International Journal of Mathematical Archive-2(10), 2011, Page: 1963-1979
Available online through www.ijma.info ISSN 2229 – 5046

BAYESIAN AND NON-BAYESIAN ANALYSIS FOR RANDOM CHANGE POINT PROBLEM USING STANDARD COMPUTER PACKAGES

Ayman A. Mostafa* and Anis Ben Ghorbal

Department of Mathematics, Al-Imam Muhammad Ibn Saud Islamic University, P.O. BOX 90950, 11623 Riyadh (Saudi Arabia)

E-mail: aasaad@imamu.edu.sa; assghorbal@imamu.edu.sa

(Received on: 10-10-11; Accepted on: 20-10-11)

ABSTRACT

Change point problems are used in many application fields. A general approach to Bayes, hierarchical Bayes, and duplicate the results using non-Bayes approach, for random change point problem are presented. In Bayesian analysis, Markov Chin Monte Carlo (MCMC) methods have become a ubiquitous tool as the computer is more powerful. Although Spiegelhalter et al. 1996 propose code software using WinBUGS (standard Bayesian software's package freely available, Spiegelhalter et al. 2003) to detect the change point of normal datasets using WinBUGS, their results yield not investigated well. We argue that a more realistic description of random change point problem as will be resolved in this article in some details and duplicate these results from non-Bayesian perspective using some of R packages (R Development Core Team, 2011). Because of the statistical performances of the Bayes estimates for the random change points are so sensitive for the prior elicitation that we should duplicate our results using the non-Bayesian approach. Within these contexts we handle two textbook data examples to compare the results between two approaches, Bayesian and non-Bayesian statistics.

Keywords: *Bayesian analysis; Change and Joint point; Longitudinal data; Software; MCMC algorithms; Model comparison (DIC); Piecewise linear regression; breakpoint; segmented linear regression.*

MSC 2010 Mathematics Subject Classification: *62C10, 62J99, 62-07, 68N15.*

---

1. INTRODUCTION

In statistical analysis, it is of interest to check if and when a functional relationship between an explanatory and a response variable remains unchanged over the whole domain of the explanatory variable or whether the functional form changes at certain unknown points, the so-called breakpoints (or change points). The problem of change point considered with increasing interest in the literature, see e.g., *(Picard, et al., 2011)* and references therein). Thus, testing for the existence of a breakpoint is often important task *(Haybach, et al., 1997)*. The problems of identifying changes at unknown times and of estimating the location of changes is known as 'the joint point problem' or 'the change point problem'. The technique of joint point regression (also known as broke stick, hockey stick regression, piecewise regression, change of phase, or segmented) that explains the relationship between two variables by means of a segmented linear regression constrained to be continuous everywhere, in particular in those places where the slope of the regression function changes can detect the breakpoints everywhere in the relationship between the two variables. Applications of this technique are widely applied in many areas of interest. For example, in survival data analysis (see e.g., *(Jensen, et al., 2008)*) or incidence series in epidemiological studies (see e.g., *(Czaikowski, et al., 2008)*). In these applications, the number (if any) and the location of the changes in trends (known as change points or joint points) are usually unknown, the main goal being to assess their existence and determine their location. Historically, methods such as control charts have been used to detect changes in data; however, these charts often fall short in detecting change if the change in the data is subtle. The standard change point problem in regression models consists of two main problems. The first is testing the null hypothesis that no change in regimes has taken place against the alternative that observations were generated by two (or possibly more) distinct regression equations, and the second problem is estimating the two regimes that gave rise to the data that desired to detect the change points. There are two common models. First are models where the regression line is continuous at the break point and models where the regression line can be discontinuous. In this paper, we only consider the continuous case. The standard method is that fitting piecewise regression models are prone to yield different parameter estimates depending on the assigned starting values.

**Corresponding author: Ayman A. Mostafa*, *E-mail: aasaad@imamu.edu.sa*

This problem has a long history (see e.g., *(Lerman, 1980)*; *(Seber, et al., 2003)*; *(Toms, et al., 2003)*; and others). The change point model starts with the broken stick model seen earlier, i.e.

$$y = \beta_1 + \beta_2 (x)^- + \beta_3 (x - c)^+ + \varepsilon \qquad (1)$$

Where y is the response variable, x is the covariate, and c is the change point where the break occurs, β_1 is the intercept, β_2 is the slope before the change point c and β_3 is the difference in slope after the change point. The slope after the change point is $\beta_2 + \beta_3$. The variable $(x - c)^+$ is a derived variable which takes the value of zero for values of x less than c and the values $(x - c)$ for values of x greater than c, this is called a linear spline. When testing whether for the existence of the change point, the unknown change point is often a parameter present only under the alternative. The standard change point problem in regression models consists of

(1) Testing the null hypothesis that no change in regimes has taken place against the alternative that observations were generated by two (or possibly more) distinct regression equations and;
(2) Estimating the two regimes that gave rise to the data.

In statistical inference, there are two broad categories of interpretations of probability: Bayesian inference and non-Bayesian (frequentist) inference (see for details (Hall, 2011)). Hall introduces article, examples and a Package in R (R Development Core Team, 2011) as a handy summary for experienced Bayesians. From frequentist point of view, the hypothesis of interest is to test the null hypothesis that H_0: $\beta_3 = 0$, which indicates no change in slope between $x < c$ and $x > c$. If the value of c is specified in advance, ordinary least-squares can be used to fit the model. Most computer packages can easily fit this model like JMP software (see e.g. http://www.jmp.com/applications/statistics/). There has been some overlap between estimation and testing. For example, Kosorok and Song (Kosorok, et al., 2007) test for the existence of a change point by trying to find the supremum of the score process and use Monte Carlo procedures to evaluate the properties of the test statistic. But perhaps testing and estimation should be separated, at least procedurally. Henderson (Henderson, 1990) demonstrates that reasonable estimates for the change point that come from testing procedures (such as maximizing a statistic over a range of potential values for the change point) might not be sufficient under the alternative hypothesis. In other words, there might be more information about the change point from the data than is contained in the estimates of the change point that are obtained from maximizing test statistics. There are two common models. First are models where the regression line is continuous at the change point and models where the regression line can be discontinuous (see for details (Hawkins, 1980)). In this paper, we only consider the continuous case. In many cases the location of the break is not known, and it is of interest to estimate the break point as well. The problems of identifying changes at unknown times and of estimating the location of changes is known as 'the change point problem'. Numerous methodological approaches have been implemented in examining change point models, from the two points of view Bayesian and non-Bayesian (frequentist). Maximum-likelihood estimation (see e.g., (Fotopoulos, et al., 2001)), Bayesian estimation (Barry, et al., 1993), isotonic regression (see, e.g., (Wu, et al., 2001)), piecewise regression (see (Toms, et al., 2003)), quasi-likelihood (see e.g., (Braun, et al., 2000)) and non-parametric regression (see (Loader, 1996)) are among the methods which have been applied to resolving challenges in change point problem. Grid-searching approaches have also been used to examine the change point problem. A review of the literature (as of 2008) especially as it applies to regression problem is available at: http://biostats.bepress.com/cobra/ps/art44/. This model is appropriate where there is an unexpected transition at the break point, but a smooth transition may be more realistic for some data. The drawback of model (1) appears clearly in Bayesian approach, which is that convergence problems can occur in locating c when the data are sparse around the neighborhood of c. The so-called bent-cable model is also used to describe natural phenomena which exhibit a potentially sharp change in slope. The model comprises two linear segments, joined smoothly by a quadratic bend. The bent-cable regression model was recently developed by (Chiu, et al., 2006). The bent-cable regression model fits a smooth transition between the two linear parts of the model. The model comprises two linear segments, joined smoothly by a quadratic bend. The class of bent cables includes, as a limiting case, the popular piecewise-linear model (with a sharp knot), and otherwise known as the broken stick. The latter is also applicable to regression models where the x variable is time and auto-correlation may be present (see (Chiu, et al., 2010)). There are many segmentation algorithms are based on Bayesian inference or on the penalized maximum likelihood principle which can be viewed as a Bayesian strategy with appropriate prior knowledge. Although frequentist procedures for change point analysis estimate specific locations of change points, the Bayesian procedure offers a probability distribution - the probability of a change point at each location in a sequence. To assist the reader, we present briefly a basic method for detecting the change points using the both perspectives Bayesian and non-Bayesian. We then describe the Bayesian approach using the free software WinBUGS and we duplicate this detection using some of methods in non-Bayesian methods. The Bayesian perspective may be necessary are the following:

(1) Models with many random effects (easily set up and estimate difficult models);
(2) Models using exotic probability distributions (most probability models become exotic when coupled with random effects, e.g., there currently are limited frequentist tools available for fitting negative binomial regressions with random effects.);
(3) There are many parameters to estimate;
(4) Models in which the estimated variances of some of the random effects are small and/or the number of replicates is small.

