USDOT Tier 1
University Transportation Center
on Improving Rail Transportation
Infrastructure Sustainability and Durability

Final Report UD-9

APPROXIMATE BAYESIAN COMPUTATION FOR RAILWAY TRACK GEOMETRY PARAMETER ESTIMATION

By

Grace Ashley
Graduate Student
Department of Civil and Environmental Engineering
University of Delaware
Newark, DE

and

Nii Attoh-Okine, Ph.D., P.E., F. ASCE, Snr Member IEEE
Department of Civil and Environmental Engineering
University of Delaware
Newark, DE

September 2020

Grant Number: 69A3551747132

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

ABSTRACT

The quality of track geometry is directly linked to vehicle safety, reliability and ride quality. The performance of track is therefore considerably hindered when track geometry indicators deviate from the specified and approved limits due to loads and continuous usage. Information obtained from the analysis of track geometry data can inform the prompt application of preventive and corrective maintenance measures like tamping, to increase the lifespan of the track and provide higher train speeds, optimizing track performance. Recently, there has been the application of Bayesian statistical methods in track degradation models. However, most models rely heavily on likelihood functions which are intractable. The aim of this paper is to apply Approximate Bayesian Computation (ABC), also known as the likelihood-free method, in estimating Track Quality Indices (TQIs) which are essential for track degradation modeling. ABC is applied using methods like the rejection algorithm and Markov Chain Monte Carlo (MCMC). In ABC, it is essential that summary statistics are computed from the observed data followed by the simulation of summary statistics for different parameter values. Two ABC-MCMC algorithms were used for parameter estimation in this paper. Although ABC is computationally expensive, it was successfully applied in TQI estimation in this paper.

Keywords: Railway, Track Geometry, Track Degradation, Approximate Bayesian Computation, Bayesian Statistics, Markov Chain Monte Carlo.

CONTENTS

INTRODUCTION

The modern society recognizes the railway transportation system as an important and invaluable element of its structure. Research has shown that the global railway system market is estimated to grow at a Compound annual growth rate of 5.2%, from USD 26.0 billion in 2019 to USD 35.3 billion by 2025 [1]. This growth is due to the increase in demand for passenger and goods transportation. The demand also comes with the need to transport the passengers and goods at higher speeds and in larger numbers. The speed and weight of modern trains have a direct impact on static and dynamic forces that influence track geometry and therefore, track degradation. The quality of track geometry is directly linked to vehicle safety, reliability and ride quality. The performance of track is therefore considerably hindered when track geometry indicators deviate from the specified and approved limits due to loads and continuous usage. Information obtained from the analysis of track geometry data can inform the prompt application of preventive and corrective maintenance measures like tamping, to increase the lifespan of the track and provide higher train speeds, optimizing track performance. Recently, there has been the application of Bayesian statistical methods in track degradation models. However, most models rely heavily on likelihood functions which are not available. The purpose of this paper is to demonstrate the application of Approximate Bayesian Computation (ABC), also known as the likelihood-free method, in predicting Track Quality Indices (TQIs) for track degradation modeling.

LITERATURE REVIEW

Track Geometry

The Federal Railroad Administration (FRA) Automated Track Inspection Program (ATIP) is in charge of collecting track geometry, ride quality and additional track related data on the United States rail transportation network [2]. Track geometry defects can be detected by visual inspection, in line with FRA requirements and these inspections could be performed biweekly for mainline in Class 4 and Class 5 tracks. Track geometry cars are also used to collect data on track geometry parameters and the data is stored in a Track Data Management System (TDMS) which has functions of track data manipulation, data viewing, record keeping, and route scheduling [2].
Track geometry is measured regularly in order to identify any defects that could compromise safety and good ride quality. Track Geometry may be defined as the three-dimensional geometry of track layouts and related measurements used in design, construction and maintenance of railroad tracks [1]. The information collected through track geometry measurement is used in maintenance planning, reduction of costs and reduction in the risk of emergency maintenance. The main track geometry parameters measured are Gage, Alignment, Cant/ Crosslevel, Surface/Profile and Twist. FIGURE 1 shows how the various parameters are measured.

- *Gage:* It is measured between the rail heads at right-angles to the rails five-eighths of an inch below the top of the rail head. The limits of the gage differ according to the class of track. For instance for Class 2 and 3 track, the gauge must be at least 4’ 8” and not more than 4’ 10”.
- *Alignment:* is done by using a predefined length of string line to measure along the gauge side of the reference rail. It is the distance from the midpoint of the string line to the gauge of the reference rail. The design horizontal alignment for tangent track is zero (perfect

straight line on the horizontal layout). In the United States the design horizontal alignment on the curved track is 1 inch for each degree of curvature.

- *Cant/ Crosslevel*: It is the difference in height between two rails. Maximum cant is usually regulated to control the unloading of the high rail wheels at low speeds.
- *Surface/ Profile*: The Surface/ Profile is the surface uniformity in the vertical plane. The measurement of uniformity is done using a predefined length of string line (normally the same length used in horizontal alignment) along the track. If the midpoint of the measurement has higher elevation, it is called hump deviation. On the other hand, if the midpoint has lower elevation, it is called dip deviation.
- Twist: It is the difference in crosslevel of any two points within the specific distance along the track.

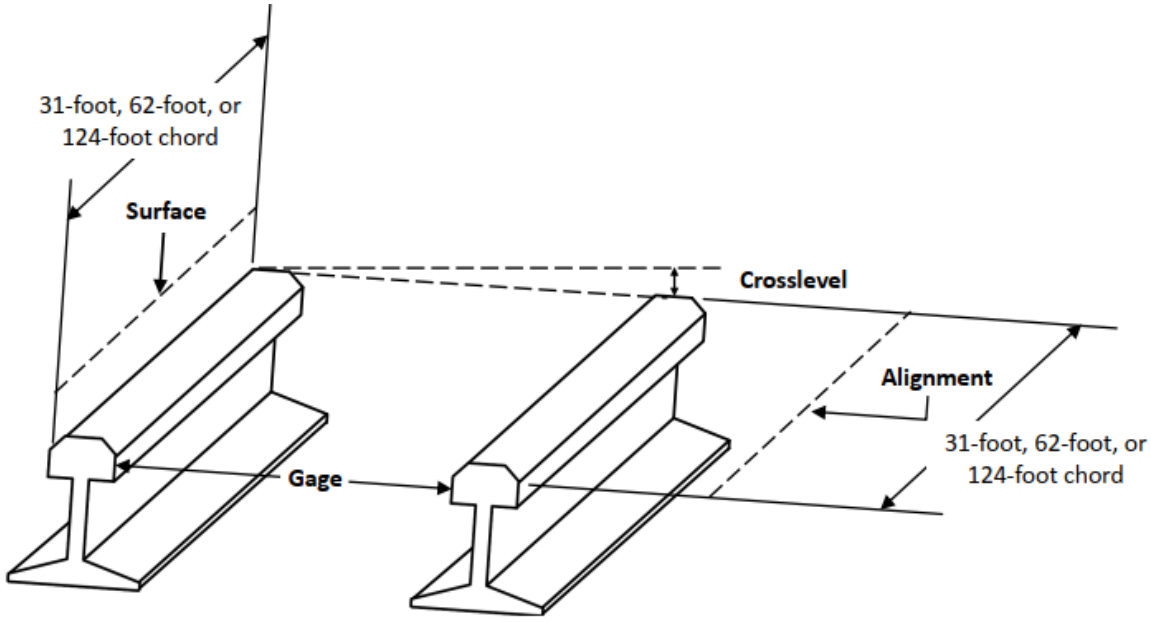

FIGURE 1 Some track geometry parameters [3]

Track Quality Index (TQI)

Track is divided into numerous short segments in order to assess track condition and also for the purpose of obtaining different statistical measures from them. The statistics obtained are then summed up and used to assess the overall quality of the track segment. The measures used to assess the track quality, particularly the deviations from the standard measures are called Track Quality Indices (TQIs). TQIs are used to assess railway track performance indicators, design interventions to improve track quality, and to compare and contrast track performances before and after the interventions are been applied [4]. There are eight main statistics used worldwide to assess the quality of track segments [4]. These statistics are presented below.

SD Index

SD index is made of different standard deviations for each parameter. Each different standard deviation is related to a unique track quality parameter and is calculated from measured values of the parameter over a track segment using Equation 1[4]. Higher values of the SD index indicate that track segments are in bad condition. The formulation of the SD index is

$$\sigma_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_{ij}^2 - \bar{x}_i^2)}, \quad (1)$$
$$\bar{x} = \sum_{j=1}^{n}\frac{x_{ij}}{n},$$

where,

σ_i = standard deviation of a quality parameter in mm
x_{ij} = Measurement value in mm for the parameter at the *j*th sampling point in the track segment
n = number of sampling points in the track segment.

Q Index
In the Netherlands, ProRail converts the SD index into a universal form across different classes of tracks, as shown in Equation 2[4]. The Q index ranges from 0 to 10. The higher the Q index value, the better the track quality of a 200 m long track segment.

$$N = 10 * 0.675^{\frac{\sigma_i}{\sigma_i^{80}}} \quad (2)$$

Where,

N = Q index for a quality parameter over a 200 m long track segment,
σ_i = standard deviation for the quality parameter,
σ_i^{80} = 80th percentile of standard deviations for 200 m long segments in a maintenance section ranging in length from 5 to 10 km).

