Bayesian Parameter Inference of Explosive Yields Using Markov Chain Monte Carlo Techniques

John Burkhardt [a*]

[a]*United States Naval Academy, Annapolis, Maryland 21402, USA*

Abstract

A Bayesian parameter inference problem is conducted to estimate the explosive yield of the first atomic explosion at Trinity in New Mexico. The first of its kind, the study advances understanding of fireball dynamics and provides an improved method for the determination of explosive yield. Using fireball radius-time data taken from archival film footage of the explosion and a physical model for the expansion characteristics of the resulting fireball, a yield estimate is made. Bayesian results from the Markov chain indicate that the estimated parameters are consistent with previous calculation except for the critical parameter that modifies the independent time variable. This unique result finds that this parameter deviates in a statistically significant way from previous predictions. Use of the Bayesian parameter estimates computed is found to greatly improve the ability of the fireball model to predict the observed data. In addition, parameter correlations are computed from the Markov chain and discussed. As a result, the method used increases basic understanding of fireball dynamics and provides an improved method for the determination of explosive yields.

* Corresponding author.
E-mail address:

1. Introduction

G.I. Taylor famously estimated the classified yield, E, of the 1945 nuclear explosion at Trinity, New Mexico, using nothing but images publicly released by the U.S. Army into the open media [1] (Fig. 1).
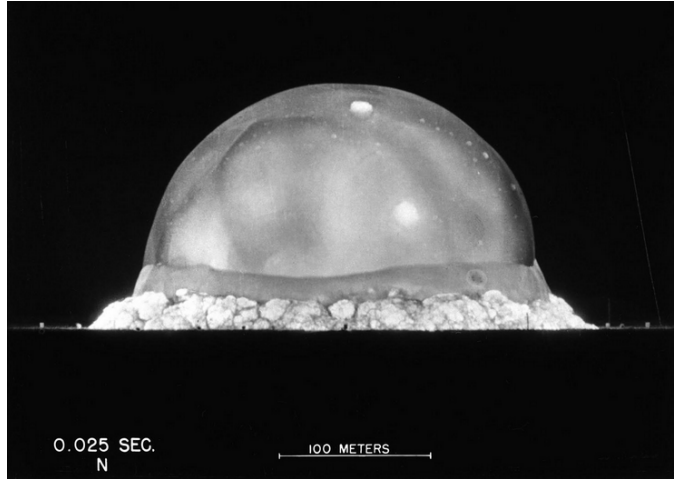

Fig. 1. A single movie frame showing the fireball created by the first nuclear explosion at Trinity, New Mexico, including time stamp and length scale.

Published with both a time stamp and length scale, Taylor was able to use the images, in conjunction with his theoretical time-dependent model of the hydrodynamics fireball radius [2, 3], to provide an accurate estimate of the weapon's explosive yield of 16.8 kT. His hydrodynamic model predicted the fireball radius, R, as a function of time, t, in terms of a non-dimensional constant he computed, S, the estimated undisturbed density of air at Trinity, ρ_0, and the unknown explosive yield, E,

$$R(t; E) = St^{2/5}E^{1/5}\rho_0^{-1/5} \tag{1}$$

This result, independently presented by other authors [4] established the unique dependence of the fireball's radius on the 2/5th power of time. The approach pioneered by Taylor to estimate explosive yield has since been applied numerous times and validated using independent measurements of explosive yields.

Taylor's, and all subsequent analyses, considered the time-radius model to depend on only one parameter, E, assuming fixed values for the exponents and model parameters S and ρ_0. In contrast to this one-parameter approach, the current study considers all of the parameter in equation 1, other than the measured values R, t and ρ_0, as unknown parameters to be determined. In addition, this study employs a Bayesian approach to nonlinear regression for determination of the unknown parameters and provides a detailed Bayesian analysis of the Trinity data using Markov chain Monte Carlo (MCMC) techniques. These results are compared with the original model, equation 1.

Ultimately, it is the objective of this Bayesian analysis is to determine the joint probability density function (pdf) of these free parameters, most specifically yield, E, so as to better match and interpret the experimental data available. Such an analysis will offer greater insight into the parameters values as well as their interdependence. The specific definitions chosen for these parameters are defined in equation 2,

$$R(t) = St^a E^b \rho_0^{-c} \tag{2}$$

Well-known nonlinear regression techniques [5, 6] based on frequentist statistics exist are capable of determining these unknown parameters using the measured time-radius data. The frequentist approach, however, provides only a limited amount of information about the parameters. This is because traditional regression analysis assumes the parameters to be fixed, exact values, while the measured data are random variables dependent on these parameters. While useful, this statistical model of the data provides only limited information about the parameters. Specifically, this approach provides directly point estimates of the parameters. Higher order statistical moments of the parameters, such as standard deviations, allowing for the generation of confidence intervals, can also be computed but require further assumptions about the distribution of error in the data. In addition, no information is typically returned about inter-parameter relationships, such as statistical correlations, which would lend greater insight into the parameters and their effects on the fireball.

In contrast, Bayesian regression assumes that both the parameters and the measured radii are random variables. In this study, the statistical model used for the data $\boldsymbol{R}_i$ assumes this randomness results from an additive noise term whose statistics can be determined by the Bayesian analysis. Consequently, the Bayesian model requires a sixth parameter to characterize the statistics of R, to which we will assign the random variable ϵ. For convenience the model parameters will be grouped in the parameter set $\boldsymbol{\theta} = \{S, E, a, b, c, \epsilon\}$.

Assuming that the parameters are random variables, rather than fixed quantities, allows us now to ask directly about their statistics, not just mean values and confidence intervals. Specifically, we are now considering the joint statistics of the parameters, θ, conditioned on the measured data set, R. This is significant because the parameters' resulting joint pdf contains all of the parameters' statistical information including all higher order statistical moments and correlations.

