November 2006

BAYESIAN ANALYSIS OF PULSE TRAINS WITH HIDDEN MISSINGNESS

Melinda Hock
Naval Research Laboratory
Washington, DC 20375

Refik Soyer*
Department of Decision Sciences
The George Washington University
Washington, DC 20052

Abstract

In this paper we present a Bayesian approach for analysis of a pulse train that is corrupted by noise and missing pulses at unknown locations. The existence of missing pulses at unknown locations complicates the analysis and model selection process. This type of *hidden missingness* in the pulse data is different than the usual missing observations problem that arise in time-series analysis where standard methodology is available. We develop Bayesian methodology for dealing with the hidden missingness. Our development is based on Markov chain Monte Carlo methods and involves both inference and model selection.

1. Introduction

The problem considered in this paper is motivated by electronic warfare applications where a receiver tracks pulses emitted by a radar [see for example, Clarkson, Howard and Mareels (1996)]. The time interval between two pulses emitted by the radar is defined as a pulse repetition interval (PRI). Many radars are designed so that the sequence of PRIs varies in an identifiable pattern. The time-series of the PRIs is referred to as a *pulse train*. Our objective in this paper is to develop Bayesian methods for

*Contact author: Department of Decision Sciences, Funger Hall 415, The George Washington University, Washington, DC 20052, USA. E-mail: soyer@gwu.edu

analysis of pulse trains and for identification of the type of PRI modulation present on an isolated pulse train. There are many possible types of PRI modulations. In this paper we consider staggered PRI modulation (several different PRIs in a repeating pattern) and jittered PRI modulation (random variation in PRI about a mean value).

In many applications due to the complexity of the environment, sensitivity of the receiver may be reduced and this results in missing pulses in the observed data. As pointed out by Clarkson, Howard and Mareels (1996), for certain types of receivers this is a very common problem and there exists very little work on estimation of the period of such pulse trains. The presence of missing pulses complicates the analysis and identification of the PRI modulation type because the location of missing pulses is unknown. This type of *hidden missingness* in the pulse data is different than the usual missing observations problems that arise in statistics [see for example, Little and Rubin (1987)] or more specifically in time-series analysis [see for example, West and Harrison (1997)] where methodologies are available. Thus, there is a need for development of new methodology for dealing with the hidden missingness. Since it is never known whether a pulse (or a number of pulses) has been missed by the receiver in an interval, the methodology should be able to infer about missing pulses based on the observed data.

In this paper, a formal Bayesian approach is presented to describe hidden missingness in pulse trains with two PRI modulation types. More specifically the stagger and jitter modulation types are considered and their Bayesian analyses are developed with hidden missingness using Markov chain Monte Carlo methods. Bayesian model selection methods also are developed to identify the appropriate PRI modulation type based on an observed pulse train subjected to hidden missingness.

Bayesian analysis of pulse trains with jitter type PRI modulation with hidden missingness presents a structure similar to the *hidden Markov models* which were considered by a host of authors such as Robert, Celeux and Diebolt (1993) for analyzing mixture models and by Chib (1996) for analyzing change-points in time-series models.

The hidden Markov type models have been used also in signal processing literature by Doucet and Duvaut (1997) where authors considered two-component mixtures. None of these earlier applications of hidden Markov models have considered the issue of hidden missingness. The Bayesian analysis developed here for stagger type PRI modulation with hidden missingness represents a contribution to the Bayesian spectral time-series literature. Previous Bayesian developments in spectral time-series analysis like Bretthorst (1988) and West (1995) have not dealt with the hidden missingness issue. Furthermore, a Bayesian model selection approach is presented for time-series with hidden missingness based on marginal likelihoods as discussed in Kass and Raftery (1995) and Chib (1995). For the stagger and jitter type PRI modulation, the marginal likelihoods under hidden missingness are computed using approximations suggested by Chib (1995). Adoption of such methods to time-series with hidden missingness and their implementation, to the best of our knowledge, are new. Thus, the novelty of our work is in the new application as well as the adoption of the Markov chain Monte Carlo methods and the Bayesian model selection approach of Chib (1995) to this application.

A synopsis of our paper is as follows: in Section 2 we present the notation and preliminaries and discuss the hidden missingness for the two PRI modulation types. In so doing, we also present a Markov chain model to describe the missingness structure. In Section 3, we develop the Bayesian inference for the models using Markov chain Monte Carlo methods, more specifically using a Gibbs sampler. The Bayesian model selection approach is discussed in Section 4 and an illustration of the developed methodology using simulated data is presented in Section 5. Conclusions and directions for future research are given in Section 6.

2. PRI Modulations and Modeling Hidden Missingness

2.1 Notation and Preliminaries

The data available for the analysis consists of a batch of observed PRI values. The pulse train of say, n, observed PRIs are obtained by taking the difference between consecutive measured pulse time of arrivals (TOAs). Let $\{\tau_0, \tau_1, \ldots, \tau_n\}$ denote the observed TOA sequence and $\{z_1, \ldots, z_n\}$ denote the corresponding observed PRI sequence such that $z_i = \tau_i - \tau_{i-1}$ for $i = 1, \ldots, n$. Both the TOA values and the observed PRI values are affected by missing pulses and noise.

Since there are missing pulses, to distinguish between the PRI values transmitted by the radar and the values observed by the receiver, we denote the transmitted PRI values by the sequence $\{y_{k_i - j}\}$, for $i = 1, \ldots, n$ and $j = 0, \ldots, m_i$ where m_i is the number of missing pulses between the $(i-1)st$ and ith observed pulses. In $\{y_{k_i - j}\}$, $(k_i - j)$ denotes the transmitted pulse index whereas i denotes the observed pulse index such that $(k_i - j) \geq i$ with equality holding only for the case of no missing observations in the pulse train. Also, note that $k_n \geq n$. Similarly, the TOA sequence associated with the transmitted pulses will be denoted by $\{t_{k_i - j}\}$,$i = 1$,...,n, and $j = 0, \ldots, m_i$, where $y_{k_i - j} = t_{k_i - j} - t_{k_i - j - 1}$. If there are no missing pulses in a given interval, the observed PRI value of the interval will be equivalent to the corresponding PRI value from the transmitted pulse train, that is, $z_i = y_{k_i}$.

If a pulse is missing in an interval from the incoming data stream, the next observed PRI value represents the sum of two consecutive PRI values in the transmitted sequence rather than a single PRI value. Similarly, if two consecutive pulses are missing from the incoming data stream, the following observed PRI value represents the sum of three consecutive PRI values. In many applications the value of m_i is small, typically 1 or 2. In general when missing pulses corrupt the data set, the observed PRIs are expressed as:

$$z_i = \sum_{j=0}^{m_i} y_{k_i - j} \, . \tag{1}$$

A sum of two or more transmitted PRI values will be referred to as *aggregate data*. For example, in Figure 1, the transmitted sequence consists of $\{y_1, y_2, y_3, y_4, y_5\}$. The observed sequence has a missing pulse in the third interval and as a result the observed sequence is obtained as $z_1 = y_1$, $z_2 = y_2$, $z_3 = y_3 + y_4$ and $z_4 = y_5$.

Although pulse width is a parameter that is commonly measured along with PRI, in our application only PRI is available. Therefore, pulse width is not used in the model and in the following development. Our only assumption about pulse width is that it is less than the PRI. More specifically, TOAs are measured at the leading edge of each pulse and PRI is defined as the time difference between adjacent TOAs.

