Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/060301

A New Approach to Bayesian Change Point Detection Using Lifting Wavelet Transform

Arunendu Chatterjee[*]

Department of Mathematics, University of Wisconsin-River Falls, 410 S. 3rd Street, River Falls, WI 54022, USA

Received: 31 Jan. 2017, Revised: 15 Jun. 2017, Accepted: 21 Jun. 2017
Published online: 1 Nov. 2017

Abstract: In this paper, we use wavelets in a Bayesian context to identify changes in the pattern of data collected over time in the presence of noise and missing observations in the data. A Bayesian analysis based on the wavelet coefficients applying lifting is discussed to identify change points. Based on a simulation study, recommendations are made on the choice of lifting wavelet coefficients in the presence of noise and missing observations using an adaptive lifting technique. We apply our algorithm to a real data problem where change points are already known to illustrate our recommendations.

Keywords: Adaptive lifting, Bayesian method, change points, simulations, wavelets.

1 Introduction

Change point is the problem of identifying sudden change at a particular time point in the pattern of the data collected over time. Change point problem has many diverse applications not only in statistics but also in other disciplines such as hydrology, climatology etc. where change point problem occurs regularly. There are various aspects to the change point problem, namely detection of change point, estimation of the time at which the change occurred and finally, modeling the data before and after change. A substantial literature exists on models that combine detection, estimation and modeling using a statistical framework. In this paper, we discuss a Bayesian procedure to detect change points in the presence of missing or irregular data points. Modeling change points may become complicated in the presence of missing data. In this paper we discuss a detection method using second generation wavelet transforms.

In statistics literature, Discrete wavelet transform (DWT) has been used to detect change points but DWT can not be used to detect change points in the presence of missing data. We introduce second generation wavelet transform technique or lifting technique to detect change points. The use of lifting technique to detect change points is a new and timely procedure.

We introduce an algorithm based on lifting transform to detect change points. Our method is easy to implement and can be applied to any data size. Bayesian procedure is used to find the posterior distribution of the change point and the position of the change point can be determined by the mode of the skewed posterior distribution.

In Section 2, we provide a brief review of the existing methodologies for studying change point problems, and wavelet analysis is discussed in Section 3 in detail. We discuss lifting transform in Bayesian framework in Section 4 followed by computational procedure and some results relating to the application of lifting in change point detection in Section 5 and 6 respectively. Section 7 concludes the paper.

2 Overview of Change point methods

A change point is the time at which some feature of the distribution of a variable changes; the most common features usually considered are changes in the mean structure in the form of shifts in trends, or changes in the variance structure. Detection of change points is a complicated problem in practice as neither the occurrence nor the possible multiplicity of change points is known.

[*] Corresponding author e-mail: arunendu.chatterjee@uwrf.edu

The change point problem was originally addressed in Bayesian statistics by Smith (1975), followed by Carter and Blight (1981). Bayesian methods were applied considering single or multiple changes in conjunction with a known or an unknown number of change points. Gelfand et al. (1990) considered a known number of change points and discussed Bayesian analysis of a variety of normal data models, including regression and ANOVA, which allowed some unequal variances. Stephens (1994) carried out a Bayesian analysis of a multiple change point problem where the number of change points was assumed to be known, but the times of occurrence of the change points remained unknown. Examples of such approaches using a known number of change points include Carlin et al.(1992), Tanner (1993) and Rasmussen (2001). Other authors considered the problem in such a way that the data series contained at most one change point (i.e., the authors emphasized the presence of zero or one change point in the data series).

3 Wavelets and Detection of Change points

Wavelets are functions that satisfy certain mathematical properties and since the resulting wavelet transforms are localized in time and space, they can be used to detect sharp changes in discontinuous functions. The main purpose here is to estimate the number, locations and sizes of a function's abrupt changes. The general idea is to detect a change point by using a wavelet approach. Jumps are identified by "unusual" behavior in the wavelet coefficients at high resolution levels at the corresponding location. Use of wavelet in change point analysis was first introduced by Wang (1995).

Bayesian methods were introduced by Richwine (1996) followed by Ogden and Lynch (1998). A prior distribution was placed on a location of the change point and the posterior distribution of the change point was formulated on the basis of the estimated wavelet coefficients.

In describing the use of wavelets for detection of change points, we first present a brief introduction to wavelet transforms.

3.1 Overview of Wavelets

Wavelets are special functions consisting of dilation and translation indices. Larger values of dilation indices correspond to higher frequency components, and larger values of translation indices correspond to rightward shifts. In practice we use a discrete wavelet transform (DWT) to map a data vector $y = (y_1, ..., y_n)$, for $n = 2^J$, to a vector of wavelet coefficients $w = (w_1, ..., w_n)$ via an orthogonal matrix W. Choice of wavelet functions determines W. Since higher frequency components occur for larger values of dilation, detection of change points involves examination of these higher frequency coefficients in w.

Computation of the discrete wavelet transform is carried out using the popular Mallat's pyramid algorithm which consists of low-pass and high-pass filters through which, at each stage, the input values of the function are decimated. When the data are of size $n = 2^J$ the DWT requires J levels of decomposition. Denoting $n_j = n/2^j$, the output of a DWT is a set of 'detail' coefficients $d_j = (d_{j,1}, d_{j,2}, ..., d_{j,n_j})$ at levels $j = 1, 2, ..., J$ along with 'smooth' coefficients $s_J = (s_{J,1}, s_{J,2}, ..., s_{J,nJ})$, corresponding to the high-pass and low-pass filters, respectively. The detail wavelets coefficients contain the high-frequency content and are used in the change point detection procedure.