The objectives of this analysis are to provide improved estimates of the parameters which impact the fireball radius model (equation 2) and their interdependencies, through which an improved understanding of fireball dynamics can be developed as well as an improved method for the determination of explosive yields. Model parameter estimates and their complete statistics will be determined using a Markov chain Monte Carlo (MCMC) simulation whose output provides an estimate of the parameter's joint pdf. MCMC simulations provide this estimate by individually sampling numerous parameter sets from the space of all possible parameter combinations in proportion to their likelihood of occurrence using Bayes Theorem.

2. Nonlinear Regression

Parameter inference, and regression analysis in particular, represents a core application of statistical science [7, 8, 9]. Its development using a frequentist approach is described below and, importantly, its similarities and differences with Bayesian regression are discussed. Despite their similarities, the differences between frequentist and Bayesian regression are significant and important to understand.

The objective of all regression analysis is, given a system model with unknown parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)$, $\boldsymbol{\Theta} \subset \mathbb{R}^p$, and often noisy observations, is to determine estimates of the model parameters, $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_p)$, as well as statistical information about their expected variation and possible interdependence. This primarily occurs for one of two reasons: to develop a model relationship between a dependent variable $\boldsymbol{y} = (y_1 ... y_n) \subset \mathbb{R}^n$, and an independent variable, $\boldsymbol{x} = (x_1 ... x_n), x \subset \mathbb{R}^n$, for the purpose of predicting system behavior for other, unobserved, dependent variable values, or; to gain insight into the characteristics of the observed system through interpretation of the determined parameter values and their statistical characteristics.

The inference of the parameters $\boldsymbol{\theta}$ using regression techniques is based on either a frequentist or Bayesian statistical foundation [10]. In the frequentist approach probabilities are thought of as limiting frequencies based on a hypothetical, infinite or asymptotic sample. In contrast, the Bayesian approach views probabilities as

subjective and are interpreted as conditioned on the information available [11]. In either approach, often the primary aim is to provide point estimates or "best guesses" for the unknown parameters, $\widehat{\boldsymbol{\theta}}$, and corresponding interval estimates that provide reasonable ranges for the unknown parameters given the data.

While having similar objectives, the frequentist and Bayesian approaches to regression are fundamentally dissimilar. The frequentist framework for regression assumes the model parameters, $\boldsymbol{\theta}$, are unknown but fixed, nonrandom quantities. Consequently, no probabilistic conclusions can be drawn about frequentist parameters. The estimated parameter values are still uncertain but this uncertainty arises from the random nature of the observations, $\boldsymbol{y}$. Probabilistically, the frequentist approach models the observations, $\boldsymbol{y}$, as elements of a random variable Y, dependent on the model parameters $\boldsymbol{\theta}$, a relationship which can be defined by the conditional probability density function, $p(\boldsymbol{y}|\boldsymbol{\theta})$. While useful, this statistical model of the data is limited in the information it can provide. Treating the unknown parameters as fixed constants masks much of the information about their variability and interdependence.

In contrast to the frequentist approach, Bayesian regression assumes that both the parameters, $\boldsymbol{\theta}$, and the measured observations, $\boldsymbol{y}$, are random variables. This assumption allows us to naturally ask directly about the statistics of the parameters, not point and interval estimates. Statistically, we are now considering the joint statistics of the parameters, $\boldsymbol{\theta}$, conditioned on the observations, $\boldsymbol{y}$, represented by the joint conditional probability density function, $p(\boldsymbol{\theta}|\boldsymbol{y})$. This is significant because the probability density function., $p(\boldsymbol{\theta}|\boldsymbol{y})$, contains all of the parameters' statistical information including all higher order statistical moments and interdependences.

Several formulations of both the frequentist and Bayesian approach to regression are available, however, the most encountered techniques employ likelihood functions. Likelihood functions are functions of the model parameters, given specific observed data, that measure the plausibility that a system with the specified parameters generated the observed data. It is functionally identical to the frequentist description of the data, $p(\boldsymbol{y}|\boldsymbol{\theta})$, with the important difference that the likelihood, L, is a function of the random variable $\boldsymbol{\theta}$ not the observed data $\boldsymbol{y}$. To emphasize this dependence the likelihood is written as an explicit function of $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta})$$

or, for both computational and mathematical convenience, as the log likelihood,

$$l(\boldsymbol{\theta}) = log\big(L(\boldsymbol{\theta})\big)$$

Development of the likelihood function is identical for both the frequentist and Bayesian approaches. Consider a nonlinear regression model

$$y_i = g(x_i, \boldsymbol{\theta}) + e_i \quad i = 1 \dots n$$

where g is a known function of the known, nonrandom, independent variable, $\boldsymbol{x} = (x_1, \dots, x_n)$ and the random model parameter $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)$. In addition, $\boldsymbol{y} = (y, \dots, y_n)$ is the dependent random variable and $\boldsymbol{e} = (e_1, \dots, e_n)$ is the random error associated with the observations, $\boldsymbol{y}$. If the random errors are assumed independent and identically distributed (i.i.d.) as Normal, $N(0, \sigma^2)$

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\left\{ -\frac{\big(Y_i - g(x_i, \boldsymbol{\theta})\big)^2}{2\sigma^2} \right\} = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n exp\left\{ -\sum_{i=1}^{n} \frac{\big(Y_i - g(x_i, \boldsymbol{\theta})\big)^2}{2\sigma^2} \right\}$$

and the log likelihood is,

$$l(\boldsymbol{\theta}) = -nlog(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - g(x_i, \boldsymbol{\theta})\right)^2 \tag{3}$$

From equation 3 it can be seen that maximizing the likelihood is the same as minimizing the residual square error or $RSE(\boldsymbol{\theta})$,

$$RSE(\boldsymbol{\theta}) = \sum_{i=1}^{n}\left(Y_i - g(x_i, \boldsymbol{\theta})\right)^2 \tag{4}$$

So, under the i.i.d. Normality assumption, the frequentist maximum likelihood estimator (*MLE*) is the same as the frequentist least square estimator (*LS*). Depending on the form g takes values for the best estimate, $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{MLE} = \hat{\boldsymbol{\theta}}_{LS}$, can then be determined analytically or numerically. For a Bayesian regression the same likelihood function is minimized (equation 3) which, as with the frequentist approach, is identical to maximizing the *RSE* (equation 4).