Track Geometry Index
In order to measure the quality of a track segment, the Track geometry index (TGI_i) uses the measurement value space curve length, L_i for a quality parameter over a track segment as shown in Equation 3. A higher the value, the worse the track segment quality.

$$TGI_i = \left(\frac{L_i}{L_o} - 1\right) * 10^6$$
$$L_i = \sum_{j=1}^{n-1}\sqrt{\left(x_{i(j+1)} - x_{ij}\right)^2 + \left(y_{j+1} - y_j\right)^2}, \quad (3)$$

Where,

L_i = measurement value space curve length for a quality parameter over a track segment
L_o = length of the track segment
y_j = milepoint of the *j*th sampling point on the track segment

Track Roughness Index
In 1998, Track Roughness Index was proposed by Amtrak [4]. It is defined as the average of squared measurement value for a quality parameter over a track segment (Equation 4).

$$TRI_i = \sum_{j=1}^{n} \frac{(x_{ij})^2}{n} \quad (4)$$

CN's Track Quality Index
Canadian National Railway Company (CN) developed a measure whereby a 2nd order polynomial equation of the standard deviation of measurement values is used for a track quality parameter for track segments in order to assess its partial quality (Equation 5).

$$TQI_i = 1000 - C * \sigma_i^2 \quad (5)$$

Where:
C = a constant which has a maximum value of 700 for the main line tracks.
σ_i = standard deviation for the parameter

The overall quality assessment is obtained by calculating the average of six partial quality indices for gauge, cross level, left (right) surface, and left (right) alignment. Tracks with better quality have higher values for track quality index.

P Index
Japanese railroads use the P index for the assessment of track quality and it is calculated by finding the ratio of the number of sampling points whose quality parameter measurements fall outside $\pm$3mm to the number of all sampling points in a track segment [5]. The P index is applied to two specific lengths of track segments, 100 m and 500 m. The larger the P index, the worse the quality of the track segment.

SNCF's Mean Deviation Indices
SNCF's indices are quite different from the above mentioned indices. Standard deviations are not used in this case. Instead, the weighted moving average over a track segment is used. This is formulated in Equation 6.

$$TL^i(y_0) = \frac{1}{300} \int_{-\infty}^{y_0} \eta^i(y)\, exp\left(\frac{y-y_0}{300}\right) dy \quad (6)$$

Where,
$y_0 =$ maximum milepoint value in the track segment
$\eta^i(y)$ = measurement data for a quality parameter at the milepoint *y*.

Chinese Track Quality Index
The Chinese Track Quality Index is similar to the SD index. The national railroads, as well as the Nanjing Metro in China mainland use the sum of standard deviations of seven quality parameters to evaluate the general track quality of a track segment using Equation 7. The two lengths used for the overall track quality assessment are 200 m and 500 m. The track length of 500 m is used for high-speed railroads. The higher the TQI value, the worse the condition of the track segment.

$$TQI = \sum_{i=1}^{7} \sigma_i$$
$$\sigma_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(x_{ij}^2 - \bar{x}_i^2\right)} \quad (7)$$
$$\bar{x}_i = \sum_{j=1}^{n} \frac{x_{ij}}{n}$$

Track Degradation

Changes in track geometry occur with time after repeated traffic loading resulting in track settlement. This settlement occurs in two phases [6]:

1. Immediately after tamping, the settlement occurs rapidly until the gaps between the ballast particles have been reduced and the ballast is consolidated.
2. The second phase of settlement occurs less rapidly and there is a nearly linear relationship between settlement and time (or load).

Track geometry degradation is assessed with the aforementioned track geometry parameters. Researchers have used both statistical and mechanistic approaches to assess and predict the deterioration in track geometry. Due to the high levels of uncertainty involved in maintenance, Bayesian methods have been employed by various researchers to predict track geometry degradation thorough its life-cycle. Andrade and Teixeira [7] built up a framework that updated the uncertainty in rail track geometry degradation throughout its life-cycle. The results from their study showed that at the design stage, the uncertainty associated with degradation rates was very high, but reduced drastically as more inspection data was collected. They proposed that maintenance costs assessments related to track geometry degradation should not be assessed at the design stage based only on the prior probability distributions of the degradation model parameters. Instead, they suggested that these costs should be assessed only after a 'warm-up' period of operation based on their posterior probability distributions.

Markov Chain Monte Carlo methods have been adopted in different variations of Bayesian models for track geometry degradation for example Hierarchical Bayesian [7] and Hybrid Bayesian-Wiener models [3]. The Hierarchical Bayesian study modeled the two quality indicators related to railway track geometry degradation which are the standard deviation of longitudinal level (SDLL) defects and the standard deviation of horizontal alignment defects (SDHA) [8]. The outcome of the research was that Hierarchical Bayesian models exhibited a worse fit of SDHA compared to the quality indicator SDLL which led to the conclusion that the horizontal alignment defects appeared to be less predictable. The Hybrid Bayesian-Wiener Process was also used to formulate track geometry degradation [3]. The Wiener process sample paths simulated using the output from the adaptive Metropolis-Hastings algorithm [3] showed that the predicted sample paths were able to capture the variability of the degradation process by bounding the observed degradation data points.

Sequential Monte Carlo methods were adopted to test and rank different prognostics models and algorithms for railway track degradation [9]. The researchers proposed this Bayesian model class methodology for prognostics performance assessment where various prognostics algorithms could be thoroughly assessed and ordered based on their relative probability to predict the future degradation process.

Approximate Bayesian Computation

The Approximate Bayesian computation (ABC) method has its roots in Bayes' Theorem, $P(\theta|D) = P(D|\theta)P(\theta)/P(D)$ where $P(\theta|D)$ is the probability of the parameters given the data, $P(D|\theta)$ is the probability of the data given the parameters, $P(\theta)$ is the probability of the parameters and $P(D)$ is the probability of the data. In interpreting Bayes Theorem, the conditional probabilities represent what is known as the degree of belief. These beliefs about the parameters θ can be updated as more information or evidence is gathered about the data, D. The prior, $P(D)$, is the initial degree of belief in D. $P(\theta|D)$, the posterior, is the degree of belief having accounted for D. $P(D|\theta)/P(D)$ is the quotient that represents the support D provides for θ and the expression $P(D|\theta)$ is proportional to the likelihood function. Researchers have developed ways to compute likelihood functions using various methods that approximate these functions. Case in point, methods like h-likelihood, variational approximations, local likelihood, composite likelihood, weighted likelihood, indirect inference, quasi-likelihood and Laplacian approximations have been explored and executed. Instead of applying these methods, there exists a category of methods/ techniques that can be used to estimate the posterior while bypassing the computation of expensive likelihood functions known as Approximate Bayesian computation (ABC).

Approximate Bayesian computation (ABC) is a group of methods used for the analysis of complex stochastic models whose posterior distributions are difficult to compute because of intractable likelihood functions. ABC bypasses the computation of likelihood functions to compute a posterior that is approximate to the actual posterior distribution through simulations which involve sampling distribution [10]. Rubin introduced the ABC idea in the early 1980's in an attempt to provide bayesianly justifiable calculations for use in applied statistics [11]. He suggested that this method of sampling would be appropriate as a way around complex likelihood functions with lower dimensional data. The algorithm he suggested at the time would later be known as the ABC rejection algorithm. After its conceptual introduction, ABC was developed by Pritchard et al [12] where they did model simulations that produced an artificial data set. ABC has been applied in different disciplines like the fields of Genetics [13], Cosmology and Astrophysics [14], and Environmental Engineering [15] , and is steadily gaining popularity in Mechanical Engineering [16][17].

ABC is comparable to the Method of Least Squares but the difference lies in the objectives of both methods [18]. While the Method of Least Squares is used to find point estimates of parameters by minimizing the sum of offsets or residuals, ABC is used to estimate the posterior distribution of parameters. The ABC posterior distribution is denoted $P_{ABC}(\theta|D) = P(\theta|\delta(D,\widehat{D}) < \epsilon)$. ABC approximates the posterior with a set of simulated summary statistics. In order to bypass the likelihood function, ABC relies on a distance function $\delta(\cdot,\cdot)$ to compare statistics from the simulated data $\widehat{D}$ to the data D that was observed. This method of simulating data to compare to observed data in order to estimate model parameters is not new in the field research [19]. The ABC

algorithm works as follows: The method samples values of the parameter from the prior distribution $P(\theta)$. This prior could be for example, a binomial distribution of each parameter. The prior could be subjective, objective and informative, or noninformative. The subjectivity of a prior means that it is based on the experimenter's personal belief. A prior is said to be objective and informative if it is either based on historical data on the distribution of parameter values or based on data from experiments preceding the one being undertaken. A noninformative prior is one that expresses ignorance as to the value of the parameter. In Approximate Bayesian Computation, the posterior will mainly reflect the observational data so long as the prior is suitably noninformative. The rate of rejection of the simulated data tends to be higher if the prior is unlike the posterior thereby increasing computational costs.

When the parameter values are chosen, a dataset $\widehat{D}$(with summary statistics $\widehat{S}$) is simulated from the prior distribution which has the same number of observations as the observed data set D (with summary statistics S). This step is taken in order to make sure that the distribution and summary statistics of the simulated data $\widehat{D}$ match that of the observed data D. The summary statistics of the simulated data ($\hat{S}$) and the observed data (S) are compared by computing a distance between them represented by a predefined distance function $\delta\ (D,\widehat{D})$. If $\delta\ (D,\widehat{D})$ is small enough(less than ϵ (tolerance)), then the simulated data $\widehat{D}$ is comparable to the observed data D such that there is a high probability that the observed parameters of the data have an approximate posterior distribution $P\ (\theta|\delta\ (D,\widehat{D}) < \epsilon)$. The parameters are therefore accepted if $\delta(D,\widehat{D})$ is less than or equal to the associated threshold ϵ or rejected if the threshold is exceeded. The ABC rejection algorithm is summarized in FIGURE 2.

