An Example of Bayesian Inference in Thermal Sciences

C Balaji and T Thaseem

This article explains the method of Bayesian inference in the estimation of parameters. An account of how Bayesian inference can be adapted in a simple case of parameter retrieval in thermal sciences is explained.


(left) C Balaji is currently a Professor in the Department of Mechanical Engineering at IIT Madras. His research interests include heat transfer, atmospheric sciences, optimization, inverse problems and satellite meteorology.

(right) T Thaseem completed his MTech in IIT Madras in 2008 in the area of heat transfer. He is currently pursuing a PhD at the University of Minnesota, Minneapolis, USA.

Estimation of parameters is an important inverse problem in thermal sciences. If x is the independent variable and y is the dependent variable, determining y given x is known as the direct problem or the forward problem. However, if y is known or measured, determining the cause x that led to y is known as the inverse problem. In thermal sciences, the parameter (x) to be estimated can be the specific heat of a solid, thermal conductivity of a solid, emissivity of a surface or heat transfer coefficient at a solid–fluid interface. The dependent variable y is invariably temperature in thermal sciences. It could also be a velocity in fluid dynamics problems, concentration in problems in chemical sciences, reflectivity in optics problems, etc. In all these examples, the parameter estimation problem can be posed as an optimization problem, wherein some sort of a least square minimization is done on a residue or error.

Estimation techniques can be broadly classified as *Deterministic methods and Stochastic methods.*

Deterministic methods are invariably calculus based, while stochastic methods by definition involve some probabilistic rules associated with the procedure. Generally, multi-parameter problems are ill-posed (i.e., they do not have a unique solution) and stochastic methods fare better in terms of robustness. By robustness, we mean the ability to give reasonably accurate results even when the input data has measurement errors, i.e., data is noisy.

Keywords
Bayesian inference, parameter estimation, noise.

However, they require generation of a large amount of data for the forward calculations (in the parlance of parameter retrievals, the direct problem in also known as the forward model). Hence, stochastic techniques are normally data driven.

The widely-used stochastic techniques include Genetic Algorithms (GA), Bayesian inference, Simulated Annealing (SA) and Artificial Neural Networks (ANN). Genetic Algorithms (GA) work on the basis of natural selection and mimic the process of biological evolution. Probabilistic rules are used to choose parents from an initial set of solutions to mate and produce offsprings (new solutions). As the iterations progress, the average fitness of the population as a whole increases, and all candidate solutions will be close to the final solution (parameter set) we are seeking. Simulated annealing is a search technique where probabilistic rules are employed to decide whether a subsequent iterate is acceptable. The procedure mimics the process of annealing where we deliberately slow down the cooling process in order to obtain the final material structure of interest. The ANN method tries to mimic the learning process of the human brain and works by training a large set of forward calculations in order to correlate the output against the input. In the retrieval or testing phase, the measured output is compared with the trained data base to 'extract' the parameters of interest. In ANN, often GA, SA and sometimes even Bayesian are used to obtain the weights. The output in an ANN is typically a weighted average of the inputs and this is similar to performing non-linear regression.

Bayesian inference is again a stochastic method of estimating parameters using what is known as a statistical inference. Herein, evidence or observations or measurements are used to update or to newly infer the probability that a hypothesis may be true. The name Bayesian inference comes from the frequent use of Bayes's theorem in the inference process. See *Box* 1.

Box 1. Thomas Bayes

English theologian and mathematician Thomas Bayes (1702–1761) has greatly contributed to the field of probability and statistics. His ideas have generated lot of controversy and debate among statisticians over the years. He is known to have published two works: *Divine Benevolence*, or an *Attempt to Prove That the Principal End of the Divine Providence and Government is the Happiness of His Creatures* (1731), and *An Introduction to the Doctrine of Fluxions, and a Defence of the Mathematicians Against the Objections of the Author of the Analyst*, in his lifetime. Throughout his life, Bayes was very interested in the field of mathematics, more specifically, the area of probability and statistics. Bayes is believed to be the first to use probability inductively. He also established a mathematical basis for probability inference.