3. Bayesian Regression

Bayesian regression differs from maximum likelihood regression. Its objective is to determine the parameters' conditional pdf, $p(\boldsymbol{\theta}|y)$, which can be expressed using Bayes Law [13]

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} \tag{5}$$

where our desired distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, is the posterior distribution, $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood distribution, $p(\boldsymbol{\theta})$ is the prior distribution, and $p(\boldsymbol{y})$ is a normalization factor [14], effectively the marginal distribution of $\boldsymbol{y}$ computed by integrating the numerator over the model parameter space, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ [14]. The posterior distribution is significant in that it provides a complete description of the model parameter statistics conditioned on the given data $\boldsymbol{y}$ including all moments and correlations. The likelihood distribution characterizes the probability that parameter set $\boldsymbol{\theta}$ generated the observed data set $\boldsymbol{y}$. Existing knowledge about the parameters can be incorporated into the model through the prior distribution, $p(\boldsymbol{\theta})$, which characterizes the probability that a particular parameter set $\boldsymbol{\theta}$ will occur.

Analytical, closed form representations for the posterior distribution can be determined most directly when the posterior and prior are in the same probability distribution family [14]. Termed conjugate distributions or a conjugate prior, this choice is often driven by analytical convenience rather than modeling accuracy. The simplest example of conjugacy would be the scenario where the prior distribution, likelihood and therefore the posterior distribution are Normally distributed. Unfortunately, in nonlinear regression, conjugate distributions are rarely available beyond the linear model.

In most common regression situations, the posterior distribution is estimated numerically by sampling sets $\boldsymbol{\theta}_j$ from the distribution $p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ after statistical models for the likelihood and prior distributions are defined. The parameter values drawn from the joint probability density function are typically determined using a Markov Chain Monte Carlo (MCMC) simulation employing the Metropolis algorithm [15] or its variant the Metropolis-Hastings algorithm [16]. The algorithm works by exploring the parameter space, computing posterior values, while randomly sampling more and more likely parameter sets. The result is an overall movement of the parameter choices toward the peak of the distribution and an ultimate estimate of the most

likely parameters termed the maximum a-posteriori (*MAP*) parameters, $\widehat{\boldsymbol{\theta}}_{MAP}$. Essentially the mode of the distribution, the *MAP* is the parameter set most representative of the best fit parameters.

4. Markov Chain Monte Carlo

The objective of Bayesian regression is to determine the posterior distribution of the parameters $\boldsymbol{\theta}$ (equation 5). This is possible analytically in only rare cases. Consequently, it is customary to resort to numerical methods to approximate the posterior distribution. The most common method for accomplishing this is the Markov chain Monte Carlo method (MCMC) which provides point estimates of the posterior distribution through a random Markov sampling process. The method works by computing the posterior distribution at random locations in the parameter space $\boldsymbol{\Theta}$. This is accomplished by randomly selecting locations in the parameter space $\boldsymbol{\theta}_i$ and computing the values of the likelihood function and prior distribution at those locations. The corresponding value of the posterior distribution can then be simply calculated by multiplying the two together (equation 5). The result of the sampling is a list, or chain, of parameter sets $\boldsymbol{\theta}_i$ whose termination is determined by one of several convergence criteria.

MCMC sampling is done sequentially with the statistical distribution of the next parameter set $\boldsymbol{\theta}_{j+1}$ only dependent on the current value $\boldsymbol{\theta}_j$ resulting in the sequence's Markov characteristics. Generation of a Markov chain representative of the posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, is accomplished using the Metropolis-Hastings algorithm. The algorithm works by first generating a proposal parameter set, $\boldsymbol{\theta}_{j+1}$, by adding a random displacement, or jump, in the parameter space, $\boldsymbol{\Theta}$, to the current parameter set $\boldsymbol{\theta}_j$ using a proposal distribution. If the ratio of the proposed posterior value to the current posterior value is greater than the value of a uniformly distributed random number between 0 and 1, the parameter set $\boldsymbol{\theta}_{j+1}$ is added to the Markov chain. If lower, the proposed parameter set is rejected and the current parameter set $\boldsymbol{\theta}_j$ is added again to the end of the chain.

In addition to its simplicity, the Metropolis-Hastings algorithm is valuable because it avoids the need to calculate the Bayesian normalization factor in equation 5. This is because when deciding whether or not the magnitude of the posterior distribution of one parameter set is greater than another, the ratio of the posterior values is considered rather than their difference. This is significant because the normalization factor can be difficult to compute in problems with many parameters, even numerically. In addition, this allows for great flexibility when defining the likelihood and prior distributions because their normalization is not required.

5. Bayesian Parameter Estimation

A Bayesian analysis of the Trinity time-radius data used by Taylor (App. A) has been conducted to estimate the parameters in the time-radius fireball expansion model and their interdependence. In this Bayesian regression application [15, 16] the likelihood distribution is chosen to be the negative of the residual square error, $RSE(\boldsymbol{\theta})$. Specific to this application,

$$RSE(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(R_i - R(t_i, \boldsymbol{\theta}) \right)^2$$

where R_i is the measured fireball radius at time t_i and $R(t_i; \boldsymbol{\theta})$ is the radius predicted by the model, equation 2, at time t_i using parameter set $\boldsymbol{\theta}$. Because $\boldsymbol{\theta}$ is a random variable so is the residual sum of the square error, *RSE*. For this study the likelihood is assumed to be Normally distributed with zero mean and standard deviation ϵ,

$$RSE \sim N(0, \epsilon^2)$$

peaking when the fit is perfect, $S = 0$, and decreasing from there. Strictly speaking this is not a valid p.d.f. because the residual square error is never negative and therefore the distribution is not properly normalized. This is, however, acceptable because neither of the Metropolis algorithms require the normalization of the posterior distribution for the computation of the Markov chain and, consequently, neither do the parameter's pdfs.

