Advances and Applications in Statistics
Volume 35, Number 1, 2013, Pages 1-27
Published Online: August 2013
Available online at http://pphmj.com/journals/adas.htm
Published by Pushpa Publishing House, Allahabad, INDIA

ESTIMATION OF NAVIGATION PERFORMANCE AND OFFSET BY THE EM ALGORITHM AND THE VARIATIONAL BAYESIAN METHODS

Masato Fujita

Electronic Navigation Research Institute
7-42-23 Jindaiji-higashi
Chofu, Tokyo 182-0012, Japan

Abstract

An offset procedure allows aircraft to fly several discrete nautical miles right of the course center line without informing air traffic control authority. On the other hand, an En-route Monitoring Agency (EMA) should monitor the navigation performance of aircraft, namely, the unintended deviation from their intended paths, for safety monitoring purposes. Only data available for the EMA are surveillance data collected for the purpose of normal air traffic control practices. They are the aircraft position data which are the sum of intended offset and unintended deviation.

This paper introduces Expectation-Maximization (EM) algorithms and variational Bayesian methods to estimate the proportion of offset and the distribution of navigation performance simultaneously from the surveillance data. A numerical experiment shows that both EM algorithms and variational Bayesian methods provide the ability to estimate the parameters of the distribution model in reasonably short

© 2013 Pushpa Publishing House

2010 Mathematics Subject Classification: Primary 62F15; Secondary 62P30.

Keywords and phrases: air traffic control, navigation performance, offset, strategic lateral offset procedure, EM algorithm, variational Bayes.

Communicated by K. K. Azad

Received February 22, 2013

time (in seconds). The estimation is accurate if the initial model parameter values or the parameters of prior distributions are appropriately set.

This paper also proposes the estimation of the offset for a single observation of an aircraft.

1. Introduction

Strategic Lateral Offset Procedure (SLOP) is used by aircraft navigating an assigned route along an airway or between published navigational waypoints in an oceanic airspace. SLOP procedure allows aircraft to fly 1 nautical mile or 2 nautical miles right of the course center line for mitigating mid-air collision hazard and reducing the effect of wake turbulence without informing Air Traffic Control (ATC) authority [1].

On the other hand, separation between aircraft becomes smaller to meet the increasing air traffic demand. The reduction of separation has a potential impact to mid-air collision hazard. The comprehensive monitoring of reduced horizontal separation practices by En-route Monitoring Agencies (EMA) starts on a regional basis and the monitoring of horizontal-plane (especially lateral) navigation performance of an aircraft is conducted in this framework [2].

The EMA can collect the position data of aircraft by means of surveillance systems such as radar and ADS-C (Automatic Dependent Surveillance - Contract. See [3, 4] for details.) We ignore the measurement errors of these surveillance systems in this paper. The EMA should know the magnitude of unintended deviation from their intended path to meet the monitoring purpose. It means that the EMA should also know the magnitude of offset. However, there are no deterministic means for the EMA to know whether an aircraft flying 1 nautical mile right to the published route applies one nautical mile SLOP. It may not apply SLOP and unintentionally deviate 1 nautical mile right from its intended path, or it may apply 2 nautical miles SLOP and unintentionally deviate 1 nautical mile left from its intended path. We should know the magnitude of offset beforehand to know the exact value of unintended deviation.

As to the previous example of an aircraft flying 1 nautical mile right of its route, 1 nautical mile SLOP application is more likely than others if the aircraft has a good navigation performance. We may be able to estimate the magnitude of lateral offset if the actual navigation performance of aircraft is known and it is smaller than the increment of the offset values. It seems a circular argument.

Even though we have no means to know the correct offset value and correct magnitude of unintended deviation, it is possible to simultaneously estimate both from observation data (the magnitude of deviation from the published route center line). We model unintended deviations as a distribution whose probability density function is $f(x)$. We employ the simple assumption that the navigation performance of aircraft is not influenced by the magnitude of offset. Then the total deviation from the route center line (the sum of an unintended deviation and offset) follows the distribution whose probability density function is the following:

$$\sum_{l=1}^{L} \omega_l f(x - o_l). \tag{1}$$

Here, the number o_l denotes the magnitude offset, and ω_l is the proportion of offset execution whose magnitude is o_l.

The main purpose of this paper is to derive the methodology to estimate the parameters of the distribution model *f* and the proportion of offset ω_l in equation (1) from observation data in a reasonably short time. This problem was already discussed by [5]. Newton's method and a variable neighborhood search algorithm were used for estimating the parameters maximizing the likelihood function in this study. We employ other algorithms. Moreover, we introduce its potential operational application for estimating the offset in a single observation on the deviation from the published route center line.

Well-known criteria for finding model parameters are maximum likelihood estimation and Bayesian update, which are introduced in Section 3. We cannot find closed analytical solutions for both cases. Our distribution defined in equation (1) has the probability density function similar to that of finite mixture distributions [6]. Finite mixture distributions are used in many areas. Multi-dimensional Gaussian distributions and Bernoulli distributions are used for the clustering of data in the field of data mining, machine learning and pattern recognition [7, 8, 6]. A numerical methodology called the *Expectation-Maximization* (*EM*) *algorithm* [10, 9] for maximum likelihood estimation and variational Bayesian method [11, 7, 12] for Bayesian updates are used for the estimation of model parameters. However, these algorithms for finite mixture models are not applicable directly to our problem. We will develop an EM algorithm and a variational Bayesian method applicable to our problem in this paper.

Section 2 first gives the mathematical description of problems considered in this paper. Section 3 introduces the basic idea of the EM algorithm and the variational Bayesian method. Section 4 develops the EM algorithm and the variational Bayesian method applicable to our problem. We also discuss the methodologies to estimate the magnitude of offset from a single observation datum using the result of parameter estimations. It is discussed in Section 5. We implemented the algorithms introduced in Section 4 as a computer program. Section 6 shows the numerical experiment results. Finally, Section 7 summarizes the paper.