It is worth mentioning that some summary statistics provide more information about some parameters than others. This is why it is required that the summary statistics used in ABC should be "sufficient". This sufficiency means that the statistics should be chosen in such a way that it represents all or as much of the information about the parameter θ included in the data, D. In doing so, ABC is then able to approximate the full posterior. The posterior can then be written as $\delta\ (\theta \mid D) = \delta\ (\theta \mid S)$. In addition, to estimate the exact posterior, $\delta\ (D,\widehat{D})$ must be chosen in a way that $\delta\ (\theta \mid D) = \delta\ (\theta \mid S) \approx P\ (\theta|\delta\ (D,\widehat{D}) < \epsilon)$. The process of choosing $\delta\ (D,\widehat{D})$ could be quite complicated since it does depend on the unknown likelihood function. Nonetheless, it has been shown that depending on the statistics used for different models, the choice of $\delta\ (D,\widehat{D})$ could be quite robust [18]. Apart from the basic ABC rejection algorithm in which approximate data is simulated from a predetermined prior, there are more complex ABC algorithms that improve upon the basic rejection algorithm. In this paper, Markov Chain Monte Carlo ABC methods will be applied in the analysis of track geometry data.

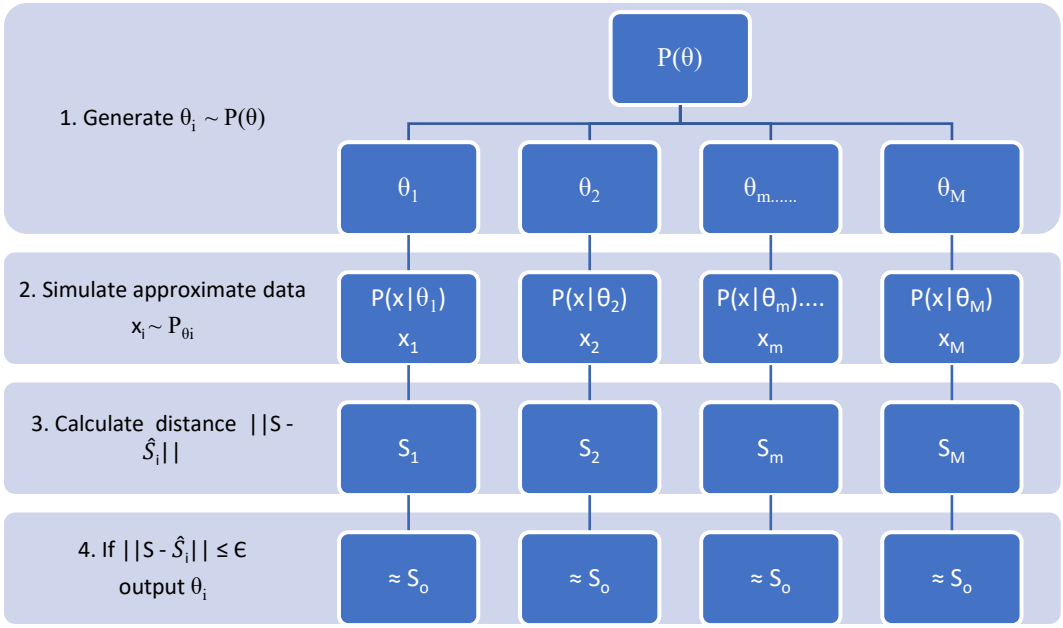

FIGURE 2 ABC Rejection Algorithm

CASE STUDY

Data preparation

The dataset used in this paper was from a US Class I railroad. For every foot of track covered by the track geometry car, data related to track geometry parameters was collected. Data for the case study was distance based track geometry data that consisted of 253602 records and 37 variables/ field names. The track geometry parameters in the dataset include:

a. Track surface, left rail 79ft space curve
b. Track surface, right rail 79ft space curve
c. Track alignment, left rail 79ft space curve
d. Track alignment, right rail 79ft space curve
e. Track alignment, left rail 31 ft chord
f. Track alignment, right rail 31 ft chord
g. Track alignment, left rail 62 ft chord
h. Track alignment, right rail 62 ft chord
i. Track alignment, left rail 124 ft chord
j. Track alignment, right rail 124 ft chord
k. Track surface, left rail 22ft chord
l. Track surface, right rail 22ft chord
m. Track surface, left rail 31ft chord
n. Track surface, right rail 31ft chord
o. Track surface, left rail 62ft chord
p. Track surface, right rail 62ft chord
q. Track surface, left rail 124ft chord
r. Track surface, right rail 124ft chord

s. Rail Cant, left rail
t. Rail Cant, right rail

The units for the above variables is mm * 100.

The initial inspection of the data prior to analysis revealed that there were no missing values or irregularities. Three 200 ft segments (named L1, L2 and L3) were subsequently selected from the data for the case study. The two variables selected for analysis were Track surface, left rail 31ft chord (SURF_31_L) and Track alignment, left rail 31 (ALIGN_31_L). For the sake of simplicity, they will be referred to as the Surface and Alignment parameters.

METHODOLOGY

ABC methods

The fundamental Approximate Bayesian Computation algorithm has been updated, improved and extended by several researchers over the years. Choosing summary statistics is an important step in ABC and the reduction of dimensionality in data increases its efficiency[20]. The curse of dimensionality subsequently comes into play when the rejection algorithm is used such that large numbers of simulations have to be run to obtain an adequate number of accepted runs, resulting in inefficiency [20]. To solve this problem, it has been suggested that ABC be combined with Markov Chain Monte Carlo (MCMC) principles [21]. MCMC allows one to characterize a distribution without knowing the mathematical properties of the distribution [22]. This is done by randomly sampling values out of the distribution. It offers the advantage of being used to draw samples from distributions even when the only thing known about the distribution is how to calculate the density for different samples [22]. This paper focuses on two methods of performing ABC-MCMC proposed by different researchers.

Marjoram Method

The first method to be discussed and used in this research was presented by Marjoram et al. [21]. They first proposed the ABC-MCMC likelihood-free approach. These researchers demonstrated that a Markov chain where simulated data is accepted when equal to the observed data D, and rejected if it is different form the observed, converges to the right posterior distribution. They suggested that for complex data sets, summary statistics should be used to replace the full data, just as in the traditional ABC. The algorithm proposed [21], which will be referred to as the Marjoram Algorithm in this paper is as follows:

Step 1. If now at θ propose a move to θ' according to a transition kernel $q(\theta \rightarrow \theta')$.
Step 2. Generate D' using model M with parameters θ'.
Step 3. If $D' = D$, go to *Step 4*, and otherwise stay at θ and return to *Step 1*.
Step 4. Calculate

$$h = h(\theta, \theta') = \min\left(1, \frac{P(\theta\prime)q(\theta' \rightarrow \theta)}{P(\theta\prime)q(\theta \rightarrow \theta\prime)}\right) \quad (8)$$

Step 5. Accept θ' with probability h and otherwise stay at θ, then return to *Step 1*.

Wegmann Method

The second technique, by Wegmann et al. [23] introduced the idea of using partial least-squares (PLS) transformation in the choice of summary statistics for ABC-MCMC. PLS regression seeks linear combinations of the original summary statistics which are simultaneously maximally decorrelated and have high correlation with the parameters [24].

PLS is comparable to Principal Component Analysis (PCA) in that instead of finding linear combinations that maximize the explained variance in the summary statistics space, PLS components are chosen to maximize the product of the variance among summary statistics and the covariance between parameters and statistics [25]. The algorithm proposed, which will be referred to as the Wegmann Algorithm in this paper is as follows [25]:

Step 1. Perform n simulations with parameters θ' randomly drawn from their priors, and each time compute their associated set of summary statistics S'.

Step 2. Compute PLS components from the n θ' and S' vectors.

Step 3. For all n simulations, transform their associated summary statistics S' into k retained PLS components, as S'_{PLS}. Transform the observed summary statistics S as S_{PLS} and compute $p_n(\delta|\, S_{\mathrm{PLS}}, \theta)$.

Step 4. Fix ϵ, estimate δ_ϵ from $p_n(\delta|\, S_{\mathrm{PLS}}, \theta)$, and set the proposal range of the parameters (for the transition kernel $q(\theta \to \theta')$ based on ρ and the variability of the parameters among the n_ϵ retained simulations.

Step 5. Start an MCMC chain of total length s from a position θ randomly chosen from the n_ϵ simulations closest to D. Set $i = 0$.

Step 6. If now at θ, propose a move to θ' according to a transition kernel $q(\theta \to \theta')$. Increment i.

Step 7. Simulate D' based on θ'. Compute the summary statistics S' and transform them into S'_{PLS}.

Step 8. If $\delta i = \| S'_{PLS} - S_{\mathrm{PLS}} \| \geq \delta_\epsilon$ stay at θ and go to *Step 10*.

Step 9. Accept θ' with probability $\min\left(1, \frac{P(\theta\prime)q(\theta' \to \theta)}{P(\theta\prime)q(\theta \to \theta\prime)}\right)$, else stay at θ.

Step 10. If $i < s$ go to *Step 6*.

In this paper, the focus is estimating the posterior distribution of TQIs for different track segments. Non-informative priors (uniform distribution) were used to minimize subjective assessments which could impact the posterior distribution. The ABC-MCMC chain was calibrated from 10,000 simulations conducted under the prior. The length of the "burn in" period for the Markov chain was 2000. The Markov Chain was allowed to "burn in" in order for it to enter a high probability region where the state of the Markov chain is more representative of the distribution being sampled.

Results and Discussion

For track lengths L1, L2 and L3, ABC-MCMC was performed using both the Marjoram and Wegmann algorithms for the Surface and Alignment parameters. The resulting Markov chains for each length and algorithm were then plotted. To visually assess model convergence, trace plots, density plots and autocorrelation plots were employed. The process of assessing the Markov chain

convergence is extremely important because valid inferences cannot be drawn if the chain does not converge. Assessing convergence of a chain could however prove difficult since chains do not converge to a fixed point but rather to a distribution. The test of stationarity in this paper is the Heidelberg and Welch diagnostic test. Subsequently, the approximated posterior distributions are presented and compared to the actual distributions with the aid of density plots. Finally, cross-validation was performed and the accuracy obtained was reported.