Bayesian Inference

In what follows, we will look at the use of Bayes's theorem to develop a technique for the estimation of parameters. This is known as Bayesian inference, which is one of the emerging techniques in solving parameter estimation problems. This method is based on using probability to represent all forms of uncertainty in the problem. The Bayes's theorem to relate the experimental data y and the parameters x for continuous random variables can be formally stated as follows:

$$P(x/y) = P(y/x) * P(x)/P(y), \tag{1}$$

where $P(x/y)$ is the posterior probability density function, $P(y/x)$ is the likelihood function or the forward model, $P(x)$ is the prior probability density function, and $P(y)$ is the normalizing constant.

The posterior probability density function (PPDF) is the probability that the variable(s) x caused the measurement vector y, while $P(y/x)$ is called the likelihood function.

The likelihood function is a concept that is of fundamental importance. For a given y, the joint probability density is given by $P(y/x)$ which in effect can be considered to be a function of x, because the data y is the same for all values of x and is also known. The likelihood function is a point function. In a typical problem in thermal sciences, y is invariably a set of temperatures and x is a parameter to be estimated or retrieved, say, emissivity or specific heat or thermal conductivity. The likelihood function is arrived at by computing the temperatures for assumed value of the parameter(s) and then evaluating the 'error' (mostly in a least squares sense) with respect to the measured temperatures. This 'error' is represented as a probability density function $P(y/x)$ by employing a suitable form, say a Gaussian or log-normal of the distribution. The $P(y/x)$ thus generated is known as the likelihood function. The solution to a forward problem will give us $y(x)$. However, because of measurement errors and because of the fact that none of the sampled values of x may exactly coincide with the final solution, $P(y/x)$ will generally be less than 1. It is instructive to mention here that $P(y/x)$ will be more for those values of x that are close to the solution (i.e., the true value of x). If the sample size (for example the number of temperature measurements) is large, then the likelihood function becomes very sharp and usually has a clear peak. The value of x at which the maximum of the likelihood function occurs is known as the Maximum Likelihood Estimate (MLE), a very potent estimator. However, the MLE is arrived at without any consideration of the 'prior' and is akin to least squares regression. Under certain conditions, it can be shown that the MLE approximately minimizes the variance.

$P(x)$ represents our prior knowledge of x. Stated explicitly, $P(x)$ represents our knowledge about x even before the measurements are made. The prior represents the subjective part of the Bayesian and can be extremely valuable when one is confronted with high-dimensional problems. A high-dimensional problem, by definition, is one in which the number of parameters is much larger compared to the size of the data vector. Such problems are mathematically ill-posed as different combinations of the parameters could, in principle, give rise to the same set of observations. Priors partially mitigate this ill-posedness by providing not only bounds on the parameters but also the differential preference to trial values of the parameters.

Priors can be broadly classified into 'objective' and 'subjective priors'. Another classification is 'informative' and 'non-informative' priors. If in a Bayesian retrieval, all trial values of the parameter to be retrieved are given equal preference or weightage, the prior becomes a 'uniform' prior. When we employ such a prior, the shape of the PPDF will be decided by the likelihood function alone. Such a prior will also, according to our classification given above, be known as an objective, non-informative prior.

A subjective but usually informative and useful prior is the Gaussian or normal prior. For example, if the parameter to be retrieved is specific heat of a solid and if we know that the material is a metal, then one can calculate the mean and standard deviation of specific heat of all metals from, say, a handbook and this will most likely serve as a good prior to help us obtain a sharper PPDF. A sharper PPDF also means that the variance in the estimated parameters is lower. If the confidence in a prior for some problem is low, then the prior automatically becomes weak in view of its large variance and the parameter estimation will be driven by the data alone. This is frequently referred to as the 'washing of the prior by the data' by statisticians. By the same token, a grossly incorrect prior when specified with a low variance may completely derail the estimation, even when the data vector is large and reliable.

