k^*$  | $\sigma(k^*)$ | $h^*$  | $\sigma(h^*)$ | Correlation | Modeling | # model evaluations |
|-------|--------|----------------|--------|----------------|-------------|----------|---------------------|
| MLE   | 0.938  | 0.109         | 1.023  | 0.024          | -0.89       | k,h      | 4                   |
| MLE   | 0.938  | 0.108         | 1.023  | 0.023          | -0.89       | 1/k,1/h  | 2                   |
| MCMC  | 0.964  | 0.110         | 1.020  | 0.023          | -0.88       | k,h      | 100                 |
| MCMC  | 0.935  | 0.114         | 1.024  | 0.022          | -0.86       | 1/k, 1/h | 100                 |
| VB    | 1.006  | 0.255         | 1.001  | 0.095          |             | k,h      | 66                  |
| VB    | 0.932  | 0.550         | 1.023  | 0.099          |             | 1/k, 1/h | 32                  |

Note, that the estimates of $k$ and $h$ are highly correlated, clearly an undesirable result. In fact, it is not possible to estimate all three parameters, $Q$, $k$, $h$, simultaneously because the sensitivities are so strongly related.

3.4. Estimated Packed Bed Parameters

Table 2 lists typical results obtained using the three methods. The estimated values of $h$ and $c_s$ from MCMC and from least squares (MLE) differ by less than 1%. The estimates of uncertainty from MLE are about one-half of the MCMC values. See Section 4 for further discussion. Figure 4 shows characteristic pdfs for $h$ and $c_s$ and it is clear that they are close to Gaussian, explaining the close agreement of the two methods. The success of the MLE depends upon these distributions being close to normal. The uncertainties obtained with MCMC come directly from the computations. On the other hand, those for MLE depend upon an analysis of the residuals. If the pdf of the parameters is a reasonably multivariate normal distribution (namely that the bounds of the priors do not significantly truncate $p(\theta|D)$), then one can use the quadratic approximation [29, pg 52] in which the negative of the inverse of the Hessian of
the likelihood combined with an estimate of the standard deviation of the measurement noise, \( \epsilon \), is the covariance matrix of the estimated parameters (see Eq. 13). If \( \sigma(\epsilon) \) is not known, then one must find a way to separate \( \epsilon \) from \( \delta \). We discuss this further in Section 4.

Figures 5 and 6 compare the measured and predicted responses at each of the thermocouple locations for the lowest and highest Reynolds numbers and we see that the discrepancy gradually increases as a function of thermocouple position. The predicted temperatures agree well at early times, but display increasing uncertainty as time increases, regardless of the thermocouple position.

### Table 2. Estimates of the Heat Transfer Coefficient and Heat Capacity

|                  | Re=900          | Re=7000         | Typical Model evaluations |
|------------------|-----------------|-----------------|---------------------------|
|                  | \( h \) | \( \sigma(h) \) | \( c_s \) | \( \sigma(c_s) \) | \( h \) | \( \sigma(h) \) | \( c_s \) | \( \sigma(c_s) \) |             |
| Heating          |                |                 |                |                |                |                 |                |                |             |
| MLE              | 60.1 | 0.09 | 618.1 | 0.18 | 254.7 | 1.09 | 641.5 | 0.73 | 10           |
| MCMC             | 60.4 | 0.19 | 618.0 | 0.36 | 254.8 | 2.17 | 641.5 | 1.49 | 20000        |
| VB               | 60.4 | 0.01 | 618.0 | 0.32 | 254.7 | 0.14 | 641.5 | 1.19 | 400          |
| Cooling          |                |                 |                |                |                |                 |                |                |             |
| MLE              | 60.4 | 0.08 | 570.5 | 0.05 | 252.7 | 1.06 | 697.4 | 2.74 | 10           |
| MCMC             | 60.1 | 0.19 | 570.5 | 0.85 | 252.7 | 2.10 | 697.4 | 6.19 | 20000        |
| VB               | 60.0 | 0.01 | 570.5 | 0.37 | 252.7 | 0.14 | 697.4 | 3.92 | 400          |

**Figure 4.** Probability Distributions for \( h \) and \( c_s \) compared to Equivalent Gaussian Distribution (generated by the MCMC software [31])
3.5. Discrepancy
Figures 5 and 6 show the measured temperatures and those predicted by the MCMC approach and the region of uncertainty based on one standard deviation. The predicted temperatures agree with those obtained using parameters obtained from the MLE analysis. We see that the difference between the data and the prediction, i.e., the discrepancy, and the uncertainty, as indicated by the shaded region, gradually increases as we move up the packed bed and that it is worse for low Reynolds numbers. We are interested in determining if there is any correlation between the discrepancies. If so, it might be possible to construct an approximate form of $\delta(t, X)$ that will improve our model and yield more realistic estimates of the uncertainty in the estimated parameters.