In the observed data $\{z_1, \ldots, z_n\}$, each data point may represent either a single PRI value or an aggregate PRI value. The number of PRI values included in a data point will be referred to as its state. The states are unobservable *latent variables* associated with each data point. We denote the states by $\{s_1, \ldots, s_n\}$. Since there are m_i missing pulses in the ith observed pulse interval, for data point i, $s_i = m_i + 1$. State $s_i = 1$ implies that the ith data point represents a single PRI value, that is, there is no missing pulse in the ith interval. The state $s_i = 2$ implies that the ith data point represents the sum of two consecutive PRI values from the transmitted pulse sequence. It follows from the above that transmitted pulse index for the ith observation is $k_i = \sum_{j=1}^{i} s_j$. For example, in Figure 1, $s_1 = s_2 = s_4 = 1$ and $s_3 = 2$ implying that $k_1 = 1$, $k_2 = 2$, $k_3 = 4$, and $k_4 = 5$.

We note that when the receiver observes a sequence of n PRIs that are corrupted by missing pulses, the actual transmitted sequence will consist of $\sum_{i=1}^{n} s_i$ PRI values, where $n \leq \sum_{i=1}^{n} s_i$.

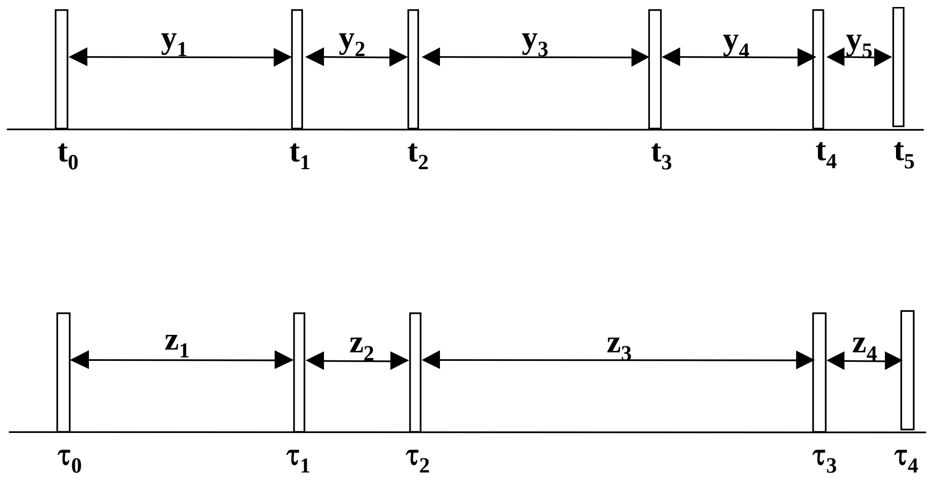

Figure 1. Effect of a single missing pulse on the observed data.

2.2 PRI Modulations and Hidden Missingness

In our development, we consider two types of PRI modulation: staggered PRI modulation and jittered PRI modulation where the latter is a special case of the former. Under the staggered PRI modulation, it is assumed that the transmitted PRIs form a discrete time-series generated by a periodic function which can be represented by a finite number of harmonics. Specifically, if a discrete time series has period p, it can be described as

$$a_0 + \sum_{i=1}^{q} a_i cos(2\pi \frac{i}{p} k) + b_i \, sin(2\pi \frac{i}{p} k) \tag{2}$$

where $k = 1,2,\ldots$ is the time index, a_i's and b_i's are the unknown coefficients and

$$q = \begin{cases} p/2 & \text{for } p \text{ even} \\ (p-1)/2 & \text{for } p \text{ odd.} \end{cases}$$

In (2) the first harmonic has frequency $1/p$ and completes its cycle in p time periods. If p is even, at most $p/2$ harmonics are required to represent the time series because the period corresponding to the $(p/2)\,th$ harmonic is 2 which is the shortest possible cycle length. If p is odd, at most $(p-1)/2$ harmonics are needed.

Thus, a staggered PRI sequence that is subjected to noise can be represented as

$$y_k = a_0 + \sum_{i=1}^{q} a_i\, cos(2\pi \frac{i}{p} k) + b_i\, sin(2\pi \frac{i}{p} k) + w_k, \tag{3}$$

where w_k's are uncorrelated noise terms which are normally distributed with mean 0 and unknown variance σ_w^2 and the last sine term reduces to zero when p is even. In what follows, we will refer to (3) as the stagger model. The jittered PRI modulation is a special case of (3) where the transmitted PRIs randomly vary around a constant mean, that is,

$$y_k = a_0 + w_k. \tag{4}$$

We will refer to (4) as the jitter model.

For an observed PRI sequence $\{z_1,\ldots,z_n\}$, corrupted with missing pulses, if the sequence of states $\{s_1,\ldots,s_n\}$ are known for all $i = 1,\ldots,n$, then using $m_i = s_i - 1$ in (1), the time-series of the observed PRIs can be aggregated as $z_i = \sum_{j=0}^{s_i-1} y_{k_i-j}$. For example, in the jitter model, (4) reduces to

$$z_i = \sum_{j=0}^{s_i-1} (a_0 + w_{k_i-j}) = s_i a_0 + u_i, \tag{5}$$

where $u_i = \sum_{j=0}^{s_i-1} w_{k_i-j}$. Note that u_i's do not share any common jitter noise terms and thus they are independent Gaussian terms with mean 0 and variance $s_i\sigma_w^2$. Similarly, for the stagger model with period p, using (3) we can write

$$z_i = \sum_{j=0}^{s_i-1} \Big[a_0 + \sum_{l=1}^{q} a_l\, cos\Big(2\pi \frac{l}{p}(k_i - j)\Big) + b_l\, sin\Big(2\pi \frac{l}{p}(k_i - j)\Big)\Big] + u_i, \tag{6}$$

where the jitter model is obtained as the special case with $q = 0$.

Thus, given the latent sequence $\{s_1,\ldots,s_n\}$, under the stagger model, the observed PRIs can be written as a linear model in matrix notation as

$$\boldsymbol{z} = \boldsymbol{X\theta} + \boldsymbol{u}, \tag{7}$$

where $\boldsymbol{z}$ is an $n \times 1$ vector of the observed PRIs corrupted with missing pulses, $\boldsymbol{u}$ is the $n \times 1$ vector of independent Gaussian noise terms, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown parameters, that is, $\boldsymbol{\theta'} = (a_0\, a_1\, b_1 \ldots a_q\, b_q)$ and $\boldsymbol{X}$ is the $n \times p$ design (regression) matrix that consists of the sine-cosine terms of (6), given by

$$\boldsymbol{X} = \begin{bmatrix} s_1 & \sum_{j=0}^{s_1-1} cos\left(2\pi \frac{1}{p}(k_1 - j)\right) & \sum_{j=0}^{s_1-1} sin\left(2\pi \frac{1}{p}(k_1 - j)\right) & \cdots & \sum_{j=0}^{s_1-1} cos\left(2\pi \frac{q}{p}(k_1 - j)\right) & \sum_{j=0}^{s_1-1} sin\left(2\pi \frac{q}{p}(k_1 - j)\right) \\ s_2 & \sum_{j=0}^{s_2-1} cos\left(2\pi \frac{1}{p}(k_2 - j)\right) & \sum_{j=0}^{s_2-1} sin\left(2\pi \frac{1}{p}(k_2 - j)\right) & \cdots & \sum_{j=0}^{s_2-1} cos\left(2\pi \frac{q}{p}(k_2 - j)\right) & \sum_{j=0}^{s_2-1} sin\left(2\pi \frac{q}{p}(k_2 - j)\right) \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ s_n & \sum_{j=0}^{s_n-1} cos\left(2\pi \frac{1}{p}(k_n - j)\right) & \sum_{j=0}^{s_n-1} sin\left(2\pi \frac{1}{p}(k_n - j)\right) & \ldots & \sum_{j=0}^{s_n-1} cos\left(2\pi \frac{q}{p}(k_n - j)\right) & \sum_{j=0}^{s_n-1} sin\left(2\pi \frac{q}{p}(k_n - j)\right) \end{bmatrix} \tag{8}$$

where the case of no missing pulses can be obtained by setting $s_i = 1$ for all $i = 1,\ldots,n$. Note that $k_i = \sum_{j=1}^{i} s_j$ and thus, the elements of the design matrix are functions of the latent variables. In the case of the jitter model the matrix reduces to one column.