2. Description of the Problem

Lateral navigation performance of aircraft, more precisely, lateral deviations from their intended path, is often modeled as a probability distribution symmetric to the vertical axis. Gaussian distributions, Laplace distributions and the mixtures of two of them are often used [13, 14]. In this paper, we will concentrate on the case where distribution models of lateral navigation performance are symmetric.

We introduce abbreviations for the sake of later use in Section 6. The NN model is the mixture distribution of two Gaussian distributions. The NDE model is the mixture distribution of a Gaussian distribution and a Laplace distribution. Finally, the DDE model is the mixture distribution of two

Laplace distributions. They are often used for the distribution models of the unintended lateral deviations of aircraft from the intended path.

In the SLOP case, the number of possible offset is 3, namely, zero nautical miles offset, one nautical mile offset and two nautical miles offset. We only consider the mixture of at most 2 distributions, namely, Gaussian distribution, Laplace distribution, NN model, NDE model and DDE model. However, these numbers 3 and 2 may change in the future operational environment. The algorithms applicable for arbitrary numbers are more useful than those applicable only for specific numbers.

We first define our terminology. We will consider the sum of a discrete random variable with a random variable following the finite mixture of an arbitrary number of 1-dimensional Gaussian distributions and Laplace distributions with zero mean. We call it an *offset mixture distribution* (of Gaussian distributions and Laplace distributions) in this paper. The discrete random variable corresponds to the offset of the aircraft. On the other hand, the finite mixture distribution represents the lateral navigation performance of the aircraft.

The probability density function of an offset mixture distribution is given by the following equation:

$$p(x \mid \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda}) = \sum_{l=1}^{L} \omega_l \left(\sum_{k=1}^{m} \pi_k \mathcal{N}(x - o_l, \sigma_k) + \sum_{k=1}^{n} \pi_{m+k} \mathcal{D}(x - o_l, \lambda_k) \right). \tag{2}$$

Here $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K)$ are finite non-negative numbers satisfying $\sum_{k=1}^{K} \pi_k = 1$, where $K = m + n$. $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, ..., \sigma_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_n)$ denote positive numbers. $\boldsymbol{o} = (o_1, o_2, ..., o_L)$ are real numbers. $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_L)$ are non-negative numbers satisfying $\sum_{l=1}^{L} \omega_l = 1$. We call ω_l an *offset mixing coefficient* in this paper. The notation $\mathcal{N}(x, \sigma)$ denotes the probability density function of a 1-dimensional Gaussian distribution with zero mean, and $\mathcal{D}(x, \lambda)$ denotes the probability density function of a Laplace distribution with zero mean:

$$\mathcal{N}(x, \sigma) = \frac{\exp(-x^2/(2\sigma^2))}{\sqrt{2\pi}\sigma}, \tag{3}$$

$$\mathcal{D}(x, \lambda) = \frac{\exp(-|x|/\lambda)}{2\lambda}. \tag{4}$$

We will consider the following generalized problem in this paper.

Problem 1. Estimate all the unknown parameters ω_l, π_k, σ_i and λ_i of an offset mixture distribution from a given data set $\boldsymbol{X} = \{x_1, x_2, ..., x_N\}$ under the following conditions:

- As to the discrete variable, a finite set $\{o_1, ..., o_L\}$ of possible realization (possible magnitudes of offset) is given, but the probability ω_l that a given number o_l is realized is unknown.
- As to the mixture distribution, all the parameters

  $$\{\pi_1, ..., \pi_{m+n}, \sigma_1, ..., \sigma_m, \lambda_1, ..., \lambda_n\}$$

  are unknown except the numbers m and n of mixed Gaussian components and Laplace components.

3. Expectation-maximization Algorithm and Variational Bayesian Method

We introduce two algorithms for estimating parameters (or posterior distribution of parameters) of distribution models.

The first algorithm is the EM algorithm [10] which is an iterative method for finding parameters maximizing the likelihood locally. A latent variable is a random variable that is not directly observed but is inferred from other random variables. An example of a latent variable is the magnitude of offset. The magnitude of offset is inferred from the total deviation from the route center line. We first need to define latent variables appropriately for the implementation EM algorithm. In the estimation of parameters, the following E-step and M-step are repeated till the likelihood converges.

E-step. Find the model parameters of the distribution of latent parameters for fixed original model parameters.

M-step. Find the parameter of the original distribution model maximizing the expected number of the log-likelihood function with respect to the fixed distribution of latent parameters.

The second algorithm is the variational Bayesian method [7] which is also an iterative method for finding the approximation of the posterior distribution of model parameters and latent variables minimizing Kullback-Leibler divergence [15].

Let $\boldsymbol{V}$ be a family of model parameters and latent variables. In Bayesian statistics, model parameters are regarded not as unknown deterministic values but as random variables. Let $p(\boldsymbol{V})$ be the probability density function of the random variables $\boldsymbol{V}$ before observation data are available. This distribution is called a *prior distribution*. The prior distribution represents the degree of confidence [16] before observation data are available. The degree of confidence changes after the observation data set $\boldsymbol{X} = \{x_1, x_2, ..., x_N\}$ is obtained. The degree of confidence after the acquisition of observation data is modeled by a probability distribution. It is called a *posterior distribution*. Let $p(\boldsymbol{V} \mid \boldsymbol{X})$ be the probability density function of the posterior distribution. The following Bayes' theorem illustrates the relationship between the prior distribution and the posterior distribution:

$$p(\boldsymbol{V} \mid \boldsymbol{X}) = \frac{p(\boldsymbol{X} \mid \boldsymbol{V})\, p(\boldsymbol{V})}{\int p(\boldsymbol{X} \mid \boldsymbol{V})\, p(\boldsymbol{V})\, d\boldsymbol{V}}. \tag{5}$$

Here, $p(\boldsymbol{X} \mid \boldsymbol{V})$ denotes the likelihood under the condition that the model parameters $\boldsymbol{V}$ are fixed. Bayesian update is the procedure for finding the posterior distribution by means of equation (5). If the analytical form of the prior distribution $p(\boldsymbol{V})$ is the same as that of the posterior distribution

$p(\boldsymbol{V} \mid \boldsymbol{X})$ and the formulae for finding the posterior distribution from the prior distribution are available, we can iteratively apply equation (5) against a series of iteratively available observation data, though the EM algorithm requires all the observation data for the estimation of parameters. A prior distribution is called a *conjugate prior distribution* if the analytical form of the prior distribution $p(\boldsymbol{V})$ is the same as that of the posterior distribution $p(\boldsymbol{V} \mid \boldsymbol{X})$.