One or more of these four conditions is typical in many fields in applications in real problems (e.g. in ecology). The only real problem from a practical change point problem with using Markov chain Monte Carlo in model estimation is not the intensive computational, but that it is difficult to assign a suitable uninformative prior for *c* (i.e. need defense of priors or sensitivity analyses of prior specification; no guarantee of Markov Chain convergence) . Therefore we try to detect the change point using the non-Bayesian approach and then we use the Bayesian approach to help us for detecting a suitable prior distribution for that change points, trying to answer the question which states that "Is MLE/frequentist approach simply Bayesian statistics in change point problem with an uninformative prior ?" . There are many packages in R which could be helping us to detect the change points (see e.g., (Killick, et al., 2011) and references herein). For example, (Erdman, et al., 2010), propose a package for the change point problem. In their work, they offer a new R implementation of the Bayesian change point procedure, which was proposed by (Barry, et al., 1993) used product partition models as distributions over possible segmentations of time-series data.

In change point problem, the statistical performances of the Bayes estimates for the random change points are so sensitive for the prior elicitation that we should duplicate our results using the non-Bayesian approach. Within these contexts, first we review the change point problem through handling two textbook data examples to compare the results between two approaches, Bayesian and non-Bayesian statistics. The two textbook examples can be considered a hierarchical linear model (HLM) or linear mixed model (LMM), or Random Coefficients model.

2. CHANGE POINT PROBLEMS AS MIXED MODELS

To illustrate the use of WinBUGS to fit statistical models in change point problems well, we introduce a standard statistical model that was historically known as mixed models. In the current version of the R environment (R Development Core Team, 2011) a number of packages can fit the change point problem, for example (Muggeo, 2011), present a general methodology of segmented relationships in regression models, (Zeileis, et al., 2011) present methodologies in Testing, monitoring and dating structural changes in (linear) regression models. Another useful package in detecting change points is available with the segclust R package to perform segmentation and segmentation/clustering as presented in the Biometrics paper (Picard, et al., 2007). (Sonderegger, 2011) offer a package which include methods with straightforward functions within R environment to help the user to detect the change point location, such as 'bent.cable', 'locally.weighted.polynomial', 'piecewise.linear' functions. These methods provide statistics for inferences, from non-Bayesian viewpoint, about whether the dependent variable *y* varies over independent variable *x* and when changes occur in the time *x*-axis. This paper follows (Spiegelhalter, et al., 1996) to show how to do the Bayesian analysis of change point models using WinBUGS and we also use some of the above mentioned in R packages in an attempt to duplicate the results of the frequentist analysis. There are a large variety of change point detection problems that present themselves in the analysis of time series and dynamical systems. We focus entirely on fitting the analysis of change point model as mixed models using WinBUGS. BUGS (Spiegelhalter, et al., 1996) is specialized software package for implementing MCMC-based analyses of full probability models in which all unknowns are treated as random variables. The BUGS project began at the Medical Research Council Biostatistics Unit in Cambridge in 1989. Since that time the software has become one of the most popular statistical modeling packages, with, at the time of writing, over 30,000 registered users of WinBUGS worldwide. BUGS have been just one part of the tremendous growth in the application of Bayesian ideas over the last 20 years, see (Lunn, et al., 2009). Prior to the widespread introduction of simulation-based methods, Bayesian ideas could only be implemented in circumstances in which solutions could be obtained in closed form in so-called conjugate analyses, or by ingenious but restricted application of numerical integration methods. The importance of the software has been acknowledged in 'An International Review of U.K. Research in Mathematics' (http://www.cms.ac.uk/irm/irm.pdf). It turns out that WinBUGS can become quite a powerful and flexible tool for Bayesian analysis, and only requires a relatively small investment on the part of the user. Once the (applied) user understands the logic of model building with WinBUGS, Bayesian analysis is conducted quite easily and many built-in features can be accessed to produce an in-depth and interactive analysis of detecting the change points. In addition, execution is reasonably fast, even of complicated models with large amounts of data and model extensions can easily be accommodated in a modular fashion. The WinBUGS software (together with a user manual) can be downloaded (the current fee is zero) from the website (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml).

Therefore, in this section we explain why one can view change point problems as mixed models and the following section will be devoted to illustrate the use and flexibility of WinBUGS for Bayesian change point modeling of the Stagnant band height data example, which has introduced by Spiegelhalter et al. (Spiegelhalter, et al., 1992). In their example, they follow the seminal paper by Carlin et al. (Carlin, et al., 1992) to describe the change linear regression problem using WinBUGS. Now; consider the regression model

$$y_{ij} = f(x_{ij}) + \varepsilon_{ij} \qquad (2)$$

where ε_{ij} are independent and identical distribution (response errors) such that $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon)$ and $f(x_{ij})$ is a smooth function; x_{ij} could be random but their distribution is left to be unspecified and could be degenerate distribution (a probability distribution of a random variable which always has the same value, e.g., x_{ij} could be equally

spaced). (Chiu, et al., 2006) give an extensive list of work using conventional continuous change point models. A popular one is the piecewise-linear (broken-stick) model, which, as discussed by these authors, can be inappropriate in the absence of clear scientific theory or evidence that supports an abrupt change as presumed by the model's kink. In contrast, as it will be outlined in this article, it is straightforward to use the WinBUGS software to change the assumptions of the model and compare the results to obtain the best one fit to the analyzed data and we recommend to duplicate the results with that methods because the prior distribution for the change point is so sensitive as we will see in the two examples given in this paper.

As an illustration of the change point problem in Fig. 1 a graphical representation of several break-points corresponding to different locations of the change points is shown. As can clearly be seen, it is assumed that we have a sequence of observations for subject i (x, y) and that c_j, $j = 1, 2, ..., m$, be $m+1$ points at the *x*-axis so that the interval $[c_0, c_m]$ contain all x_j, for simplicity, it may be assumed that the points are equidistant, i.e., $c_j - c_{j-1} = \delta$. For the change point analysis, we wish to find the $j+1$ values d_j in the *y*-axis so that the broken line defined by the $j+1$ points (c_j, d_j) 'fit' the set of points (x, y). This fitting is meant in terms of minimizing the total square error among the set of original points (x, y) and the fitted broken line (c_j, d_j) so that the square difference between *y* and $\hat{y}$ in the *y*-axis is minimum. The general form of the change-point problem is to determine the unknown location (c_j, d_j), based on an ordered sequence of observations $(x_1, x_2, ..., x_m)$. In the regression literature, the change-point model is also referred to as two-or multiple-phase regression, switching regression, segmented regression, two-stage least squares (Shaban, 1980), or broken-line regression.

To provide an illustration of the type of data will be analyzed to detect the change point location, consider the data matrix of responses as follows:

	event $1 = x_{11}$	event $2 = x_{22}$	$\cdots$	event $j = x_{ij}$	$\cdots$	event $m = x_{nm}$
subject 1	y_{11}	y_{12}	$\cdots$	y_{1j}	$\cdots$	y_{1m}
subject 2	y_{21}	y_{22}	$\cdots$	y_{2j}	$\cdots$	y_{2m}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{im}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_{n1}	y_{n2}	$\cdots$	y_{12}	$\cdots$	y_{nm}

The first subscript refers to an event (could be the time) and the second to subject. This constitutes a null model and in general we will wish to include further variables, notably age or time. Suppose we wish to express the response, as a function of time measured at each a certain event. The multivariate continuous responses y_{ij} are recorded on subjects at each event we will generally wish to model the average time relationship for each response and the covariance matrix among the responses as a function of time. This is readily done by considering the multivariate response structure as a further, lowest, level in the data hierarchy with measurements nested within events. There are several advantages to considering the joint modeling of several responses. The ability to estimate their covariance matrix as a function of time allows one to study the distribution of any function of the responses with respect to time. For example, when studying issues of prior determination it may be useful to see whether the correlation between two variables a given time apart is greater when one is the prior variable rather than the other. Likewise, it provides a general prediction procedure for one measurement, conditional on any set of observed prior measurements.

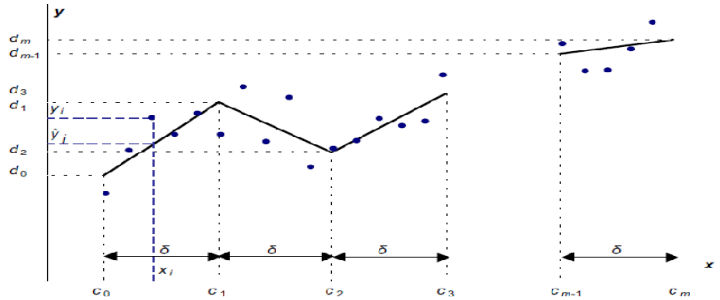

Fig. 1: Definition Sketch

3. CHANGE POINT PROBLEM AS A MIXED MODEL

This section is concerned with modelling data where measurements of one or more attributes are repeated on the same set of individuals over time (longitudinal or cross-sectional data). Statistical analysis for cross-sectional or longitudinal data with data missing in no monotonic patterns is very difficult. For analyzing longitudinal data with no ignorable and no monotone missing responses, a full likelihood method is complicated algebraically, especially with the change point problem and often requires intensive computation, especially when there are many follow-up times. As an alternative, in this section we explain a uniform and straightforward procedure using WinBUGS software to deal with cross-sectional or longitudinal with observations, which could be missing, with in arbitrary non monotonic patterns, estimating regression parameters with the change point problem using available observations based on the regression model of interest and parametric working models. The model specification will be developed for the case where a single or multiple continuous measurements are made on several occasions for a sample. This will then be extended to consider the case of multiple measurements at each time point and mention will be made of extensions to latent variable models and to continuous response data. To begin with we look at the simple, restricted, data structure where there are a fixed number of measurement occasions and each individual has a measurement at each occasion.

