Bayesian Approach Implementation on Quick Access Recorder Data for Estimating Parameters and Model Validation

Javensius Sembiring[a]*, Lukas Höhndorf[a], and Florian Holzapfel[a]
Institute of Flight System Dynamics TUM, München, Germany

Abstract: This paper presents the implementation of Bayesian inference on Quick Access Recorder data for parameter estimation purpose. Posterior density is sampled by employing Markov Chain Monte Carlo method. The reason for employing the Bayesian inference, instead of classical method such as Maximum Likelihood is because the data used in this paper has more uncertainties than the data obtained from a flight testing. These uncertainties come from the facts that Quick Access Recorder data obtained from untailored flight maneuvers, variables are measured/recorded at low and different sampling rates, control inputs such as elevator, rudder, aileron are not optimized, and flight is performed based on daily operational activities (wind and turbulence might disturb the measured variables). Results show that this approach is capable of capturing the uncertainties in the data since the estimated parameters are presented in the distribution forms. The flight data used as a case study are obtained from Airbus 320 Quick Access Recorder device. Some parameters to be estimated in this study consist of thrust and the effect of spoiler and flap deflection on lift and drag coefficient during approach phase.

Keywords: QAR Data, Bayesian Inference, Model Validation, MCMC
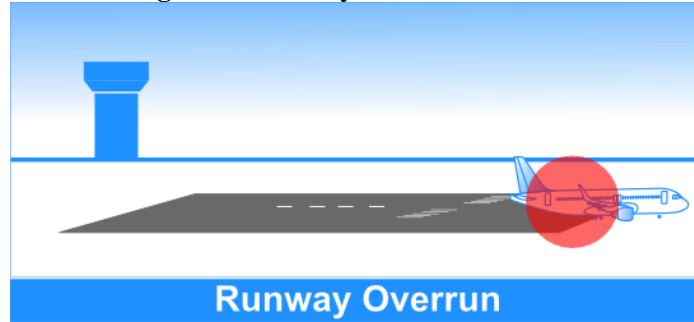
1. INTRODUCTION

Flight safety is one main issue in aviation area. Many efforts are conducted to improve safety level both from aircraft design and operational side. From operational point of view, airlines are interested in improving safety by utilizing flight data which are recorded during daily operational flight. The flight data called Quick Access Recorder (QAR) data is monitored or analyzed by flight safety crew and the result will be given to the related parties as a feedback. To deal with huge daily flight data, the flight safety department is provided with Flight Data Monitoring (FDM) program which is available commercially. These FDM programs work based on recorded parameter, analytical, and simple computation only. For instance, if there is a runway overrun incident occur, some possible parameters to investigate are spoiler deployment, thrust reverser, brakes – *are they working properly or not?*, and other related contributing factors to the incident (Figure 1).

However, sometimes these recorded parameters do not provide the flight safety crew with enough information to determine the cause of the incident. Parameter estimation technique comes into the picture by providing more parameters to be investigated in which these parameters are not recorded/measured directly in QAR data. As example of the incident mentioned above, the additional parameter which might be estimated is runway friction coefficient. This parameter is not recorded in QAR device but can be estimated by employing the parameter estimation method. Not only parameters during ground phase but also parameters during air-phase can be estimated such as lift and drag coefficient increment/decrement due to flap or spoiler deflection during approach phase. Implementation of the parameter estimation technique along with current FDM program would give a great benefit to FDM crew since more parameters are obtained and the cause of incident can be revealed

---

* Corresponding author: javensius.sembiring@tum.de

with more solid foundation. In this paper, the Bayesian inference is employed as estimation technique which presents the estimated parameter in a distribution form. The distribution of posterior density is sampled by using Markov Chain Monte Carlo (MCMC) method.

Figure 1. Runway Overrun Incident

The following sections of this paper are organized as follows: Section 2 introduces the problem formulation. In particular, the flight phases and related parameters to be estimated are thoroughly introduced. Section 3 briefly reviews the Bayesian method as well as technique for sampling the posterior density. This is followed by Section 4 which presents the implementation and result obtained by implementing the Bayesian inference on QAR data. Finally, Section 5 draws the conclusion for the paper.