In both the stagger and the jitter models where the PRIs are corrupted by missing pulses, the aggregate data can be constructed given the latent sequence $\{s_i\}, i = 1,\ldots n$. As a result, all predictions for future PRIs will be conditional on the latent variables describing the missingness structure. Thus, describing the probabilistic structure of the latent sequence is essential for modeling and forecasting of the PRIs. One strategy to model $\{s_i\}$ is to assume that this categorical time series is independent over time, but this is rather restrictive for modeling purposes since it ignores any possible correlation in the s_i's over time. A more flexible and a reasonable modeling strategy for $\{s_i\}$ is to assume that the latent sequence evolves according to a first-order Markov chain, with transition matrix $\boldsymbol{P}$, that is,

$$p(s_i|s_{i-1},s_{i-2},\ldots) = p(s_i|s_{i-1}), \tag{9}$$

where $\boldsymbol{P} = \{p_{kj}\}, k, j = 1, 2, \ldots, M+1$, is the transition probability matrix of the Markov chain, that is, $p_{kj} = Pr(s_i = j | s_{i-1} = k)$. We assume that the Markov chain is time-homogeneous. Note that in our setup M represents the upper bound for the number of missing pulses in an interval. As previously noted M is a small number for typical applications and thus the dimension of the transition matrix will not be high. We note that a higher-order Markov model can also be specified in (9) and the development presented in Sections 3 and 4 can be modified accordingly.

If the observed PRIs follow the stagger model of (6) with the Markov chain structure on the latent sequence $\{s_i\}$, we can write

$$p(z_i | \boldsymbol{\theta}, \sigma_w^2, \boldsymbol{S}_{i-1}) = \sum_{l=1}^{M+1} p(z_i | \boldsymbol{\theta}, \sigma_w^2, \boldsymbol{S}_{i-1}, s_i = l) p(s_i = l | s_{i-1}), \tag{10}$$

where $\boldsymbol{S}_{i-1} = (s_1, \ldots, s_{i-1})$ and $(z_i | \boldsymbol{\theta}, \sigma_w^2, \boldsymbol{S}_{i-1}, s_i = l)$ follows the normal model given by (7). This type of model where the latent sequence is described by a Markov chain is referred to as a *hidden Markov* model. Examples of such models can be found in engineering, econometrics and statistics literature. For example, hidden Markov models have been used in engineering literature for speech recognition [Rabiner and Juang (1985)], in econometrics for modeling change points of a time-series [Chib (1996)] and in statistics for analysis of mixture models [Robert, Celeux and Diebolt (1993)].

3. Bayesian Analysis

For the stagger model with period p, the observed PRIs are given by (6). The Bayesian approach requires that we specify our uncertainty about the unknown coefficient vector $\boldsymbol{\theta}$ and the unknown variance σ_w^2 in the linear model setup (7) by a prior probability distribution. Under the normality of u_i's, which are uncorrelated noise terms with zero mean and variance $s_i \sigma_w^2$, a conjugate prior for (7) is given by the *normal-*

gamma distribution

$$\boldsymbol{\theta}|\phi \sim \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{m},\boldsymbol{V}/\phi), \tag{11a}$$

$$\phi \sim \mathcal{G}(\phi|\, d/2,c/2). \tag{11b}$$

In (11) $\phi = 1/\sigma_w^2$ is the unknown precision of u_i's, $\boldsymbol{m}$ is a $p \times 1$ specified mean vector and $\boldsymbol{V}$ is a specified variance-covariance matrix of $\boldsymbol{\theta}$ and d and c are specified prior parameters for the distribution of ϕ. The joint distribution of $\boldsymbol{\theta}$ and ϕ which is obtained as the product of the (11a) and (11b) is known as the normal-gamma distribution and implies that the marginal prior of $\boldsymbol{\theta}$ is a multivariate student-t density with degrees of freedom $(d + p)$, mean vector $\boldsymbol{m}$ and scale matrix $\frac{c}{d}\boldsymbol{V}$. It is assumed that apriori $(\boldsymbol{\theta}, \phi)$ are independent of the latent variables s_i's.

For the observed PRI sequence $\boldsymbol{z}' = (z_1 \ldots z_n)$, given $(\boldsymbol{\theta},\phi)$ and the latent variables $\boldsymbol{S}_n = (s_1,\ldots,s_n)$, the distribution of $\boldsymbol{z}$ will be a multivariate normal density given by

$$(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n) \sim \mathcal{N}(\boldsymbol{z}|\,\boldsymbol{X\theta},\boldsymbol{U}/\phi), \tag{12}$$

where $\boldsymbol{X}$ is given by (8) and

$$\boldsymbol{U} = \begin{bmatrix} s_1 & 0 & \cdots\cdots & 0 \\ 0 & s_2 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & s_n \end{bmatrix},$$

implying that z_i's have different variances. Given the latent sequence $\boldsymbol{S}_n$, the Bayesian analysis of the model can be made using conjugate analysis to obtain the posterior distribution $p(\boldsymbol{\theta},\phi|\boldsymbol{S}_n,\boldsymbol{z})$. Since $\boldsymbol{S}_n$ is not observed, the Bayesian analysis requires the joint posterior distribution $p(\boldsymbol{\theta},\phi,\boldsymbol{S}_n|\boldsymbol{z})$ given by

$$p(\boldsymbol{\theta},\phi,\boldsymbol{S}_n|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)\, p(\boldsymbol{\theta},\phi)\; p(\boldsymbol{S}_n) \tag{13}$$

due to the prior independence of $(\boldsymbol{\theta},\phi)$ and $\boldsymbol{S}_n$. Since $\boldsymbol{S}_n$ follows a Markov chain with unknown transition matrix $\boldsymbol{P}$, the joint posterior distribution required for the Bayesian analysis is

$$p(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P}|\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)\, p(\boldsymbol{\theta},\phi)\; p(\boldsymbol{S}_n|\boldsymbol{P})\, p(\boldsymbol{P}), \tag{14}$$

where $p(\boldsymbol{P})$ is the prior distribution for the transition matrix $\boldsymbol{P}$ of the Markov chain. It is assumed that $\boldsymbol{P}_i$, the ith row of $\boldsymbol{P}$ follows a Dirichlet distribution

$$p(\boldsymbol{P}_i) \propto \prod_{j=1}^{M+1} p_{ij}^{\alpha_{ij}-1}\ , i = 1,\ldots,M+1 \tag{15}$$

with parameters α_{ij} denoted as $\boldsymbol{P}_i \sim \mathcal{D}(\alpha_{ij}; j = 1,\ldots,M+1)$. Also, $\boldsymbol{P}_i$'s are assumed to be independent for all i and furthermore $\boldsymbol{P}$ is assumed to be independent of $(\boldsymbol{\theta},\phi)$.

Under any choice of prior distributions, the joint posterior distribution $p(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P}|\boldsymbol{z})$ in (14) can not be evaluated analytically. Such evaluation is also computationally infeasible due to the presence of latent sequence $\boldsymbol{S}_n$. Each latent variable s_i may belong to one of the $M+1$ categories implying that there are $(M+1)^n$ possible latent sequences for n observed pulses. Thus, in what follows, a Markov chain Monte Carlo approach or more specifically a Gibbs sampler will be presented to generate samples from the joint posterior distribution (14). The attractive feature of the Gibbs sampler is that it enables the generation of samples from the posterior distributions without having to obtain the exact distributional forms. This is achieved by successive drawings from the *full conditional distributions* of $\boldsymbol{\theta}$, ϕ, $\boldsymbol{S}_n$ and $\boldsymbol{P}$ given $\boldsymbol{z}$. For a more detailed discussion of the Gibbs sampler and other related Monte Carlo methods, see Gelfand and Smith (1990).