It is hard to find the true posterior distribution $p(\boldsymbol{V} \mid \boldsymbol{X})$ by means of equation (5) in our problem. The variational Bayesian method gives the methodology for finding an approximation distribution model $q(\boldsymbol{V})$ of the posterior distribution $p(\boldsymbol{V} \mid \boldsymbol{X})$. The assumption that parameters and latent variables are divided into mutually independent disjoint groups is employed for the derivation of the variational Bayesian method. Assume that $\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_s)$ is a division into mutually independent disjoint groups. In other words, the approximation distribution model $q(\boldsymbol{V})$ of the posterior distribution is given by the following formula:

$$q(\boldsymbol{V}) = \prod_{i=1}^{s} q(\boldsymbol{V}_i). \tag{6}$$

The distribution model $q(\boldsymbol{V}_i)$ satisfying the following equation minimizes Kullback-Leibler divergence when the distributions of other parameters in $\boldsymbol{V}$ are fixed [7]:

$$\ln q(\boldsymbol{V}_i) = \mathbf{E}_{i \neq j}[\ln p(\boldsymbol{X}, \boldsymbol{V})] + \text{constant}. \tag{7}$$

Here, $\mathbf{E}_{i \neq j}[\cdot]$ denotes the expectation with respect to all the variables other than $\boldsymbol{V}_i$. The notation $p(\boldsymbol{X}, \boldsymbol{V})$ denotes the joint probability of $\boldsymbol{X}$ and $\boldsymbol{V}$. Formula (7) is repeatedly utilized till convergence. The lower bound $\mathcal{L}(q)$ given by

$$\mathcal{L}(q) = \int q(\boldsymbol{V}) \ln\left\{\frac{p(\boldsymbol{X}, \boldsymbol{V})}{q(\boldsymbol{V})}\right\} d\boldsymbol{V} \tag{8}$$

is often utilized for the check of convergence because $\mathcal{L}(q)$ is easier to evaluate than Kullback-Leibler divergence

$$KL(q\,||\,p) = -\int q(\boldsymbol{V})\ln\left\{\frac{p(\boldsymbol{V}\mid\boldsymbol{X})}{q(\boldsymbol{V})}\right\}d\boldsymbol{V}, \tag{9}$$

and the sum of $\mathcal{L}(q)$ with Kullback-Leibler divergence $KL(q\,||\,p)$ is always the constant number $\int p(\boldsymbol{X}\,|\boldsymbol{V})\,p(\boldsymbol{V})\,d\boldsymbol{V}$, independent of the distribution model $q(\boldsymbol{V})$.

Note that both the EM algorithm and the variational Bayesian method may estimate the model parameters which give a local minimum, but not global minimum of evaluation functions.

The main difficulties for developing EM algorithms and variational Bayesian methods lie on the development of an appropriate definition of latent variables, the discovery of conjugate prior distribution models and the development of an appropriate assumption on the division into mutually independent disjoint groups. In fact, the variational Bayesian method initially developed by the author, which requires less memory consumption than the algorithm in this paper, did not converge to the appropriate value.

Section 4 introduces the EM algorithm and the variational Bayesian method developed by the author.

4. Algorithm for the Estimation of Model Parameters

The purpose of this section is to propose the EM algorithm and the variational Bayesian method for solving Problem 1. Subsection 4.2 summarizes the EM algorithm for offset mixture models and Subsection 4.3 summarizes variational Bayesian method for them.

4.1. Latent variable

We need to define an appropriate latent variable for the development of the algorithms. In our algorithms, the latent variable $\boldsymbol{Z} = (z_{ikl})$ is defined as follows. The variable z_{ikl} takes the value 1 if the *i*th observation data x_i is generated from the model experiencing the offset o_l and generated by kth Gaussian/Laplace component, and 0 otherwise.

4.2. Expectation-maximization

The following is an EM algorithm for an offset mixture model (2). The input and output parameters are described in Table 1. We omit the derivation of this algorithm.

Table 1. Input/output parameters of EM algorithm

Parameters	Description
Input	
$\boldsymbol{X} = \{x_1, x_2, \dots, x_N\}$	Observation data
$\boldsymbol{o} = (o_1, o_2, \dots, o_L)$	Possible realizations of the discrete random variable (magnitudes of offset)
m	The number of Gaussian components mixed in $p(\boldsymbol{X} \mid \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$
n	The number of Laplace components mixed in $p(\boldsymbol{X} \mid \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$
ε	Small positive number used for defining the stop condition of the algorithm
Output	
$\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_L)$	Offset mixing coefficients
$\pi = (\pi_1, \pi_2, \dots, \pi_K)$	Mixing coefficients
σ_k	Standard deviations of Gaussian distributions $(k = 1, \dots, m)$
λ_k	Scale parameters of Laplace distributions $(k = m + 1, \dots, K)$

ALGORITHM

Initialization. Initialize variables $\boldsymbol{\omega}$, $\boldsymbol{\pi}$, $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_m)$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$. Set $\mathcal{L} = -\infty$.