4. Gaussian Process-Kriging
Figure 7 compares the residuals at low and high Reynolds numbers. It is clear that the residuals do not have the desired properties to enable us to separate the effect of the model errors from...
the experimental noise to estimate the uncertainty in the estimated properties. The estimated uncertainties given in Table 2 are unrealistically small, an effect of the large amount of data. Recall that in the Least Squares approach uncertainties are inversely proportional to the square root of the number of readings if the noise is uncorrelated. For our data the residuals, Fig.7 clearly provide no information about the degree of correlation. Furthermore, we are unlikely to be able to express the residuals in a closed form that would represent the discrepancy function. Naturally the best approach would be to consider a more complex model which would include radial conduction with consideration of the wall behavior and local velocity variations. As noted in the introduction, this is not a realistic approach. In particular it would be difficult to determine which effect was the cause of the behavior shown in the figure. Instead, we are interested in seeing if there is any statistical relationship between the residuals and in particular if we could extract the measurement noise by

$$\epsilon = res - \delta_e$$

where $\delta_e$ is a smoothed representation of the residual.

We create $\delta_e$ using a method known as Gaussian Process. For example, let a response be a function of two variables, $y(\theta_1, \theta_2)$, and consider that over a range of these variables that the responses at individual points are closely related to each other, i.e., there exists a statistical relationship between the responses. Kriging, developed by the South African engineer D. G. Krige [8, 33], assumes that the response and covariance can be expressed as

$$y(s) = \mu(s) + \delta(s)$$

$$\text{var}(\delta(s_1) - \delta(s_2)) = 2\gamma(s_1 - s_2)$$

where $s$ represents a specific point $\theta_1, \theta_2$, $\mu(s)$ represents a large scale trend and $2\gamma$ is termed the variogram. $\delta(s)$ is a zero mean stationary random process. The variogram would either be chosen from a number of simple forms or derived from the difference between measured data and $\mu(s)$. In this way, values of $y(s)$ could be obtained efficiently by interpolation.

![Figure 7. Residuals for Re=240 and 7300](image-url)

In 2001 Kennedy and O’Hagan [18] used this idea to represent the variability of the computed response of $y(\theta_1, \theta_2)$. They proposed that an expensive-to-compute simulation model could be used to compute $y$ at a defined set of points, $s = \theta_1, \theta_2$, and $\delta(s)$ would represent the response at points nearby, i.e., an inexpensive interpolation model to account for the approximation errors associated with the surrogate model. In contrast to ‘kriging’, their model was based on a Gaussian distribution, the so called ‘squared exponential’ where $\sigma$ represents the length scale of correlation between the data values.
\[
exp\left(-\frac{1}{2} \frac{|\mathbf{s}_1 - \mathbf{s}_2|^2}{\sigma^2}\right)
\]  

(12)

where \( |\mathbf{s}_1 - \mathbf{s}_2| \) is the distance between \( \mathbf{s}_1 \) and \( \mathbf{s}_2 \). See [25] for a useful description of the method.

In principle, the method is easy to apply, particularly for the small number of data points usually employed. For a large number of values, as exists in the residuals which are evaluated as shown in Figure 7, the derived matrices are very poorly conditioned and special techniques are needed. Fortunately, there are free versions of the algorithms available on the web that run under MATLAB [35, 36] and MATLAB itself has a kriging toolbox [21], all of which can handle a very large number of values, of the order of thousands. [However, these different programs may give slightly different results, but usually with inconsequential differences, based upon the specific optimization algorithm used.]

Figure 8 displays the result of kriging the residuals associated with one of the thermocouples. Using these smoothed functions (also known as ‘trend lines’ in kriging) to estimate \( \epsilon \) we find that this estimated noise has a zero mean and is uncorrelated and its standard deviation is consistent with experiments conducted to evaluate the noise of the data acquisition system, \( \sigma(da_q) = 0.04 \text{C} \).

The uncertainties listed in Table 2 for MLE were computed based on Laplace’s approximation [29] and assuming an isotropic noise, \( \sigma(noise) \), using

\[
cov(\hat{\Theta}) = \sigma(noise)^2 (S^T S)^{-1}
\]  

(13)

where \( S \) is the sensitivity of the model to the parameters, \( \Theta \), and \( \sigma(noise) \) is the combined effect of data acquisition noise, \( \sigma(da_q) = 0.04 \text{C} \), and the calibration noise, \( \sigma(calibration) = 0.24 \text{C} \).