A classical approach towards detecting a change point is to choose a level j and examine the corresponding detail coefficients, d_j. Specifically, one can choose a threshold value based on the sample size and an estimated sample variance, compare the coefficients in d_j to this threshold value and decide whether any coefficients are 'large' enough to indicate the existence of a change point. Such a procedure is similar to wavelet-based outlier detection procedures in Wang (1995) and Bilen and Huzurbazar (2002).

All the applications discussed above suggest a very essential first step, that is, to transform the data into empirical wavelet coefficients through the discrete wavelet transform (DWT). DWT has its own limitations. Such a transformation is only possible under the following conditions:

(a) the time points (i.e. t_i's of equation (1)) should be equally spaced, and

(b) total number of observations n should be a power of 2.

Various methods have been proposed for adjusting irregularly spaced data. Cai and Brown (1998) proposed a method which takes into account the irregularity by using the correspondence $t_i = H^{-1}(i/n)$, where H is a strictly increasing function which needs to be estimated. The motivation was to use the approximation $t_{(i)}$ to be distributed as $E(t_{(i)}) = i/(n+1)$, and the observations $(i/(n+1), f_i)$ are considered as alternates instead of (t_i, f_i). Kovac and Silverman (2000) made a mapping of irregularly spaced data, f, to a regular grid, $\tilde{f}$, by linear interpolation of the original noisy values; $\tilde{f} = Rf$, where the matrix R describes the interpolation. To simultaneously handle the choice of wavelets, primary resolution

level and threshold for estimating the true function, Nason (2002) developed a fast cross-validation algorithm for irregular grids using the Kovac-Silverman procedure. Antoniadis and Fan (2001) formulated a penalized least squares problem in terms of unknown wavelet coefficients of $g(i/n)$. Assuming a regular grid and $n = 2^J$, and imposing certain conditions on the penalty function, they arrived at a unique solution. They also introduced a new universal threshold which produced estimators with smaller risk than that of universal threshold.

Another approach is to address the shortcoming of DWT. The most popular of these methods are lifting technique. We describe method of lifting in the next section.

3.2 Adaptive Lifting

Lifting transform is relatively new in statistics literature and there are no applications of lifting in change point detection. The algorithm was first introduced by Sweldens (1996) which facilitated for a wavelet construction of non-standard data, including irregular data on a grid. Jansen (2004) introduced a new lifting algorithm which was modified by Nunes et al. (2006). An imputation method based on lifting was later introduced by Heaton and Silverman (2008). Lifting essentially consists of three steps: splitting the data, predicting the removed data and updating the remaining data. In this section we describe the different lifting algorithms.

As a first step, the data is subsampled into two disjoint subsets: points representing the even positions in the grid, and points representing the odd positions in the grid (Swelden 1996). Jansen (2004) and Jansen et al. (2001) proposed generating just one wavelet coefficient at each step. Nunes et al. (2006) used a flexible lifting scheme that suggests removing one coefficient at a time in order to build up adaptive prediction steps and embedding them into the lifting algorithm.

The second step is to predict the function values corresponding to the odd positions by using polynomial regression of the corresponding function values to the even positions. The error in prediction (the difference between the true and predicted functional values of odd positions) is then quantified in a vector referred to as the set of wavelet (or detail) coefficients.

In the final step, the function values of the even positions are updated by using linear combinations of the current function values of the even positions and the vector of detail coefficients obtained at the previous step.

The split-predict-update steps can be re-iterated on the updated data set, and the initial data set is replaced by the remaining updated subsample (which reproduces the coarse scale features of the initial signal). The detail coefficients are accumulated throughout this process. This is similar to the DWT method, discussed above, which replaces the initial signal by a set of scaling and wavelet coefficients. The procedure is easily inverted by undoing the update stage, the prediction stage and then merging the subsamples.

Jansen (2004) and Jansen et al. (2001) introduced the concept of lifting just one coefficient in each step. The split step of the lifting algorithm suggests choosing a point which can be removed. The odd/even split, however, poses some problems in higher dimensions. Jansen (2004) and Jansen et al. (2001) proposed removing the points in an order guided by the configuration, namely, that the points belonging to denser areas be removed first. Again, each location is supposed to be associated with an interval values: the shorter the interval, higher the densely sampled area around the location. Once a point has been selected for removal, identify its set of neighbors. The next step is to predict scaling coefficients by using regression over the neighboring locations. The prediction error will be the detail coefficient corresponding to that location. In the update step, only the function-values of the neighboring points are updated by using a linear combination with the detail coefficients. At this stage, the lengths of the intervals associated with the neighboring points are also getting updated, accounting for the decreasing number of scaling points within the range of same interval. The procedure is then repeated on the updated data set, and with each repetition a new wavelet coefficient is added. Hence after say $(n-L)$ removals, we have L scaling coefficients and $(n-L)$ wavelet coefficients.