In the stagger model, the implementation of the Gibbs sampler requires the full conditional distributions $p(\boldsymbol{\theta}|\phi, \boldsymbol{S}_n, \boldsymbol{P}, \boldsymbol{z})$, $p(\phi|\,\boldsymbol{\theta}, \boldsymbol{S}_n, \boldsymbol{P},\boldsymbol{z})$, $p(\boldsymbol{P}|\boldsymbol{\theta}, \phi, \boldsymbol{S}_n, \boldsymbol{z})$, and $p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z})$ for $i = 1,\ldots,n$, where $\boldsymbol{S}^{-i} = \{s_j|\, j \neq i,\, j = 1,2,....,n\}$.

3.1 Full Conditional Distributions of $\boldsymbol{S}_n$ and $\boldsymbol{P}$

From (14) it can be seen that, given $\boldsymbol{S}_n$, $\boldsymbol{z}$ does not depend on the transition matrix $\boldsymbol{P}$. Thus, learning about $\boldsymbol{P}$ is through the updating of $\boldsymbol{S}_n$ based on observed data $\boldsymbol{z}$. This implies that the full conditional distribution of $\boldsymbol{P}$ is $p(\boldsymbol{P}|\boldsymbol{\theta}, \phi, \boldsymbol{S}_n, \boldsymbol{z}) = p(\boldsymbol{P}|\boldsymbol{S}_n)$. Furthermore, given $\boldsymbol{S}_n$ the full conditional distributions of $\boldsymbol{\theta}$ and ϕ are not dependent on $\boldsymbol{P}$, that is, $p(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n,\boldsymbol{P},\boldsymbol{z}) = p(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n,\boldsymbol{z})$ and $p(\phi|\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{P},\boldsymbol{z}) = p(\phi|\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z})$.

The latent sequence $\boldsymbol{S}_n$ follows a Markov chain with a state space of dimension $M+1$, where M is the maximum possible number of missing pulses in an interval. We can write down the full conditional distribution of s_i as

$$p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z}) \propto p(\boldsymbol{z}|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P})\, p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P}). \tag{16}$$

As pointed out above the first term on the right hand side of the (16) does not depend on the transition matrix $\boldsymbol{P}$, that is, $p(\boldsymbol{z}|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P}) = p(\boldsymbol{z}|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i})$. Furthermore, due to the conditional independence of z_i's given s_i, $\boldsymbol{\theta},\phi$, and $\boldsymbol{S}^{-i}$, we can write

$$p(\boldsymbol{z}|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i}) = \prod_{j=1}^{n} p(z_j|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i}).$$

It follows from (6) that in the above $p(z_j|s_i,\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i}) = p(z_j|\boldsymbol{\theta},\phi, \boldsymbol{S}_j)$ where $\boldsymbol{S}_j = (s_1,s_2,\ldots,s_j)$. Thus, for $j < i$ observed z_j's do not provide any information about s_i and the full conditional distribution of s_i reduces to

$$p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z}) \propto \prod_{j=i}^{n} p(z_j|\boldsymbol{\theta},\phi, \boldsymbol{S}_j)\, p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P}). \tag{17}$$

Due to the prior independence of $(\boldsymbol{\theta},\phi)$, and s_i's, the second term on the right-hand side of (17) can be written as $p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P}) = p(s_i|\boldsymbol{S}^{-i},\boldsymbol{P})$ which yields

$$p(s_i|\boldsymbol{\theta},\phi, \boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z}) \propto \prod_{j=i}^{n} p(z_j|\boldsymbol{\theta},\phi, \boldsymbol{S}_j)\, p(s_i|\boldsymbol{S}^{-i},\boldsymbol{P}). \tag{18}$$

Note that $p(s_i|\,\boldsymbol{S}^{-i},\boldsymbol{P})$ is proportional to $p(\boldsymbol{S}_n|\boldsymbol{P})$ which can be obtained as

$$p(\boldsymbol{S}_n|\boldsymbol{P}) = p(s_1,s_2,\ldots s_n|\boldsymbol{P}) = p(s_1|\boldsymbol{P})\,p(s_2|s_1,\boldsymbol{P})\cdots\cdots p(s_n|s_{n-1},\boldsymbol{P})$$

using the Markov property. Thus, for $2 \leq i \leq n-1$

$$p(s_i|\,\boldsymbol{S}^{-i},\boldsymbol{P}) \propto\; p(s_i|s_{i-1},\boldsymbol{P})\,p(s_{i+1}|s_i,\boldsymbol{P}) \tag{19}$$

and (18) reduces to

$$p(s_i|\boldsymbol{\theta},\phi,\boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z}) \;\propto p(s_i|s_{i-1},\boldsymbol{P})\,p(s_{i+1}|s_i,\boldsymbol{P})\prod_{j=i}^{n} p(z_j|\boldsymbol{\theta},\phi,\boldsymbol{S}_j). \tag{20}$$

In (20) the terms $p(z_j|\boldsymbol{\theta},\phi,\boldsymbol{S}_j)$ in the product are normal densities as implied by (6). Since s_i's are discrete random variables (20) can be easily normalized by summing over all possible values of s_i. Similar forms can also be obtained for the special cases of $i=1$ and $i=n$.

For example, if $n=4$, then the full conditional distribution of s_2 is given by

$$p(s_2|\boldsymbol{\theta},\phi,s_1,s_3,s_4,\boldsymbol{P},\boldsymbol{z}) \;\propto p(s_2|s_1,\boldsymbol{P})\,p(s_3|s_2,\boldsymbol{P})\,p(z_2|\boldsymbol{\theta},\phi,s_1,s_2)$$
$$\times\, p(z_3|\boldsymbol{\theta},\phi,s_1,s_2,s_3)\,p(z_4|\boldsymbol{\theta},\phi,s_1,s_2,s_3,s_4).$$

For the case of the jitter model where $\boldsymbol{\theta}=a_0$ is a scalar, the full conditional of s_i reduces to

$$p(s_i|a_0,\phi,\boldsymbol{S}^{-i},\boldsymbol{P},\boldsymbol{z}) \propto p(s_i|s_{i-1},\boldsymbol{P})\,p(s_{i+1}|s_i,\boldsymbol{P})\;p(z_i|a_0,\phi,s_i), \tag{21}$$

where $p(z_i|a_0,\phi,s_i)$ is a normal density with mean $s_i a_0$ and variance s_i/ϕ as implied by (5). Unlike the stagger model, in this case full conditional of s_i depends only on the neighboring latent variables s_{i-1} and s_{i+1}. This is due to the fact that in the jitter model only the current observation z_i gives information on s_i as shown by the likelihood term $p(z_i|a_0,\phi,s_i)$ in (21).

As previously pointed out, the full conditional distribution of the transition matrix $\boldsymbol{P}$ is dependent only on $\boldsymbol{S}_n$, that is, $p(\boldsymbol{P}|\boldsymbol{\theta},\,\phi,\,\boldsymbol{S}_n,\,\boldsymbol{z}) = p(\boldsymbol{P}|\,\boldsymbol{S}_n)$. Given the independent Dirichlet priors on each row $\boldsymbol{P}_i \sim \mathcal{D}(\alpha_{ij}; j = 1,\ldots,M+1)$, the full conditional distribution of each row can be obtained as a Dirichlet of the form

$$(\boldsymbol{P}_i|\,\boldsymbol{S}_n) \sim \mathcal{D}(\alpha_{ij} + \sum_{k=1}^{n} 1(s_{k+1} = j, s_k = i)\,; j = 1,\ldots,M+1), \tag{22}$$

where $1(\bullet)$ is an indicator function and given $\boldsymbol{S}_n$, the $\boldsymbol{P}_i$'s, $i = 1,\ldots M+1$, are also independent.