E-step. Evaluate $r_{i,k,l}$ and $s_{i,k,l}$ by means of the following equations:

$$r_{i,k,l} = \frac{\omega_l \pi_k \mathcal{N}(x_i, \sigma_k)}{\sum_{l=1}^{L}\left(\sum_{k'=1}^{m} \pi_{k'} \mathcal{N}(x_i - o_l, \sigma_{k'}) + \sum_{k'=1}^{n} \pi_{m+k'} \mathcal{D}(x_i - o_l, \sigma_{k'})\right)}, \quad (10)$$

$$s_{i,k,l} = \frac{\omega_l \pi_{k+m} \mathcal{D}(x_i, \lambda_k)}{\sum_{l=1}^{L}\left(\sum_{k'=1}^{m} \pi_{k'} \mathcal{N}(x_i - o_l, \sigma_{k'}) + \sum_{k'=1}^{n} \pi_{m+k'} \mathcal{D}(x_i - o_l, \sigma_{k'})\right)}. \quad (11)$$

M-step. Update $\boldsymbol{\omega}$, $\boldsymbol{\pi}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ by means of the following equations:

$$\omega_l = \frac{\sum_{i=1}^{N}\left(\sum_{k=1}^{m} r_{i,k,l} + \sum_{k=1}^{n} s_{i,k,l}\right)}{N}, \quad (12)$$

$$\pi_k = \begin{cases} M_k / N & (k = 1, \ldots, m), \\ N_{k-m} / N & (k = m+1, \ldots, n), \end{cases} \quad (13)$$

$$\sigma_k = \sqrt{\frac{\sum_{i=1}^{N} \sum_{l=1}^{L} r_{i,k} (x_i - o_l)^2}{M_k}}, \quad (14)$$

$$\lambda_k = \frac{\sum_{i=1}^{N} \sum_{l=1}^{L} s_{i,k} | x_i - o_l |}{N_k}. \quad (15)$$

Here, $M_k = \sum_{i=1}^{N} \sum_{l=1}^{L} r_{i,k,l}$ and $N_k = \sum_{i=1}^{N} \sum_{l=1}^{L} s_{i,k,l}$.

Evaluation of log-likelihood. Evaluate the log-likelihood $L(q)$ by means of the following equation. If $L(q) - \mathcal{L} > \varepsilon$, set $\mathcal{L} = L(q)$ and return back to E-step:

$$L(q) = \sum_{i=1}^{N} \ln\left(\sum_{l=1}^{L} \omega_l \left(\sum_{k=1}^{m} \pi_k \mathcal{N}(x_i - o_l, \sigma_k) + \sum_{k=1}^{m} \pi_{m+k} \mathcal{D}(x_i - o_l, \sigma_k)\right)\right). \tag{16}$$

4.3. Variational Bayesian method

We introduce a variational Bayesian method for the offset mixture model (2). A new variable $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_K)$ defined by

$$\eta_k = \begin{cases} \dfrac{1}{\sigma_k^2} & (1 \le k \le m), \\ \dfrac{1}{\lambda_{k-m}} & (m+1 \le k \le K) \end{cases} \tag{17}$$

is introduced. Table 2 describes the assumption on the prior distribution of parameters. The approximations $q(\boldsymbol{V})$ of the posterior distributions $p(\boldsymbol{V} \mid \boldsymbol{X})$ also belong to the same families as those in Table 2. The input and output parameters of the variational Bayesian method are described in Table 3.

The following is the algorithm of the variational Bayesian method for offset mixture distribution:

ALGORITHM

<u>Initialization.</u> Set $\boldsymbol{p} = \boldsymbol{p}_0$, $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, $a_k = a_{0,k}$, $b_k = b_{0,k}$ $(k = 1, ..., K)$ and $\mathcal{L} = -\infty$.

<u>E-step.</u> Evaluate $r_{i,k,l}$ by means of the following equations. Here, $\psi(\cdot)$ is a digamma function, $\hat{\boldsymbol{\alpha}} = \sum_{k=1}^{K} \alpha_k$ and $\hat{\boldsymbol{p}} = \sum_{l=1}^{L} p_l$:

$$\ln \rho_{i,k,l} = \psi(\alpha_k) - \psi(\hat{\boldsymbol{\alpha}}) + \psi(p_l) - \psi(\hat{\boldsymbol{p}})$$

$$+\begin{cases}\frac{1}{2}\left(\psi(a_k)-\ln b_k-\ln 2\pi-\frac{a_k}{b_k}(x_i-o_l)^2\right) & \text{if } k\le m,\\ \psi(a_k)-\ln b_k-\ln 2-\frac{a_k}{b_k}|x_i-o_l| & \text{if } k>m,\end{cases} \tag{18}$$

$$r_{i,k,l}=\frac{\rho_{i,k,l}}{\sum_{k=1}^{K}\sum_{l=1}^{L}\rho_{i,k,l}}. \tag{19}$$

M-step. Update $\boldsymbol{p}=(p_1,\ p_2,\ ...,\ p_K)$, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,...,\alpha_K)$, α_k and b_k by means of the following equations:

$$\boldsymbol{p}=(p_1,\ p_2,\ ...,\ p_K), \tag{20}$$

$$p_l=p_{0,l}+T_l, \tag{21}$$

$$\boldsymbol{\alpha}=(\alpha_1,\alpha_2,...,\alpha_K), \tag{22}$$

$$\alpha_k=\alpha_{0,k}+R_k, \tag{23}$$

$$a_k=a_{0,k}+\begin{cases}\frac{R_k}{2} & 1\le k\le m,\\ R_k & m+1\le k\le K,\end{cases} \tag{24}$$

$$b_k=b_{0,k}+\begin{cases}\sum_{i=1}^{N}\sum_{l=1}^{L}\frac{r_{i,k,l}(x_i-o_l)^2}{2} & 1\le k\le m,\\ \sum_{i=1}^{N}\sum_{l=1}^{L}r_{i,k,l}|x_i-o_l| & m+1\le k\le K.\end{cases} \tag{25}$$

Here, $R_k=\sum_{i=1}^{N}\sum_{l=1}^{L}r_{i,k,l}$ and $T_l=\sum_{i=1}^{N}\sum_{k=1}^{K}r_{i,k,l}$.