Application to a Heat Transfer Problem

Now, we will turn our attention to an engineering problem, more specifically a heat transfer problem. The goal of this example is to illustrate how the Bayesian inference can be used in the retrieval of parameters, an important inverse problem is science and engineering.

Consider a first order system. A 3mm thick mild steel plate having a 10cm × 15cm cross- section is suspended in quiescent air at 30° C. The plate is initially at 100° C and it now undergoes cooling. The inverse problem boils down to determining the heat transfer coefficient h (which finds a place in the definition of the time constant τ) once a transient response of temperature vs. time is available. We would now like to examine the possibility of using Bayesian inference to determine the time constant τ, once a measured temperature–time history is available.

The first step is to convert the general Bayes's theorem to a form suitable for use in an inverse problem like the one discussed above.

Before doing this we will look at the heat transfer problem under consideration, which is known as the forward problem in the parlance of parameter estimation.

$$mC_{\mathrm{p}}\frac{\mathrm{d}T}{\mathrm{d}t} = -hA(T - T_\infty)\,, \tag{2}$$

where m: mass of the plate (kg), C_{p}: specific heat (kJ/kg), h: heat transfer coefficient (W/m^{2}K), A: surface area (m^2), T_∞: ambient temperature (K), T: temperature (K) of the plate, t: time (s).

Let us define a temperature excess θ as follows

$$\theta = T - T_\infty\ , \tag{3}$$

$$mC_{\mathrm{p}}\frac{\mathrm{d}\theta}{\mathrm{d}t} = -hA\theta\ . \tag{4}$$

On integration with the initial condition that at t=0, $\theta = \theta_{\mathrm{i}}$, we have

$$\frac{\theta}{\theta_{\mathrm{i}}} = e^{-t/\tau}\,, \tag{5}$$

where τ is the time constant of the system given by mC_{p}/hA. Denoting $\theta/\theta_{\mathrm{i}}$ as ϕ, we have

$$\phi = e^{-t/\tau}\ . \tag{6}$$

If we were to conduct an experiment we will obtain the transient response of the plate, i.e., $\phi\ (t)$. The inverse problem is the determination of τ once t is available. It is evident from (3) that for every value of t, we can retrieve or obtain a value of τ. Not all of these will be the same, consequent upon inbuilt errors. Hence, one has to obtain τ in such a way that $\sum(\phi_{\mathrm{meas}}(t) - \phi_{\mathrm{simulated}}(t))^2$ is minimized. We will now look at the modified form of the Bayes's theorem for use in this problem.

Bayesian Retrieval Algorithm

The Bayesian retrieval algorithm uses a database of pre-calculated temperature profiles for many possible values of the time constant and integrates them over the points in the database with the Bayes's theorem. Bayesian inversion methods formally add priors (discussed earlier) to that provided by the measurements to obtain a well-posed retrieval and corresponding uncertainty estimate [2]. Bayes's theorem (equation (1)) can be stated mathematically for the retrieval problem, in the following form for convenience:

$$p_{\mathrm{post}}(x|y) = \frac{p_{\mathrm{f}}(y|x)p_{\mathrm{pr}}(x)}{\int p_{\mathrm{r}}(y|x)p_{\mathrm{pr}}(x)\mathrm{d}x}\,, \tag{7}$$

where x represents the state vector (τ) and y represents the vector of observations (temperatures at various values of time). $p_{\mathrm{p}}(x)$ is the prior probability density function (PDF) of the state x, $p_{\mathrm{f}}(y|x)$ is the conditional probability density function of the measurements given the state vector, and $\mathrm{p}_{\mathrm{post}}(x|y)$ is the posterior probability density function of the state vector. Equation (7) is the mathematical representation of the Bayes's theorem as explained earlier. Because we are referring to a continuous rather than a discrete case, the $\sum$ sign usually seen in the definition of Bayes's theorem is replaced by the $\int$ sign.

The retrieved parameter x_{ret} is calculated by integrating over the posterior PDF in order to determine the mean state:

$$x_{\mathrm{ret}} = \frac{\int x p_{\mathrm{f}}(y|x)p_{\mathrm{pr}}(x)\mathrm{d}x}{\int p_{\mathrm{r}}(y|x)p_{\mathrm{pr}}(x)\mathrm{d}x} \tag{8}$$

This is a Monte Carlo integration (*Box* 2) because the database points are chosen randomly.