Nunes et al. (2006) proposed a modification to the LOCAAT (Lifting One Coefficient At a Time) lifting scheme which is called “Adaptive Lifting”. This algorithm is primarily based on lifting one coefficient at a time without imposing any restriction on a choice of prediction or a choice of neighborhood at the beginning of the procedure. The adaptive lifting scheme is called “adaptive” because at each step a choice of prediction or a choice of neighborhood is made. Two types of neighbor choices are used in adaptive lifting. We can choose symmetrical neighbors that is, same number of neighbors on the left and right of the removed points or we can choose closest neighbors to the removed points irrespective of which side they lie. This algorithm is computationally efficient than the LOCAAT lifting schemes suggested by Jansen et al (2001).

We know that the DWT scale is a discrete dyadic quantity. In “one coefficient at a time” lifting, this scale becomes more of a continuous type. In this adaptive lifting scheme the finest level contains half of the wavelet coefficients, the next coarser level represents a quarter and so on.

A few modifications of the resulting coefficients are needed to apply adaptive lifting in Bayesian change point detection procedure. In the framework of nonparametric regression, use of wavelets $\varepsilon_{j,k}$ follows independent normal distribution

with constant variance. This may not be true in case of lifting. In fact, in this case of lifting there will be a correlation between coefficients and different coefficients will have different variances. To overcome this, we normalize the resulting coefficients by multiplying $[diag(\tilde{W}\tilde{W}^T)]^{-1/2}$ for each step where $\tilde{W}$ is the the matrix associated with the transform.

4 Lifting Transformation in Bayesian framework

In this section, we derive the marginal posterior distribution of change point using lifting-based wavelet coefficients. From our discussion in previous section we know that the lifting coefficients are not independently distributed with constant variance as we expect in case of wavelet coefficients derived from DWT. If we assume initial independence of data then the resulting lifting coefficients are independent but do not have homogeneous variances. To make the variance constant, we have to normalize coefficients at each step by multiplying a transformation matrix. Even if the initial data has a covariance structure, normalizing lifting coefficients at each step by multiplying a transformation matrix will marginally affect detection of change points.

In the general change point problem, as mentioned by Ogden and Lynch (1998) $Y_1, ..., Y_n$ are the ordered observed data such that $Y_i \sim N(f(i/n), \sigma^2)$ for some function f defined on [0,1]. Hence, f has the form

$$f(u) = \begin{cases} \mu & \text{if } 0 \le u < \tau \\ \mu + \triangle & \text{if } \tau \le u < 1 \end{cases} \tag{1}$$

for some change point $\tau \in [0,1]$. Our objective is to estimate the parameters τ, μ, $\triangle$ and σ^2. Estimation of change-point τ is important here. Later we will choose suitable priors for μ, $\triangle$ and σ^2.

In previous section we defined lifting transformation. In the lifting-based procedure after $(n-L)$ removals, we have L scaling coefficients and $(n-L)$ wavelet coefficients. These lifting coefficients are independent but with a non constant variance. To overcome this, we normalize the resulting coefficients by multiplying $[diag(\tilde{W}\tilde{W}^T)]^{-1/2}$ for each step where $\tilde{W}$ is the matrix associated with the transform. After normalization we denote these transformed empirical lifting coefficients as w. Since the empirical lifting coefficients are independent with constant variance σ^2, the joint distribution can be written as:

$$p(w|\tau, \triangle, \sigma^2) = \prod_j \prod_k p(w_{j,k}|\tau, \triangle, \sigma^2)$$

where w represents the vector of coefficients. As before, the posterior distribution of the parameter is

$$p(\tau, \triangle, \sigma^2|w) \propto \prod_j \prod_k p(w_{j,k}|\tau, \triangle, \sigma^2) \cdot \pi(\tau, \triangle, \sigma^2)$$

for a joint prior distribution π on $\tau, \triangle$ and σ^2. The distribution of a single lifting-based wavelet coefficient was taken to be

$$w_{j,k}|\tau, \triangle, \sigma^2 \sim N(\triangle q_{j,k}(\tau), \sigma^2/n).$$

Ogden and Lynch (1998) defined the mean function $q_{j,k}(\tau)$ as

$$q_{j,k}(\tau) = 2^{j/2}/n \times \begin{cases} 2^{-j}k - [n\tau] & \text{if } 2^{-j}k \le \tau < 2^{-j}(k+1)/2 \\ -(n2^{-j}(k+1) - [n\tau]) & \text{if } 2^{-j}(k+1)/2 \le \tau < 2^{-j}(k+1) \end{cases} \tag{2}$$

for $\tau \in [0,1]$. The $q_{j,k}$ function is continuous and piecewise linear, shaped like an inverted hat; it is zero outside the support of the corresponding wavelet and goes to a peak at the midpoint in support of $\psi_{j,k}$. To get the values of $q_{j,k}$ for all j and k we can apply lifting scheme to the vector of 0 and 1 ($[n\tau] - 1$ zeros and $n - [n\tau] + 1$ ones).

Hence, the posterior density will be

$$p(\tau, \triangle, \sigma^2|w) \propto \sigma^{-(n-1)/2} \pi(\tau, \triangle, \sigma^2) \times exp(-\sum_j \sum_k (w_{j,k} - \triangle q_{j,k}(\tau))^2/(2\sigma^2)) \tag{3}$$

The analytical result may vary according to the choice of the prior distribution π on $\tau, \triangle$ and σ^2.