2. PROBLEM FORMULATION

The number of parameters to be estimated is related to the postulated model and flight phase of the aircraft. In this paper, the approach flight (air phase) is selected as flight phase to be investigated (see Figure 1 and Table 1). During the approach phase, parameters such as lift and drag coefficient increment/decrement due to flap or spoiler deflection are estimated. Along with these aerodynamic parameters, thrust produced by aircraft engine is also estimated during this phase. The mathematical formulation during the flight phase is postulated as a linear relation as shown in the following:

$$a_x = \frac{1}{m}(\overline{q}S \sin \alpha\, C_L - \overline{q}S \cos \alpha\, C_D + \delta_T T) \tag{1}$$

where $a_x, m, \alpha, \overline{q}, \delta_T, S$ consecutively denote the acceleration along longitudinal axes, mass of the aircraft, angle of attack, dynamic pressure $\left(\frac{1}{2}\rho v^2\right)$, throttle input, and wing area. All these parameters are obtained from QAR data except S parameter which is obtained from A320 technical data [3].

Aerodynamic coefficients denoted by C_L (lift coefficient), and C_D (drag coefficient) as well as thrust (T) are parameters to be estimated. The aerodynamic coefficients during the selected flight phase are affected by flap and spoiler deflection and are modeled as incremental changes in the lift and drag coefficients, i.e., as ΔC_{LF}, ΔC_{LS}, ΔC_{DF}, and ΔC_{DS} (the subscript F and S denote flap and spoiler deflection) [1]. The effects of flap and spoiler deflection on lift and drag coefficient are investigated for three different flap settings, namely $\delta_F = 0$, 15, and 35 degrees and four different spoiler settings, i.e., $\delta_S = 0$, 18, 22, and 27 degrees.

Figure 1: Selected Approach Flight Phase (Flap, Spoiler Deflection and Longitudinal Acceleration Plots)
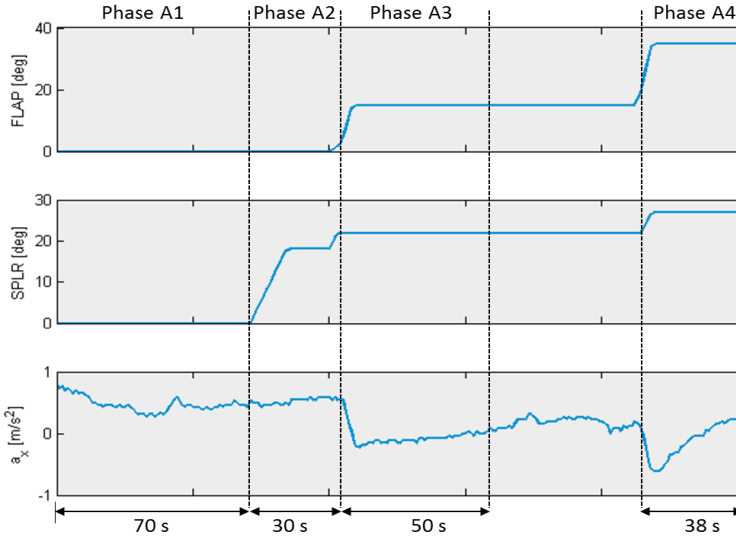
Table 1: Selected Flight Phase Description

Phase	Description
A1	FLAP = 0 SPLR = 0 Duration = 70 s
A2	FLAP = 0 SPLR = 18 Duration = 30 s
A3	FLAP = 15 SPLR = 22 Duration = 50 s
A4	FLAP = 35 SPLR = 27 Duration = 38 s

3. BAYESIAN INFERENCE AND POSTERIOR SAMPLING METHOD

3.1. The Bayes Formula

In the Bayesian context, the probability is represented as distribution of possible values. This approach is based on prior and likelihood distributions of parameters. The prior distribution describes our belief about the problem beforehand (subjective judgment). The likelihood represents the probabilities of observing a certain set of parameter values. Both of these distributions are updated to a posterior distribution, which represents the parameter distribution given on the observed data, formulated as follows:

$$p(\boldsymbol{\theta}|\boldsymbol{Z}) = \frac{p(\boldsymbol{Z}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{\int p(\boldsymbol{Z}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{2}$$