Evaluation of lower bound. Evaluate the lower bound $\mathcal{L}(q)$ by means of the following equation. If $\mathcal{L}(q)-\mathcal{L}>\varepsilon$, set $\mathcal{L}=\mathcal{L}(q)$ and return back to E-step:

$$\mathcal{L}(q) = -\sum_{i=1}^{N}\sum_{k=1}^{K}\sum_{l=1}^{L} r_{i,k,l} \ln r_{i,k,l} - \frac{\ln 2\pi}{2}\sum_{k=1}^{m} R_k - \ln 2 \sum_{k=m+1}^{K} R_k$$
$$+ \ln\Gamma\left(\sum_{l=1}^{L} p_{0,l}\right) - \ln\Gamma\left(\sum_{l=1}^{L} p_l\right) - \sum_{l=1}^{L}(\ln\Gamma(p_{0,l}) - \ln\Gamma(p_l))$$
$$+ \ln\Gamma\left(\sum_{k=1}^{K} \alpha_{0,k}\right) - \ln\Gamma\left(\sum_{k=1}^{K} \alpha_k\right) - \sum_{k=1}^{K}(\ln\Gamma(\alpha_{0,k}) - \ln\Gamma(\alpha_k))$$
$$+ \sum_{k=1}^{K}(a_{0,k}\ln b_{0,k} - a_k \ln b_k) - \sum_{k=1}^{K}(\ln\Gamma(a_{0,k}) - \ln\Gamma(a_k)). \quad (26)$$

We give a short summary on the derivation of this algorithm. We assumed that the latent variable $\boldsymbol{Z}$ and the other parameters $(\boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\eta})$ are independent. We used equation (7) for the derivation of E-step and M-step. E-step is derived from the assumption that $(\boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\eta})$ are fixed. On the other hand, M-step is derived from the assumption that $\boldsymbol{Z}$ is fixed. The derivation shows that $q(\boldsymbol{Z})$ is of the form

$$q(\boldsymbol{Z}) = \prod_{i=1}^{N}\prod_{k=1}^{K}\prod_{l=1}^{L} r_{i,k,l}^{z_{ikl}}. \quad (27)$$

Table 2. Assumption on the prior distribution of parameters

Parameter	Prior distribution
$\boldsymbol{\pi}$	Dirichlet distribution $\mathrm{Dir}(\boldsymbol{\pi} \mid \boldsymbol{\alpha_0})$
η_k	Gamma distribution $\mathrm{Gam}(\eta_k \mid a_{0,k}, b_{0,k})$
$\boldsymbol{\omega}$	Dirichlet distribution $\mathrm{Dir}(\boldsymbol{\pi} \mid \boldsymbol{p_0})$, where $\boldsymbol{p_0} = (p_{0,1}, p_{0,2}, \ldots, p_{0,K})$

Table 3. Input/output parameters of variational Bayesian method

Parameter	Description
Input	
$\boldsymbol{X} = \{x_1, x_2, ..., x_N\}$	Observation data
$\boldsymbol{o} = (o_1, o_2, ..., o_L)$	Possible realizations of the discrete random variable (magnitudes of offset)
m	The number of Gaussian components mixed in $p(\boldsymbol{X} \mid \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$
n	The number of Laplace components mixed in $p(\boldsymbol{X} \mid \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$
$\boldsymbol{p_0}$	Parameters of the prior distribution $\mathrm{Dir}(\boldsymbol{\omega} \mid \boldsymbol{p_0})$ of the offset mixing coefficients $\boldsymbol{\omega}$
$\boldsymbol{\alpha_0}$	Parameters of the prior distribution $\mathrm{Dir}(\boldsymbol{\pi} \mid \boldsymbol{\alpha_0})$ of the mixing coefficients $\boldsymbol{\pi}$
$a_{0,k}$ and $b_{0,k}$	Parameters of the prior distribution $\mathrm{Gam}(\eta_k \mid a_{0,k}, b_{0,k})$ of precision parameters $\boldsymbol{\eta}$ of Gaussian/Laplace distribution $(k = 1, ..., K)$
ε	Small positive number used for defining the stop condition of the algorithm
Output	
$\boldsymbol{p}$	Parameters of the posterior distribution $\mathrm{Dir}(\boldsymbol{\omega} \mid \boldsymbol{p})$ of the offset mixing coefficients $\boldsymbol{\omega}$
$\boldsymbol{\alpha}$	Parameters of the posterior distribution $\mathrm{Dir}(\boldsymbol{\pi} \mid \boldsymbol{\alpha})$ of the mixing coefficients $\boldsymbol{\pi}$
a_k and b_k	Parameters of the posterior distribution $\mathrm{Gam}(\eta_k \mid a_k, b_k)$ of precision parameters $\boldsymbol{\eta}$ of Gaussian/Laplace distribution $(k = 1, ..., K)$

5. Estimation of Offset

We discuss the problem to estimate the magnitude of offset in a single observation x of deviation from the route center line. We assume that the output of the algorithms discussed in Section 4 is already available. The probability $\Pr(\mathcal{O}(x) = o_l)$ that the magnitude of offset equals to o_l is first discussed in this section. Here, $\mathcal{O}(x)$ denotes the latent random variable which represents the magnitude of offset in the observation x. It does not make sense for frequentists (non-Bayesian) to consider such probability because the magnitude of applied offset is a deterministic single value even though the ground does not know its value. However, the estimated magnitude of offset may be useful information for air traffic controllers, even if the estimation is a Bayesian probability, which only represents the degree of confidence.

We first consider the case where we only have point estimates of parameters, for instance, the case where we apply the EM algorithm for the estimation of model parameters. We will identify $p(\boldsymbol{Z} \mid x, \boldsymbol{\omega}, \boldsymbol{\pi}, \boldsymbol{\sigma}, \boldsymbol{\lambda})$ with the probability that the magnitude of offset equals to o_l and the Gaussian/Laplace component used for the generation of this datum x is the kth component. Here, $\boldsymbol{Z}$ is the latent variable defined in Section 4. The probability $\Pr(\mathcal{O}(x) = o_l)$ is given by the following formula:

$$\frac{\omega_l\left(\sum_{k=1}^{m} \pi_k \mathcal{N}(x - o_l, \sigma_k) + \sum_{k=1}^{n} \pi_{m+k} \mathcal{D}(x - o_l, \lambda_k)\right)}{\sum_{l=1}^{L} \omega_l\left(\sum_{k=1}^{m} \pi_k \mathcal{N}(x - o_l, \sigma_k) + \sum_{k=1}^{n} \pi_{m+k} \mathcal{D}(x - o_l, \lambda_k)\right)}. \tag{28}$$

We next consider the case where a posterior distribution for each parameter is available, for instance, the case where we apply the variational Bayesian algorithm for the estimation of model parameters. In this case, we simply apply the variational Bayesian algorithm in Subsection 4.3 for a single observation x. The distribution $p(\boldsymbol{Z})$ of $\boldsymbol{Z}$ gives the probability that the magnitude of applied offset equals to o_l and the Gaussian/Laplace component used for the generation of this datum is the *k*th component, say $r_{1,k,l}$. Therefore, the probability $\Pr(\mathcal{O}(x) = o_l)$ is simply $\sum_{k=1}^{K} r_{1,k,l}$.