3.1 Comparing Bayesian and frequentist results in change point problem

To compare between Bayesian and frequentist results in change point problem, our basic strategy in this Section will be how to find non-Bayesian change point and then turn to find Bayesian change point estimates numerically using dedicated Bayesian software WinBUGS that implement Markov chain Monte Carlo (MCMC) methodology. BUGS is an acronym for 'Bayesian inference Using Gibbs Sampling'. Gibbs sampling is a specific MCMC method. The Gibbs sampler yields a Markov chain whose stationary distribution is the posterior distribution.

This will be illustrated using an example concerned with the prediction of log flow rate of water down an inclined channel (in g. /cm.sec.) given a series of the log of the height of the stagnant surface layer (in cm) for a sample of $N = 29$ different surfactants, taken from (Bacon, et al., 1971), who report that data during the investigation of the behavior of stagnant surface layer height in a controlled flow of water. In this case, one of our response measurements, flow rate of water down an inclined channel, is made at the level of the subject and the height of the stagnant surface layer is made at the event level. Muggeo (Muggeo, 2011) used such data in his package 'segmented' within the R Environment. These data is given in Table 1. Like R packages, WinBUGS is capable of handling missing observations. In Table 1, we present the missed observation as WinBUGS team writes 'NA'. NA is the one of the few non-numbers that we could include in the table without generating an error during write the code program for the both R packages and WinBUGS.

3.2 Stagnant band height data example, from non-Bayesian perspective

The graph in Fig. 1 and the estimated degrees of freedom both indicate a slight deviation from linearity (glmmBUGS package is used to estimate degrees of freedom; the glmmBUGS package is a bridging tool between generalized linear mixed models (GLMM's) in R and the WinBUGS language, see (Patrick, 2010), for more details). The graph suggests that there may be two regions as it was outlined in (1). Further, based on the Fig. 2 & **Error! Reference source not found.**, we deduce that the change point for the data occur at $c \approx 0$. The largest log-likelihood was achieved using a breakpoint of $c \approx 0$. It is useful to plot the log-likelihood as a function of *c* yielding what we call a profile log-likelihood, as in **Error! Reference source not found.**. From Bayesian approach, these data were analyzed by (Spiegelhalter, et al., 1996), who identify the change point for these data and they do not take the repeated measures in their consideration. Further, they consider the problem as a non-hierarchical (standard linear regression with a change point).

As a simple demonstration that the change point could be handled well using WinBUGS well, inference on the change points will be described in the following example. Denote by y_{ij} and x_{ij} the log flow rate of water down an inclined channel (in g. /cm.sec.) and the height of the stagnant surface layer (in cm) for rate i at time j (in this example $x_{ij} = x_j$). (Spiegelhalter, et al., 1996) do not take the repeated measures in their considerations and consider that there is only one change point, *c*. Now suppose we have a probability model for these data. Take the j^{th} measurement on the i^{th} height and following formulation by (Spiegelhalter, et al., 1996), the regression model is as follows:

$$\begin{cases} y_{ij} = m(x_j) \\ m(x_j) = \beta_1 + \beta_2 (x_{(j)})^- + \beta_3 (x_{(j)} - c)^+ \\ y_{ij} \sim N(\mu_{ij}, \sigma^2) \end{cases} \quad (3)$$

Equation (3) is the likelihood of our data; one drawback of this model is that convergence problems can occur in locating c when the data are sparse around the neighborhood of c_j. Khan (Khan, 2010) developed the bent-cable regression using MCMC which could be used and studied by the same method in this paper. In general, the statistical model for this framework can be written as follows:

$$\begin{cases} y_{ij} = m(x_{(j)}) + \varepsilon_{ij} \\ m(x_j) = \beta_1 + \beta_2 (x_{(j)} - c_j)_- + \beta_3 (x_{(j)} - c_j)_+ + b_j \\ y_{ij} \sim N(\mu_{ij}, \sigma^2) \\ b_j \sim N(0, \sigma_b^2) \\ \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \end{cases} \tag{4}$$

Here, the continuous predictor $x_{(j)}$ is the times-ordered of observations x_j and y_{ij} is the corresponding vector of response, $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$, such that the total number of observations is $N = \sum_{i=1}^{n} \sum_{j=1}^{m} n_{ij}$. The number n_i is the within-subject sample size; it is the number of observations on subject *i*. Without loss of generality, in this paper, we will consider the sum of n_i is equal to m. Therefore if m is equal to one, then we have univariate. Hence, we have a sample of *n* subjects with *i* indexing subjects and running from 1 up to *n*. The total number of observations is then $n \times m$. Also, we suppose that there exist m change points, c_j occurs through $x_{(1)} < x_{(2)} < ... < x_{(m)}$ such that the mean of the change is constant between two changes and different from a change to another. Let c_j be the change point at for the observation. Here $(x_{(j)} - c_j)_-$ is equal to $(x_{(j)} - c_j)$ if $x_{(j)} < c_j$ and $(x_{(j)} - c_j)_+$ is equal to $(x_{(j)} - c_j)$ if $x_{(j)} \geq c_j$. $\beta = (\beta_1, \beta_2, \beta_3)$ be the vector of fixed effects. β_2 and β_3 represent the overall slope before and after the change-points K_j and β_1 is the all over intercept. Let ω_j be the random effects for observation j, and ε_{ij} be independent errors. Hence, the parameters of the model (4) are β, $\{c_j\}_{j=1}^{m}$, σ_b^2, σ_b^2, σ_ε^2. Technically the $\{b_j\}_{j=1}^{m}$ are not parameters but instead are considered to be unobserved random variables for which we form 'predictions' instead of 'estimates'. β, b_j and ε_{ij} are all assumed independent. Hence, we have a sequence of independent observations $\{x_j, y_{ij}\}_{j=1}^{m}$, $i = 1, 2, ..., n$ satisfying model (4). In this example, it is assumed that y_{ij} distributed as normal (also called 'Gaussian') distribution with mean and variance given by

$$\begin{cases} E(y_{ij}) = \mu_{ij} = \beta_1 + \beta_2 (x_j - c_j)_- + \beta_3 (x_j - c_j)_+ \\ V(y_{ij}) = \tau \end{cases} \tag{5}$$

Here, (5) is the expected and variance of the change in curve fitting between the change-points.

$(y_{ij} | c_j)$ in (4) follows a mixed-effects model (or just mixed models), and are often appropriate for representing clustered, and therefore dependent, data – arising, for example, when data are collected hierarchically, when observations are taken on related individuals (such as siblings), or when data are gathered over time on the same individuals. Combining all n subject-specific regression models above into one model to get the common formulation of the linear mixed model for clustered data is the Laird-Ware (Laird, et al., 1982) model as follows:

$$Y = X\beta + Zb + \varepsilon \tag{6}$$

where X is a model matrix for the fixed effects, includes baseline covariates, possibly time-varying covariates, and the before change point and after change point vectors, β is a vector of fixed and unknown parameters, Z is a model matrix for the random effects, b is a vector of random effects, and $\boldsymbol{\varepsilon}$ is a vector of random errors. Linear mixed models may be expressed in different but equivalent forms. For example, the Laird-Ware model may also be expressed in terms of a particular cluster by including a subscript i, denoting the linear mixed model for the i^{th} cluster ($i = 1, 2, ..., n$).Alternatively but equivalently, in matrix form, (6) could be written as

$$\begin{cases} y_i = X_i\beta + Z_i b_i + \varepsilon_i \\ b_i \sim N_q(0,\Psi) \\ \varepsilon_i \sim N_n(0,\sigma^2 I_m) \end{cases} \tag{7}$$

where

- y_i is the $m \times 1$ observation vector for the i^{th} subject;
- x_i is the $m \times p$ model matrix for the fixed effects within subject *i*;
- β is the $p \times 1$ a vector of population-average regression coefficients called fixed effects;
- Z_i is the $m \times q$ model matrix for the random effects for observations between subject *i*;
- b_i is the $q \times 1$ is a vector of subject-specific regression coefficients for cluster *i*, the b_i describe how the evolution of the i^{th} cluster deviates from the average evolution in the population;
- ε_i is the $m \times 1$ vector of residual (errors) for observations in cluster *i*;
- Ψ is the $m \times q$ covariance matrix for the random effects;
- $\sigma^2 I_m$ is the $m \times m$ covariance matrix for the errors in cluster *i* with I_m is the order- *m* identity matrix.