Trace Plots

The trace plots in the top row were produced by applying the Marjoram Algorithm to lengths L1, L2 and L3, in that order. Figure 3A contains the trace plots for the Surface parameter. The trace plots on the bottom row were also obtained by applying the Wegmann Algorithm to lengths L1, L2 and L3, in that order. Figure 3B, contains the trace plots for the Alignment parameter. As was done for the Surface Parameter, the trace plots in the top row were produced by applying the Marjoram Algorithm to lengths L1, L2 and L3, in that order. The trace plots on the bottom row were also obtained by applying the Wegmann Algorithm to lengths L1, L2 and L3, in that order. MCMC was implemented such that its stationary distribution reflects the posterior distribution of the parameter vector θ, which in this case are Surface and Alignment. As mentioned earlier, there was a burn in period, a period in which the initial iterations were removed from the simulation process so that the remaining chain can be used to infer the parameters. A visual inspection of the chain shows that the trace plots for the Marjoram Algorithm for both the Surface and Alignment parameters were less stationary than the trace plots for the Wegmann Algorithm.

Histograms

The associated histograms for the trace plots are presented in Figure 4A and Figure 4B. Figure 4A contains histograms for the Surface parameter. Those in the top row of were produced by applying the Marjoram Algorithm to lengths L1, L2 and L3, in that order.

The histograms on the bottom row were also obtained by applying the Wegmann Algorithm to lengths L1, L2 and L3, in that order. The presentation of the resulting histograms in Figure 4B follows the same format as Figure 4A except that Figure 4B represents histograms for the Alignment parameter. For all the histograms, the true values true values were generally slightly left skewed in the posterior distribution.