In practice, this integral is often replaced by a sum over the cases in the database. One advantage of the Bayesian framework is that the uncertainties in retrieved parameters are naturally defined by the variance of the posterior PDF:

$$\sigma_x^2 = \frac{\int (x - x_{\mathrm{ret}})^2 p_{\mathrm{f}}(y|x)p_{\mathrm{pr}}(x)\mathrm{d}x}{\int p_{\mathrm{r}}(y|x)p_{\mathrm{pr}}(x)\mathrm{d}x}\,, \tag{9}$$

Box 2. Monte Carlo Integration

The Monte Carlo method solves a problem by generating suitable random numbers and determining the fraction of the numbers that obey some properties. The method is useful for obtaining numerical solutions to problems which are too complicated to solve analytically. It was named by S Ulam, who in 1946 became the first mathematician to dignify this approach with a name, in honor of a relative having a propensity to gamble (Hoffman, *The Man Who Loved Only Numbers: The Story of Paul Erdos and the Search for Mathematial Truth*, Hyperion, New York, p.239, 1998). Nicolas Metropolis also made important contributions to the development of such methods. In order to integrate a function over a complicated domain D, Monte Carlo integration picks random points over some simple domain D' which is a superset of D, checks whether each point is within D, and estimates the area of D' (volume, n-dimensional content, etc.) as the area of D' multiplied by the fraction of points falling within D'.

where σ_x is the one sigma error bar in x and x_{ret} is the mean x of the posterior PDF. Another popular estimate is the maximum *a posteriori* (MAP) which is the value of the parameter x at which the PPDF becomes a maximum. In this study, we have used the mean also known as expectation to report the results of the parameter estimation process.

The conditional PDF $p_{\mathrm{f}}(y|x)$ or the likelihood function is the probability density of the measurement vector given a state vector. The forward PDFs of the measurement vector are assumed to be normally distributed about the simulated vector for each observation.

$$p_{\mathrm{f}}(y/x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\varphi_{\mathrm{meas}} - \varphi_{\mathrm{simulated}})^2}{2\sigma^2}\right). \tag{10}$$

It is assumed that the uncertainty σ is due to the measurement errors. Since the conditional distribution is effectively zero if the measurement vector is far from the observed one, the Bayesian algorithm interpolates between those points in the database that are reasonable matches to the observations.

As already discussed, the prior probability density function is a way of introducing other known information about the parameters. In addition, the prior distribution gives a good indication of how to choose the samples or cases for the database. In fact, by choosing the database points in parameter space x_{i} distributed according to the prior PDF, the Monte Carlo integration simplifies to

$$x_{\mathrm{ret}} = \frac{\sum_{\mathrm{i}} x_{\mathrm{i}}\, p_{\mathrm{f}}\left(y|\, x_{\mathrm{i}}\right)}{\sum_{\mathrm{i}} p_{\mathrm{f}}\left(y|\, x_{\mathrm{i}}\right)} \quad . \tag{11}$$

The sums in the Bayesian retrieval algorithm span all points in the database. Including all the database points wastes computation time in calculating the conditional PDF for points where the PDF is effectively zero (i.e., where the measurement is far from that of the database case). The conditional PDF is the product of exponentials, and thus, is also the exponential of χ^2:

$$x_{\mathrm{ret}} = \frac{\sum_{\mathrm{i}} x_{\mathrm{i}} \ \exp\left[-(1/2)\chi_{\mathrm{i}}^2\right]}{\sum_{\mathrm{i}} \exp\left[-(1/2)\chi_{\mathrm{i}}^2\right]}; \quad x_{\mathrm{i}} \text{ from } p_{\mathrm{p}}(x), \tag{12}$$

where χ^2 is a measure of the difference between the measurement vector and database simulated vector.