For the prior distributions it is argued that the individual parameter values will be distributed smoothly and symmetrically about their mean value. As a result, the parameters are assumed to be distributed as a multivariate Normal distribution with mean, $\overline{\boldsymbol{\theta}}_{pr}$ where the subscript *pr* indicates prior information,

$$\overline{\boldsymbol{\theta}}_{pr} = \left\{ \overline{S}_{pr}, \overline{a}_{pr}, \overline{b}_{pr}, \overline{c}_{pr}, \overline{E}_{pr}, \overline{\epsilon}_{pr} \right\} \tag{6}$$

Because few details are known about the statistical relationship between the parameters a priori, their distributions are assumed to be statistically independent of one another and, consequently, uncorrelated, resulting in an assumed, diagonal, prior covariance matrix, Σ_{pr},

$$\Sigma_{pr} = \begin{bmatrix} \sigma^2_{prS} & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_{pra} & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{prb} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_{prc} & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_{prE} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_{pr\epsilon} \end{bmatrix} \tag{7}$$

As with the statistical model of the residual square error, *RSE*, the prior distribution is also defined for unrealistic negative values. This inconsistency is unimportant, however, for the same reason given for *RSE*.

6. Markov Chain Monte Carlo Computations

The Markov chain simulations presented were generated using MATLAB and some elementary functions from its Statistics and Machine Learning Toolbox. All computations took no more than a few minutes when performed on a standard desktop computer. The algorithms employed are straight forward and required no special software or hardware beyond that listed above. While MATLAB was used, the calculations could be easily executed using any number of other software packages or applications. The details of the Metropolis-Hastings algorithm and its implementation are discussed below.

The fundamental objective of the algorithm is to sample the model's parameter space $\boldsymbol{\Theta}$ and compute the posterior distribution. Each sample of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ drawn from the parameter space contains six values, one for each parameter in the set. Individually, the values of the parameters $\boldsymbol{\theta}_j$ drawn occur with the statistical characteristics of the underling joint conditional density function (posterior distribution) $p(\boldsymbol{\theta}|R)$. This includes the frequency with which individual values are observed as well as any correlations that exist between them. Therefore, drawing samples, $j = 1 \dots L$, provides an estimate of the underlying joint probability density. Theory predicts that as the number of samples approaches infinity, $L \to \infty$, the estimated parameter distribution will

converge to the theoretical distribution, $p(\boldsymbol{\theta}|R)$. Consequently, the result of the sampling is an estimate of $\boldsymbol{\theta}$'s six-dimensional joint pdf for the parameters S, E, a, b, c, and ϵ.

In this study, the proposal distribution is assumed to be a six-dimensional, multivariate, Normal distribution with the mean value being the current parameter set $\boldsymbol{\theta}_j$ and with covariance matrix, Σ_p, characterizing the parameter jump statistics and their correlations. As with the prior distribution, the proposal's covariance is unknown. Therefore, out of necessity, the individual parameters are assumed to be statistically independent, resulting in a diagonal proposal covariance matrix, where the subscript p indicates proposal information,

$$\Sigma_p = \begin{bmatrix} \sigma_{pS}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{pa}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{pb}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{pc}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{pE}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{p\epsilon}^2 \end{bmatrix} \tag{8}$$

This deficiency is improved upon by updating the proposal covariance adaptively using a portion of the Markov chain itself, as described below. Updating in this way results in a fully populated proposal covariance matrix during the later stage of the Markov chain. As a result, parameter proposals are more likely to be accepted, improving the efficiency of the algorithm's search for the peak of the posterior distribution.

In total, the Markov chain can be displayed as a series of staircase traces, each tracking the values of a single parameter in the chain. The result at each step in the Markov chain is a series of six numbers, one for each of the parameters. An example of the initial 6000 steps of a posterior distribution's Markov chain is shown in Fig. 2. Accepted proposal parameter sets are indicated by horizontal steps while retained current parameter sets results in vertical steps.
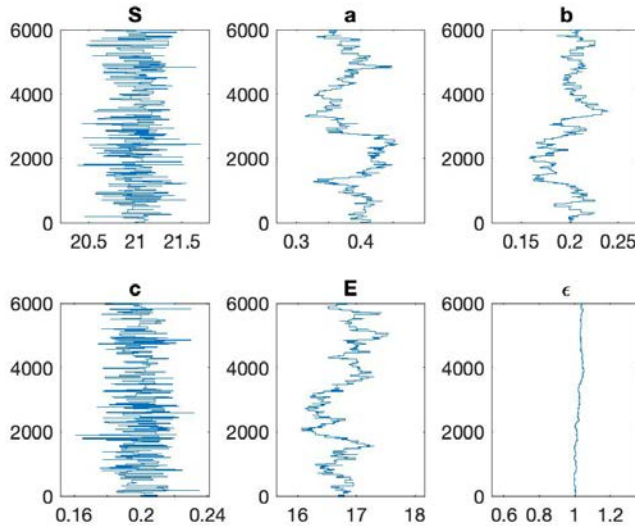

Fig. 2. Initial 6000 steps of Markov chain.