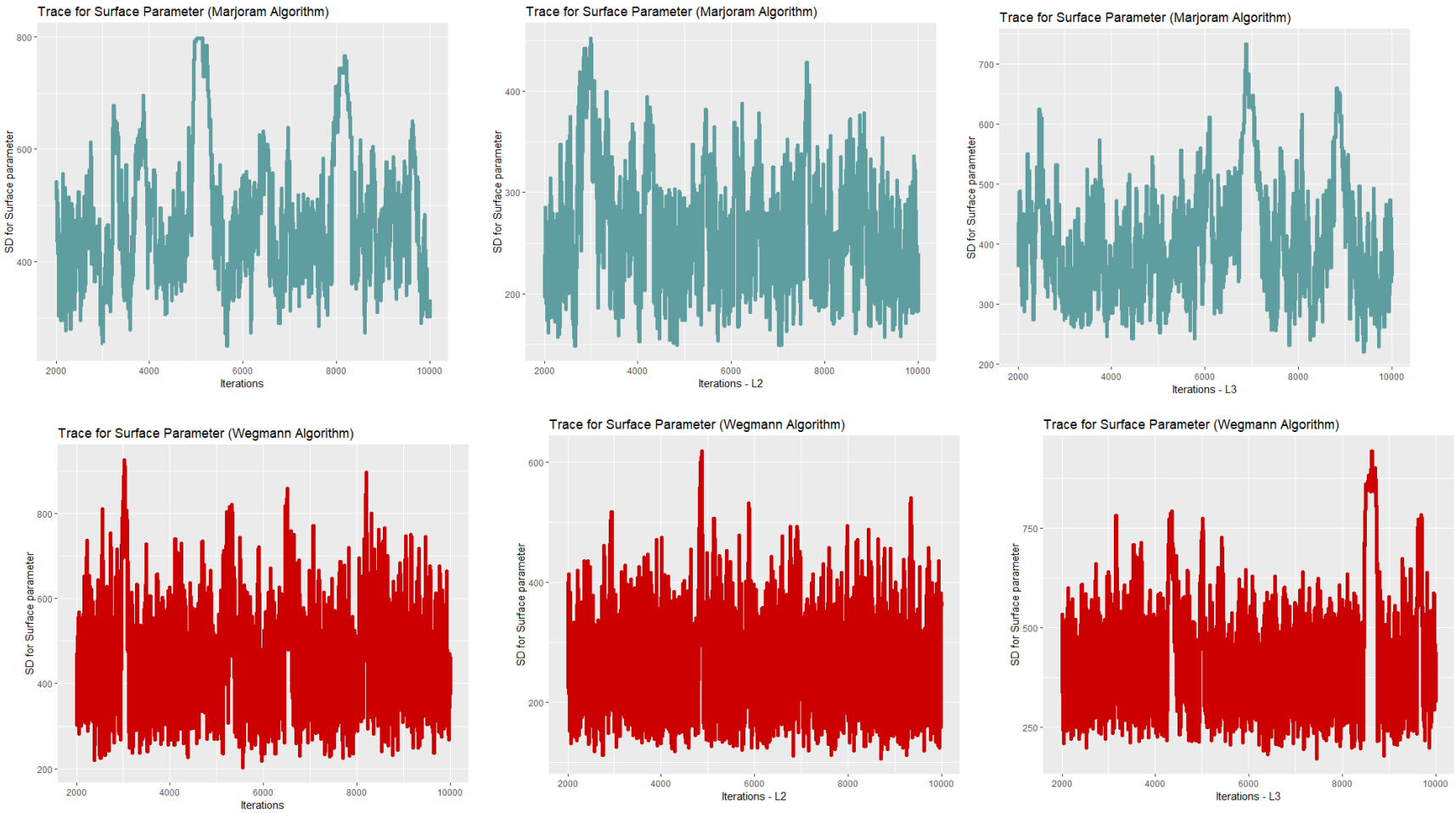

Figure 3A Trace plots for Surface Parameter

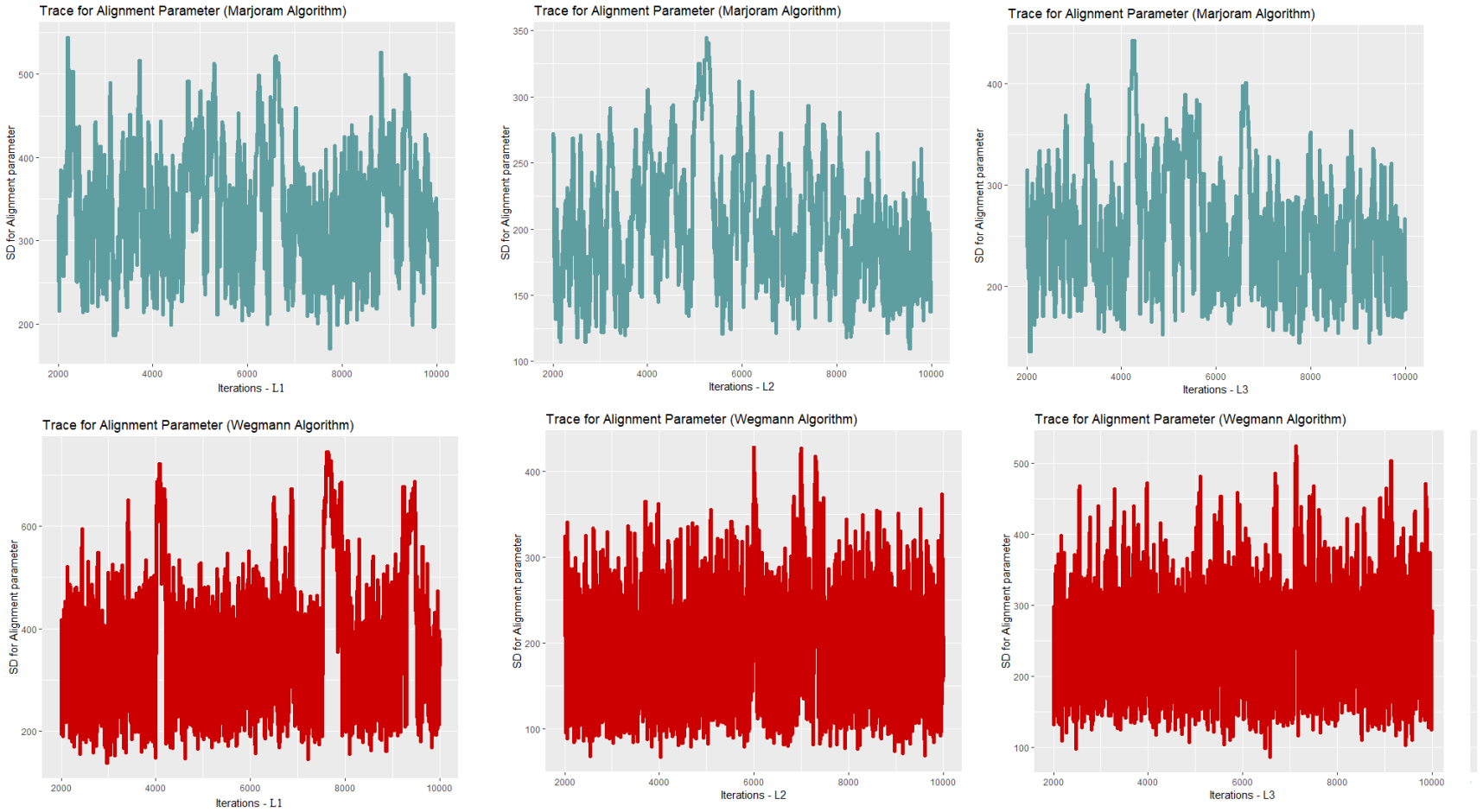

Figure 3B Trace plots for Alignment Parameter

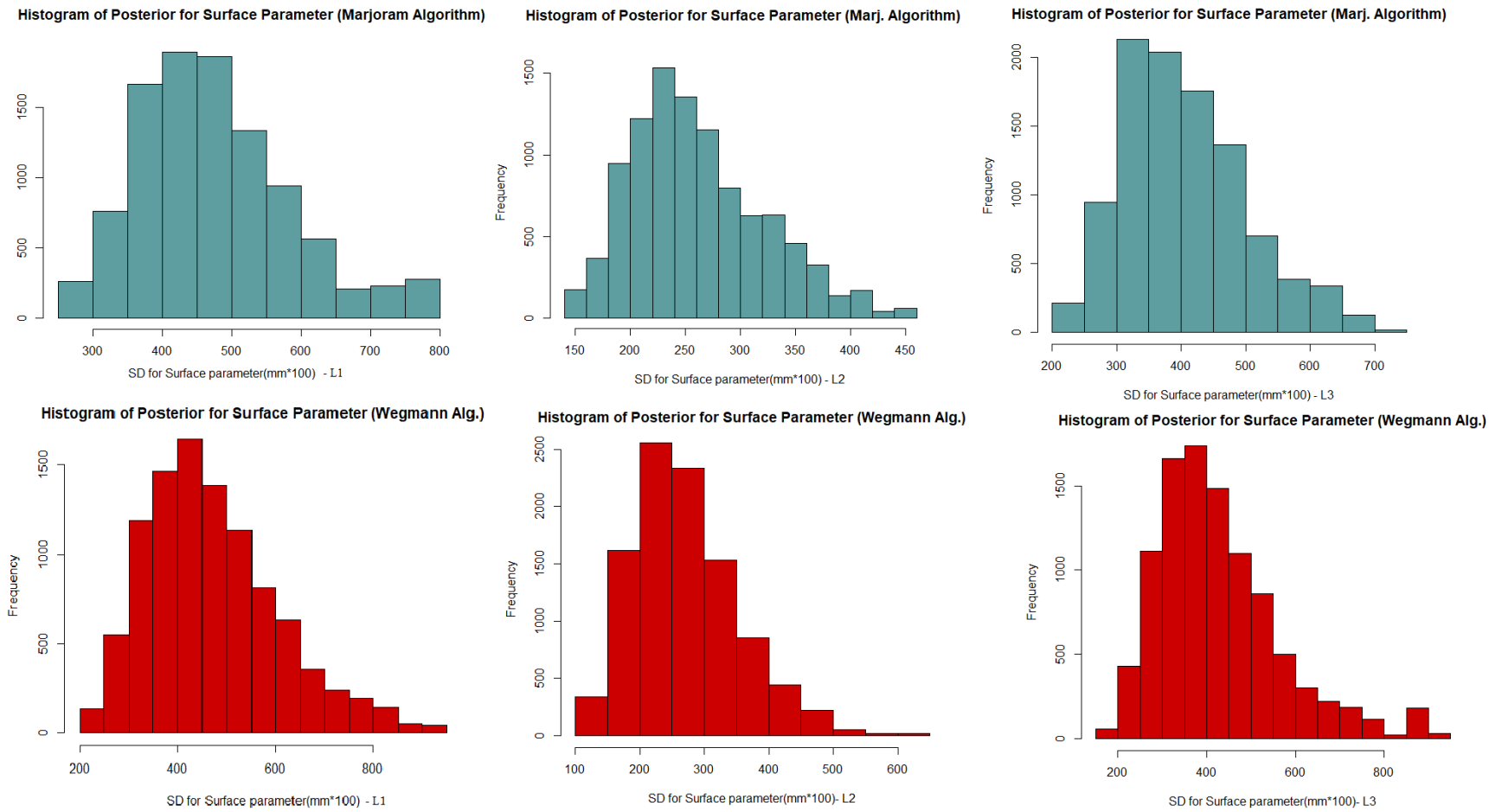

Figure 4A Histograms for Surface Parameter

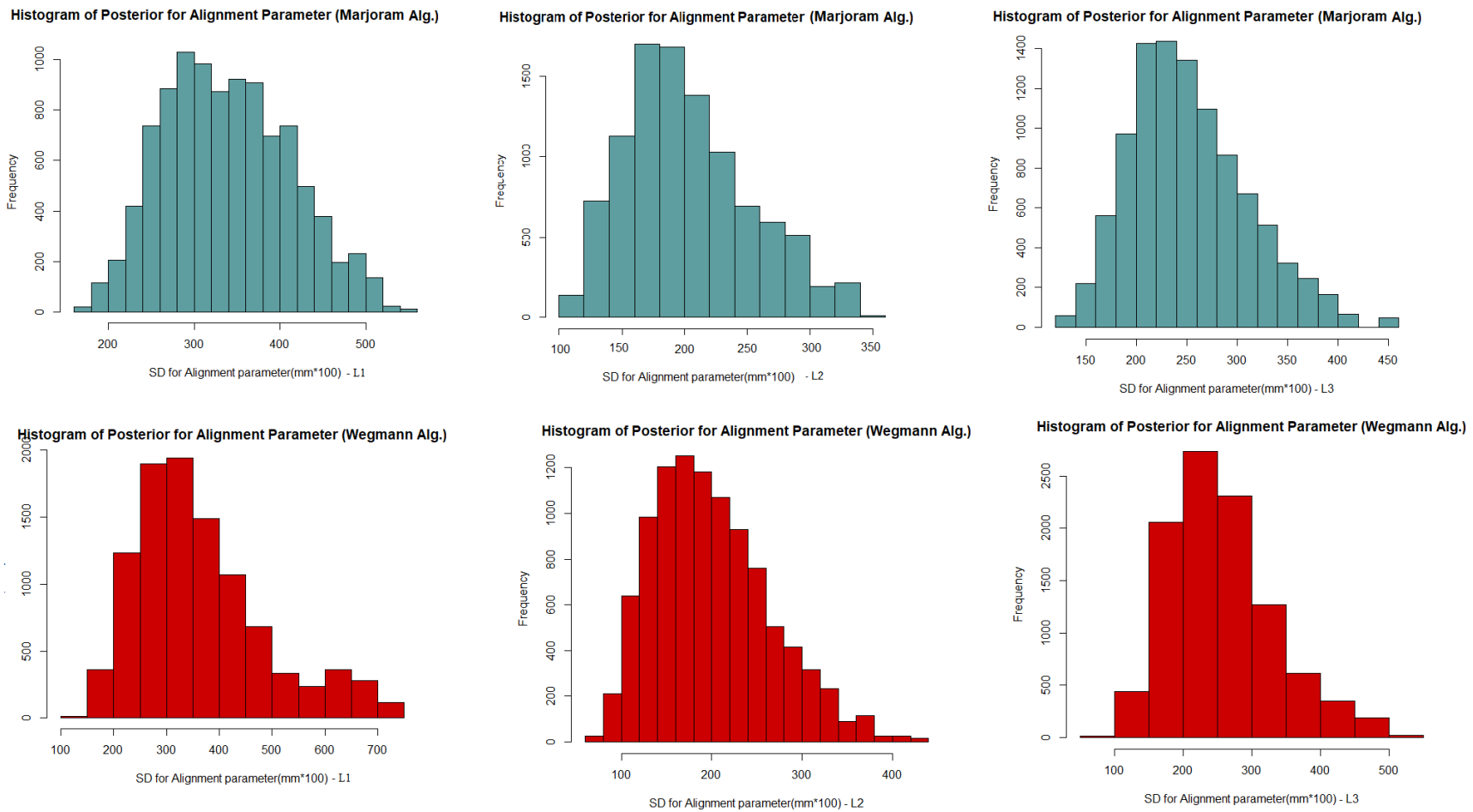

Figure 4B Histograms for Alignment Parameter

Autocorrelation Plots

Autocorrelation plots were used to detect if there was any non-randomness in the chains. The plots were used to identify the correlation (found on y-axis) of samples for each step of the chain with prior estimates of that same variable (Surface or Alignment), lagged by 500 iterations (found on x-axis). Figure 5A and Figure 5B are the autocorrelation plots for the Surface parameter and Alignment parameter respectively. The autocorrelation plots follow the same arrangement as the trace plots and histograms for each parameter. In Figure 5A(Surface parameter), for lengths L1, L2 and L3, the autocorrelation plots for the Marjoram algorithm generally declined more slowly than the plots for the Wegmann algorithm. It can be inferred that the samples from the stationary distributions obtained using the Wegmann Algorithm were less reliant on initial values in the chain than the distributions obtained using the Marjoram Algorithm.