$$\chi^2 = \sum (\phi_{\mathrm{meas}} - \phi_{\mathrm{simulated}})^2/\sigma^2. \tag{13}$$

For issues related to alternative approaches for Bayesian inference and speeding up of the algorithm, the reader is advised to consult [1-5]. For the problem under consideration, first the forward problem is solved with an assumed value of the heat transfer coefficient h as 5 W/m^{2}K resulting in a τ of 1060 s. The measurement error σ is assumed to be 0.1, which is reasonable for the system under consideration. For the purposes of algorithm testing, it would suffice if we determine a temperature time history of the plate with this value of time constant for a time interval, say 0 to 2500 s. This would serve as the 'measured' temperature history. Once the basic procedure is in place, Gaussian noise may be added to the temperatures and the effect of noise on the retrievals can be studied.

For the purpose of demonstrating the Bayesian retrieval for this problem, temperature-time profiles were simulated for 11 values of the time constant ranging from 500s to 1500s in steps of 100s. *Figure* 1 shows some representative plots along with the 'measured' data.

Table 1 gives the results obtained by using the formulation discussed above.

Therefore $\tau = 948.46/0.8874 = 1068.8$s. From *Table* 1, it is clear that the retrieved value of the time constant τ (1068.8) is very close to the actual value of 1060 s. Column 4 gives the likelihood function. It can be seen that whenever the assumed value of τ is close to the true value, the likelihood function has a high value, as expected.

It can be seen that the Bayesian inference gives a weighted mean of all the samples, with higher weights being assigned to those samples that match the observations more closely. Furthermore, with just 10 simulated profiles, and no special *a priori*

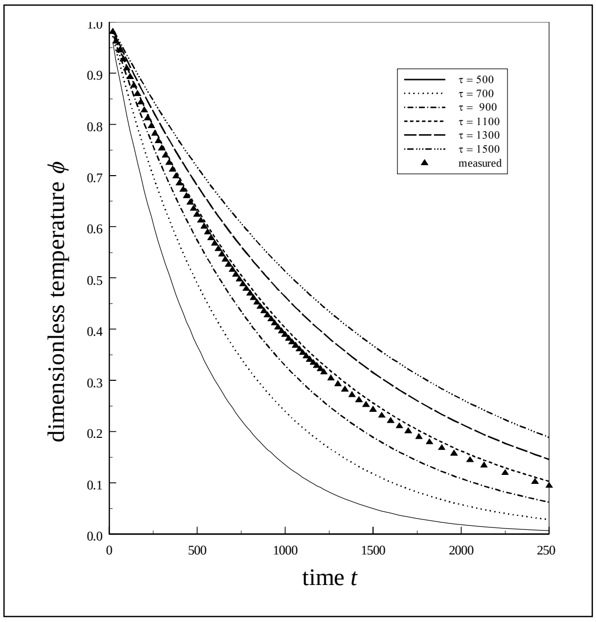


Figure 1. Simulated temperature profiles for the problem of cooling of the first order system. Also shown is the 'measured' profile (without the addition of the Gaussian noise).

distribution modeling, we are able to retrieve the parameter of interest with good accuracy.

As mentioned earlier, column 4 of *Table* 1 shows the Posterior Probability Density Function (PPDF) of various values of the time constant τ for the set of measurements θ vs. t. It will be instructive to look at the variation of the posterior PDF

S.No	τ_i, s	χ^2	$\exp(-\chi^2/2)$	$\tau_i \exp(-\chi^2/2)$
1	500	382.2	-	-
2	600	228.9	-	-
3	700	124.9	-	-
4	800	58.4	-	-
5	900	19.9	4.82 E-05	0.0433
6	1000	2.52	0.283	283.1
7	1100	1.02	0.601	661.2
8	1200	11.35	3.43E-03	4.12
9	1300	30.49	2.39E-07	3.1E-04
10	1400	56.1	-	-
11	1500	86.5	-	-
		Σ	0.8874	948.46

Table 1. Numerical results for the example problem. The normalizing factor $1/\sqrt{2\pi}$ that appears in the original formulation has been omitted from both the denominator and numerator for convenience).