Each empirical wavelet coefficient contains information regarding the changes of a function in a localized region. For this reason, we use more localized or higher level coefficients to compute the posterior density. The posterior density is

thus computed using only the wavelet coefficients with dilation index $j \geq j_0$. Ogden and Lynch (1999) used Jeffreys' non-informative joint prior for $\tau, \triangle$ and σ such that

$$\pi(\tau, \triangle, \sigma^2) \propto 1/\sigma.$$

Thus the joint posterior density is given by

$$p(\tau, \triangle, \sigma^2|w) \propto \sigma^{-(n+1)/2} \times \\ exp(-\sum_{j\geq j_0}\sum_k (w_{j,k} - \triangle\, q_{j,k}(\tau))^2/(2\sigma^2))$$

Integrating out $\triangle$ and σ, the marginal posterior of change point τ will be (see Appendix for detailed derivation)

$$p(\tau|w) \propto C^{-1/2}(A - B^2/C)^{-(n+9)/4} \tag{4}$$

where $A = \sum_{j\geq j_0}\sum_k w_{j,k}^2, B = \sum_{j\geq j_0}\sum_k w_{j,k} q_{j,k}(\tau)$, and $C = \sum_{j\geq j_0}\sum_k q_{j,k}^2(\tau)$. Mode of above posterior probability distribution of τ will give us the change point.

5 Computation

Computation of marginal posterior probabilities with lifting-based procedures involves three steps: computing lifting coefficients from the original data, computing lifting coefficients from the mean function and finally, computing posterior probabilities using both lifting coefficients in previous two steps.

Step 1: As mentioned before, adaptive lifting does not follow assumptions of independence and the total number of observations need not to be a power of 2. Also in section 3.4 we described artificial leveling of detail coefficients which we use in this step of our computation. At this initial step we use adaptive lifting technique to denoise the data and to get the lifting coefficients simultaneously. Denoising of the signal involves the estimation of noise variances of artificial levels. We first use the transformation matrix to normalize the detail coefficients produced from the lifting transformation. The coefficients are divided into artificial levels and the coarsest level is used to estimate the noise variances of the coefficients. A particular package "adlift" in R can be modified to produce relevant results.

There are several options for modify the function as necessary. We can choose neighboring points at each step over which the regression is performed. The neighbors are chosen symmetrically on both sides around the removal point; otherwise closest neighbor is chosen. We can also select the resolution that conveys the number of scaling coefficients to be kept in the final representation of the initial signal.

The output of this function denotes the denoised lifting coefficients with artificial assignment of detail coefficients.

Step 2: The mean function $q_{j,k}(\tau)$ is a continuous and piecewise linear function, as mentioned in section 4. To get the values of $q_{j,k}(\tau)$ for all j and k, we can apply lifting procedure to the vector of 0 and 1 ($[n\tau] - 1$ zeros and $n - [n\tau] + 1$ ones). We do not denoise the resulting lifting coefficients but normalize them. Those normalized lifting coefficients are used to estimate change points.

Step 3: We calculate posterior probabilities using lifting coefficients obtained from both the original data and $q_{j,k}(\tau)$ for all j and k. Since, there is no hard and fast rule to select the number of coefficients, and it is not possible to prove theoretically optimum number of coefficients needed to detect change points under different conditions, one can use simulation technique to come up with some right numbers of coefficients.

5.1 Simulations

Statistical simulation studies provide powerful tools for the analysis of many mathematical models and real data problems when analytical solution is not possible. To recommend a suitable choice of lifting-based coefficients, it is essential to explore different cases of variable sample size, variable noise variances, variable jump sizes and variable time series structures. The results of the study will provide us a guideline to choose lifting-based coefficients.

As mentioned earlier, there is no specific rule of selecting a specific number of coefficients for detection of change points. Due to lifting scheme of resolution L, we have L smooth and $(n-L)$ detail coefficients. These $(n-L)$ detail coefficients are assigned with artificial levels into fine and coarse levels. To detect change points we use only the fine level detail coefficients.

In our simulation framework we will choose $n = 2^m$ where $m = 7$ to make a comparison between DWT and lifting. We simulate data for 4% of missing observations present in the dataset which will simulate a dataset with minimal

number of data points missing. For a strong noise in the dataset and a correlated variance structure with high percentage of missing observations, DWT does not perform well in terms of catching change point. We perform simulation under similar conditions prevailing under lifting coefficients to see how well it performs under such circumstances. We choose different sets of detail coefficients to find an optimum number of coefficients that can detect change points. We introduce number of data points missing from minimum number of data points missing (4%) to maximum number of data points missing (14%). We simulate data from AR(1) covariance structure and instead of modeling the AR(1) structure to detect change points we select lifting coefficients that can detect change points. We increase sample sizes to $m = 8$ *and* 9 to study the possible effects on lifting coefficients. In lifting based scheme we can select the resolution level L. We choose $L = 2$ for the simulation.

6 Results

We decided on 1500 simulations. The principle was that for a fixed set of lifting coefficients, the percentage of detection should stay constant even if we increase the number of simulations. Hence, we run 1500 simulations for different choices of sample sizes, resolution levels and different sets of lifting coefficients. Following are the discussions of different cases of simulations performed here.