By defining characteristic of a longitudinal study, which is that individuals are measured repeatedly through time, the linear mixed effects model in (6) or (7) provide an effective approach to longitudinal data analysis. There are a lot of longitudinal data in a variety of disciplines such as social science, medical science, economics, agriculture and industry. Longitudinal data require special statistical methods because the set of observations on the same individual tends to be inter-correlated. This correlation must be taken into account to draw valid scientific inference.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
x	-1.39	-1.08	-0.94	-0.8	-0.63	-0.25	-0.12	0.01	0.11	0.25	0.34	0.44	0.59	0.7	0.85	0.99	1.19
y	1.12	0.99	0.92	0.9	0.81	0.65	0.6	0.51	0.44	0.33	0.25	0.13	-0.01	-0.13	-0.3	-0.46	-0.65
	1.12	1.03	NA	NA	0.83	0.67	0.59	NA	0.43	0.3	0.24	NA	NA	-0.14	-0.33	-0.43	NA
	NA	NA	NA	NA	NA	NA	NA	NA	0.43	NA	NA	NA	NA	NA	NA	NA	NA

Table 1: Stagnant band height data; *x*: log (flow rate in g./cm.sec.), *y* : log (band height in cm.)
Source: Carlin et al. (1990) (Data contains between 1 and 3 observations per *x* "ragged array")

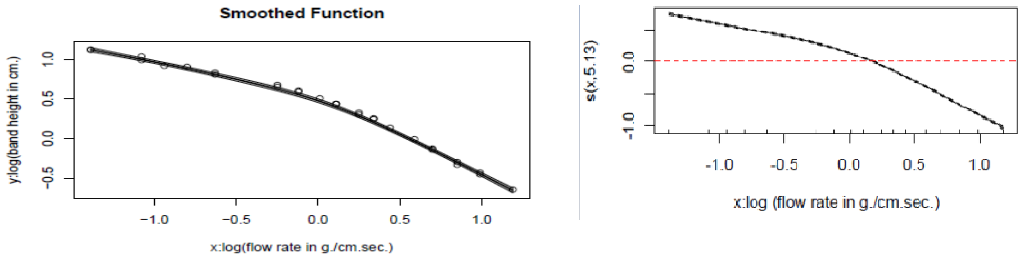

Fig. 2: Scatterplot, estimated degrees of freedom with a smoothed function (solid) estimated fit to the stagnant data

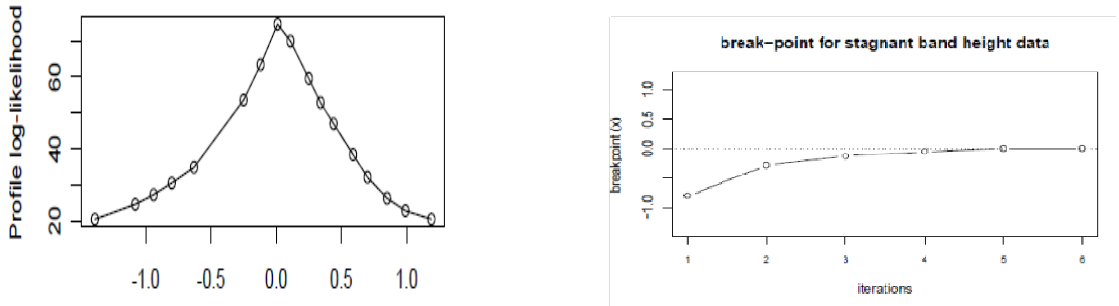

Fig. 3: break-point and profile log-likelihood for the breakpoint estimated fit to the stagnant data.

3.3 Stagnant band height data example, from Bayesian perspective

We begin by using the WinBUGS programming language to formulate the likelihood of our data (4), under our chosen model (5), along with prior distributions for the model parameters. To complete the Bayesian specification, suitable priors have to be assigned to all model parameters in the likelihood function. The following priors were used,

$$\begin{cases} \beta_1, \beta_2, \beta_3 \sim \mathrm{N}(0, 10^6) \\ \sigma_b^{-2}, \sigma_\varepsilon^{-2} \sim \mathrm{Gamma}(0.001, 0.001) \end{cases} \tag{8}$$

The second parameter of the normal distribution in (8) is the variance. In many applications a normal prior distribution centered at zero with standard error equal to 1000 is sufficiently improper (non-informative, objective or default) prior. If there are reasons to suspect, either using alternative estimation methods or prior knowledge, that the true parameter is in another region of the space, then the prior should be adjusted accordingly. The parameterization of the Gamma distribution in (8) with shape parameter a_0 and scale parameter a_1 is chosen so that its mean is $a_0 / a_1 = 1$ and its variance is $a_0 / a_1^2 = 1000$. Two other possible prior distributions for variance components, the first one is the uniform prior on $[0, L]$, where L is chosen very large, and the second is also the uniform distribution on $[-L, L]$ for the log of variance. The practitioner should try a few dispersed priors in a sensitivity analysis to the choice of priors.

We now describe the WinBUGS program that follows closely the description of the Bayesian parametric model in equations (4), (5) and (8). The entire program is provided in Appendix A.1. While this program was designed for the stagnant band height data, it can be used for other general mixed effect-effects model with change points with just minor adjustments. Many features of this program will be repeated in the next example in the next Section and changes will be described, as needed. Equation (4) is the likelihood part of the model (see details in Appendix A.1.). In WinBUGS, models are expressed in code through the distributions of the observations and parameters together with their independence structure. Equation (4) is the likelihood part of the model.

Based on the information we give to WinBUGS software (see Appendix A.1) set up the appropriate Gibbs sampler (or some alternative MCMC method) that yields Markov chains that we in turn will track. Convergence to the posterior distributions was assessed by using two different initial values of model parameters and visually inspecting several chains corresponding to the model parameters in Table 1. Convergence was attained in less than 10, 000 simulations, but we discarded the first 30,000 burn-in simulations in. For inference we used 20, 000 simulations and a thinning to every 25th draw and then, we monitored all parameters of the model. WinBUGS has a number of tools for reporting the posterior distribution. A simple summary (Table 2) can be generated showing posterior mean, median and standard deviation with a 95% posterior credible interval, for the quantities of primary interest in change point problem. Parameter names are related in an obvious way to the model in both equations (3) and (4). A fuller picture of the posterior distribution can be provided using the density option in the Sample Monitor Tool, within the WinBUGS software, which draws a kernel density estimate of the posterior distribution of any chosen parameter, as in **Error! Reference source not found.**.

Table 2: WinBUGS output for the for stagnant band height data (see Appendix A.1): posterior statistics for parameters of main interest in change point problem

parameter	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
beta[1]	0.618	0.03435	0.001323	0.5481	0.6187	0.6832	30001	40000
beta[2]	0.3064	0.03413	0.001268	0.2439	0.3052	0.3762	30001	40000
beta[3]	−1.058	0.05029	0.001885	−1.164	−1.057	−0.9644	30001	40000
average.c	−0.0023	0.01284	3.80E-04	−0.03028	−0.0019	0.02374	30001	40000

Table 2: WinBUGS output for the for stagnant band height data, see Appendix A. 1: posterior statistics for parameters of main interest in change point problem

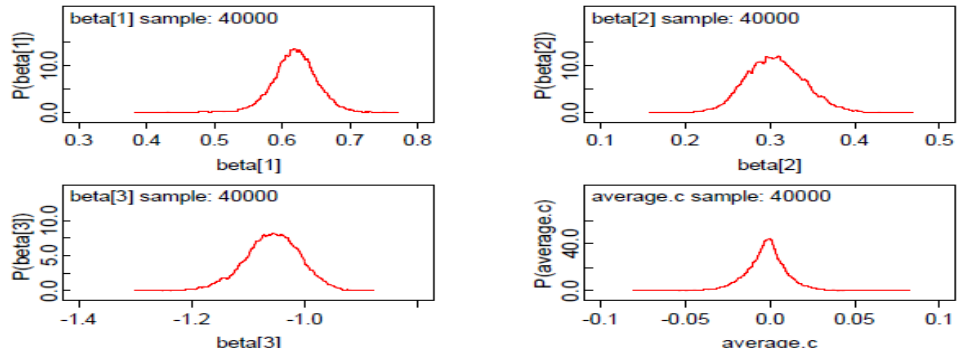

Fig. 3: WinBUGS output for the for stagnant band height data: posterior densities for parameters in Table 2

The autocorrelation function for the chain of each dimensions parameter (as shown in Fig. 3) can also indicate of the posterior distribution that is mixing slowly. Slow mixing is often associated with high posterior correlation between parameters. The plots indicate that all parameters are mixing well with autocorrelation vanishing before 100 lags in each case. The correlation tool (the correlation option in the inference menu within WinBUGS software) can produce scatterplots of every parameter against every other parameter to indicate correlation or a correlation coefficient can be estimated from the current output. Graphical representations of the posterior distribution can indicate problems with the performance of the MCMC algorithm. More sophisticated methods for convergence detection can be implemented in R environment, see (Plummer, et al., 2011). WinBUGS produces output that is formatted for direct use with such package and allows the behavior of the chain to be investigated using some popular functions and statistical tests.