![Figure 8. Residuals based on a Gaussian Process](image)

5. Conclusions

Because the posterior distributions, Figure 4, are close to normal both the MLE and VB estimates are in good agreement with the MCMC values. Clearly both of these approaches are superior to MCMC in terms of the number of evaluations of the model needed. The MCMC required of the order of 20000 evaluations of the model, of which about half were in the burn in period and thus discarded. In contrast, the MLE needed in the order of 10 evaluations, and the VB about 400 (about 100 if only the heating or cooling data, but not both, were used). However,
if the posterior were multi-modal or non-normal then, 1) while the MLE might identify the most likely value of the parameter, its estimate of the covariance will be significantly in error, 2) the VB approach will need a different form of $q(\theta)$ and a more complicated expression for $F(q)$.

In terms of the estimated parameters, all three methods give equivalent results. The MCMC method yields uncertainty estimates without any input from the analyst, in contrast to MLE. While the uncertainties vary, they refer only to the confidence in the estimate and do not reflect the inherent uncertainty associated with experiment. Table 3 gives relevant information for the set of experiments.

We had hoped that an analysis of the residuals might give us a way to estimate the discrepancy, $\delta(t, TC)$. Clearly from Figure 7 the discrepancy is a strong function of the Reynolds number and thermocouple position, but when using the Gaussian process to analyze them, we were unable to determine any features that would enable us to form any formulaic representation of $\delta(t, TC)$. As a consequence, it is not possible to separate $\delta$ and $\epsilon$ and to get reasonable estimates of the uncertainties of the estimated parameters as shown by the unrealistic values listed in Table 2.

Table 3. Uncertainties of Experimental Results based on Data from Figure 3

|        | $\sigma(\hat{h}-h_{\text{Mills}})$ | $\sigma(\hat{h}-h_{\text{trend}})$ | $\sigma(\hat{c}_s)/\text{mean}(c_s)$ |
|--------|---------------------------------|---------------------------------|---------------------------------|
| Heating| 0.15                            | 0.14                            | 0.12                            |
| Cooling| 0.08                            | 0.06                            | 0.08                            |
| Both   | 0.12                            | 0.11                            | 0.11                            |

Regardless of the model discrepancy, the uncertainties listed in Table 3 are all of the order of 10%. The estimates $\hat{h}$ behaved the same for both the heating and cooling analyses. In contrast, most estimates of the specific heat gained from the cooling data were smaller than that from the heating data and the standard deviations are much larger, Table 2. We attribute this to the initial temperatures for each phase. When heating, all the marble temperatures are initially constant and equal to the ambient temperature. For the cooling phase, the upper section of the bed was still heating while the lower section was cooling, introducing a sizeable transient condition. In some experiments, heating was carried out long enough that the entire bed came to a constant temperature and the differences in the specific heat was less than 1%.