In principle, the chain shown contains all of the statistical information about the parameters including mean values and correlations. This information is very difficult to ascertain from the Markov chain directly so typically the statistics of the individual parameters are displayed using their marginal distributions in the form of histograms. The marginal distribution histograms for the 6000 step Markov chain in Fig. 2 are shown in Fig. 3.
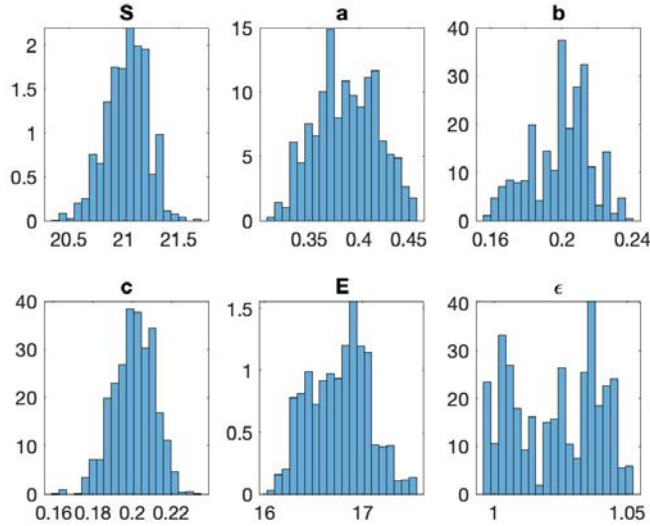

Fig. 3. Marginal distribution histograms of the 6000 step Markov chain.

Correlation information, is even more difficult to ascertain from a visual inspection of the Markov chain and is better understood through inspection of the chain's covariance and correlation matrices.

7. Metropolis Algorithm Implementation

The technique outlined above has, with great consistency, been used to estimate the posterior distribution of the parameter set, $\boldsymbol{\theta} = \{S, E, a, b, c, \epsilon\}$, and therefore Taylor's model parameters (2), using the fireball data associated with a number of nuclear explosions. To demonstrate the technique the results for the analysis of the Trinity explosion are presented. The total radius-time data set available is composed of 25 radius-time pairs (App. A).

Implementing the technique first required the selection of mean values and variances for each of the prior and proposal Normal distributions. Because of the known accuracy of Taylor's prediction, the selection of mean values for the explosion-independent parameters in the prior model, S, a, b, c, can be accurately estimated using Taylor's theory, (equations 1, 2) while an estimate of the explosive yield can use Taylor's original estimate of 16.8 kT (Table 1). The prior means have additional importance. Being our best guess for the parameter values they are used to initialize the Metropolis algorithm.

Table 1. Taylor's parameter predictions used as prior means

$\bar{S}_{pr}$	$\bar{a}_{pr}$	$\bar{b}_{pr}$	$\bar{c}_{pr}$	$\bar{E}_{pr}$	$\bar{\epsilon}_{pr}$
21	0.4	0.2	0.2	16.8	50

Unlike mean values, little is known about the parameter's variances and correlations. For this reason, the parameters are assumed initially statistically independent and therefore uncorrelated. In addition, as in many instances, prior variance values were chosen based on intuition (Table 2).

Table 2. Prior variances.

σ^2_{prS}	σ^2_{pra}	σ^2_{prb}	σ^2_{prc}	σ^2_{prE}	$\sigma^2_{pr\epsilon}$
1.0E0	1.0E-4	2.5E-5	1.0E-4	4.0E0	2.5E1

While the mean value of the proposal distribution is known, chosen as the current parameter set, the proposal variances were determined using scientific judgement and trial and error (Table 3). Realistically, each proposed parameter set should be reasonably "near" the current parameter set. For this reason, a trial set of proposal variances were chosen then varied systematically until a usable, stable, Markov chain was produced. The final proposal variances used in the MCMC simulation were,

Table 3. Proposal variances.

σ^2_{pS}	σ^2_{pa}	σ^2_{pb}	σ^2_{pc}	σ^2_{pE}	$\sigma^2_{p\epsilon}$
1.0E0	1.6E-3	4.0E-4	4.0E-4	4.0E0	1.0E0

Uncertainty in these parameters, and their effect on simulation results, was of concern. For this reason, the sensitivity of the posterior distribution to changes in the Metropolis algorithm's starting point was investigated. In the study, the Metropolis algorithm's initial parameter set was randomly perturbed a maximum of 50% from their initial values. The resulting Markov chains were recorded and shown to converge. An example from the study, for parameter *a* is shown in Fig. 4 with the red dots at step zero indicating ten randomly chosen initial parameter values. In every instance, each similarly perturbed Markov chain parameter converged to a common distribution. The dramatic change in the character of Markov chain component at step 100,000 is due to an adaptive change in the proposal covariance matrix which is discussed in more detail below.
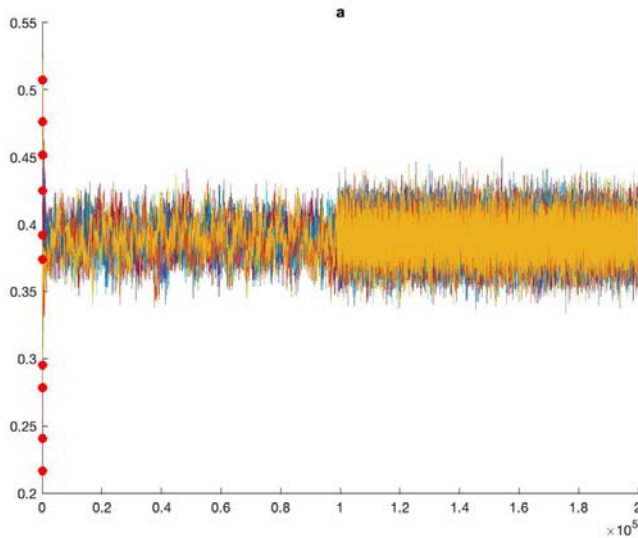

Fig. 4. Parameter convergence example.

8. Results

Using fireball radius-time data from the Trinity explosion, and the prior and proposal data defined above, an estimate of the parameter's posterior distribution was generated. To accomplish this three 200,000 step Markov chains were independently generated. During the simulation, each chain was divided into four segments. The first three segments were discarded and were only executed to improve the quality of the fourth segment which was retained for analysis.