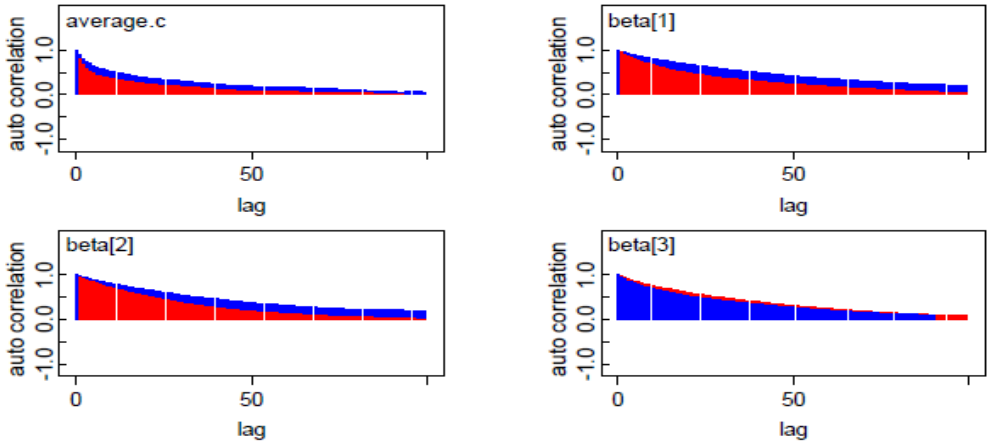

Fig. 4: WinBUGS output for the for stagnant band height data: autocorrelation functions of the chain

WinBUGS automatically implements the DIC (Spiegelhalter, et al., 2003) model comparison criterion. This is a portable information criterion quantity that trades off goodness-of-fit against a model complexity penalty. In hierarchical models, deciding the model complexity may be difficult and the method estimates the 'effective number of parameter', denoted here by P_D. $\overline{D}$ is the posterior mean of the deviance (–2 × log likelihood) and $\hat{D}$ is a plug-in estimate of the latter based on the posterior mean of the parameters. The DIC is computed as DIC = $\overline{D} + P_D = \hat{D} + 2$ P_D. Lower values of the criterion indicate better fitting models. Table 3 Table 3 records the values computed, in the format given by WinBUGS. For our purposes here, we will focus only on the DIC value. The method was designed to be easy to implement using a sample from the posterior distribution and the interested reader is directed to (Spiegelhalter, et al., 2003) for a lively discussion of its merits and its relation to the more usual Bayes factor.

	$\overline{D}$	$\hat{D}$	DIC	pD
y	–78.67	–89.13	–68.22	10.46
total	–78.67	–89.13	–68.22	10.46

Table 3: WinBUGS output for the for stagnant band height data: DIC with random effect $\overline{D}$ = post.mean of –2log L;

$\hat{D}$ = –2Log L at post.mean of stochastic nodes

There are many parameters of interest in code program in Appendix A.1, which may be useful in implementation of hierarchical Bayesian models in WinBUGS. For example, Variance Partition Coefficient (VPC) (see e.g., (Kostoulas, et al., 2009)) can be calculated. Sensitivity to priors for the change point should be studied well. There are various other options for displaying the posterior distributions for other parameters. For example, the compare…menu item, in WinBUGS Software, brings up the comparison Tool that draw caterpillar plot. When p.pred (see the code WinBUGS, in Appendix A.1) was monitored, it is found that the y[2,14] is the outlier, from Table 1, that is at $d = 14$, we have $\hat{y} = -0.14$.

The posterior distribution for the change points is plotted in Fig. 5. Finally;

Table *4* give estimates of other parameters of interest an Fig. 6 gives us our final most important of interest her, the change point problem.

parameter	Mean	SD	MC error	2.5%	Median	97.5%	Start	Sample
R2	0.9848	5.28E-03	1.69E-04	0.9723	0.9857	0.9925	30001	40000
VPC	0.7812	0.2209	0.005431	0.1761	0.8669	0.9877	30001	40000
y.pred[2,14]	-0.04563	0.08184	0.001824	-0.2163	-0.0421	0.1074	30001	40000

Table 4: WinBUGS output for the for stagnant band height data: Some of posterior statistics for the other parameters in code WinBUGS program in Appendix A.1

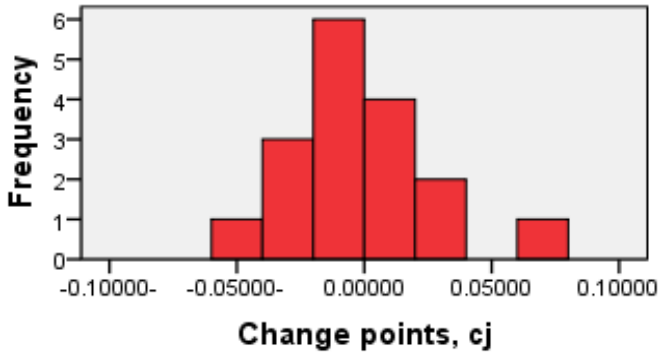

Fig. 5: Plot of the posterior mean values of the change points c_j

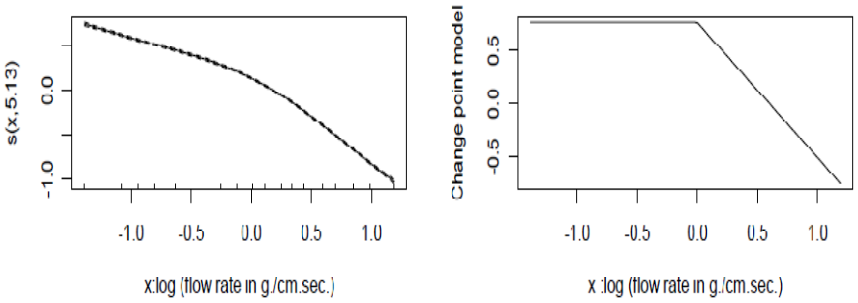

Fig. 6: Estimated change point model with breakpoint $c = -0.002$

4. THE CORONARY SINUS POTASSIUM DATA

We now consider the coronary sinus potassium concentration data measured on 36 dogs published by (Grizzle, et al., 1969) and (Wang, 1998). We repeat the above proposed analysis to show that the flexibility of R packages with WinBUGS for detecting the change points easily. The measurements on each dog were taken every 2 minutes from 1 to 13 minute (7 observations per dog) after occlusion (i.e., an artificially induced heart attack). The 36 were divided into 4 treatment groups, which will be treated as a categorical predictor in regression equation. The coronary sinus potassium concentration data are an example of longitudinal data.

4.1 The Coronary sinus potassium data example, from non-Bayesian perspective

That is, they are repeated measurements taken on the same experimental units over time. A plot of the coronary sinus potassium concentration versus time points measurements were made (see Fig. 6 Fig. 7, for the four groups and Fig. 7, for the first group only) by subject shows there is considerable variation between subjects in both the intercept and the slope of the linear trend. In such data we have the following variables:

- *y*: is a numeric vector denote to measurements of coronary sinus potassium concentrations;
- group: a vector of group index for the four groups of dogs;
- time (*x*): a numeric vector of time points measurements were made. All methods in packages in R environment have been used (see Fig. 7 and 8) , prove that 'The threshold change point occur at: 0.7395216' i.e., change point occurs at time equal to 0.73 approximately (see e.g., Sonderegger, 2011).

Figure 6: Estimated change point model with breakpoint $c_1 = 0.7395216$

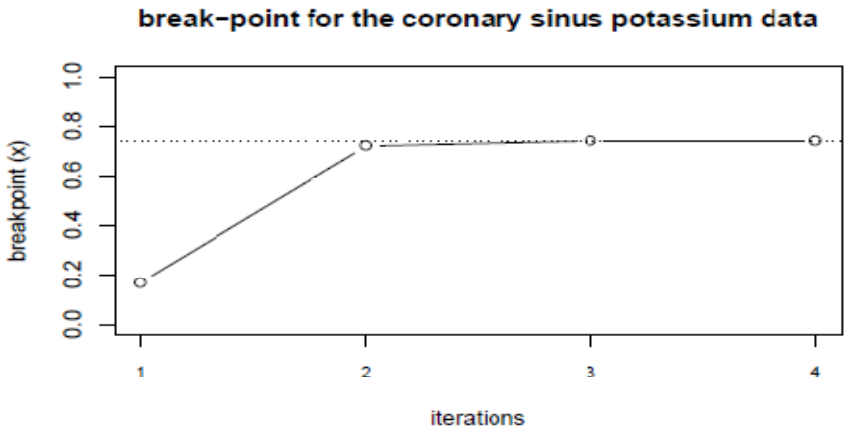

Fig. 7: Estimated change point model with breakpoint $c_1 = 0.7395216$

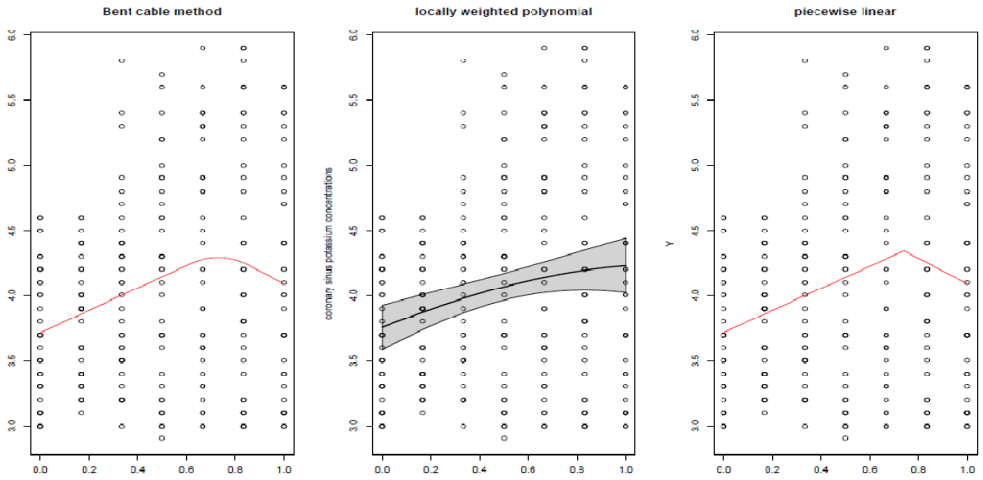

Fig. 8: Scatterplot, Bent cable method, locally weighted polynomial and piecewise linear methods for the coronary sinus potassium concentration data