(a) Comparison between DWT and lifting:

For 1500 simulations we first compare DWT-based coefficients and lifting-based coefficients. In this case our sample size is $n = 2^7$ and the resolution level is two. There are no missing observations. We consider $d7$, $d6$ and $d5$ wavelet coefficients to calculate posterior probabilities. We divide our grid into equal spaced 128 points. If five highest posterior probabilities match with any of five points on the grid (i.e. actual change point and two nearest values on both sides of the change point) we then consider it as a successful detection. For $n = 2^7$, $d7$, $d6$ and $d5$ are the finest level coefficients of DWT which should have information about jumps. For lifting-based scheme we have four artificially assigned levels of detail coefficients. Out of these 126 detail coefficients, we choose all coefficients of two finest levels and 22 of 31 coefficients from the third finest level. We simulate data with error terms for three different choices of variances. As the variance term goes up, noise in the data also goes up and detection of change points will be difficult. We carried out simulation for two jump sizes. We can expect that with the big jump size change point detection becomes relatively easier. Following is the comparison table between DWT and lifting:

Table 6.1: Comparison between DWT and Lifting

	Variance					
	0.5		1		1.5	
Size of jump	DWT	Lifting	DWT	Lifting	DWT	Lifting
1	54.1	93.7	36.3	89.8	28.9	87.4
3	98	93.7	91.9	91.1	84.9	90.7

From the above table we can conclude that lifting-based procedure coefficients are performing consistently better than DWT coefficients under the presence of different noise variances. In fact when the noise variance is high and jump size is small, DWT coefficients can detect change points in 28.9% cases compared to 97.4% cases for lifting. Hence we strongly recommend lifting-based procedure coefficients as the choice for detecting change points.

(b) Constant variance with no missing observations:

For a sample size of $n = 2^7$ we change noise variances from 0.5 to 1.5 with an increment of 0.5. We try three different sets of lifting coefficients to choose the best one for different noise levels in the data. Presented in Table 6.2 are the percentages of detection for different noise variances and jump sizes. The first set of coefficients are the same as described in (a). For the second set of coefficients we only consider all coefficients of the two finest levels. The last set of coefficients are chosen arbitrarily using part of the three finest level detail coefficients. For a constant noise variance we note the percentages of successful detection of change points for these three different choices of coefficients. Following are the percentages of successful change point detection under different scenario.

Table 6.2: Percentages of detection using lifting, $n = 2^7$

	Variance					
	0.5			1		
Jump size	73:126	96:126	77:124	73:126	96:126	77:124
1	93.7	68.5	42.2	89.8	59.5	37.8
3	93.7	64.9	43.1	91.1	63.8	41.3

Variance			
	1.5		
Jump size	73:126	96:126	77:124
1	87.4	55.6	35.3
3	90.7	62.7	42.1

Table 6.3: Percentages of detection using lifting, $n = 2^8$

Variance						
	0.5			1		
Jump size	148:254	195:254	158:252	148:254	195:254	158:252
1	91.3	62.8	40.2	86.8	57	41.1
3	92	57.9	43.1	81	55.3	37.2

Variance			
	1.5		
Jump size	148:254	195:254	158:252
1	85.2	54.2	39.2
3	78.3	51	37

Table 6.4: Percentages of detection using lifting, $n = 2^9$

Variance						
	0.5			1		
Jump size	298:510	392:510	320:508	298:510	392:510	320:508
1	94.5	70.2	41.9	89.1	68.1	41
3	96	69.2	45.1	90.4	67.7	44.8

Variance			
	1.5		
Jump size	298:510	392:510	320:508
1	87.6	66.9	40.6
3	90.3	65.8	44.2

Table 6.2 shows that the detection of change points seems to be difficult when the jump size is small. If we consider the first set of lifting coefficients at the finest level i.e. 73 to 126 which accounts for 45% of all coefficients and low noise level, the percentage of detection becomes 93.7% for both the jump sizes considered. The same set of coefficients perform better compared to other choices of coefficients in the case of high error variance. We may thus conclude that for both high and low noise-level in the data we can choose 45% of the finest level of detail coefficients which are able to detect change points for at least 87.4% cases.

If we increase the sample size to $n = 2^8$, 45% of the finest level detail coefficients can detect change points 85% for both the jump sizes (Table 6.3). This percentage, however, marginally decreases (i.e. 82%) with the increase in sample size from $n = 2^8$ to $n = 2^9$ (Table 6.4). It may be mentioned in this context that The only drawback of the lifting based procedure takes much more run time for each simulation higher than DWT. For example, DWT takes approximately 10 minutes to run 1500 simulations. With a processor speed of 2.8 GHz, it takes 5 hours of CPU time to run a simulation of 1500 lifting procedures when $n = 2^7$. The CPU time increases with the increase of sample size n. For $n = 2^8$ and $n = 2^9$, CPU time is approximately 8 hours for 1500 simulations.

From the above observations we can now conclude that about 45% of the finest level lifting coefficients may be considered as the best choice for detecting change points irrespective of any jump size and presence of noise in the data.

(c) Constant variances with 4% missing data:

We now consider the cases of missing data. We choose 4% of the simulated data missing at random. This means that initially, we intent to study the simulation results of lifting based coefficients when small number of data points are missing. These results may be compared with the situation of no missing observations. Thus, following the same procedure of lifting described in (b), we have presented in Tables 6.5, 6.6 and 6.7 the percentages of detection of change points with 4% missing data by different sets of lifting coefficients for $n = 2^7$, $n = 2^8$ and $n = 2^9$ respectively.