The first segment, conservatively chosen and accounting for 25% of the chain (50,000 steps), termed "burn-in," was a series of steps within which the chain was allowed to stabilize and reach some approximate equilibrium. The second segment (tune-in), accounting for 24.5% of the chain (49,000 steps), was used to generate an estimate of the chain's covariance. The estimated covariance was then used as the Metropolis algorithm's proposal covariance during the third and fourth segments of the chain. The third segment was a short 0.5% (1000 step) delay to allow any transients in the chain resulting from the covariance switch to dissipate. The fourth segment, accounting for the remaining 50% of the chain (collection), used the estimated covariance from the tune-in segment to generate improved, correlated steps. The three separate collection segments, one from each of the chains, were then combined to produce a composite Markov chain. The combination of the individual chain segments was justified by comparing their consistency within and across each chain, confirming that each chain had converged to the same distribution. The consistency check was performed by computing the chain segments' Gelman-Rubin statistic [14], which indicated convergence.

An example of a complete 200,000 step chain showing the burn-in, tune-in and collection segments separated by dashed lines is shown in Fig. 5. Early segments in several of the parameter's chain contributions show significant variation, indicating the utility of the burn-in period, while the change in the characteristics of the Markov chain resulting from the updated covariance matrix segment are dramatic and improve the Markov chain's exploration of the parameter space.
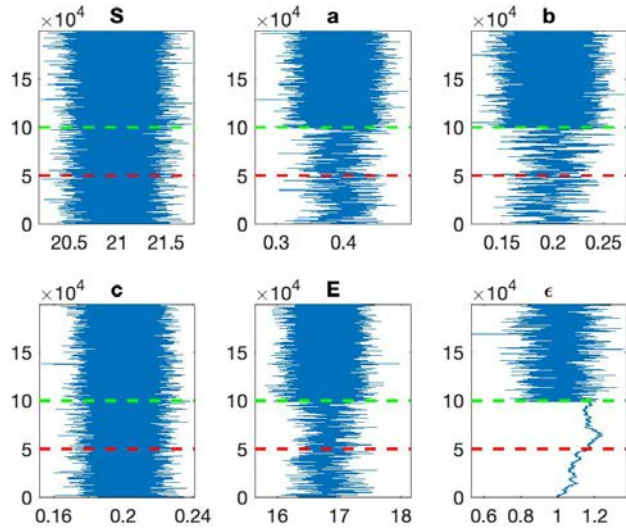

Fig. 5. One complete 200,000 step Markov chain showing burn-in, up to red dashed line, tune-in, between the red and green dashed lines, and collection, from the green dashed line until the end of the chain segment.

An example of a composite Markov chain constructed from three converged collection segments is shown in Fig. 6. Each parameter's contribution to the composite Markov chain was then used to generate their marginal

distribution histogram, also shown in Fig. 6. Superimposed in red on the marginal distribution histograms are the parameters' prior distributions.
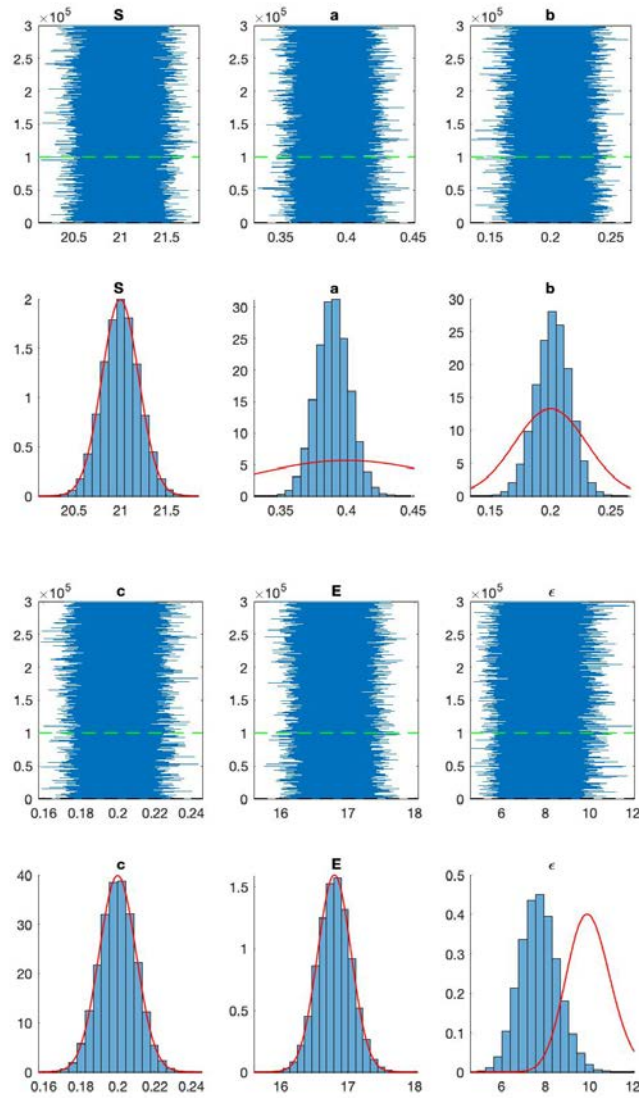

Fig. 6. Composite, 300,000 step Markov chain and their marginal distribution histograms. The red lines show the prior distribution proposed for each of the parameters.

A summary of the parameter statistics for the composite Markov chain, determined from the marginal distribution histograms, and their existing prior/estimated values, are shown in Table 4. A maximum a-posteriori (MAP) set is provided as well. The MAP set is the parameter set whose values occur in combination to maximize the posterior. Because of the large data set the 95% confidence intervals for the different parameters, determined from the Markov chain, are essentially zero to the precision shown.