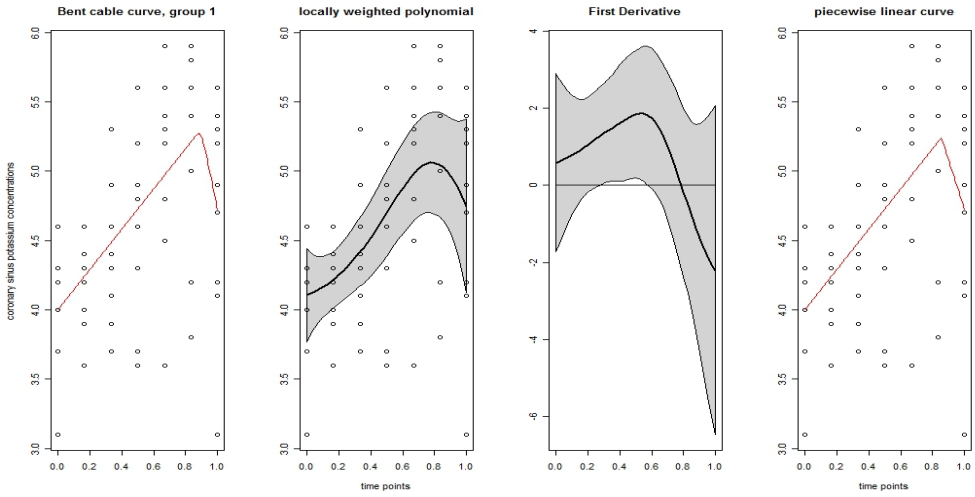

Fig. 9: Scatterplot, Bent cable method, locally weighted polynomial and piecewise linear methods for the coronary sinus potassium concentration data, for the first group

4,2 The Coronary sinus potassium data example, from Bayesian perspective

In this section, we show how to implement such a complex model in WinBUGS using model (4). After running the code software in the Appendix A. 2, it is straightforward to monitor the main nodes of interest, and repeat the analysis as the previous example. Let us, now show the importance of the hierarchical model in the change point problem. It is noted from

Table 5 that

- the variability in the intercept in the hierarchical model;
- how the residual variance (σ^2) is reduced when random effects;
- The effect of hierarchical model on the change point problem, i.e., when working with the non-hierarchical model and choose a suitable prior for the change point of interest, the results are the same from both hierarchical and non-Bayesian perspectives.

non-hierarchical model			hierarchical model		
β_1	4.755	(4.572, 4.935)	η_α	4.767	(4.388,5.133)
			σ^2_α	0.2605	(0.1471,0.422)
β_2	−1.413	(−3.966, −0.3934)	β_2	−1.073	(−2.407, −0.4882)
c1	0.6713	(0.3182, 0.9785)	c1	0.7296	(0.4566, 0.9707)
Δ_2	−1.072	(−1.289, −0.8513)	Δ_2	−1.068	(−1.577, −0.5522)
Δ_3	–0.4794	(–0.7089, –0.2493)	Δ_3	–0.4741	(–0.9831, 0.08042)
Δ_4	–0.6841	(–0.907, –0.4586)	Δ_4	–0.6831	(–1.179, –0.164)
σ^2	0.409	(0.3424, 0.4877)	σ^2	0.1853	(0.1533, 0.2238)

Table 5: Posterior mean (95% credible interval) for the non-hierarchical and hierarchical models fitted to the coronary sinus potassium concentration data

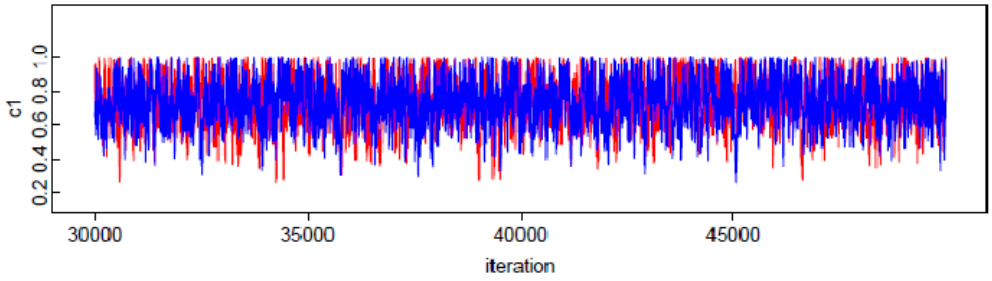

Fig. 10: Trace plots of the Markov chains for the change point, in case hierarchical model

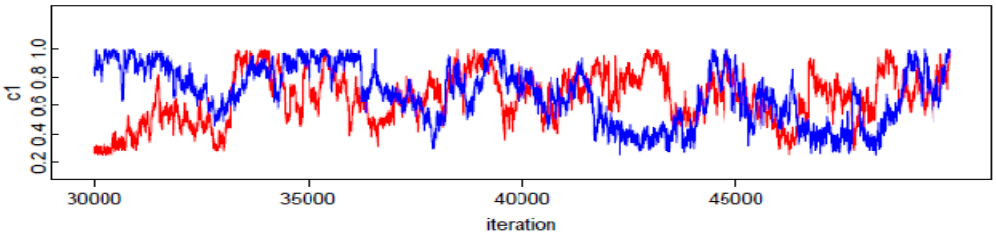

Fig. 11: Trace plots of the Markov chains for the change point, in case non-hierarchical model

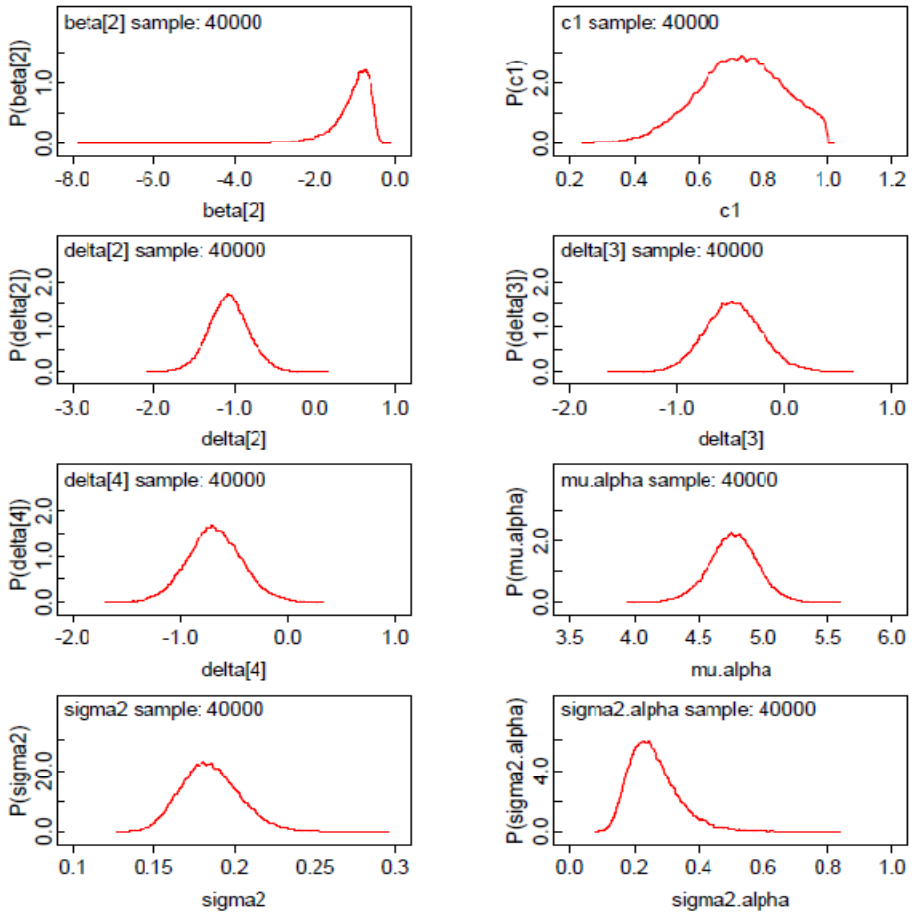

Fig. 12: WinBUGS output for the for stagnant band height data: posterior densities for parameters in hierarchical model, in

Table **5**

Fig. 10 shows plots of the two Markov chains for the change point parameter, c_1. Observe that trajectories are thoroughly interdigitating indicating that the chains have mixed fairly well. The fact that the two chains are traversing the same range of *y*-values provides evidence that the chains are now sampling from the posterior distribution of c_1. The fact that the individual chains are jumping around a lot indicates very little autocorrelation in the chains. This is good because it implies that the returned observations are nearly independent meaning that our effective sample size is close to the actual sample size, on the contrary for these, see

Fig. 11, in case of no-hierarchical model, which is not mixing well. Finally; Fig. 12 gives us a fuller picture of the posterior distribution of the kernel density estimate of the posterior distribution of the chosen parameter in

Table 5 for the non-hierarchical model.