where, $p(\theta|Z)$, $p(data|\theta)$, $p(\theta)$, and $\int p(data|\theta) \cdot p(\theta)d\theta$ consecutively denote posterior, likelihood, prior and normalizing constant, while Z and θ denote data and unknown parameters consecutively. The Bayesian solution for parameter estimation is the posterior distribution of parameters (conditional probability of unknown parameters given the data). This posterior distribution is the distribution we are interested in knowing since it represents the distribution directly of the unknown parameter. In this paper, the prior of unknown parameter is assumed to be *uninformative prior*, i.e. $p(\theta) = 1$, whereas the likehood is formulated as:

$$p(Z_i|\theta) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} e^{-\frac{1}{2}(Z_i - y(t_i;\theta)^T C^{-1}(Z_i - y(t_i;\theta)/\sigma^2} \tag{3}$$

Variable $y(t_i; \theta)$ denotes the postulated model which depends on the unknown parameters. The likelihood formulation in equation (3) is based on assumptions that the measurement error is distributed as Gaussian with mean zero and covariance C, that is $\epsilon \sim N(0, C)$. Furthermore, if the measurement error ($\epsilon_i = Z_i - y(t_i; \theta)$) is assumed to be independent and normally distributed, that is $\epsilon_i \sim N(0, \sigma^2)$ and $\epsilon \sim N(0, \sigma^2 I)$, equation (3) is simplified to (4):

$$p(Z_i|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(Z_i - y(t_i;\theta)^2/\sigma^2} \tag{4}$$

or in the combined likelihood of all the measurements can be written as a product:

$$p(Z|\theta) = \prod_{i=1}^{n} p(Z_i|\theta) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(SS_\theta)^2/\sigma^2} \tag{5}$$

where $SS_\theta = \sum_i^n (Z_i - y(t_i; \theta))^2$. By combining the prior and likelihood as defined above, the posterior up to the normalizing constant can be written as:

$$p(\theta|Z) \propto p(Z|\theta) \tag{6}$$

Equation (6) is the distribution of interest which will be sampled using technique described in section 3.2 below.

3.2. Posterior Sampling Method

Sampling from posterior density function is performed by employing Markov Chain Monte Carlo method. Metropolis algorithm is used as acceptance/rejection criteria, while random walk along with component wise method is used for proposing the candidate parameter. The Metropolis acceptance/rejection criterion is defined in equation (7) below.

$$\alpha(\theta^{i-1}, \theta^*) = min\left\{1, \frac{p(\theta^*)}{p(\theta^{i-1})}\right\} \tag{7}$$

The candidate parameter (θ^*) will be accepted if $(\theta^{i-1}, \theta^*) > u$, where $u \sim U[0,1]$. The pseudo-code of this algorithm is presented in Section 4.

3.3. Initial Parameter and Covariance Estimation

Using Random Walk Metropolis algorithm with Gaussian proposal distribution in posterior sampling process requires a guess for the covariance matrix C as well as starting values of parameters. The starting value of parameters can be estimated by using a least square sense as formulated in equation (8).

$$\theta_0 = \min_\theta \sum_{i=1}^{n} \left(Z_i - y(t_i; \theta)\right)^2 \tag{8}$$

Taking parameters obtained from equation (8) as starting values in MCMC process avoid a long burn-in time and speed-up the convergence rate of sampling process. The parameter covariance matrix C is obtained by employing equation (9) below:

$$C \approx \sigma^2 [X^T X]^{-1} \tag{9}$$

where σ^2 and X consecutively represent variance of the residual and Jacobian matrix calculated at θ_0. In details, the step obtaining the initial parameter and covariance matrix can be found in [4].

3.4. Model Validation

Model validation is performed by comparing model output and measurement. In the context of Bayesian inference, the model output is described as predictive distributions of model output. The predictions of model output are naturally based on the posterior distribution of estimated parameters, as defined as [2]:

$$p(Z^*|Z) = \int p(Z^*, \theta|Z)d\theta = \int p(Z^*|\theta)\, p(\theta|Z)d\theta \tag{10}$$

Here $p(Z^*|Z)$ denotes the prediction of future observations Z^* given Z as the current one. Since the model output based on distribution of estimated parameter, the model output will be in distribution form as well as forming a confidence region. This approach is different with the classical estimation approach in which the model output is based on a single value of estimated parameter hence forms the model output without distribution [5, 6].