Table 4. Parameter statistics

Parameter	Average	St. Dev.	MAP	Prior Estimate
S	21.0±0.000	1.99E-1	21.0	21.0
a	0.390±0.000	1.30E-2	0.391	0.4
b	0.202±0.000	1.40E-2	0.199	0.2
c	0.200±0.000	1.00E-2	0.200	0.2
E	16.8±0.00	2.51E-1	16.8	16.8
ϵ	7.72±0.000	8.74E-1	7.76	10

While mean values and confidence intervals are of interest most unique to the analysis are the covariance and correlation matrices of the composite Markov chain's parameters, examples of which are shown in Tables 5 and 6.

The covariance matrix, Table 5, is of interest because the magnitude of the off-diagonal terms quantifies the degree of dependence of one parameter on another. The magnitude of each correlation term is, however, scaled by the general magnitude of the parameters considered.

Table 5. Parameters' covariance matrix.

	S	a	b	c	E	ϵ
S	3.97E-02	-1.13E-04	-5.45E-04	4.94E-06	4.13E-04	-3.29E-04
a		1.56E-04	-1.51E-04	1.95E-08	-4.98E-05	2.53E-04
b			2.04E-04	6.77E-07	-2.17E-04	-2.71E-04
c				9.92E-05	-1.68E-05	-3.71E-05
E					6.29E-02	-1.31E-03
ϵ						7.64E-01

Table 6. Parameters' correlation matrix. The largest off-diagonal terms (>1x10-1) are shown in red.

	S	a	b	c	E	ϵ
S	1.00E+00	-4.55E-02	-1.91E-01	2.49E-03	8.29E-03	-1.89E-03
a		1.00E+00	-8.46E-01	1.57E-04	-1.59E-02	2.31E-02
b			1.00E+00	4.75E-03	-6.05E-02	-2.17E-02
c				1.00E+00	-6.73E-03	-4.26E-03
E					1.00E+00	-5.98E-03
ϵ						1.00E+00

As a result, comparison between covariance terms for the purpose of determining the relative strengths of their effects from one parameter to another is difficult. For this reason, the correlation matrix, Table 6, is considered because it more directly shows the relative influence of each parameter on the others as each correlation coefficient has been normalized. The largest off-diagonal correlation term ($>1x10^{-1}$) is shown in red. Consequently, the exponents a and b and the parameters b and S are most strongly correlated. Of particular interest is the lack of a strong correlation between E and b. Strictly speaking parameters E and b are potentially unidentifiable, in combination producing a single value. For this reason, it was expected that there would be an anti-correlation between the two parameters which was not observed.

9. Discussion

In general, the Bayesian MAP parameter values (Table 4) are consistent with Taylor's theoretical predictions (Table 1) except for the parameter a which is found to different significantly from Taylor's prediction. With a value of $a_{MAP} = 0.391$ the Bayesian estimate differs from Taylor's prediction of 0.4 by 0.009 (2.3%). While not large in absolute terms or relative to the parameter's Markov chain standard deviation σ (0.692σ) the value is estimated with considerable precision strongly supporting the observed difference.

The significance of a's deviation from Taylor's model prediction is most dramatically demonstrated by plotting the theoretical model predictions for R vs. t (equation 2) using both Taylor's parameter predictions (Table 1) and the Bayesian MAP values (Table 4) versus the observed data (App A). The resulting plot (Fig. 7) shows a good match between both models and the observed data at early times but significant deviations between Taylor's model and the observed data at late times. In contrast, the model using the MAP parameter values much more closely follows the observed data for all times. This deviation is not unexpected as the parameter a is the exponent of the independent time variable, t.
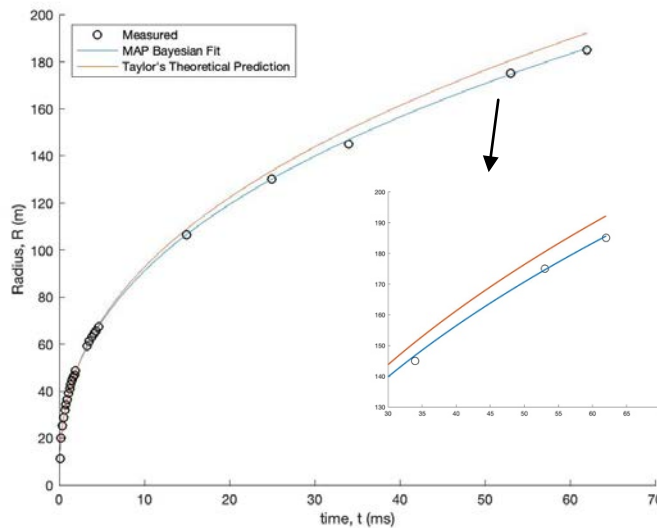

Fig. 7. Radius versus time data and model predictions showing both the MAP and mean fit (nearly indistinguishable).

This indicates the importance of the parameter a in the radius-time model and shows the dramatically improved fit to the measured data provided by the model using a_{MAP}. This difference is also significant in that the value for a, traditionally assumed to be 0.4, is the basis for the most commonly used methods for explosive yield determination. In addition, the lower value predicted for a_{MAP} is supported by the more detailed fireball expansion theory of Porzel [19] which predicts a to be 0.372.