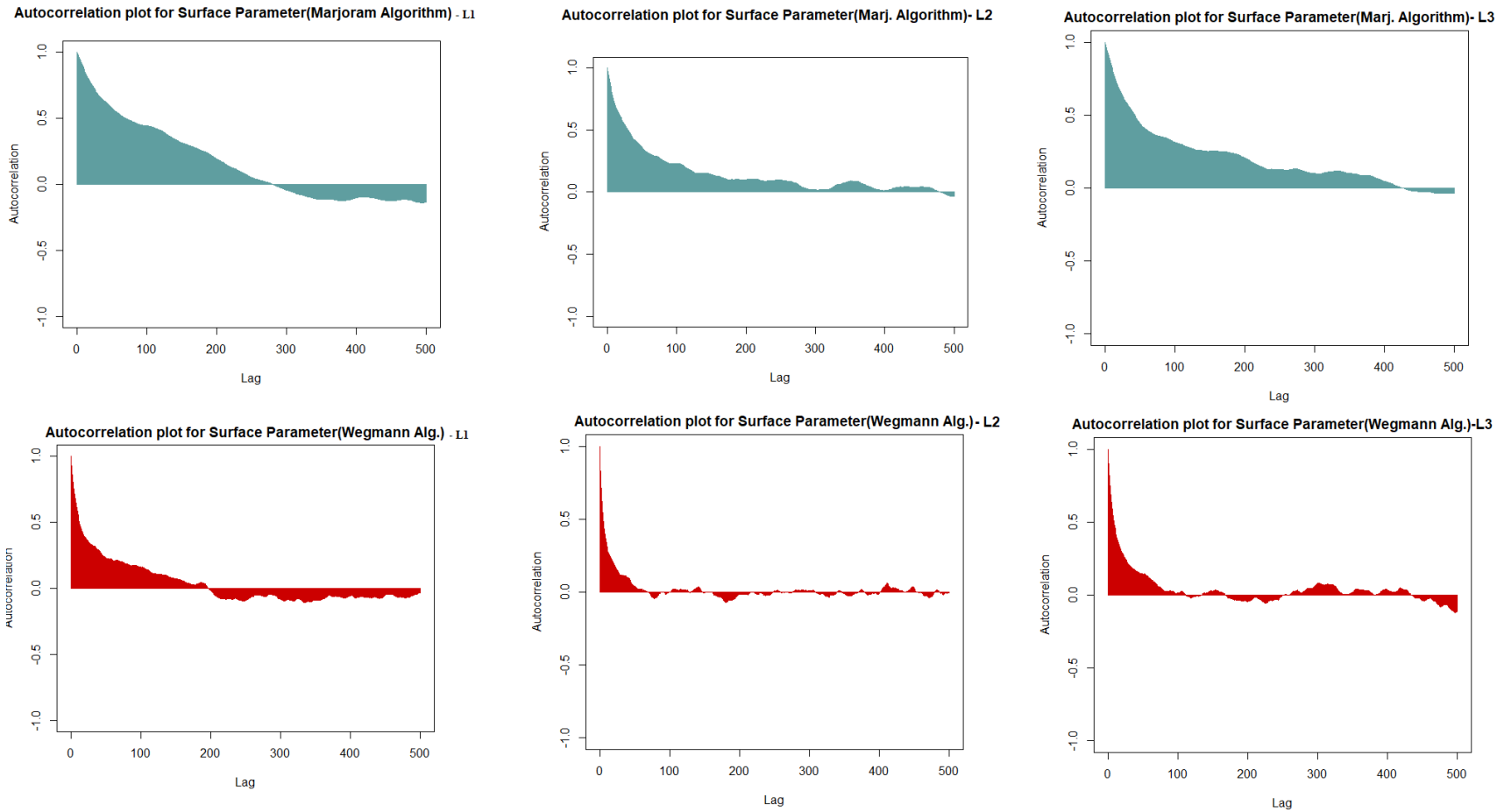

Figure 5A Surface Autocorrelation plots

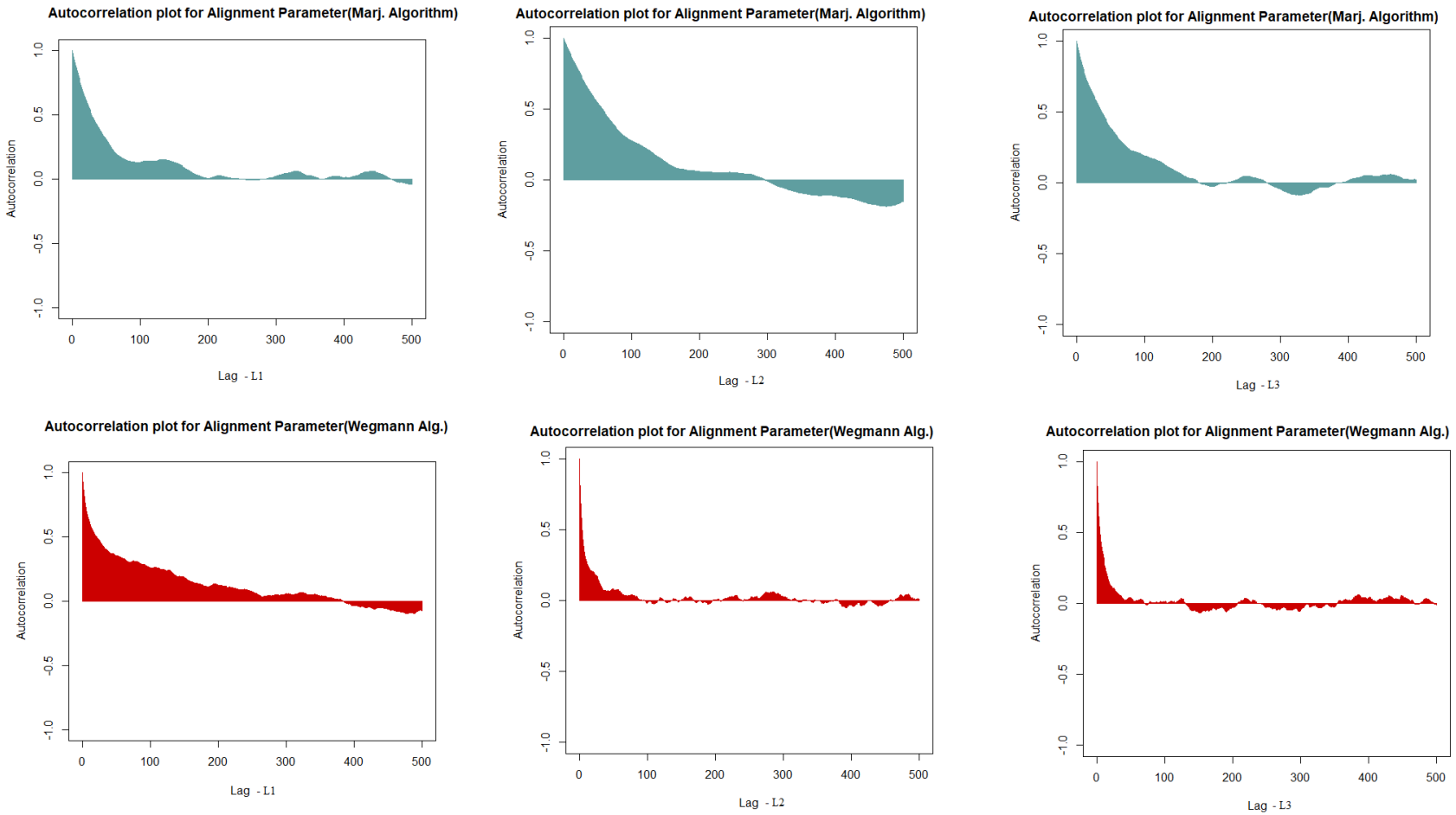

Figure 5B Alignment Autocorrelation plots

Heidelberg and Welch Diagnostics

The Heidelberg and Welch diagnostic is in two parts. In the first part of this test, a test statistic is calculated based on the Cramer-von Mises statistic. This is used to assess the Markov chain by testing the hypothesis that the chain comes from a stationary process. Specifically, the test statistic is used to accept or reject the null hypothesis that the Markov chain is from a stationary distribution. In this test, a p-value of ≤ 0.05 means that we reject the hypothesis that the samples have the same mean. The second part of this test is the half-width test which determines whether the Markov chain sample size can be used to adequately approximate the mean values of the parameters under study given a particular confidence interval (95% CI in this case). TABLE 1 and TABLE 2 contain the results for the Heidelberg and Welch Diagnostic test performed for the surface and alignment parameters respectively. All the Markov chains from the analysis passed the stationarity and half-width tests. Noticeably, the half-widths for the Wegmann Algorithm across the different track lengths were generally wider than those for the Marjoram Algorithm.

TABLE 1 Heidelberg and Welch Diagnostics for Surface Parameter

SURFACE					
Heidelberg and Welch Diagnostic					
	ABC Algorithm	**Stationarity test(ST)**	**ST p-value**	**Halfwidth test**	**Halfwidth**
L1	Marjoram	passed	0.870	passed	28.32
	Wegmann	passed	0.051	passed	22.86
L2	Marjoram	passed	0.321	passed	13.1
	Wegmann	passed	0.604	passed	8.118
L3	Marjoram	passed	0.136	passed	20.61
	Wegmann	passed	0.232	passed	14.05

TABLE 2 Heidelberg and Welch Diagnostics for Alignment Parameter

ALIGNMENT					
Heidelberg and Welch Diagnostic					
	ABC Algorithm	**Stationarity test(ST)**	**ST p-value**	**Halfwidth test**	**Halfwidth**
L1	Marjoram	passed	0.77	passed	14.31
	Wegmann	passed	0.12	passed	23.7
L2	Marjoram	passed	0.45	passed	22.9
	Wegmann	passed	0.23	passed	6.823
L3	Marjoram	passed	0.41	passed	12.04
	Wegmann	passed	0.784	passed	7.02

Kernel Density Plots and MCMC SE

To approximate the posterior distribution, kernel density plots used to summarize the sampled values that define the stationary distribution of values [26] . In the density plot, the value with the most support from the data and the prior is found at the apex (the maximum a posteriori) and is the mode of the distribution. With this plot, summary statistics like posterior mean and posterior median can be obtained quite effortlessly. The main function of kernel density estimation is to smooth over the samples and provide an estimate of the posterior distribution [26]. In FIGURE 6A and FIGURE 6B, the kernel density plots obtained from using the Marjoram and Wegmann Algorithms to predict Surface and Alignment parameters respectively were superimposed. The

dashed green line represents the true value of the TQI for each track segment in Figure 6A. The dashed black line represents the true value of the TQI for each track segment in Figure 6B. The Marjoram and Wegmann density plots can be identified using the color key provided on the plots. Plots for the surface parameter revealed that the Marjoram Algorithm produced estimates that were closer to the actual parameter values for L2 and L3. Also, plots for the Alignment parameter reveal that the Marjoram Algorithm produced estimates that were closer to the actual parameter values for L1 and L3. In general, the Wegmann Algorithm produced lower estimates than the Marjoram Algorithm. TABLE 3 and TABLE 4 contain the estimated TQI values from analysis of the Surface and Alignment parameters respectively.

The Markov Chain Monte Carlo Standard Error (MCMC SE) gives an estimate of the inaccuracy of Monte Carlo samples, regarding the expectation of posterior samples, from Markov chain Monte Carlo (MCMC) algorithms used in this research. Alternatively, it can be said that MCMC SE quantifies the uncertainty in the posterior mean estimate. MCMC SE approaches zero as the number of independent posterior samples also approaches infinity. MCMC SE is a standard deviation around the posterior mean of the samples resulting from the uncertainty related to using the ABC-MCMC algorithms. The errors are approximately equal for each track segment analyzed and can be found in TABLE 3 and TABLE 4 for the Surface and Alignment parameters respectively.