3.2 Full Conditional Distributions of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$

As pointed out in the above, the full conditional distributions of $\boldsymbol{\theta}$ and ϕ do not depend on $\boldsymbol{P}$ given $\boldsymbol{S}_n$. Thus, we need $p(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n,\boldsymbol{z})$ and $p(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z})$. It can be shown that the full conditional of $\boldsymbol{\theta}$ is

$$(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n,\boldsymbol{z}) \sim \mathcal{N}(\boldsymbol{\theta}|\,\boldsymbol{m}^*,\boldsymbol{V}^*/\phi), \tag{23}$$

where $\boldsymbol{m}^*$ and $\boldsymbol{V}^*$ are given in the appendix. Note that the results given in the appendix are similar to the Bayesian analysis presented by Gelman et. al. (1995, pp. 255) for regression models with unequal variances or with *heteroskedasticity*.

The full conditional distribution of ϕ, $p(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z})$ is given by

$$p(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)\,p(\phi|\boldsymbol{\theta}) \propto p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)\,p(\boldsymbol{\theta}|\phi)p(\phi). \tag{24}$$

Using standard conjugate analysis, it can be shown that the full conditional distribution is a gamma density, that is, $(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z}) \sim \mathcal{G}(\phi\,|\,d^*/2, c^*/2)$, where d^* and c^* are given in the appendix.

4. Model Comparison

In Bayesian paradigm, model comparison/selection is typically based on Bayes factors which are obtained as the ratio of marginal likelihoods under two competing models; see Kass and Raftery (1995) for a comprehensive review. For example, if we have two altenative models associated with two different periods in the stagger model, then the Bayes factor is obtained as the ratio of marginal likelihood $p^1(\boldsymbol{z})$ to $p^2(\boldsymbol{z})$ where $p^i(\boldsymbol{z})$ is the marginal likelihood under model i. In many problems $p^i(\boldsymbol{z})$ is not available in an analytical form and its evaluation using posterior Monte Carlo samples is not a trivial task. Thus, various alternatives to marginal likelihoods have been suggested in the literature for model selection using Monte Carlo samples; see Gelfand (1996) for a recent review.

However, in certain problems where a Gibbs sampler is used and all the full conditional distributions are known, it is possible to approximate the marginal likelihoods from the posterior samples using the approach introduced by Chib (1995). We note that there are other model selection approaches such as the one presented in Green (1995). This approach which is based on reversible jump Markov chain Monte Carlo methods provides posterior probabilities of the candidate model. Another approach that provides posterior model probabilities using Markov chain Monte Carlo is presented in Carlin and Chib (1995). A comprehensive review of these and other approaches is given in Han and Carlin (2001).

As discussed in Section 3, in Bayesian analyses of the stagger and jitter models with hidden missingness all the full conditionals are known. Thus, in what follows the approach proposed by Chib (1995) will be adopted to our problem. Chib's approach is based on two ideas. First, the marginal likelihood for a particular model is expressed as

$$p(\boldsymbol{z}) = \frac{p(\boldsymbol{z}|\Theta)p(\Theta)}{p(\Theta|\boldsymbol{z})}, \tag{25}$$

where Θ is a vector of parameters. Secondly, as pointed out by Chib the above holds for any value of Θ, say Θ^*, and the value of posterior density $p(\Theta^*|\boldsymbol{z})$ can be estimated by $\widehat{p}(\Theta^*|\boldsymbol{z})$ using Monte Carlo samples. Since $p(\boldsymbol{z}|\Theta^*)$ and $p(\Theta^*)$ can be evaluated at Θ^*, the log marginal likelihood can be estimated as

$$ln\,\widehat{p}(\boldsymbol{z}) = ln\,p(\boldsymbol{z}|\Theta^*) \ + ln\,p(\Theta^*) - \ \widehat{p}(\Theta^*|\boldsymbol{z}). \tag{26}$$

In evaluating (26), the only term which is not readily available is $\widehat{p}(\Theta^*|\boldsymbol{z})$, but as shown in Chib (1995) this can be obtained using the outputs from the Gibbs sampler.

In our case Θ consists of $(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P})$ and using independence assumptions we can write the marginal likelihood as

$$p(\boldsymbol{z}) = \frac{p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)\,p(\boldsymbol{\theta}|\phi)p(\phi)\,p(\boldsymbol{S}_n|\boldsymbol{P})\,p(\boldsymbol{P})}{p(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P}|\boldsymbol{z})}\,. \tag{27}$$

All the terms in the numerator of (27) can be evaluated at $(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P}) = (\boldsymbol{\theta}^*,\phi^*,\boldsymbol{S}_n^*,\boldsymbol{P}^*)$. We note that (27) holds for any value of $(\boldsymbol{\theta},\phi,\boldsymbol{S}_n,\boldsymbol{P})$, but, as pointed out in Chib (1995), (27) can be more accurately approximated by evaluating it at a high density point. Thus, the posterior modes, that can be easily approximated from the Gibbs output, will be used in $(\boldsymbol{\theta}^*,\phi^*,\boldsymbol{S}_n^*,\boldsymbol{P}^*)$.

To approximate $p(\boldsymbol{z})$ we need to obtain $p(\boldsymbol{\theta}^*,\phi^*,\boldsymbol{S}_n^*,\boldsymbol{P}^*|\boldsymbol{z})$ which is not immediately available. Using multiplication rule and the conditional independence of $(\boldsymbol{\theta}, \phi)$ with transition matrix $\boldsymbol{P}$ given $\boldsymbol{S}_n$, $p(\boldsymbol{\theta}^*,\phi^*,\boldsymbol{S}_n^*,\boldsymbol{P}^*|\boldsymbol{z})$ is given by

$$p(\boldsymbol{\theta}^*,\phi^*,\boldsymbol{S}_n^*,\boldsymbol{P}^*|\boldsymbol{z}) = p(\boldsymbol{S}_n^*|\boldsymbol{z})\,p(\boldsymbol{P}^*|\ \boldsymbol{S}_n^*)p(\phi^*|\,\boldsymbol{S}_n^*,\boldsymbol{z})\,p(\boldsymbol{\theta}^*|\phi^*,\boldsymbol{S}_n^*,\boldsymbol{z}), \tag{28}$$

where the term $p(\boldsymbol{P}^*|\ \boldsymbol{S}_n^*)$ is the product of independent Dirichlet densities of (22) and $p(\boldsymbol{\theta}^*|\phi^*,\boldsymbol{S}_n^*,\boldsymbol{z})$ is the full conditional of $\boldsymbol{\theta}$ given by (23). The third term on the right hand side of (28), $p(\phi^*|\,\boldsymbol{S}_n^*,\boldsymbol{z})$, can be obtained as

$$p(\phi^*|\,\boldsymbol{S}_n^*,\boldsymbol{z}) = \int p(\phi^*|\,\boldsymbol{\theta},\boldsymbol{S}_n^*,\boldsymbol{z})\,p(\boldsymbol{\theta}|\boldsymbol{S}_n^*,\boldsymbol{z})\,d\boldsymbol{\theta}. \tag{29}$$

Note that the posterior samples obtained using the Gibbs sampler are from the posterior density $p(\boldsymbol{\theta}|\boldsymbol{z})$ and not from $p(\boldsymbol{\theta}|\boldsymbol{S}_n^*,\boldsymbol{z})$. However, as suggested in Chib (1995), if we continue to sample for additional G' iterations using conditional densities $p(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n^*,\boldsymbol{z})$ and $p(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n^*,\boldsymbol{z})$, then we can obtain a Monte Carlo estimate as

$$p(\phi^*|\,\boldsymbol{S}_n^*,\boldsymbol{z}) \approx \frac{1}{G'}\sum_{g=1}^{G'} p(\phi^*|\,\boldsymbol{\theta}^{(g)},\boldsymbol{S}_n^*,\boldsymbol{z}) \tag{30}$$

where $\boldsymbol{\theta}^{(g)}$ are samples from $p(\boldsymbol{\theta}|\phi,\boldsymbol{S}_n^*,\boldsymbol{z})$. In our particular case this step can be avoided by updating $\boldsymbol{\theta}$ and ϕ as a block to obtain $p(\boldsymbol{\theta},\phi|\boldsymbol{S}_n,\boldsymbol{z})$, since $(\boldsymbol{\theta},\phi|\boldsymbol{S}_n,\boldsymbol{z})$ follows a normal-gamma density. Thus, the only term we need to evaluate is $p(\boldsymbol{S}_n^*|\boldsymbol{z})$.