4. IMPLEMENTATION AND RESULTS

4.1. Implementation

The pseudo code of Random Walk Metropolis algorithm is presented below. It shows the initial value of parameter and covariance are first estimated in the least square sense. After the initialization, the posterior sampling process is done iteratively until reaching the maximum number of samples. The candidate parameters will be accepted/rejected based on criterion defined in equation (7). The accepted parameters are then stored and form the posterior distribution.

Random Walk Metropolis pseudo code with estimated initial parameter and covariance

1. Initialize
- Choose $\theta^0 = \arg\min_{\theta \in Q} \sum_{j=1}^{n}\big(z_j - y(t_i, \theta)\big)^2 = \arg\min_{\theta \in Q} SS_\theta$ using optimization routine.
- Compute $SS_{\theta^0} = \sum_{j=1}^{n}\big(z_j - y(t_i, \theta^0)\big)^2$
- Estimate error variance $\sigma^2 = \frac{1}{n-p}\sum_{j=1}^{n}\big(z_j - y(t_i, \theta^0)\big)^2$
- Construct covariance matrix $C = \sigma^2[X(\theta^0)^T \cdot X(\theta^0)]^{-1}$
- Compute $R = \text{cholesky decomposition}(C)$

2. Do sampling
- for i = 1, 2, ...n (n = number of samples)

    # sample $z \sim N(0,1)$

    # construct candidate $\theta^* = \theta^{i-1} + Rz$

    # sample $u_\alpha \sim U(0,1)$

    # compute $SS_{\theta^*} = \sum_{j=1}^{n}\big(z_j - y(t_i, \theta^*)\big)^2$

    # compute $\alpha(\theta^*, \theta^{i-1}) = min\left(1, e^{-[SS_{\theta^*} - SS_{\theta^{i-1}}]/2\sigma^2}\right)$

    if $u_\alpha < \alpha$

        # set $\theta^i = \theta^*$, $SS_{\theta^i} = SS_{\theta^*}$

    else

        # set $\theta^i = \theta^{i-1}$

    endif

  endfor

The pseudo code of Random Walk Metropolis algorithm shown above is then implemented based on the postulated model and the selected flight phase as defined in Section 2. Number of samples is set to 80,000 with acceptance ratio for each of parameters varies between 3% - 18%. The related results are shown in the Figure 3 to 6 below.

During Phase A1, the thrust varies in range between 0.8 – 1 kN, lift coefficient deflection varies in range between 0.4 – 0.5, and drag coefficient has value distributed in range 0.03 – 0.05. Both of these aerodynamic coefficients are obtained with no flap and spoiler deflection. The predictive model output distribution is also presented along with measurement (left side). From Figure 3, it shows that the predictive model output match with a good agreement with that of the measurement.

In Phase A2, the flap deflection is still in the same state as Phase A1 but spoiler now deflected to 18 degrees. This spoiler deflection decreases the lift coefficient and increases the drag coefficient. These effects can be seen in Figure 4, i.e. the lift coefficient decreases to value between 0.24 – 0.28, whereas drag coefficient increases in range 0.08 – 0.1. Predictive output plot is also presented along with measurement (Figure 4, left side), but the distribution of predictive plot is wider than that of Phase A1. This indicates the estimated parameters in Phase A2 have more uncertainties than the parameters estimated in Phase A1. The change in thrust parameter is not caused by flap/spoiler deflection but by the flight condition and throttle command from pilot.