Table 6.5: Percentages of detection when 4% data missing, $n = 2^7$

Variance						
	0.5		1		1.5	
Jump size	68:121	91:121	68:121	91:121	68:121	91:121
1	84.9	65.7	80.6	53.2	79.6	48.3
3	83.1	61.5	81.8	60.8	80.2	61.2

Table 6.6: Percentages of detection when 4% missing, $n = 2^8$

Variance						
	0.5		1		1.5	
Jump size	138:244	185:244	138:244	185:244	138:244	185:244
1	87.2	59.7	87.5	58.1	86.8	57.7
3	89.5	62.6	88.2	62.7	89.2	60.1

Table 6.7: Percentages of detection when 4% missing, $n = 2^9$

Variance						
	0.5		1		1.5	
Jump size	277:489	371:489	277:489	371:489	277:489	371:489
1	90.6	68.3	87.5	68.1	87.2	67.2
3	92.5	70.0	90.1	67.6	88.6	64.2

Comparison of results presented in Table 6.2 and Table 6.5 shows that the detection of change points is not easy when 4% of the observations are missing compared to no missing observations. If we consider the first set of lifting coefficients at the finest level i.e. 68 to 121, accounting for 45% of all coefficients and low noise level, the change point detection is possible for 85% cases for both the jump sizes. Same set of coefficients perform better compared to other choice of coefficients when the error variance is high. Interestingly if we increase sample sizes, percentage of detection remains almost same. It is, however, true that the percentages of detection in the case of missing data are relatively lower compared to the situation of no missing observations, as observed in Tables 6.2, 6.3 and 6.4.

(d) AR(1) variance structure with no missing data:

Keeping in mind the problem of change point detection in time series data an attempt has been made in our simulation study, we generate data from a simple time series structure as autoregressive (1) or AR(1). Following are few cases where we recommend lifting-based procedure to detect change points under AR(1) structure.

Correlated variance structure is very common in real data problems. To study the detection procedure under correlated variance structure, we simulate three different AR coefficients 0.2, 0.6 and 0.85 denoted by α. There are no missing observations in the dataset. We use the same sets of lifting coefficients shown in Table 6.2 and choose the best one for different AR coefficients. Note that the higher the value of α means the stronger correlation structure in the data. Noise variance is constant at 1 for these sets of simulations. Simulation results are presented in Tables 6.8, 6.9 and 6.10.

Table 6.8: Percentages of detection with AR(1) structure, no missing data, $n = 2^7$

α						
	0.2		0.6		0.85	
Jump size	68:121	91:121	68:121	91:121	68:121	91:121
1	81.3	82.3	82.9	83.7	81.7	81.1
3	87.1	75.6	87.7	75.2	88.3	74.9

Table 6.9:Percentages of detection with AR(1) structure, no missing data, $n = 2^8$

α						
	0.2		0.6		0.85	
Jump size	148:254	195:254	148:254	195:254	148:254	195:254
1	86.2	82.6	87.0	83.6	83.4	82.2
3	92.4	77.0	89.6	76.4	90.1	78.8

Table 6.10:Percentages of detection with AR(1) structure, no missing data, $n = 2^9$

α						
	0.2		0.6		0.85	
Jump size	298:510	392:510	298:510	392:510	298:510	392:510
1	87.1	68.5	88.5	70.2	87.6	66.7
3	90.6	73.2	91.1	73.4	91.5	72.6

It is interesting to note from Tables 6.8, 6.9 and 6.10 that the percentages of detection remain the same across different values of α for the same set of coefficients. If we consider the first set of lifting coefficients at the finest level i.e. 68 to 121 which accounts for 45% of all coefficients and having low noise level, the detection of change points becomes 87% cases irrespective of any jump size. It is thus clear that irrespective of high or low values of AR(1) coefficients, we can choose 45% of the finest level of detail coefficients which successfully catch the change points for at least 82% cases. It is

also seen from the same tables percentage of detection remains same across different sample sizes. If we compare results between constant variance and a variance structure like AR(1), we find that irrespective of jump sizes and sample sizes, the percentage of detection drops 6% on an average due to the introduction of a variance structure across all the jump and sample sizes. Undoubtedly, our method is simplistic but at the same time effective in comparison to model the AR structure. Apparently, the effect of the presence of variance structure affects minimally to detect change point.

Now, with a processor speed of 2.8 GHz, it takes more than 5 hours of CPU time to run a simulation of 1500 lifting procedures when $n = 2^7$. For $n = 2^8$ and $n = 2^9$, CPU time is few hours more. For DWT, irrespective of sample sizes, it takes approximately 10 minutes to run 1500 simulations.

(e) AR(1) variance structure with 4% missing data:
Now, incorporating 4% missing observations in the dataset we simulate three different AR coefficients 0.2, 0.6 and 0.85 from AR(1) variance structure. We use three different sets of lifting coefficients as shown in Table 6.11 and choose the best one from different AR coefficients. Higher value of α means the stronger correlation structure in the data. Noise variance is constant at 1 for these sets of simulations. Results are presented in Tables 6.11, 6.12 and 6.13.