We can consider various algorithms for the estimation of the magnitude of offset o_l from a single observation datum x using the probability $\Pr(\mathcal{O}(x) = o_l)$. A simple algorithm is to choose o_l maximizing the probability $\Pr(\mathcal{O}(x) = o_l)$. However, this algorithm returns some value o_l even if the probability $\Pr(\mathcal{O}(x) = o_l)$ is small. Information with weak confidence may create a confusion in the real operational environment. This algorithm does not seem appropriate.

An alternative algorithm is to choose the value o_l satisfying the following condition:

$$\Pr(\mathcal{O}(x) = o_l) > T. \tag{29}$$

Here, T is the pre-determined threshold. This algorithm returns the output 'unknown' when no o_l satisfies the above condition. We will investigate the accuracy of this algorithm through a numerical experiment in Subsection 6.2.

6. Numerical Experiment

We implemented the algorithms discussed in Section 4 as a Java program. This section introduces the result of a numerical experiment. A computer having a single Intel Pentium 4 CPU (2.8GHz) was used for the experiment.

Offset mixture distributions whose finite mixture distribution component

$$\sum_{k=1}^{m} \pi_k \mathcal{N}(x - o_l,\, \sigma_k) + \sum_{k=1}^{n} \pi_{m+k} \mathcal{D}(x - o_l,\, \lambda_k)$$

is an NN model, NDE model and DDE model are called an *offset NN model*, an *offset NDE model* and an *offset DDE model* for short.

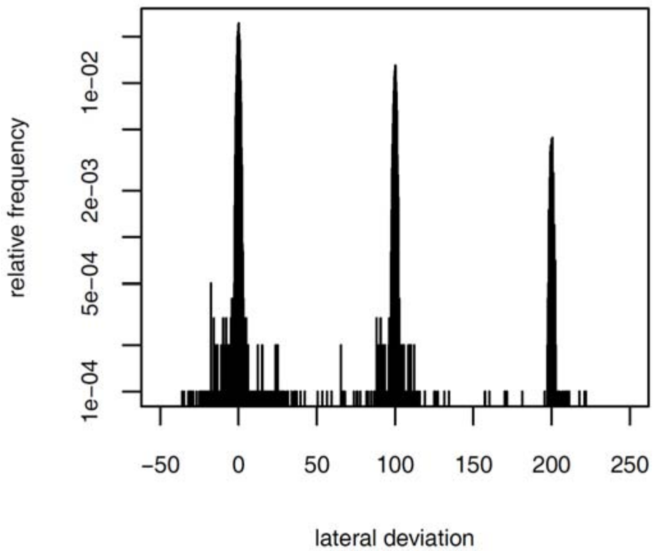

Figure 1. Histogram of offset mixture distribution (logarithmic scale in vertical axis), large offset increment case.

6.1. Estimation of model parameters

Consider the case where the scale parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ are much smaller than the increment of offset values. Figure 1 illustrates the relative frequency of pseudo-random numbers generated from an offset NDE model in logarithmic scale. The model parameters of these distributions are $L = 3$, $o_1 = 0$, $o_2 = 100$, $o_3 = 200$, $\omega_1 = 0.6$, $\omega_2 = 0.3$, $\omega_3 = 0.1$, $m = 1$, $n = 1$, $\pi_1 = 0.95$, $\pi_2 = 0.05$, $\sigma_1 = 1$ and $\lambda_1 = 10$ in the notations defined in Section 2.

The pseudo-random numbers which have a fixed magnitude of offset make a mountain-shape around the given magnitude in the graph. Almost all data are concentrated around the possible magnitudes of offset 0, 100 and 200. Very few are on the intermediate region among offset values. The magnitude of offset chosen by aircraft is intuitively trivial for many observation data, and we can estimate the unintended deviation for each observation with strong confidence. It is relatively easy to estimate the model parameters in this case. The numerical experiment showed the significant coincidence of the model parameter estimations by our algorithm with the true parameter values even in the case where the initial parameter values used in the algorithms are far from the true values.

The navigation performance requirement for aircraft flying in oceanic airspace is not so severe [17], and the scale parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ may not be smaller than the increment of offset values. Figure 2 also illustrates the relative frequency of pseudo-random numbers generated from an offset NDE model in logarithmic scale. In this case, the model parameters of these distributions are $L = 3$, $o_1 = 0$, $o_2 = 1$, $o_3 = 2$, $\omega_1 = 0.6$, $\omega_2 = 0.3$, $\omega_3 = 0.1$, $m = 1$, $n = 1$, $\pi_1 = 0.95$, $\pi_2 = 0.05$, $\sigma_1 = 0.1$ and $\lambda_1 = 1$. These parameters look more realistic than previous ones.