**Figure 2: Phase A1 Predictive Model Output and
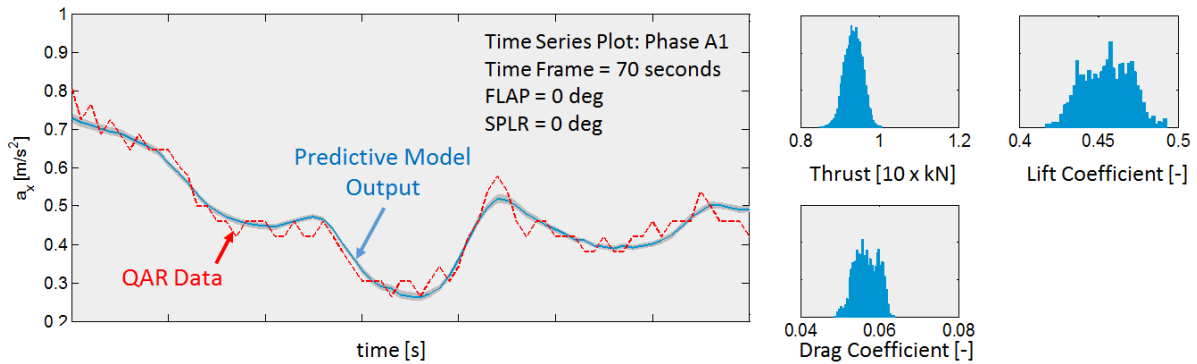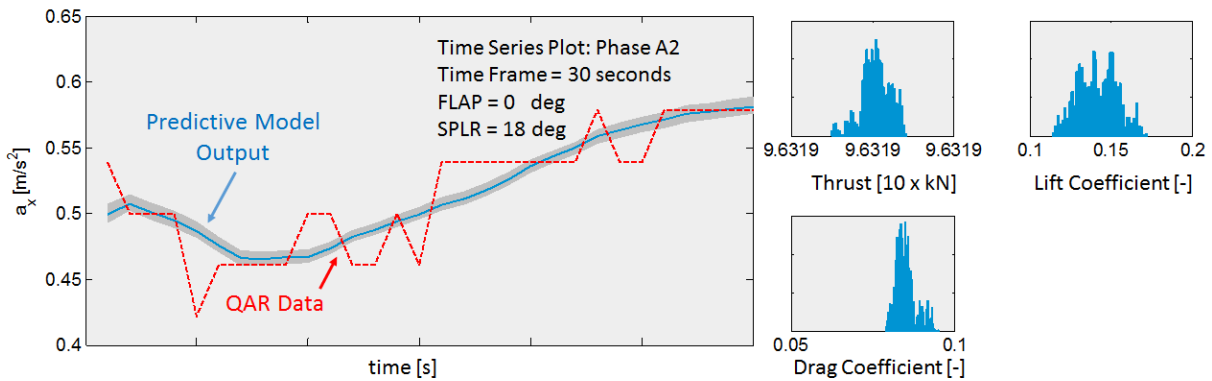Related Estimated Parameters Distribution**

**Figure 4: Phase A2 Predictive Model Output and
Related Estimated Parameters Distribution**

During Phase A3, both flap and spoiler are in deflection state of 15 and 18 degrees consecutively. This deflection affects both lift and drag coefficient. The lift coefficient increases in range 0.24 – 0.28, whereas the drag coefficient varies in range 0.125 – 0.135 as shown in Figure 5 (right side). The predictive plot distribution of model output is also plotted together with the measurement. From Figure 5, it can be seen that the trend of the measurement can be captured by the model output.

**Figure 5: Phase A3 Predictive Model Output and
Related Estimated Parameters Distribution**
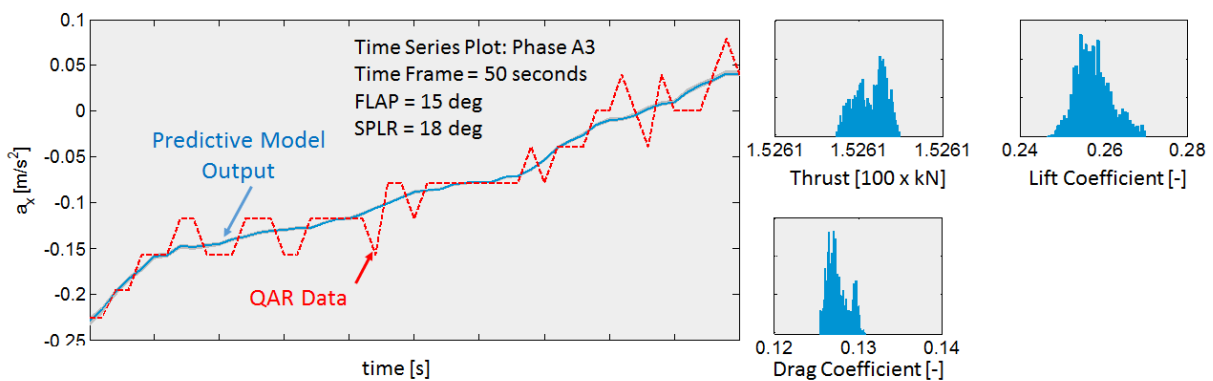
The results during Phase A4, presented in Figure 6 below, shows that the lift coefficient is now increasing with values around 0.8. This increment might come mostly due to the flap deflection (35 degrees). On the other hand, the drag coefficient slightly change and remain in value around 0.09 – 0.1. The predictive model output distribution is also plotted together with measurement as shown in Figure