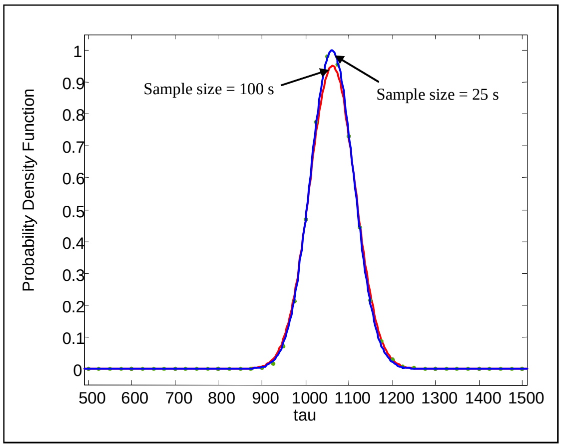


***Figure 2** Posterior PDF against τ for 2 different sampling sizes of τ : 100s and 25s.*

with respect to the sample size. In the table, 11 values of τ have been used. Suppose we were to use, say, 41 values of τ, i.e., τ is in steps of 25 s starting from 500 s and going up to 1500 s; one would expect the PDF to become taller and sharper. *Figure* 2 confirms this.

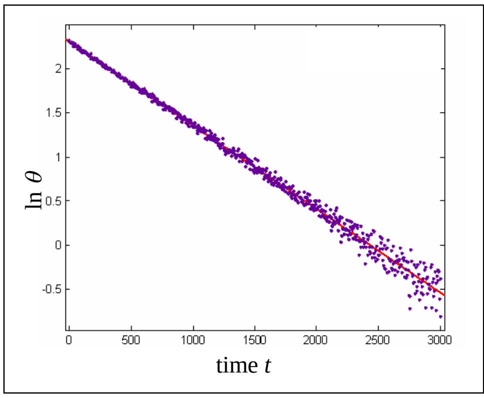
From *Figure* 2, it is clear that when the sampling size for τ obtained with step size of 25 s is used, the PPDF obtained is taller and sharper than when a step size of 100 s is used.

Retrieval Using Linear Regression

Naturally, a question might be raised on why the method of least squares cannot be used in the cooling case discussed above. It seems perfectly logical to plot log (θ) against time t. The inverse of the slope should give the value of the retrieved parameter (time constant in this case). In the problem discussed above, we can use linear regression analysis to retrieve the parameter. Rewriting (6), and taking logarithms on both the sides, we get the following equation:

$$\ln\theta = \ln\theta_{\mathrm{i}} - t/\tau. \tag{14}$$

Now, if ln (θ) is plotted against time t, the reciprocal of the slope of the curve gives us an estimate of τ (See *Figure* 3). From the figure, it is clear that a good plot is possible and the retrieved parameter is much closer to the actual value. We get a value of 1057 s for the time constant. In this case by taking the logarithm a linear equation is obtained and this results in an accurate estimate of the parameter. This method gives accurate results when the error is on the lower side. When the noise in measurement measurement exceeds 5%, the parameter retrieved using this method deviates from the real answer. Hence the Bayesian



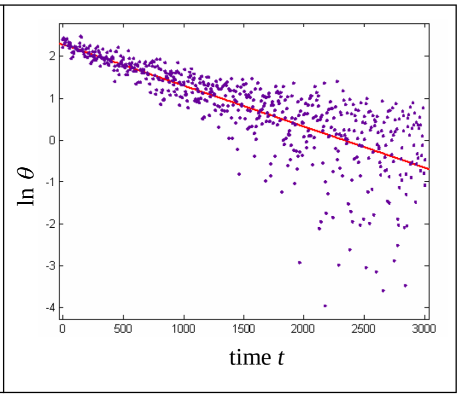


Figure 3 (left). Parameter retrieval using linear regression of the first order system discussed above (noise not included).

Figure 4 (right). Parameter retrieval using linear regression of the first order system discussed in the text (noise included).

scores over this method, as a measurement error of 5% is common. This is an empirical observation based on the results for a few representative values of the experimental errors.