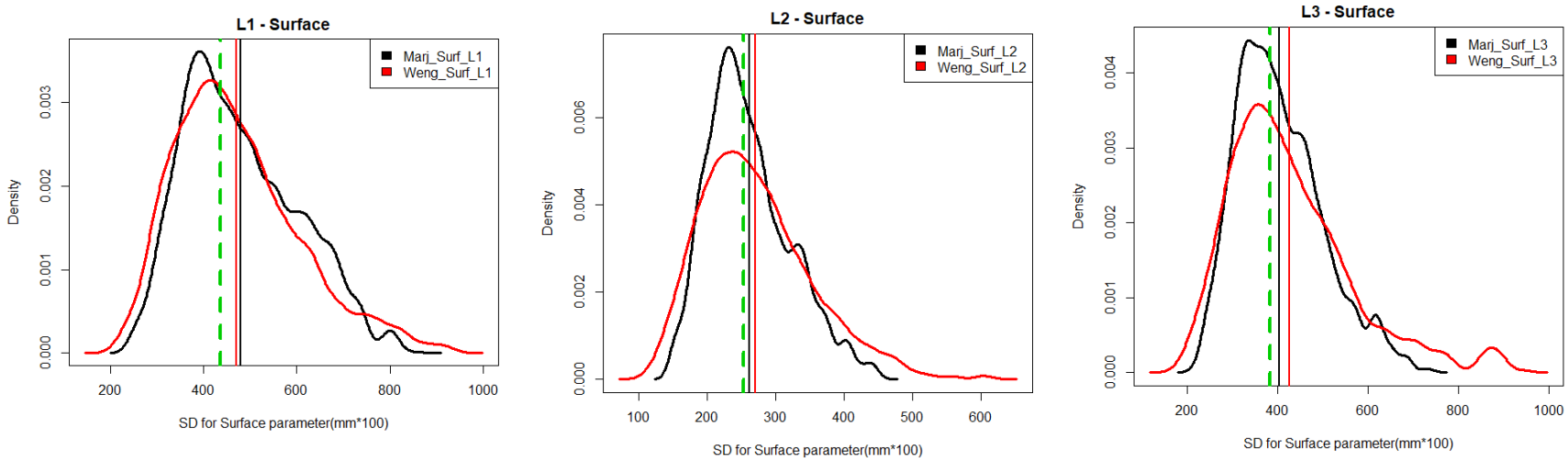

FIGURE 6A Combined kernel density plots for Marjoram and Wegmann Algorithms (Surface)

TABLE 3 MCMC SE and TQI values for Surface Parameter

Surface			
	ABC Algorithm	**Mean of TQI (mm*100)**	**MCMC SE**
L1	Marjoram	469.0	20.6
	Wegmann	470.0	20.6
L2	Marjoram	264.2	11.2
	Wegmann	270.1	11.2
L3	Marjoram	407.1	17.6
	Wegmann	414.1	17.8

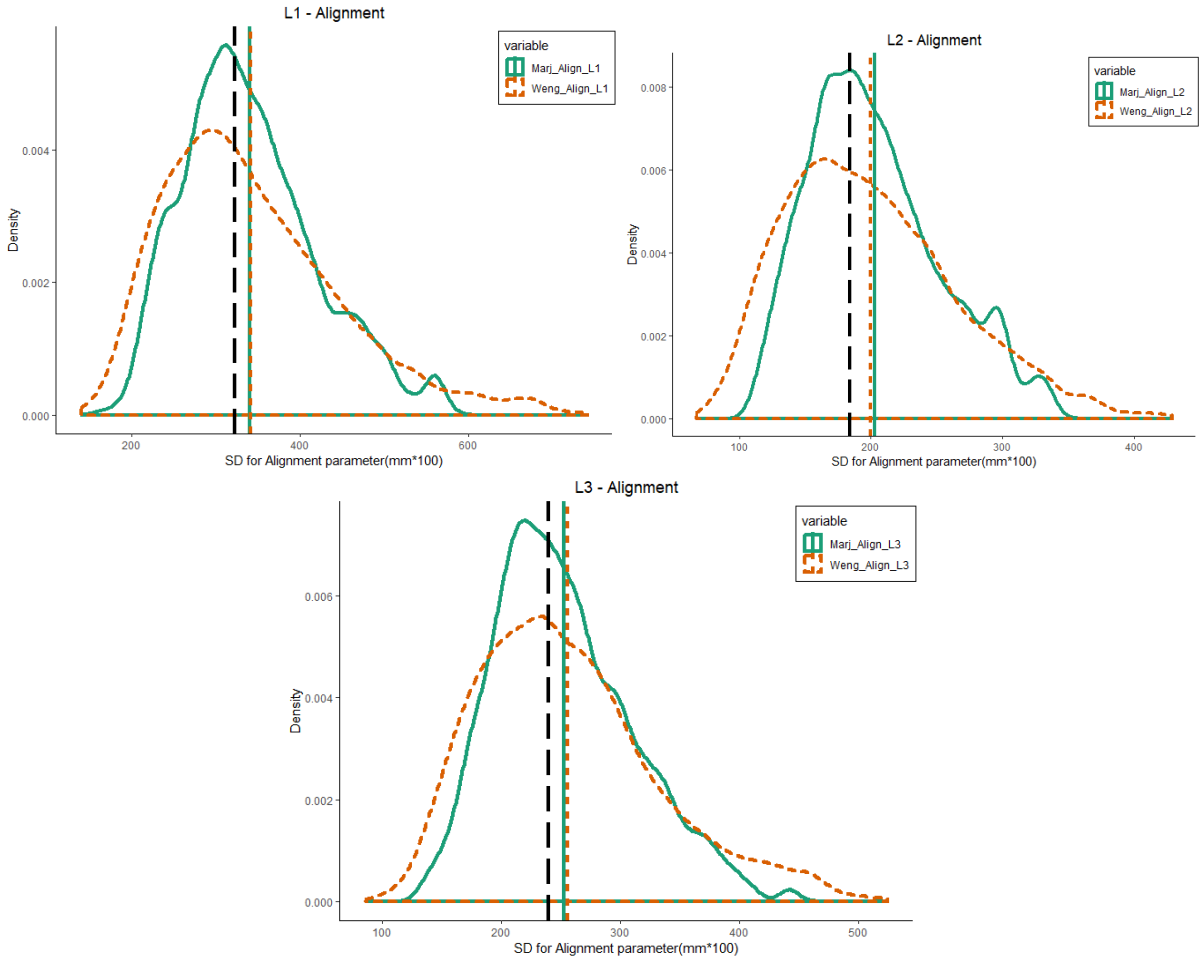

FIGURE 6B Combined kernel density plots for Marjoram and Wegmann Algorithms (Alignment)

TABLE 4 MCMC SE and TQI values for Alignment Parameter

Alignment			
	ABC Algorithm	**Mean of TQI (mm*100)**	**MCMC SE**
L1	Marjoram	339.7	14.4
	Wegmann	341.1	15.9
L2	Marjoram	203.0	8.7
	Wegmann	199.7	8.2
L3	Marjoram	254.3	10.5
	Wegmann	257.0	10.9

Accuracy and Credible Interval for TQI estimates

In ABC, the summary statistics of the simulated data ($\hat{S}$) and the observed data (S) are compared by calculating a distance between them. If this distance is less than a selected tolerance ϵ, the parameter values related to the summary statistics are retained and are used to determine the posterior distribution. In this analysis, the accuracy of estimation related to different tolerance rates was calculated (Figure 7). Calculations indicated that a tolerance of 0.01 used in TQI estimation for the Alignment parameter gave the highest accuracy of 90%. In addition, a tolerance of 0.005 used in TQI estimation for the Surface parameter gave the highest accuracy of 86%.

The uncertainty related to the parameter estimates was quantified by computing the 89% credible interval, which is an important concept in Bayesian statistics. In other words, the interval within which the TQI estimates fall with an 89% probability was estimated and results have been compiled in Table 5. The 89% credible interval was used because it has been determined by McElreath [27] to be more stable than the default 95% that is used in frequentist statistics. Table 5 contains the lower and upper confidence intervals computed for TQI estimates for produced by both Marjoram and Wegmann algorithms. It is observed that for each length of track (L1, L2 or L3), narrower credible intervals were related to the Marjoram algorithm estimates while wider credible intervals were associated with the Wegmann algorithm estimates. Narrower credible intervals indicate higher precision hence it can be inferred that the application of the Marjoram algorithm produces more precise estimates.

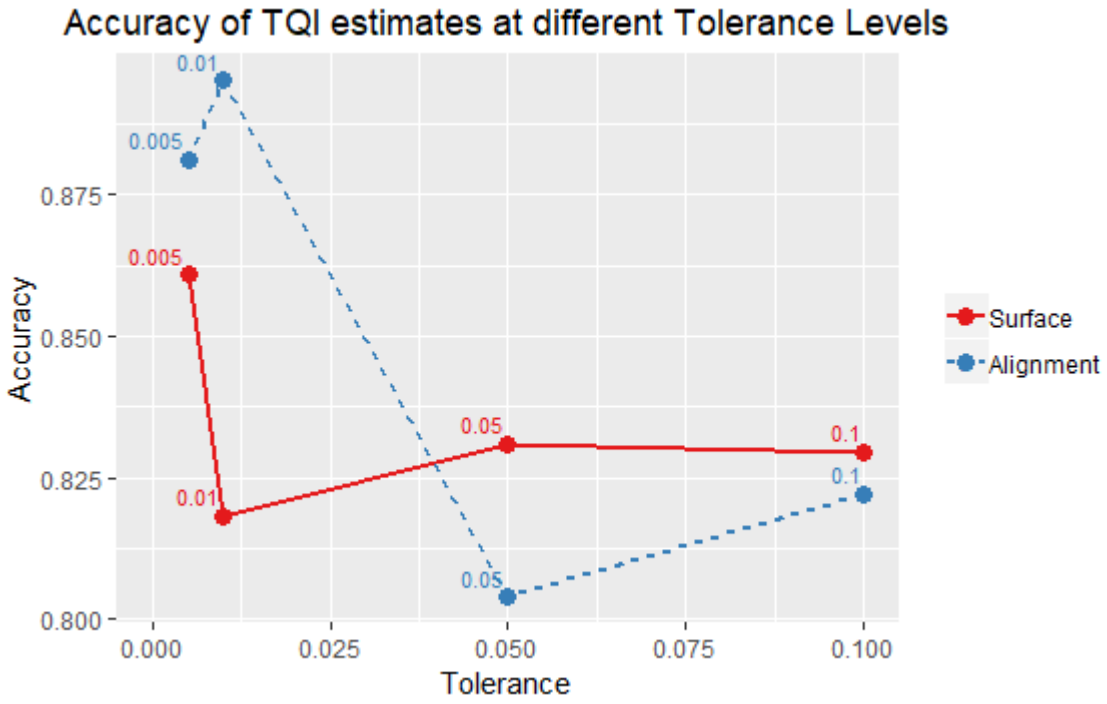

FIGURE 7 Accuracy of TQI Estimates at Difference Tolerance Levels.