While the point-wise MAP estimates provide insight into the optimum parameter values, Bayesian parameter variation is still present due to the expected variation in the parameter set along the Markov chain. An indication of this variability is shown in Fig. 8 where 50 radius-time model predictions, generated using 50 parameter sets chosen randomly from the composite Markov chain, are shown, Fig. 8. This plot, while showing considerable variation is centered around the MAP model prediction and observed Trinity data set analyzed.
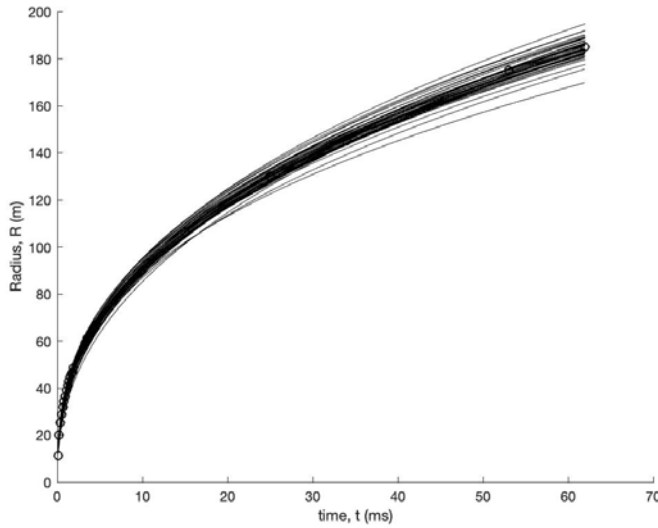
Fig. 8. Radius versus time data and model predictions for 50 random parameter sets chosen from the composite Markov chain.

Parameter's covariance and correlation matrices, as derived from the Markov chain, are presented in Table 5 and Table 6. In general, correlation coefficients are low, typically less than $\pm 10^{-2}$. Two correlation coefficients are, however, considerably larger in magnitude than the others (shown in red in Table 6). The values indicate unusually large anti-correlations between parameters b and the parameters S and a. The anti-correlation between b and S seems the less interesting of the two as the parameter b is the exponent of E (equation 2), combining to produce what is effectively a multiplicative constant, as is the parameter S. They vary inversely with respect to each other in effect canceling each other out. The more provocative anti-correlation is the one between parameters a and b. The parameter a, as previously discussed, is particularly important as the exponent of the independent time variable, t. The significance of its anti-correlation with b is, however, unclear at this time.

Also, of interest, and interpretable from the data provided, is the degree to which the data, versus the prior distribution, influences the Markov chain. This interplay can be observed by comparing each parameter's marginal distribution histogram to the assumed prior distribution. Looking at Fig. 6 it can be concluded that the posterior marginal distributions of parameters S, c and E are dominated by the prior information because of the excellent match between the MCMC parameter marginal histograms and the prior distributions (red lines). The opposite is true of a, b and ϵ where the Fig. 6 shows that the final posterior marginal distributions differ significantly from the assumed prior distribution. This indicates that the data had a strong influence on these elements of the Markov chain. That a and b are strongly influenced by the data and strongly anti-correlated is intriguing but any connection is impossible to infer given the data at hand.

10. Conclusion

The Bayesian MAP parameter estimates computed support Taylor's radius-time model prediction except for significant deviations in the parameter a. While the deviation is small, its impact is found to be significant for the prediction of fireball radii at late times. This is largely due to a's role as the exponent of the independent time variable, t. In addition, parameters a and b were found to be significantly anti-correlated and, as indicated

by the Markov chain, strongly influenced by the data set analyzed. The implication of these results in unclear and will require addition analysis to explain.

Acknowledgement

References

[1] Mack JE. U.S. Atomic Energy Commission. Semi-popular motion-picture record of the Trinity explosion. Oak Ridge, TN: Atomic Energy Commission; 1946.

[2] Taylor GI. The Formation of a Blast Wave by a Very Intense Explosion. I. Theoretical Discussion Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences 1950; 201:159-174.

[3] Taylor GI. The Formation of a Blast Wave by a Very Intense Explosion. II. The Atomic Explosion of 1945 Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences 1950; 201:175-186.

[4] Bethe HA and Fuchs K. Measurement of Nuclear Bomb Efficiency by Observation of the Ball of Fire at Early Stages Los Alamos Scientific Laboratory, Tech. Rep. LA-516; 1946.

[5] Gallant AR. Nonlinear regression. The American Statistician 1975; 29:73-81.

[6] Kass RE. Nonlinear regression analysis and its applications. Journal of the American Statistical Association 1990; 85:594-596.

[7] Casella G., and Berger R. Statistical inference. Belmont, CA: Brooks/Cole Cengage Learning; 2017.

[8] Boos D., and Stefanski L. Essential statistical inference: Theory and methods (Springer texts in statistics, v. 120). Dordrecht: Springer; 2012.

[9] Wasserman L. All of statistics: A concise course in statistical inference (Corrected second print. ed., Springer texts in statistics). New York: Springer; 2010.

[10] Fahrmeir L. Kneib T, Lang S, and Marx,B. Regression. Berlin: Springer-Verlag; 2007.

[11] Wakefield J. Bayesian and frequentist regression methods (Springer series in statistics). New York: Springer; 2013.

[12] Seber G.A, and Wild CJ. Nonlinear regression analysis. New York: John Wiley and Sons; 1989.

[13] Bernardo JM, and Smith AF. Bayesian theory. New York: John Wiley and Sons; 2009.

[14] Gelman,A, Carlin JB, Stern HS, and Rubin DB. Bayesian data analysis. Chapman and Hall/CRC; 2013.

[15] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, and Teller E. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 1953; 21, 1087.

[16] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970; 57:97-109.

[17] Box GE, and Tiao GC. Bayesian inference in statistical analysis (Vol. 40). New York: John Wiley & Sons; 2011.

[18] Denison DG, Holmes CC., Mallick BK., and Smith AF. Bayesian methods for nonlinear classification and regression (Vol. 386). New York: John Wiley and Sons; 2002.

[19] Porzel FB, Rate of Growth of Atomic Fireballs Los Alamos Scientific Laboratory, Tech. Rep. LA-1214, Unclassified, Distribution Unlimited; 1951.

Authors' Profiles

John Burkhardt is a Mechanical Engineering Professor at the U.S. Naval Academy in Annapolis, Maryland. He received his Ph.D. and M.S. degrees in Theoretical and Applied Mechanics from the University of Illinois, Urbana-Champaign. His bachelor of engineering degree in Civil Engineering was awarded by The Cooper Union in New York City.