Note that using the multiplication rule we can write

$$p(\boldsymbol{S}_n^*|\boldsymbol{z}) = p(s_1^*|\boldsymbol{z})\,p(s_2^*|s_1^*,\boldsymbol{z})\cdots\cdots p(s_i^*|\boldsymbol{S}_{i-1}^*,\boldsymbol{z})\cdots\cdots p(s_n^*|\boldsymbol{S}_{n-1}^*,\boldsymbol{z}), \tag{31}$$

where the first term $p(s_1^*|\boldsymbol{z})$ can be estimated from the draws available from the Gibbs sampler as

$$p(s_1^*|\boldsymbol{z}) \approx \frac{1}{G}\sum_{g=1}^{G} p(s_1^*|\,\boldsymbol{\theta}^{(g)},\phi^{(g)},(\boldsymbol{S}^{-i})^{(g)},\boldsymbol{P}^{(g)},\boldsymbol{z}). \tag{32}$$

Evaluation of the remaining densities requires additional sampling. For a general term $p(s_i^*|\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$ which is given by

$$p(s_i^*|\boldsymbol{S}_{i-1}^*,\boldsymbol{z}) = \int p(s_i^*|\,\boldsymbol{\theta},\phi,\boldsymbol{S}_{l>i},\boldsymbol{P},\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$$
$$p(\boldsymbol{\theta},\phi,\boldsymbol{S}_{l>i},\boldsymbol{P}|\boldsymbol{S}_{i-1}^*,\boldsymbol{z})\,d\boldsymbol{\theta}\,d\phi\,d\boldsymbol{S}_{l>i}\,d\boldsymbol{P}, \tag{33}$$

where $\boldsymbol{S}_{l>i} = \{s_l; l > i\}$, we need to continue sampling from full conditionals of $(\boldsymbol{\theta},\phi,s_i,\boldsymbol{S}_{l>i},\boldsymbol{P})$ given $(\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$. In other words, additional sampling will use the full conditional distributions:

$p(\boldsymbol{\theta},\phi|s_i,\boldsymbol{S}_{l>i},\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$,$p(\boldsymbol{P}|s_i,\boldsymbol{S}_{l>i},\boldsymbol{S}_{i-1}^*)$, $p(s_i|\boldsymbol{\theta},\phi,\boldsymbol{S}_{l>i},\boldsymbol{P},\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$, and $p(s_j|\boldsymbol{\theta},\phi,s_i,\boldsymbol{S}_{l>i}^{-j},\boldsymbol{P},\boldsymbol{S}_{i-1}^*,\boldsymbol{z})$, $j = i+1,\ldots,n$.

Then (33) can be approximated as

$$p(s_i^*|\boldsymbol{S}_{i-1}^*,\boldsymbol{z}) \approx \frac{1}{G'}\sum_{g=1}^{G'} p(s_i^*|\,\boldsymbol{\theta}^{(g)},\phi^{(g)},(\boldsymbol{S}_{l>i})^{(g)},\boldsymbol{P}^{(g)},\boldsymbol{S}_{i-1}^*,\boldsymbol{z}), \tag{34}$$

where $(\boldsymbol{\theta}^{(g)},\phi^{(g)},(\boldsymbol{S}^{-i})^{(g)},\boldsymbol{P}^{(g)})$ represents samples from $p(\boldsymbol{\theta},\phi,\boldsymbol{S}_{l>i},\boldsymbol{P}|\boldsymbol{S}_{i-1}^{*},\boldsymbol{z})$. This completes all the terms needed to compute (28) and to approximate the marginal likelihood (27).

We note that for each given model, which may represent a specific period p in (6), the marginal likelihood (27) can be approximated and these are compared to find the model which is most supported by the data.

5. Example Using Simulated Data

Even though the problem and methodology presented in previous sections are motivated by a real application, there is no available published real data. Thus, we will illustrate the implementation of the Bayesian methodologies of Sections 3 and 4 using simulated data sets.

To illustrate the performance of the model comparison approach of Section 4 we consider different values of the noise variance $\sigma_w^2 = 1/\phi$ in the stagger model given by (3). The simulated data set consists of a period-3 staggered PRI sequence with stagger elements 200, 300, 500 contaminated with noise. The actual values of the coefficients in (3) for this data are $a_0 = 333.33$, $a_1 = -133.33$, and $b_1 = 115.47$. There are eight missing TOA values in the sequence resulting in aggregate data points at indices 12, 18, 45, 54, 65, 71, and 80. The aggregate data point at index 45 is composed of three PRI values and thus represents two missing pulses. All the other aggregate data points are composed of two PRI values and thus, are associated with a single missing pulse. Each data set consists of 90 observed PRI's implying that the actual number of transmitted PRIs is 98.

In Figure 2 we illustrate a simulated data set with noise standard deviation $\sigma_w = 25$. We note that because of the wide variation in PRI values in the sequence, some of these aggregate data points cannot be identified by magnitude. For example, the aggregate data points at indices 18, 54, and 65 can not be identified based on the PRI value at those indices.

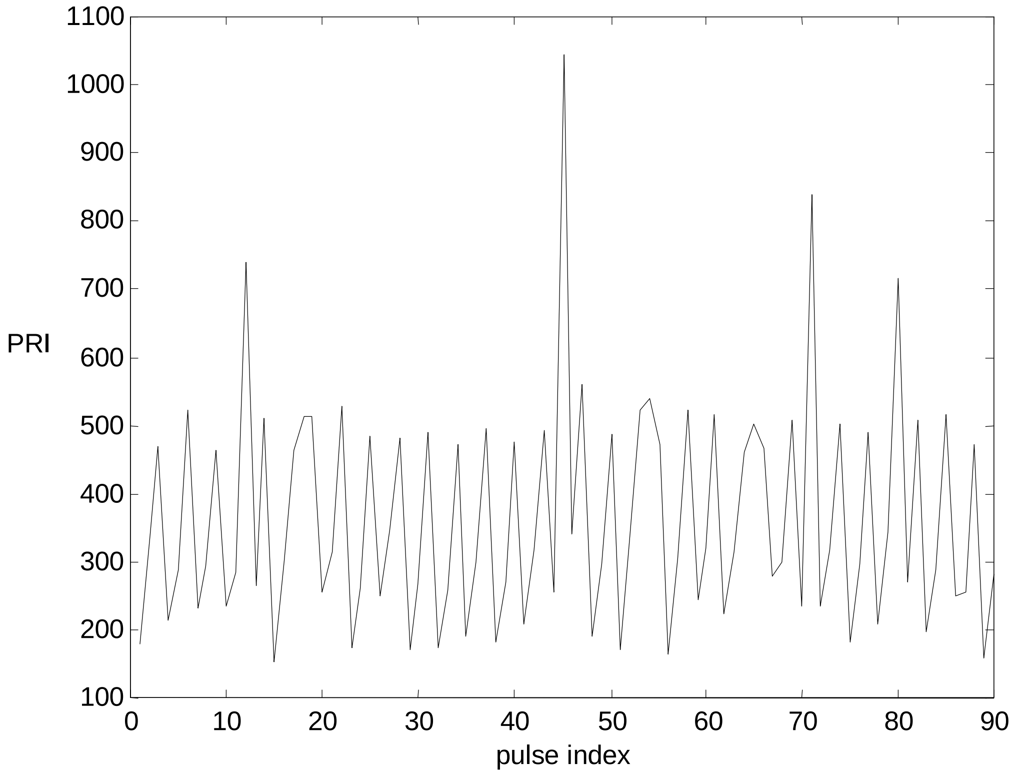

Figure 2. Stagger-3 data set with seven aggregate data points.