Table 5 89% Credible Interval (CI) for TQI Estimates

		Alignment		Surface	
		Lower CI	*Upper CI*	*Lower CI*	*Upper CI*
L1	*Marjoram*	203	450	307	624
	Wegmann	184	489	257	711
L2	*Marjoram*	122	280	181	373
	Wegmann	101	298	140	398
L3	*Marjoram*	164	340	242	532
	Wegmann	137	368	215	611

CONCLUDING REMARKS

The ABC method was successfully applied in this paper for the estimation of Track Quality Indices (TQIs), which are essential components in the modelling of track degradation. The ABC-MCMC method was applied using different algorithms referred to in this paper as the Wegmann and Marjoram algorithms. Although initial visual inspection of the diagnostic plots indicated that the Wegmann Algorithm had better convergence than the Marjoram Algorithm, the kernel density plots revealed that the Marjoram Algorithm produced more accurate posterior predictions. This was confirmed when the 89% CI was computed, hence it was concluded that the Marjoram algorithm is best used for TQI estimation since it produces more precise estimates. This research has brought to light the fact that to obtain high accuracy of estimates the Surface and Alignment

parameters, tolerances of 0.005 and 0.01 must be used respectively. Although TQI estimation using ABC is computationally expensive, it allows for the reasonable estimation of the statistic. In addition, it provides the advantage of potentially estimating probabilities associated with obtaining certain high TQI values for each track segment. Hence, abnormally large parameter values can be detected and maintenance activities can be undertaken if needed. This tackles the issue of uncertainty related to the behavior of track geometry parameters. Professionals can therefore make informed decisions on which track segments to prioritize depending on the probability of obtaining TQI estimates beyond standard thresholds, optimizing time and financial resources, and improving track safety.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding and support of the US Department of Transportation and the University Transportation Center program. They would also like to acknowledge its UTC team, RAILTEAM, with lead University of Nevada at Las Vegas and partner Virginia Tech.

REFERENCES

[1] MarketsandMarkets, Railway System Market by System Type (Auxiliary Power, HVAC, Propulsion, On-board Vehicle Control, Train Information & Train Safety), Transit Type (Conventional & Rapid), Application (Passenger & Freight Transportation),& Region-Global Forecast to 2025, *AT 7019*, 2019. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/railway-system-market-203815831.html?gclid=EAIaIQobChMI3vf-2a_a4wIVQVqGCh1RxgSjEAAYASAAEgJuCvD_BwE. [Accessed: 04-May-2019].

[2] A. L. Clouse, J. Bloom, and S. Cook, Track Inspection into the 21St Century, 2006.

[3] S. A. Galvan Nunez, Hybrid Bayesian-Wiener process in track geometry degradation analysis, University of Delaware, 2017.

[4] A. R. B. Berawi, R. Calçada, R. Delgado, and C. Vale, Evaluating Track Geometrical Quality through Different Methodologies, *Int. J. Technol.*, vol. 1, no. 1, pp. 38–47, 2010.

[5] R.-K. Liu, P. Xu, Z.-Z. Sun, C. Zou, and Q.-X. Sun, Establishment of Track Quality Index Standard Recommendations for Beijing Metro, *Discret. Dyn. Nat. Soc.*, vol. 2015, pp. 1–9, Mar. 2015.

[6] T. Dahlberg, Some railroad settlement models—A critical review, *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit*, vol. 215, no. 4, pp. 289–300, Jul. 2001.

[7] A. R. Andrade and P. F. Teixeira, A Bayesian model to assess rail track geometry degradation through its life-cycle, *Res. Transp. Econ.*, vol. 36, no. 1, pp. 1–8, Sep. 2012.

[8] A. R. Andrade and P. F. Teixeira, Statistical modelling of railway track geometry degradation using Hierarchical Bayesian models, *Reliab. Eng. Syst. Saf.*, vol. 142, pp. 169–183, Oct. 2015.

[9] J. Chiachío, M. Chiachío, D. Prescott, and J. Andrews, A Bayesian assessment for railway track geometry degradation prognostics, *Proc. Eur. Conf. PHM Soc. PHM Soc. Eur. Conf. 4 . PHM Soc. New York.*, pp. 1–7, 2018.

[10] S. Paul, Erina; Basu, Approximate Bayesian Computation in Nonparametric Bayesian Models, Northern Illinois University, 2017.

[11] D. B. Rubin, Bayesianly Justifiable and Relevant Frequency Calculations for the Applied

Statistician, *Ann. Stat.*, vol. 12, no. 4, pp. 1151–1172, 2007.
[12] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman, Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites, *Mol. Biol. Evol*, vol. 16, no. 12, pp. 1791–1798, 1999.
[13] A. L. Kursanov, Metabolism and the Transport of Organic Substances in the Phloem, *Adv. Bot. Res.*, vol. 1, no. C, pp. 209–278, 1963.
[14] J. Akeret, A. Refregier, A. Amara, S. Seehars, and C. Hasner, Approximate Bayesian computation for forward modeling in Cosmology, *J. Cosmol. Astropart. Phys.*, vol. 2015, no. 8, 2015.
[15] X. Wu, L. Marshall, and A. Sharma, Approximation Bayesian computation for calibration of water quality models considering the impact of observation uncertainty, *Am. Geophys. Union, Fall Meet. 2018*, pp. 24–26, 2018.
[16] A. Ben Abdessalem, N. Dervilis, D. Wagg, and K. Worden, Model selection and parameter estimation in structural dynamics using approximate Bayesian computation, *Mech. Syst. Signal Process.*, vol. 99, pp. 306–325, Jan. 2018.
[17] J. D. Christopher, Approximate Bayesian Computation for Parameter Estimation in Complex Thermal-Fluid Systems, 2018.
[18] B. M. Turner and T. Van Zandt, A tutorial on approximate Bayesian computation, *J. Math. Psychol.*, vol. 56, no. 2, pp. 69–85, 2012.
[19] R. M. Nosofsky, D. R. Little, C. Donkin, and M. Fific, Short-term memory scanning viewed as exemplar-based categorization., *Psychol. Rev.*, vol. 118, no. 2, pp. 280–315, 2011.
[20] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder, Approximate Bayesian computational methods, *Stat Comput*, vol. 22, pp. 1167–1180, 2012.
[21] P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré, Markov chain Monte Carlo without likelihoods, 2003.
[22] D. Van Ravenzwaaij, P. Cassey, and S. D. Brown, A simple introduction to Markov Chain Monte-Carlo sampling, *Psychon Bull Rev*, vol. 25, pp. 143–154, 2018.
[23] D. Wegmann, C. Leuenberger, and L. Excoffier, Efficient Approximate Bayesian Computation Coupled With Markov Chain Monte Carlo Without Likelihood.
[24] S. Aeschbacher, M. A. Beaumont, and A. Futschik, A Novel Approach for Choosing Summary Statistics in Approximate Bayesian Computation, *Genetics*, vol. 192, no. 3, pp. 1027–1047, Nov. 2012.
[25] A. Kousathanas, P. Duchen, and D. Wegmann, A Guide to General-Purpose Approximate Bayesian Computation Software, Jun. 2018.
[26] G. Hamra, R. MacLehose, and D. Richardson, Markov Chain Monte Carlo: an introduction for epidemiologists, *Int. J. Epidemiol.*, vol. 42, no. 2, pp. 627–634, Apr. 2013.
[27] R. McElreath, Statistical rethinking: A bayesian course with examples in r and stan., *Chapman; Hall/CRC.*, 2015.

ABOUT THE AUTHORS

Grace Ashley

Grace Ashley is a PhD candidate in the Civil and Environmental Engineering Department at the University of Delaware. She obtained her Master's degree in Civil Engineering from Louisiana State University. She served as the Vice President of the Institute of Transportation Engineers (LSU student Chapter) and is the current Civil Engineering Department Representative for Women in Engineering (WiE). She is a reviewer for Transportation Research Part F: Psychology and Behavior and ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. She is presently a member of Prof. Attoh-Okine's research group. She is currently specializing in Railway Engineering for her doctoral studies and her interests lie in the Application of Advanced Data Techniques in the Analysis of Track Geometry Data.

Nii O. Attoh-Okine, Ph.D., P.E., F. ASCE, Snr Member IEEE

Nii O. Attoh-Okine, Professor of Civil and Environmental Engineering, and Electrical and Computer Engineering. He is also the Interim Academic Director of the University of Delaware Cybersecurity Initiative. In the last couple of years, he has authored two books which are defining the direction of research across disciplines: a) Resilience Engineering: Models and Analysis and b) Big Data and Differential Privacy in Railway Track Engineering. He is a founding associate editor for ASCE/ASME Journal of Risk and Uncertainty Analysis. He has served as an Associate Editor on the four ASCE Journals. Attoh-Okine is currently a member of a group of researchers from the United States and Japan working on Smart Cities and various cyber issues related to the Tokyo 2020 Olympic Games.