5. CONCLUDING REMARKS

The following remarks are concluded from the above Sections:

(1) The best strategy for having success in fitting change point problem is to use the approaches, Bayesian and non-Bayesian;
(2) The flexibility of free software, WinBUGS with many packages of R to handle the change point problem clearly and easy. BUGS code looks just like R code but with important differences, which must be studied well (

(3) The easy way to handle WinBUGS with any other packages in R environment, which include about 3065 available contributed packages such that it is increasing weekly, in the all fields to deal with the statistics, in general (i.e. it easy to run WinBUGS from R).

(4) A landmark in the development of Bayesian analysis is the 2009 book by Ntzoufras (simply titled 'Bayesian Modeling Using WinBUGS'). As far as we know, it is the only book-length treatment on the topic to date. For instance, in their Book (Ch. 9) survey the idea of the basic concepts in Bayesian hierarchical models, hierarchical models, generalized linear mixed model formulation, with simple illustrative examples. While there is much to commend this work, for the most part it is relatively complex to implement. This may be an impediment to practitioners who wish to use such techniques in change point problem.

(5) The prior distribution for the change point using such technique in this paper needed further research.

REFERENCES

[1] **Bacon D. W. and Watts D. G.** Estimating the transition between two intersecting lines [Journal] // Biometrika. - 1971. - Vol. 58. - pp. 525-534.

[2] **Barry D. and Hartigan J. A.** A Bayesian Analysis for Change Point Problems. [Journal] // Journal of the American Statistical Association. - 1993. - 3 : Vol. 35. - pp. 309-319.

[3] **Braun J. V, Braun R. K. and Müller H. G.** Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation [Journal] // Biometrica. - 2000. - 2 : Vol. 87. - pp. 301-314.

[4] **Carlin B. P., Gelfand A. E. and Smith A. F. M.** Hierarchical Bayesian analysis of changepoint problem [Journal] // Applied Statistics. - 1992. - Vol. 41. - pp. 389-407.

[5] **Chiu G. S. and Lockhart R. L.** Bent-cable regression with auto-regressive noise [Journal] // Canadian Journal of Statistics. - 2010. - Vol. 38. - pp. 386-407.

[6] **Chiu G., Lockhart R. and Routledge R.** Bent-cable regression theory and applications [Journal] // Journal of the American Statistical Association. - 2006. - 474 : Vol. 101. - pp. 542-553.

[7] **Czaikowski M., Gill R. and Rempala G.** Model selection in logistic joinpoint regression with applications to analyzing cohort mortality patterns [Journal] // Statistics in Medicine. - 2008. - 9 : Vol. 27. - pp. 1508–1526.

[8] **Erdman C. and Emerson J. W.** A Package for performing a Bayesian analysis of change point problems [Online] // R Package version 2.2.0. - 2010. - http://cran.r-project.org/web/packages/bcp/.

[9] **Fotopoulos S. and Jandhyala V.** Maximum likelihood estimation of a change-point for exponentially distributed random variables [Journal] // Statistics & Probability Letters. - [s.l.] : 51, 2001. - Vol. 4. - pp. 423-429.

[10] **Grizzle J. E. and Allan D. M.** Analysis of dose and dose response curves [Journal] // Biometrics. - 1969. - pp. 357-381.

[11] **Hall B.** LaplacesDemon: An R Package for Bayesian Inference [Online] // R package version 11.10.03. - 2011. - http://cran.r-project.org/web/packages/LaplacesDemon/index.html.

[12] **Hawkins D. M.** A Note on Continuous and Discontinuous Segmented Regressions [Journal] // Technometrics. - 1980. - pp. 443-444.

[13] **Haybach G. and Kuechenhoff H.** Testing for a breakpoint in two-phase linear and logistic regression models [Online] // Discussion Paper 77 / ed. (SFB386) The Collaborative Research Center. - 1997. - http://www.stat.unimuenchen.de/sfb386/papers/dp1997.html.

[14] **Henderson R.** A problem with the likelihood ratio test for a change-point hazard rate model [Journal] // Biometrika. - 1990. - Vol. 77. - pp. 835-843.

[15] **Jensen U. and Lütkebohmert C.** A Cox-type regression model with change-points in the covariates [Journal] // Lifetime Data Analysis. - 2008. - 3 : Vol. 14. - pp. 267-285.

[16] **Khan S. A.** Flexible bent-cable models for mixture longitudinal data [Report] : Ph.D. Thesis / University of Waterloo. - Waterloo, Canada : [s.n.], 2010.

[17] **Killick R. and Eckley I. A.** Package 'changepoint' [Online] // R package version 0.4. - 2011. - http://cran.r-project.org/web/packages/changepoint/index.html.

[18] **Kosorok M. R. and Song R.** Inference under right censoring for transformation models with a change-point based on a covariate threshold [Journal] // The Annals of Statistics. - 2007. - Vol. 35. - pp. 957-989.

[19] **Kostoulas P. [et al.]** Bayesian estimation of variance partition coefficients adjusted for imperfect test sensitivity and specificity [Journal] // Preventive Veterinary Medicine. - 2009. - 3-4 : Vol. 89. - pp. 155-162.

[20] **Laird N. M. and Ware J. H.** Random-effects models for longitudinal data [Journal] // Biometrics. - 1982. - Vol. 38. - pp. 93-974.

[21] **Lerman P. M.** Fitting segmented regression models by grid search [Journal] // Applied Statistics. - 1980. - Vol. 29. - pp. 77-84.

[22] **Loader C. R.** Change point estimation using nonparametric regression [Journal] // The Annals of Statistics. - 1996. - 4 : Vol. 24. - pp. 1667-1678.

[23] **Lunn D. J. [et al.]** The BUGS project: Evolution, critique and future directions [Journal] // Statistics in Medicine. - 2009. - Vol. 28. - pp. 3049-3067.

[24] **Muggeo V. M. R.** Segmented: Segmented relationships in regression models [Online] // R package version 0.2-8. - 2011. - ttp://cran.r-project.org/web/packages/segmented/index.html.

[25] **Ntzoufras I.** Bayesian Modeling Using WinBUGS [Book]. - New York : John Wiley and Sons, 2009.

[26] **Patrick B.** Generalized Linear Mixed Models and Spatial Models with WinBUGS, BRugs, or OpenBUGS [Online] // R package version 1.9. - 2010. - http://cran.r-project.org/web/packages/glmmBUGS/index.html.

[27] **Picard F. [et al.]** A segmentation/clustering model for the analysis of array CGH Data [Journal] // Biometrics. - 2007. - Vol. 63. - pp. 758-766.

[28] **Picard F. [et al.]** Joint segmentation of multivariate Gaussian processes using mixed linear models [Journal] // Computational Statistics and Data Analysis. - 2011. - Vol. 55. - pp. 1160-1170.

[29] **Plummer M. [et al.]** Output analysis and diagnostics for MCMC [Online] // R Package version 0.14-4. - 2011. - http://cran.r-project.org/web/packages/coda/index.html.

[30] **R Development Core Team R** R: A language and environment for statistical computing [Online] // R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. - 2011. - http://www.R-project.org/. - ISBN 3-900051-07-0.

[31] **Seber G. A. F. and Wild C. J.** Nonlinear regression [Book] / ed. Hoboken. - New Jersey : Wiley, 2003.

[32] **Shaban S.** Change point problem and two-phase regression: An annotated bibliography [Journal] // International Statistical Review. - 1980. - Vol. 48. - pp. 83-93.

[33] **Sonderegger D.** SiZer: Significant Zero Crossings [Online] // R package version 0.1-4. - 2011. - http://cran.rproject..

[34] **Spiegelhalter D. [et al.]** BUGS 0.5: Examples Volume 1 [Report] / Institute of Public Health. - Cambridge, UK : MRC Biostatistics, 1996.

[35] **Spiegelhalter D. [et al.]** WinBUGS User Manual, Version 1.4, MRC Biostatistics Unit [Online] // Institute of Public Health and Department of Epidemiology and Public Health, Imperial College School of Medicine, UK. - Institute of Public Health and Department of Epidemiology and Public Health, 2003. - http://www.mrc-bsu.cam.ac.uk/bugs/.

[36] **Toms J. D. and Lesperance M. L.** Piecewise regression: a tool for identifying ecological thresholds [Journal] // Ecology. - 2003. - Vol. 84. - pp. 2034-2041.