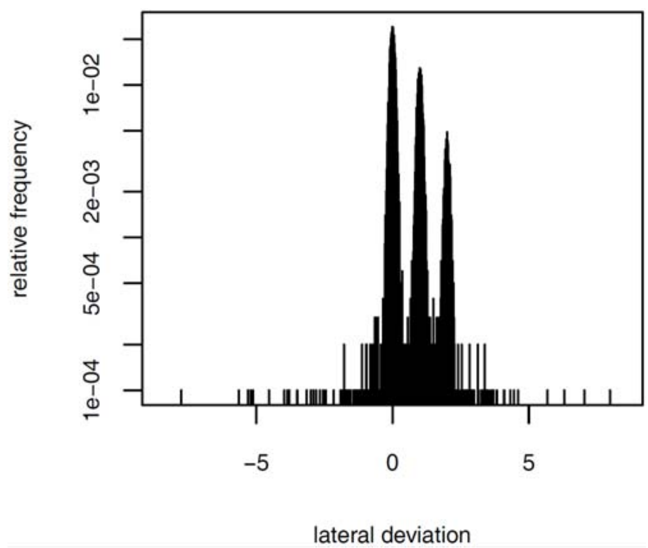

Figure 2. Histogram of offset mixture distribution (logarithmic scale in vertical axis), small offset increment case.

The feet of the two mountains in Figure 1 seem not to overlap. On the contrary, the feet in Figure 2 overlap. It makes it hard to estimate the magnitude of offset of a datum in the intermediate region in the latter case, and as a result, the estimation of model parameters becomes hard. We will apply our algorithms to this difficult case in this numerical experiment.

We consider an offset NN model, an offset NDE model and an offset DDE model in the numerical experiment. We generate 10,000 pseudo-random numbers for each model, and estimate the model parameters by means of the EM algorithm and the variational Bayesian method. The number of available observations is generally determined by operational environments and the duration of observation periods. Experiences show that it is several thousands, and the number 10,000 is not so unrealistic. Two computations (#1 and #2) were conducted for the variational Bayesian parameter estimation in different settings. We employed a prior distribution which is far from the true distribution in the first case. In the latter case, we employed a prior distribution whose point estimation coincides with the model parameters of the original models used for the generation of pseudo-random numbers.

Table 4 summarizes the result of model parameter estimations. The possible magnitudes of offset are 0, 1 and 2 in this experiment. The 'original dist.' column summarizes model parameters of the distribution generating pseudo-random numbers. We employed MAP (maximum a posteriori probability) estimates for the point estimation of model parameters from their posterior distribution in the variational Bayesian case. In MAP estimation, the point estimation of a model parameter is the mode of the posterior distribution.

Table 5 and Table 6 give the parameters of prior and posterior distributions used for the numerical experiments 'variational Bayes #1 and #2' in Table 4.

The computation times are in 1-8 seconds in all the cases. The estimations of offset NDE model parameters by means of the EM algorithm are far from the true values. It is because the EM algorithm is trapped into a local minimum. The software gave a reasonable estimation when we ran the same software against the same data with appropriate initial parameter values. The EM algorithm should be run several times changing initial parameter values to avoid being trapped into a local minimum. Table 4 also shows that the EM algorithm gives good estimations unless it is trapped into local minima.

Table 4. Parameter estimation result

Parameter \ Algorithm	Original dist.	EM initial	EM result	Variational Bayes #1	Variational Bayes #2
Offset NN model					
ω_1	0.6	0.5	0.5974	0.5974	0.5974
ω_2	0.3	0.3	0.3011	0.3016	0.3011
ω_3	0.1	0.2	0.1016	0.1011	0.1015
π_1	0.95	0.5	0.9594	0.9580	0.9546
π_2	0.05	0.5	0.0456	0.0420	0.0454
σ_1	0.1	10	0.1016	0.1012	0.0996
σ_2	1	20	1.0077	1.1830	1.0101
Log-likelihood	–1820.8		–1819.3	–1825.2	–1819.2
Execution time			4.209 sec	1.608 sec	1.348 sec
Offset NDE model					
ω_1	0.6	0.5	0.5945	0.5966	0.5967
ω_2	0.3	0.3	0.3065	0.3057	0.3055
ω_3	0.1	0.2	0.0990	0.0977	0.0977
π_1	0.95	0.5	0.0209	0.9580	0.9543
π_2	0.05	0.5	0.9791	0.0420	0.0457
σ_1	0.1	10	2.2112	0.1011	0.0995
λ_1	1	20	0.0845	1.2151	1.0420
Log-likelihood	–1912.5		–2254.9	–1914.5	–1910.2
Execution time			2.368 sec	1.456 sec	1.198 sec
Offset DDE model					
ω_1	0.6	0.5	0.5958	0.5952	0.5958
ω_2	0.3	0.3	0.3033	0.3036	0.3033

ω_3	0.1	0.2	0.1010	0.1011	0.1009
π_1	0.95	0.5	0.9499	0.9596	0.9504
π_2	0.05	0.5	0.0501	0.0404	0.0496
λ_1	0.1	10	0.0996	0.1013	0.0997
λ_2	1	20	1.0256	1.2808	1.0319
Log-likelihood	–4251.8		–4251.8	–4254.9	–4251.3
Execution time			7.081 sec	2.419 sec	2.859 sec

Table 5. Parameters of prior distributions and posterior distributions #1

Model / Parameter	Offset NN model	Offset NDE model	Offset DDE model
Prior distribution			
$p_{0,1}$	0.5	0.5	0.5
$p_{0,2}$	0.3	0.3	0.3
$p_{0,3}$	0.2	0.2	0.2
$\alpha_{0,1}$	0.8	0.8	0.8
$\alpha_{0,2}$	0.2	0.2	0.2
$a_{0,1}$	1	1	1
$b_{0,1}$	1	1	1
$a_{0,2}$	10	10	10
$b_{0,2}$	50	50	50
Posterior distribution			
p_1	5972.83	5965.20	5951.96
p_2	3015.68	3057.01	3036.26
p_3	1011.50	977.79	1011.78

α_1	9574.41	9578.98	9595.55
α_2	420.59	421.02	404.45
a_1	4790.30	4790.09	9595.75
b_1	49.08	48.99	972.39
a_2	220.20	430.82	414.25
b_2	306.78	522.28	529.30
Lower-bound	–1950.21	–2043.48	–4297.95

On the other hand, Table 4 shows the sensitivity of choice of prior distributions in variational Bayesian method. We can get a good estimation when the prior distributions are appropriately set. A heuristic solution for choosing prior distribution is to run EM algorithm first and determine the model parameters of prior distributions based on the output of EM algorithm.