Table 6.11:Percentages of detection with AR(1), 4% missing data, $n = 2^7$

	α					
	0.2		0.6		0.85	
Jump size	68:121	91:121	68:121	91:121	68:121	91:121
1	84.2	72.7	84.2	74	82.5	73.6
3	89.5	79.5	88.3	79.4	88.9	78.4

Table 6.12: Percentages of detection with AR(1), 4% missing data, $n = 2^8$

	α					
	0.2		0.6		0.85	
Jump size	138:244	185:244	138:244	185:244	138:244	185:244
1	79.4	78.9	80.3	78.2	77.6	72.4
3	85.3	75.7	84.6	75.4	83.2	72.5

Table 6.13: Percentages of detection with AR(1), 4% missing data, $n = 2^9$

	α					
	0.2		0.6		0.85	
Jump size	277:489	371:489	277:489	371:489	277:489	371:489
1	83.4	76.3	83.6	75.8	83.1	72.4
3	87.1	79.5	86.7	78.1	86.5	79.2

Based on the results presented in Tables 6.11, 12 and 13, we find that the detection of change points become once again difficult if we consider 4% missing data and small jump size. However, the percentages of detection remain the same across different values of α for the same set of coefficients. Now, considering the first set of lifting coefficients at the finest level i.e. 68 to 121 which accounts for 45% of all coefficients and having low noise level, the detection of change points is possible for 84% cases for the jump sizes considered here. If we change sample size to $n = 2^8$ the detection percentage is still close to 80%. For $n = 2^9$ detection percentage increases marginally to 83%. If we compare our results with constant variance case, irrespective of any sample size, percentage of detection drops 6% on an average for considering AR(1) variance structure.

(f) AR(1)variance structure with 14% missing data:
If we enlarge the percentage of missing data (14%) and simulate three different AR coefficients 0.2, 0.6 and 0.85 from AR(1) variance structure, we find that the detection of change points do not provide any fruitful result in the case of small jump size.

Table 6.14:Percentages of detection, AR(1) structure, 14% missing data, $n = 2^7$

	α					
	0.2		0.6		0.85	
Jump size	55:108	78:108	55:108	78:108	55:108	78:108
1	29.6	58.9	28.8	58.4	31.1	56.9
3	34	61.4	34.7	61.1	34	57.7

Table 6.15:Percentages of detection, AR(1) structure, 14% missing data, $n = 2^8$

	α					
	0.2		0.6		0.85	
Jump size	124:218	159:218	124:218	159:218	124:218	159:218
1	24.5	46.7	24.8	45.9	23.8	44.7
3	25.8	47.2	25.3	46.1	25.1	45.2

Table 6.16:Percentages of detection, AR(1) structure, 14% missing data, $n = 2^9$

	α					
	0.2		0.6		0.85	
Jump size	249:440	321:440	249:440	321:440		
1	30.1	49.7	29.7	49.3	29.1	48.8
3	33.8	51.4	32.1	50.8	28.6	49.1

For $n = 2^7$, considering the first set of lifting coefficients at the finest level i.e. 55 to 108 which accounts for 45% of all coefficients and having low noise level, the detection of change points is possible only for 32% cases irrespective of any jump size. The percentages of detection is found to be moderately high if we take the second set of lifting coefficients i.e. 78 to 108 for both the jump sizes. If we increase sample sizes from $n = 2^7$ to $n = 2^8$ and $n = 2^9$ no striking difference of the results from the earlier one is noticed. Thus, percentage of detection remains more or less same across sample size. (Tables 6.14, 6.15 and 6.16).

From the above observations Hence, we can draw conclusion that 45% of all coefficients can detect change points irrespective of any jump size and error variance. In the presence of missing data those set of coefficients work well if a huge number of data points are not missing. If a considerable number of data points are missing, 32% of detail coefficients work better than 45% of all detail coefficients.

We may now turn to provide a real data example to substantiate our observations based on simulation results.

7 St. Lawrence Streamflow Data

Our case study, for purposes of comparison, uses annual streamflow data from the St. Lawrence River at Ogdensbourg, New York from the years 1860 to 1950. A description of the data is given in Rasmussen (2001). Here we note that using lifting-based coefficients, we found the mode of the posterior distribution for the time of the change point to be 1891. For this example, the original data had 90 observations. We do not need to augment it to 2^J to apply wavelet transform. Lifting transform can perform with any number of data points. Using 45% of fine detail coefficients we get our change point as 1891. Plot of the posterior pdf (Figure 1) shows multiple modes. Lifting-based coefficients are detecting the correct change point. It is not unusual to have a posterior pdf with multiple modes using lifting-based coefficients. There are two possible reasons for the multiple modes in the plot. Our procedure is useful for finding single change point in the data set but if there exists multiple change points, we can not be able to find them. Secondly, lifting based coefficients are based on polynomial regression which does not take into account the time dependence of the data which may contribute to the calculation of lifting coefficients and hence to the posterior pdf.

8 Conclusion

In this paper, we have investigated the choice of lifting coefficients in the context of detecting change points. We find that the detection of change points depends on the choice of lifting coefficients. We simulate time series data with AR(1) structure in the presence of missing observations and also with no missing observation. Lifting-based coefficients work when number of missing observations are not large. We presented some results based on sample size $n = 2^7$. Further increase of sample size to $n = 2^8$ or $n = 2^9$ we find that the results are not different from the smaller sample size. For a large number of missing observations with moderately high noise in the data, lifting performs moderately well. Overall, lifting coefficients perform very well in terms of finding change points for noisy data and data with smaller number of missing observations.