Furthermore, in the case where more than one parameter is to be retrieved,the least squares method cannot be used. The disadvantage in this method can be clearly seen in *Figure* 4. The method actually fails when the error in measurement is high.

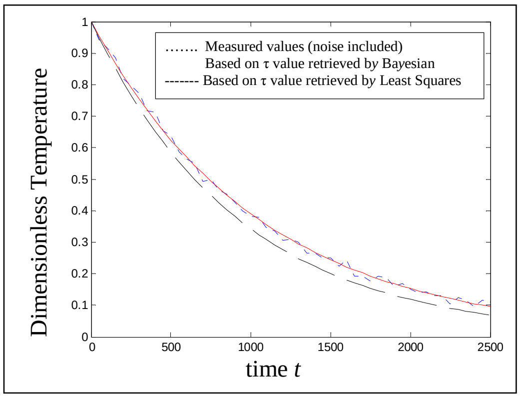
Retrieval Using Linear Regression, with Noise in the Measured Values

From *Figure* 4, it is evident that this method fails to retrieve the parameter. In fact it was observed that this method failed when the error was above 0.05. The noise introduced was based on the error in measurement and hence the measurement error had a direct bearing on the noise. Furthermore, the retrieved value of the time constant τ was 933 s for an initial temperature difference of 50 K.

Retrieval Using Bayesian, with Noise in the Measured Values

Bayesian inference was again used to retrieve the parameter after the addition of noise (normal distribution) on the measured values. The retrieved value of the time constant is used to plot the dimensionless temperature against time in *Figure* 5. Bayesian can thus be used to successfully retrieve the time constant even when error is introduced in the measured values. Using Bayesian retrieval, the time constant τ turned out to be 1050 s which is fairly accurate. Hence,

Figure 5. Temperature profiles plotted using the τ values retrieved by Bayesian and Least Squares. Also, shown is the 'measured' profile with the associated noise in the measurements.



Bayesian inference is robust and works well even when the input data is quite noisy.

Important questions that can arise here are: *Where is the randomness? It looks like non-linear least squares*! The answer to these questions is that the randomness is inbuilt in our choice of (i) computing for times 0–2500 s and evaluating the 'error' at 50 s time interval and (ii) the choice of τ values for which we ran our forward calculations. The latter is known as 'sampling' and is a subject of serious study and research in Bayesian statistics. In a sense, we have employed a crude *a priori* and arbitrary sampling, so to speak. Even so, as the problem we have considered was simple, our choices did not affect the final results.

However, as the problem gets more complex and multiple variables are involved, the number of forward profiles required for a satisfactory retrieval will increase substantially. Furthermore, a careful consideration of constructing the *a priori* distribution may be essential.

Suggested Reading

[1] S A McFarlane, K F Evans and A S Ackerman, A Bayesian Algorithm for the Retrieval of Liquid Water Cloud Properties from Microwave Radiometer and Millimeter Radar Data, *J. Geophys. Res*., Vol.107, (D16), p.4317, 2002.

[2] D S Sivia, *Data Analysis: A Bayesian Tutorial*, Clarendon, Oxford, England, 1996.

[3] J M Bernardo and A F M Smith, *Bayesian Theory*, Wiley, New York, 1994.

[4] M Deiveegan, C Balaji, and S P Venkateshan, Comparison of Various Methods for Simultaneous Retrieval of Surface Emissivities and Gas Properties in Gray Participating Media, *Journal of Heat Transfer*, Vol.128, pp. 829–837, August 2006.

[5] M Deiveegan, V Swaminathan, C Balaji, and S P Venkateshan, A Bayesian Algorithm for the Retrieval of Geophysical Parameter in the Atmosphere, *13th International Heat Transfer Conference*, Sydney, Australia, 2006.

Address for Correspondence: C Balaji and T Thaseem, Department of Mechanical Engineering, Indian Institute of Technology Madras, Chennai 600 036, India. Email: balaji@iitm.ac.in