6.2. Estimation of offset

We discussed methodologies for the estimation of applied offset in Section 5. We generated one thousand pseudo-random numbers from an offset NN model, an offset NDE model and an offset DDE model, respectively. The estimated model parameters and the distributions of model parameters in Table 4, Table 5 and Table 6 were used for this study. (The parameter re-estimated with appropriate initial values was used only in the 'Offset NDE model vs EM algorithm' case.) Table 7 shows the accuracy of the estimation against the given threshold values. The accuracy of estimation seems high, and the algorithm seems effective.

Table 6. Parameters of prior distributions and posterior distributions #2

Parameter \ Model	Offset NN model	Offset NDE model	Offset DDE model
Prior distribution			
$p_{0,1}$	0.6	0.6	0.6
$p_{0,2}$	0.3	0.3	0.3
$p_{0,3}$	0.1	0.1	0.1
$\alpha_{0,1}$	0.95	0.95	0.95
$\alpha_{0,2}$	0.05	0.05	0.05
$a_{0,1}$	100	100	10
$b_{0,1}$	1	1	1
$a_{0,2}$	10	10	10
$b_{0,2}$	10	10	10
Posterior distribution			
p_1	5973.52	5966.68	5957.43
p_2	3010.82	3055.35	3032.84
p_3	1015.66	977.98	1009.73
α_1	9544.77	9541.92	9502.97
α_2	455.23	458.08	497.03
a_1	4871.91	4870.49	9512.02
b_1	48.33	48.17	948.06
a_2	237.59	468.03	506.98
b_2	241.38	486.64	522.11
Lower-bound	–1838.20	–1929.00	–4272.55

Table 7. Summary of offset probability estimation

Method \ Threshold	99%	95%	90%	70%	50%
Offset NN model					
EM algo.	72.2%	94.7%	96.3%	98.5%	98.8%
Variational Bayes # 1	74.2%	94.5%	96.4%	98.4%	98.8%
Variational Bayes # 2	72.3%	94.7%	96.4%	98.5%	98.8%
Offset NDE model					
EM algo.	89.2%	96.5%	96.6%	97.7%	98.1%
Variational Bayes # 1	79.7%	95.4%	96.1%	97.4%	97.7%
Variational Bayes # 2	78.4%	95.2%	95.9%	97.4%	97.7%
Offset DDE model					
EM algo.	66.9%	92.6%	95.8%	98.0%	98.2%
Variational Bayes # 1	72.6%	94.3%	96.3%	98.0%	98.2%
Variational Bayes # 2	67.2%	92.7%	95.8%	98.0%	98.2%

7. Conclusion

We introduced EM algorithms and variational Bayesian methods for navigation performance and offset simultaneously in Section 4. Section 5 remarked that the same algorithm can be applied to the problem of estimating the offset from a single observation when Bayesian distributions of model parameters are available. We also gave a formula for the case where a point estimation of model parameters is only available. Numerical experiments in Section 6 showed that computations finish in seconds. The estimations of model parameters by means of EM algorithms and variational Bayesian methods are accurate when the initial parameters or the model parameters of prior distributions are appropriately set. They also showed the applicability of the algorithm for estimating the offset discussed in Section 5.

Acknowledgement

The author appreciates Claus Gwiggner, Sakae Nagaoka and Naoki Kanada for reviewing the draft papers and giving useful comments.

References

[1] Federal Aviation Administration (FAA), Aeronautical Information Publication, 21st eds.: Federal Aviation Administration, 2011.

[2] International Civil Aviation Organization (ICAO), Asia/Pacific Region En-route Monitoring Agency (EMA) Handbook, 2nd eds.: International Civil Aviation Organization, 2011.

[3] Aeronautical Radio, Incorporated (ARINC), Automatic Dependent Surveillance (ADS), ARINC Characteristic 745-2: Aeronautical Radio, Incorporated, 1993.

[4] RTCA, Minimum Operational Performance Standards for Airborne Automatic Dependent Surveillance (ADS) Equipment, RTCA DO-212: RTCA, 1992.

[5] R. Mori, Identifying the ratio of aircraft applying SLOP by statistical modeling of lateral deviation, Trans. Japan Soc. Aeronaut. Space Sci. 54(183) (2011), 30-36.

[6] G. J. Mclachlan and D. Peel, Finite Mixture Model, Wiley, New York, 2000.

[7] C. M. Bishop, Pattern Recognition and Machine Learning, Springer-Verlag, New York, 2006.

[8] P. F. Lazarsfeld and N. W. Henry, Latent Structure Analysis, Houghton Mifflin, 1968.

[9] G. J. Mclachlan and T. Krishnan, The EM Algorithm and its Extensions, Wiley, New York, 1997.

[10] E. M. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. 39B(1) (1977), 1-38.

[11] H. Attias, Inferring parameters and structure of latent variable models by variational Bayes, K. B. Laskey and H. Prade, eds., Uncertainty in Artificial Intelligence: Proceedings of Fifth Conference, Morgan Kaufmann, 1999, pp. 21-30.

[12] M. Sevensén and C. M. Bishop, Robust Bayesian mixture modelling, Neurocomputing 64 (2004), 235-252.

[13] D. A. Hsu, Long-tailed distributions for position errors in navigation, Applied Statistics 28(1) (1979), 62-72.

[14] T. Brewer-Dougherty, B. Colomosca, E. A. Elsayed, C. Gerhardt-Falk, D. Livingston, L. Martin and C. Schroepfer, Collision risk modeling in the northern Pacific airspace under separation reduction and improvements in navigational performance, ATC Quarterly 14 (2006), 257-282.

[15] S. Kullback and R. A. Leibler, On information and sufficiency, Ann. Math. Stat. 22(1) (1951), 79-86.

[16] R. T. Cox, Probability, frequency and reasonable expectation, Amer. J. Phys. 14(1) (1946), 1-13.

[17] International Civil Aviation Organization (ICAO), Performance-based Navigation (PBN) Manual, 3rd eds.: International Civil Aviation Organization, 2008.