6 (left side). It shows that they are in a good agreement with small spread of the predictive distribution of model output. This indicates that the estimated parameters have a high accuracy in the postulated model being employed.

**Figure 6: Phase A4 Predictive Model Output and
Related Estimated Parameters Distribution**
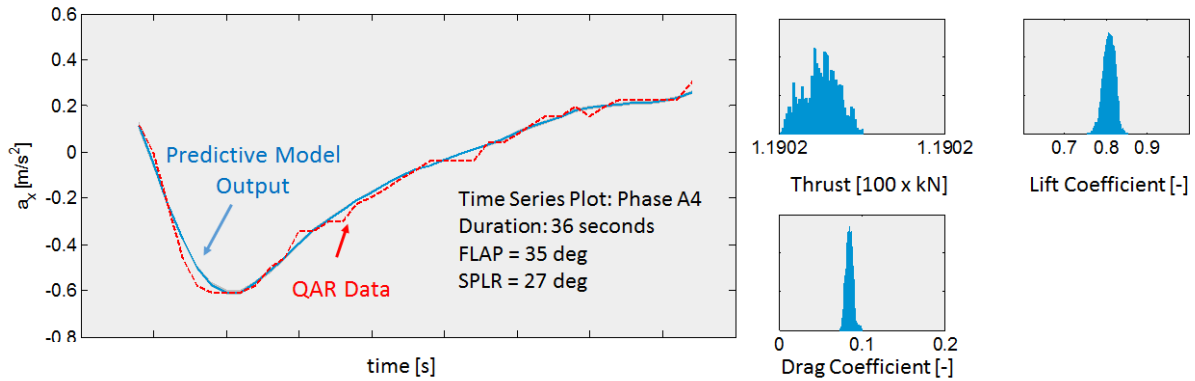
5. CONCLUSIONS

The Bayesian inference along with MCMC technique are successfully implemented on Quick Access Recorder data. The estimated parameters are represented as distribution form which gives information about the uncertainties in parameters. The algorithm implemented in this study opens the possibility to be integrated into current FDM program hence can extend the capability and functionality of the program. The benefit of this implementation provides parameters which are not recorded/measured in QAR device. These estimated parameters can be used as additional information by FDM crew to investigate a specific event or incident so that the cause of incident can be determined with more solid foundation. The output of this work is also used as an input in one of active research related to incident prediction in Institute of Flight System Dynamics, TUM.

References

[1] Jategaonkar, R.V., Mönnich, W., Fischenberg, D., and Krag, B., "*Identification of Speed Brake, Air-Drop, and Landing Gear Effects from Flight Data*", Journal of Aircraft, Volume 32, No. 2, (1997).

[2] Laine, Marko, "*Adaptive MCMC Methods with Applications in Environmental and Geophysical Models*", Dissertation, Finnish Meteorological Institute, 2008, Helsinki.

[3] http://www.airbus.com/aircraftfamilies/passengeraircraft/a320family/a320/specifications/ Feb. 2 2014.

[4] Solonen, Antti, "Monte Carlo Methods in Parameter Estimation of Nonlinear Models", Master's Thesis, Lappeenranta University of Technology, 2006, Lappeenranta.

[5] Jategaonkar, R.V., *Flight Vehicle System Identification – A Time Domain Methodology*, AIAA Progress in Astronautics and Aeronautics, AIAA, 2006, New York.

[6] Klein, V., and Morelli, E., *Aircraft System Identification: Theory and Practice*, AIAA Progress in Astronautics and Aeronautics, AIAA, 2006, New York.

[7] Bolstad, William M., *Understanding Computational Bayesian Statistics*, John Wiley & Sons, 2010, New Jersey.