We have analyzed the data using four different candidate models: jitter model and stagger with periods 3, 5, and 6. In analyzing the data, in all cases, we have used noninformative priors for the coefficient vector $\boldsymbol{\theta}$ and noise precision ϕ. For the prior of $\boldsymbol{\theta}$, in (11a) we have specified $\boldsymbol{m}$ as zero vector and $\boldsymbol{V}$ as a diagonal matrix with diagonal components are all equal to 1×10^{13}. For prior of ϕ, in (11b) we have specified $d = c = 1 \times 10^{-6}$.

In the analysis the maximum number of missing observations in an interval was assumed to be $M = 2$ implying a 3×3 Markov transition matrix. For each row of the transition matrix, using (15), a Dirichlet prior was assigned with parameters $(90\ 9\ 1)$

implying a prior expectation of 10% missing pulses in the series. We note that there maybe prior information available on what percent of pulses are missing (but not on their location) and such information can be used to specify the prior parameters of the Dirichlet distribution.

Using the above priors the Gibbs sampler was run for each candidate model following our development in Section 3. From the Gibbs sampler output, for each candidate model the marginal likelihoods were computed using the procedure presented in Section 4. This was repeated for simulated data sets with different values of the noise standard deviation σ_w. More specifically, the analyses were repeated for each model using simulated data with $\sigma_w = 1, 5, 10, 15, 20, 25,$and 30. In each simulation the data was generated from the model with period 3 and there were 8 missing pulses. In Figure 3, we present the log marginal likelihoods of the four models for the different values of the σ_w (as shown on the x-axis). The figure illustrates that our approach identifies the stagger model with period 3 as the correct model in all cases except the last one where $\sigma_w = 30$. As expected, performance deteriorates as σ_w gets large as shown by the decreasing value of the log marginal likelihood for stagger model with period 3. Similarly, the posterior distribution of σ_w under the correct model was more concentrated around the actual values of σ_w compared to the other models.

We note that the performance of the approach is dependent on its ability to infer the unknown locations of the missing pulses. In Table 1, for the stagger model with period 3 and $\sigma_w = 25$ we illustrate the posterior distributions of the latent variables s_i's associated with missing pulse locations $i = 12, 18,45, 54, 65, 71$ and 80. The posterior distribution of s_4 associated with a no missing pulse location is also shown for comparison purposes.

Each row in Table 1 represents the posterior distribution associated with the particular pulse location. We note that for s_4 which is the state variable for location 4 with no missing pulse, the posterior probability of no missing pulse ($s_4 = 1$) is 0.998. For locations $i = 12$, 71 and 80 where there is one missing pulse in each case, the corresponding posterior probabilities $p(s_i = 2|\,\boldsymbol{z})$ are all higher than 0.92. Note that these are the PRI values that are large in magnitude as can be seen from Figure 1. For locations $i = 18$, 54 and 65 with one missing pulse the probabilities $p(s_i = 2|\,\boldsymbol{z})$ are not as large as in the previous group, but they are still around 0.79-0.80. Note that these are the locations with PRI values that are not large in magnitude, but the procedure still is able to infer the missing pulse with high probability. Finally, location 45 has two missing pulses and the corresponding posterior probability $p(s_{45} = 3|\,\boldsymbol{z}) = 0.8434$. In summary, the inference procedure seems to be identifying the location of missing pulses.

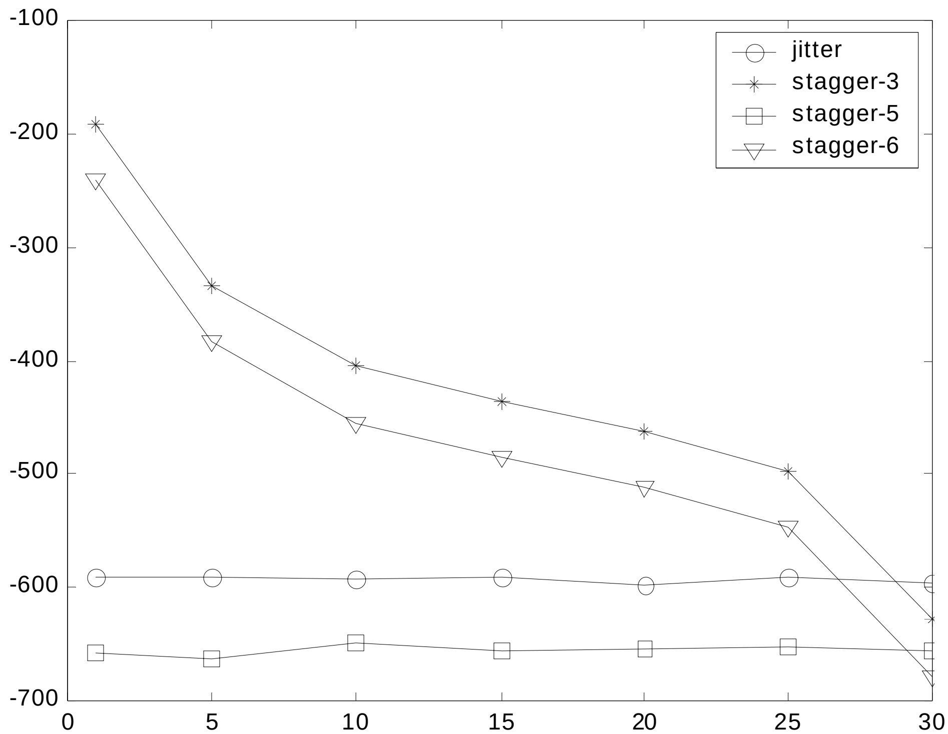

Figure 3. Plots of the Log Marginal Likelihoods versus σ_w.

Table 1
Posterior Distributions of Selected s_i's
Under Stagger Model with Period 3

i	$p(s_i = 1\| \boldsymbol{z})$	$p(s_i = 2\| \boldsymbol{z})$	$p(s_i = 3\| \boldsymbol{z})$
4	0.9980	0.0020	0.0000
12	0.0538	0.9376	0.0086
18	0.2086	0.7910	0.0040
45	0.0004	0.1562	0.8434
54	0.1972	0.8024	0.0004
65	0.2142	0.7852	0.0006
71	0.0120	0.9604	0.0276
80	0.0678	0.9254	0.0068

6. Conclusions

In this paper a Bayesian approach was developed for analysis of pulse trains corrupted by missing pulses at unknown locations. The development was motivated by electronic warfare applications where it is of interest to infer the locations of missing pulses to identify the PRI modulation type.

The implementation of the approach was demonstrated using simulated data and it was shown that the Bayesian approach performed adequately in locating the missing pulses and identfying the correct model. However, the results presented here are based on limited data sets and more experience is needed with more complicated data structures.