References

[1] Achenbach E 1995 Review: Heat and Flow Characteristics of Packed Beds Experimental Thermal and Fluid Science 10 pp 17-27
[2] Adebiyi G A, Nsofor E C, Steele W G and Jalalzadeh-Azar A A 1998 Parametric Study on the Operating Efficiencies of a Packed Bed for High-Temperature Sensible Energy Storage ASME Solar Energy Engrg. 120 pp 2-13
[3] Andrieu C, de Freitas N, Doucet A and Jordan M I 2003 An Introduction to MCMC for Machine Learning Machine Learning 50 pp 5-43
[4] Augier, F, Idoux F and Delenje J Y 2010 "Numerical simulations of transfer and transport properties inside packed beds of spherical particles" Chemical Engineering Science 65 1055-1064
[5] Behnam M, Dixon A G, Nijemeisland M and Stitt E H 2013 A New Approach to Fixed Bed Radial Heat Transfer Modeling Using Velocity Fields from Computational Fluid Dynamics Simulations Ind. Eng. Chem. Res. 52 pp 15244-15261
[6] Bergman T L, Lavine A S, Incropera F P and Dewitt D P 2011 Introduction to Heat Transfer, 6th Ed. J. Wiley and Sons, NY, page 461
[7] Cascetta M, Cau G, Puddu P and Serra F 2015 A study of a packed-bed thermal energy storage device: test rig, experimental and numerical results Energy Procedia 81 pp 987-994
[8] Cressie N A C 1993 *Statistics for Spatial Date* J. Wiley and Sons, NY
[9] Daunizeau J, Adam V and Rigoux L, 2014 VBA: a probabilistic treatment of nonlinear models for neurobiological and behavioural data. PLoS Comp Biol 10 e1003441
[10] Dixon A G and Njimeisland M 2001 CFD as a Design Tool for Fixed-Bed Reactors *Ind. Eng. Chem. Res* 40 pp 5246-5254
[11] Emery A F and Sien S-I 2017 Characterizing Uncertainties in a Thermal Storage Packed Bed Using Bayesian Inference with MCMC and Variational Approximations *Proc. 2nd Thermal and Fluids Engineering Summer Conference TFESC/IWHT-2017-17670 Las Vegas, NV
[12] Gershman S J, Hoffman M D and Blei D M 2012 Non-parametric Variational Inference *Proc. 29th International Conference on Machine Learning* Edinburgh, Scotland, UK
[13] Gillespie B M, Crandall E D and Cargerry J J 1968 Local and Average Interphase Heat Transfer Coefficients in a Randomly Packed Bed of Spheres *AIChE J.* 14 pp 483-490
[14] Gilmour J 2011 An Introduction to Bayesian Inference via Variational Approximations *Political Analysis* 19 pp 32-47
[15] Guardo A, Coussirat M, Larrayoz M A, Recasens F and Egusquiza E 2005 Influence of the Turbulence Model in CFD Modeling of Wall-to-Fluid Heat Transfer in Packed Beds *Chem. Engng. Science* 60 pp 1733-1742
[16] Guo X and Dai, R 2010 Numerical Simulation of flow and heat transfer in a random packed bed" *Particuology* 8 293-299
[17] Hansen P C, Pereyra V and Scherer G 2012 *Least Squares Data Fitting with Applications*, Project Muse Johns Hopkins University Press
[18] Kennedy M C and O’Hagan A 2001 Bayesian calibration of computer models *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 pp 425-464
[19] Kuwahara F, Shirota M and Nakayama A, 2001, A numerical study of interfacial convective heat transfer coefficient in two-energy equation model for convection in porous media, *Int. J. Heat and Mass Transfer* 44 pp 1153-1159
[20] Magniço P 2003 Hydrodynamic and Transport Properties of Packed Beds in Small Tube-to-Sphere Diameter Ratio: Pore Scale Simulation Using an Eulerian and a Lagrangian Approach *Chem. Engng. Science* 58 pp 5005-5024
[21] MATLAB www.mathworks.com, Natick, MA
[22] Mills A T 1999 *Basic Heat and Mass Transfer* Prentice-Hall NJ page 323
[23] Nsofor E C and Adebiyi G A 2001 Measurements of the Gas-Particle Convective Heat Transfer Coefficient in a Packed Bed for High-Temperature Energy Storage *Exp. Thermal Fluid Science* 24 pp 1-9
[24] O’Hagan A and Forster J 1994 *Bayesian Inference* Oxford University Press New York
[25] Ormerod J T and Wand M P 2010 Explaining Variational Approximations *American Statistician* 64 pp 140-152
[26] Pei Wen Li, Jon Van Lew, Wafaa K, Cho Lik Chan, Stephens J and OBrien J E 2011 "Transient Heat Transfer and Energy Transport in Packed Bed Thermal Storage Systems" *Developments in Heat Transfer* Dr Marco Aurelio Dos Santos Bernardes (Ed.) InTech DOI: 10.5772/20979. Available from: http://www.intechopen.com/books/developments-in-heat-transfer/transient-heat-transfer-and-energy-transport-in-packed-bed-thermal-storage-systems
[27] Riaz M 0000 "Transient Analysis of Packed Bed Thermal Storage Systems" *Solar Energy* 21 123-128
[28] Sivia D S 1997 *Data Analysis, A Bayesian Tutorial* Oxford University Press New York
[29] Thakur, N S, Saini J S and Solanki S C, 2003, Heat Transfer and friction factor correlations for packed bed solar air heater for a low porosity system *Solar Energy* 74 pp 319-329
[30] Vrugt J 2016 Markov Chain Monte Carlo Simulation using the DREAM Software Package: Theory, Concepts and MATLAB Implementation *Environmental Modeling and Simulation* 75 pp 273-316
[31] Wakao N and Kaguei S 1982 *Heat and Mass Transfer in Packed Beds* Gordon and Breach Science Publ.
[32] Webber R and Oliver M A 2007 *Geostatistics for Environmental Scientists* J. Wiley and Sons, NY
[33] Whithaker S 1972 "Forced Convection Heat Transfer Correlations for Flow in Pipes, Past Flat Plates, Single Cylinders, Single Spheres, and for Flow in Packed Beds and Tube Bundles" *AIChE Journal* 18 pp 361-371
[34] www.gaussianprocess.org last accessed 2/6/2017
[35] http://www.sumo.intec.ugent.be/ooDACE last accessed 2/6/2017