[37] **Wang Y.** Mixed effects smoothing spline analysis of variance [Journal] // J. Royal Statistical Society. - 1998. - Series B : Vol. 60. - pp. 159-174.

[38] **Wu W. B., Woodroofe M. and Mentz G.** Isotonic regression: another look at the change point problem [Journal] // Biometrika. - 2001. - 3 : Vol. 88. - pp. 793-804.

[39] **Zeileis A. [et al.]** strucchange: Testing, Monitoring and Dating structural changes [Online] // R package version 1.4-4. - 2011. - http://cran.r-project.org/web/packages/strucchange/.

#----------

Appendix A. 1: (The example in Section 3)

#----------

```
  Model; #WinBUGS Code
          {
# The number of subjects, N = n*m = 29 given observations in the data Set in Table 1 and add 22
#missed observations could be estimated from the following Model = 51, is a constant in the
program
for(i in 1 : n) {
for(j in 1:m)  {
y[i,j] ~ dnorm(mu[i,j], tau)      # likelihood for observed data
#y[i,j] ~ ddexp(mu[i,j], tau)
#link and linear predictor
#mu[i] <- alpha + beta[J[i]] * (x[j] - c)+b[j]#Spiegelhalter et al. 1996 used this link
```



```
# It is preferred to use the following link:
mu[i,j] <- beta[1] + beta[2]*(x[j]- c[j])*min(x[j]-c[j],0)+beta[3]*(x[j] - c[j])*step(x[j] - c[j])+b[i]
p.pred[i,j] <- step(y[i,j] - y.pred[i,j]) # post. mean = p-value
res[i,j]<- (y[i,j]-mu[i,j])   # residual
sres[i,j]<- (y[i,j]-mu[i,j])/sigma   #standardised residual
p.sres[i,j] <- phi(res[i,j]) # p-value
# predictive p-value
y.pred[i,j] ~ dnorm(mu[i,j], tau) # re-predict each obs.
            }
b[i]~dnorm(0,tau.b)    #x means (Random effects)
            }
for(k in 1:3) {
beta[k] ~ dnorm(0.0,1.0E-6)
            }
for(j in 1:m) {
c[j]~dnorm(0,tau.c)
            }
tau ~ dgamma(0.0001, 0.0001)
sigma <- 1 / sqrt(tau)
#prior 1 for the change points c[j]
#K[j]~dnorm(0,tau.c)I(0,) # CAUCHY TRICK
#c[j]<-abs(K[j])/sqrt(tau.c)
#Prior 2 For the change-points c[j]
#c[j]~dunif(-1.3,1.1)
s2<-1/tau;      #Is the 'level 1' variance (i.e. variance of Normal likelihood)
s2.b<-1/tau.b;  #Is the 'level 2' variance (i.e. Random effects variance)
VPC<-s2/(s2+s2.b)#intra-cluster correlation (ICC) coefficient
R2 <- 1 - pow(sd(res[1:n,1:m])/sd(y[1:n,1:m]), 2 )
tau.c~dgamma(0.0001,0.0001)
# hyperpriors on random effects mean and variance
tau.b~dgamma(0.0001,0.0001);
average.c<-mean(c[])
typical.y <- beta[1] + beta[2]*min(mean(x[])-average.c,0) + beta[3]*step(mean(x[])-average.c)
            }
DATA (LIST)
list(n = 3,m = 17,y = structure(.Data
=c(1.12,0.99,0.92,0.9,0.81,0.65,0.6,0.51,0.44,0.33,0.25,0.13,–0.01, –0.13, –0.3, –0.46, –0.65,
1.12,1.03,NA,NA,0.83,0.67,0.59,NA,0.43,0.3,0.24,NA,NA,0.14,-0.33,-0.43,NA,
NA,NA,NA,NA,NA,NA,NA,NA,0.43,NA,NA,NA,NA,NA,NA,NA,NA),.Dim = c(3, 17)),
 x = c(–1.39, –1.08, –0.94, –0.8, –0.63, –0.25,
–0.12,0.01,0.11,0.25,0.34,0.44,0.59,0.7,0.85,0.99,1.19))
INITS
list(beta = c(–0.5, –0.5, 1.0), tau.c = 8, tau=2,tau.b = 1);
list(beta = c(1,2, 2), tau.c = 5,tau=2, tau = 3,tau.b=2).
```

#----------

Appendix A. 2 (The example in Section 4)

#----------

```
Model {#WinBUGS Code

for (i in 1:36) {
for (j in 1:7)   {
# Fixed effects regression model for Dog data
y[i,j] ~ dnorm(mu[i,j], tau)
mu[i,j] <-beta[1]+beta[2]*(x[j] - c1)*min(x[j]-c1,0)+delta[group[i]]+alpha[i]
            }
#alpha[i]<-beta[1]+U[i]
alpha[i]~dnorm(mu.alpha,tau.alpha)
            }
c1~dunif(0.25,1)
#c2~dunif(0,1)
```



```
delta[1] <- 0 # set coefficient for reference category to zero
delta[2] ~ dnorm(0, 0.00001)
delta[3] ~ dnorm(0, 0.00001)
delta[4] ~ dnorm(0, 0.00001)
tau.alpha~dgamma(0.001, 0.001)
tau ~ dgamma(0.001, 0.001)
# Priors
beta[1]<-0 # corner-point constraint
beta[2] ~ dnorm(0, 0.00001)  # regression coefficients
mu.alpha~ dnorm(0, 0.00001)
sigma2 <- 1/tau  # variance of residual error
sigma2.alpha <- 1/tau.alpha  # variance of intercept
            }
```

DATA (LIST)

y[,1]	y[,2]	y[,3]	y[,4]	y[,5]	y[,6]	y[,7]	group[]
4.00	4.00	4.10	3.60	3.60	3.80	3.10	1.00
4.20	4.30	3.70	3.70	4.80	5.00	5.20	1.00
4.30	4.20	4.30	4.30	4.50	5.80	5.40	1.00
4.20	4.40	4.60	4.90	5.30	5.60	4.90	1.00
4.60	4.40	5.30	5.60	5.90	5.90	5.30	1.00
3.10	3.60	4.90	5.20	5.30	4.20	4.10	1.00
3.70	3.90	3.90	4.80	5.20	5.40	4.20	1.00
4.30	4.20	4.40	5.20	5.60	5.40	4.70	1.00
4.60	4.60	4.40	4.60	5.40	5.90	5.60	1.00
3.40	3.40	3.50	3.10	3.10	3.70	3.30	2.00
3.00	3.20	3.00	3.00	3.10	3.20	3.10	2.00
3.00	3.10	3.20	3.00	3.30	3.00	3.00	2.00
3.10	3.20	3.20	3.20	3.30	3.10	3.10	2.00
3.80	3.90	4.00	2.90	3.50	3.50	3.40	2.00
3.00	3.60	3.20	3.10	3.00	3.00	3.00	2.00
3.30	3.30	3.30	3.40	3.60	3.10	3.10	2.00
4.20	4.00	4.20	4.10	4.20	4.00	4.00	2.00
4.10	4.20	4.30	4.30	4.20	4.00	4.20	2.00
4.50	4.40	4.30	4.50	5.30	4.40	4.40	2.00
3.20	3.30	3.80	3.80	4.40	4.20	3.70	3.00
3.30	3.40	3.40	3.70	3.70	3.60	3.70	3.00
3.10	3.30	3.20	3.10	3.20	3.10	3.10	3.00
3.60	3.40	3.50	4.60	4.90	5.20	4.40	3.00
4.50	4.50	5.40	5.70	4.90	4.00	4.00	3.00
3.70	4.00	4.40	4.20	4.60	4.80	5.40	3.00
3.50	3.90	5.80	5.40	4.90	5.30	5.60	3.00
3.90	4.00	4.10	5.00	5.40	4.40	3.90	3.00
3.10	3.50	3.50	3.20	3.00	3.00	3.20	4.00
3.30	3.20	3.60	3.70	3.70	4.20	4.40	4.00
3.50	3.90	4.70	4.30	3.90	3.40	3.50	4.00
3.40	3.40	3.50	3.30	3.40	3.20	3.40	4.00
3.70	3.80	4.20	4.30	3.60	3.80	3.70	4.00
4.00	4.60	4.80	4.90	5.40	5.60	4.80	4.00
4.20	3.90	4.50	4.70	3.90	3.80	3.70	4.00
4.10	4.10	3.70	4.00	4.10	4.60	4.70	4.00
3.50	3.60	3.60	4.20	4.80	4.90	5.00	4.00

```
END
list(x=c(.000000,.166667,.333333,.500000,.666667,.833333,1.000000))
INITS
list(alpha = c(1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1, 1),
beta = c(NA,3), c1 = 0.5802, delta = c(NA, -2, 4,4), mu.alpha = 2, tau = 2, tau.alpha = 0.5)
list(alpha = c(2,2,2,2,2, 2,2,2,2,2, 2,2,2,2,2, 2,2,2,2,2, 2,2,2,2,2, 2,2,2,2,2, 2), beta = c(
NA,3), c1 = 0.52, delta = c(NA, 2, -4,-4), mu.alpha = 2, tau = 2, tau.alpha
```

****