A comparison of the model selection approach with other approaches such as the the reversible jump Markov chain Monte Carlo methods of Green (1995) remains to be seen. This is an area for further study.

An important extension of the presented model is the case where the noise terms are correlated. This situation may arise in certain applications and requires a modification of our setup and procedures.

APPENDIX: Details of the Gibbs Sampler

A.1. Full Conditional Distribution of $\boldsymbol{P}_i$

The full conditional distribution of $\boldsymbol{P}_i$ is given by

$$p(\boldsymbol{P}_i|\ \boldsymbol{S}_n) \propto \prod_{k=1}^{n} p(s_{k+1}|s_k = i)\, p(\boldsymbol{P}_i) \tag{A1}$$

where $p(\boldsymbol{P}_i)$ is the Dirichlet prior given in (15). The above can be written as

$$p(\boldsymbol{P}_i|\ \boldsymbol{S}_n) \propto \prod_{j=1}^{M+1} p_{ij}^{\alpha_{ij}-1} p_{ij}^{\sum_{k=1}^{n} 1(s_{k+1}=j, s_k=i)}, \tag{A2}$$

where $1(\bullet)$ is an indicator function that takes value 1 if its argument is correct and 0 otherwise.

A.2. Full Conditional Distributions of $\boldsymbol{\theta}$ and ϕ

In obtaining $p(\boldsymbol{\theta}|\phi, \boldsymbol{S}_n, \boldsymbol{z})$, we can write

$$p(\boldsymbol{\theta}|\phi, \boldsymbol{S}_n, \boldsymbol{z}) \propto p(\boldsymbol{z}|\boldsymbol{\theta}, \phi, \boldsymbol{S}_n)\, p(\boldsymbol{\theta}|\phi), \tag{A3}$$

where $p(\boldsymbol{\theta}|\phi)$ is given by (11a) and it does not depend on the latent vector $\boldsymbol{S}_n$ due to the prior independence assumptions. The first term on the right-hand side of (A3) is the conditional likelihood function of $\boldsymbol{\theta}$ given $(\phi, \boldsymbol{S}_n, \boldsymbol{z})$ which is given by the multivariate normal form (12) with $\boldsymbol{X}$ and $\boldsymbol{U}$ are functions of the latent sequence $\boldsymbol{S}_n$. Thus, (A3) can be written as

$$p(\boldsymbol{\theta}|\phi, \boldsymbol{S}_n, \boldsymbol{z}) \propto exp\Big(-\frac{\phi}{2}\big[(\boldsymbol{z}-\boldsymbol{X\theta})'\boldsymbol{U}^{-1}(\boldsymbol{z}-\boldsymbol{X\theta}) + (\boldsymbol{\theta}-\boldsymbol{m})'\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\boldsymbol{m})\big]\Big),$$

and, using standard Bayesian conjugate analysis [see for example, DeGroot (1970), pp. 251], it can be shown that the full conditional is normal with mean and variance are given by

$$\boldsymbol{m}^* = (\boldsymbol{V}^{-1} + \boldsymbol{X}'\boldsymbol{U}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{V}^{-1}\boldsymbol{m} + \boldsymbol{X}'\boldsymbol{U}^{-1}\boldsymbol{z}) \tag{A4}$$

and

$$\boldsymbol{V}^{*} = (\boldsymbol{V}^{-1} + \boldsymbol{X}'\boldsymbol{U}^{-1}\boldsymbol{X})^{-1}, \tag{A5}$$

repectively.

In writing down the full conditional distribution of ϕ, the first term on the right-hand side of (24), $p(\boldsymbol{z}|\boldsymbol{\theta},\phi,\boldsymbol{S}_n)$, is the conditional likelihood of ϕ which is a multivariate normal form and the remaining terms are the components of the normal-gamma prior given by (11a) and (11b). Thus, (24) can be written as

$$p(\phi|\,\boldsymbol{\theta},\boldsymbol{S}_n,\boldsymbol{z}) \propto \left[\phi^{n/2} exp\Big(-\frac{\phi}{2}\big[(\boldsymbol{z}-\boldsymbol{X\theta})'\boldsymbol{U}^{-1}(\boldsymbol{z}-\boldsymbol{X\theta})\Big)\right] \times$$

$$\left[\phi^{p/2} exp\Big(-\frac{\phi}{2}\big[(\boldsymbol{\theta}-\boldsymbol{m})'\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\boldsymbol{m})\big]\Big)\right] \times \phi^{(d/2)-1}\, exp(-\phi\frac{c}{2}), \tag{A6}$$

and it follows from above that the full conditional of ϕ is a gamma density with parameters $d^{*}/2$ and $c^{*}/2$, where $d^{*} = n+p+d$ and

$$c^{*} = \big[(\boldsymbol{z}-\boldsymbol{X\theta})'\boldsymbol{U}^{-1}(\boldsymbol{z}-\boldsymbol{X\theta}) + (\boldsymbol{\theta}-\boldsymbol{m})'\boldsymbol{V}^{-1}(\boldsymbol{\theta}-\boldsymbol{m}) + c\big].$$

References

Bretthorst, L. G. (1988). Bayesian Spectrum Analysis and Parameter Estimation, Lecture Notes in Statistics, Springer-Verlag, New York.

Carlin, B. and Chib, S. (1995). Bayesian Model Choice by Markov Chain Monte Carlo, *Journal the Royal Statistical Society*, Ser. B, Vol. 57, pp. 473-884.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output, *Journal the American Statistical Association*, Vol. 90, pp. 1313-1321.

Chib, S. (1996). Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models, *Journal of Econometrics*, Vol. 75, pp. 79-97.

Clarkson, I. V. L., Howard, S. D., and Mareels, I. M. Y. (1996). Estimating the Period of a Pulse Train from a Set of Sparse, Noisy Measurements, *Proceedings of the International Symposium on Signal Processing and its Applications*, Vol. 2, pp.885-888.

DeGroot, M. H. (1970). Optimal Statistical Decisions. McGraw Hill.

Doucet, A. and Duvaut, P. (1997). Bayesian Estimation of State-space Models Applied to Deconvolution of Bernoulli-Gaussian Processes, *Signal Processing*, Vol. 57, pp. 147-161.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, *Journal of the American Statistical Association*, Vol. 85, pp. 398-409.

Gelfand, A. E. (1996). Model Determination Using Sampling-based Methods. In Markov Chain Monte Carlo in Practice. (Eds. W. R. Gilks, S. Richardson and D. J. Spiegelhalter) pp. 145-161, Chapman & Hall.

Gelman, A., Carlin, J. B., Stern, H.S., & Rubin, D. B. (1995). Bayesian Data Analysis. Chapman and Hall.

Green, P. J. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, Vol. 82, pp. 711-732.

Han, C. and Carlin, B. (2001). MCMC Methods for Computing Bayes Factors: A Comparative Review, *Journal of the American Statistical Association*, Vol. 96, pp. 1122-1132.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors, *Journal of the American Statistical Association*, Vol. 90, No. 430, pp.773-795.

Little, R. J. A. and Rubin, D. B. (1987). Statistical Analysis with Missing Data. Wiley.

Rabiner, L. R. and Juang, B-H. (1985). Mixture Autoregressive Hidden Markov Models for Speech Signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-33, pp. 1404-1412.

Robert, C. P., Celeux, G. and Diebolt, J. (1993). Bayesian Estimation of Hidden Markov Chains: A Stochastic Implementation, *Statistics and Probability Letters*, Vol. 16, pp.77-83.

West, M. (1995). Bayesian Inference in Cyclical Component Dynamic Linear Models, *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 1301-1312.

West, M. and Harrison, J. (1997). Bayesian Forecasting and Dynamic Models, Second Edition, Springer-Verlag.