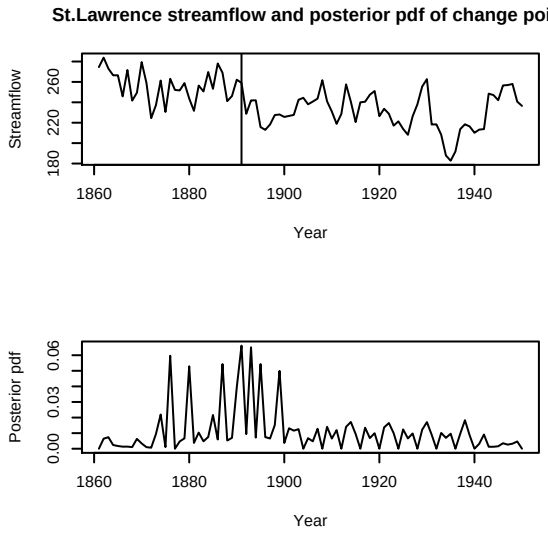

Fig. 1: Plot of the St. Lawrence streamflow data (top) with change point at year 1891 and posterior pdf (bottom)

We confirm our findings by applying them to a real dataset which had previously known change points. Lifting transformation and associated detail coefficients have never been used to detect change points in Bayesian settings or otherwise. We propose Bayesian change point detection method using lifting wavelet coefficients. We also come up with estimated percentages of lifting coefficients which can be used to detect change points in real dataset through simulations. Since change point problems occur regularly in other disciplines such as Ecology, Glaciology, Hydrology etc, it is suggested that our proposed algorithm should be used to get useful results for other fields of research. In the future we would like to apply this algorithm concerning with real data change point problems for other disciplines.

In our simulation results, we are concerned with one resolution level. However, for lifting transformation the resolution level can be changed for the detection of relatively undocumented small changes. In our simulation study, we keep the resolution level at two which is the default resolution level. For relatively small jump size and presence of noise in the data, change in resolution level may result a better percentage of detection of change points. This, however, is not true for the case of large number of missing observations with the presence of moderate noise in the dataset. In this case, the resolution level needs to be increased to increase the percentages of detection.

Change point detection is a common problem in time series data. In our case study we encounter time series data with relatively high noise. In such a situation, it is appropriate to choose AR(1) structure of simulation to examine the effectiveness of our algorithm and find that it is quite useful. We also observe that the percentage of detections remain unchanged even if we increase the correlation in variance structure. This percentage changes as and when the size of the jump or the noise in the dataset changes.

Another important observation regarding the lifting technique is that it depends on polynomial regression to get the detail wavelet coefficients. Polynomial regression may not be an appropriate technique when the real data is very noisy. A correlated data structure may influence negatively to predict the detail coefficients on the basis of polynomial regression. We may have a better time series predictive approach towards lifting technique.

9 Appendix: Calculation of the marginal posterior distribution of τ

From section 4.2 the joint posterior density is given by

$$p(\tau, \triangle, \sigma^2|w) \propto \sigma^{-(n+1)/2} \times exp(-\sum_{j\geq j_0}\sum_k (w_{j,k} - \triangle\, q_{j,k}(\tau))^2/(2\sigma^2))$$

$$= \sigma^{-(n+1)/2} \times exp((A - 2\triangle B + \triangle^2 C)/(2\sigma^2))$$

where $A = \sum_{j\geq j_0}\sum_k w_{j,k}^2, B = \sum_{j\geq j_0}\sum_k w_{j,k} q_{j,k}(\tau)$, and $C = \sum_{j\geq j_0}\sum_k q_{j,k}^2(\tau)$.

Ogden and Lynch (1999) used the same joint posterior density to derive the marginal posterior distribution for τ. But in the final expression, after integrating out $\triangle$ and σ, they came up with a different number in the exponent. We redid the integration and our exponent is different from what they came up with. But using our final expression we get sensible answer for all the change point problems. Following is the detailed derivation.

$$p(\tau|w) \propto \int_0^\infty \int_{-\infty}^\infty \sigma^{-(n+1)/2} exp(-C/(2\sigma^2)(\triangle^2 - 2\triangle B/C + A/C))\, d\triangle\, d\sigma$$

$$= \int_0^\infty \sigma^{-(n+1)/2} \int_{-\infty}^\infty exp(-C/(2\sigma^2)[(\triangle - B/C)^2 - B^2/C^2 + A/C]\, d\triangle\, d\sigma$$

$$= \int_0^\infty \sigma^{-(n+1)/2} \int_{-\infty}^\infty exp(-C/(2\sigma^2)(\triangle - B/C)^2 exp((-1/(2\sigma^2)(A-B^2)/C)\, d\triangle\, d\sigma$$

$$= \int_0^\infty \sigma^{-(n+1)/2} exp((-1/(2\sigma^2)(A-B^2)/C) \int_{-\infty}^\infty exp(-C/(2\sigma^2)(\triangle - B/C)^2\, d\triangle\, d\sigma$$

$$= \int_0^\infty \sigma^{-(n+1)/2} exp((-1/(2\sigma^2)(A-B^2)/C)(\frac{\sigma}{\sqrt{C}})(\sqrt{(2\pi)}) \int_{-\infty}^\infty (\frac{1}{\sqrt{(2\pi)}(\frac{\sigma}{\sqrt{C}})}) exp(-C/(2\sigma^2)(\triangle - B/C)^2\, d\triangle\, d\sigma$$

Using property of the normal distribution the inside integral is 1 and choosing $k = \frac{1}{2}(A - B^2)/C$ we can write

$$p(\tau|w) \propto 1/\sqrt{C} \int_0^\infty \sigma^{-(n-1)/2} exp((-k/\sigma^2))d\sigma$$

Substituting $m = k/\sigma^2$ and calculating the Jacobian we get

$$p(\tau|w) \propto 1/\sqrt{C} \int_0^\infty (k/m)^{(n-1)/4} exp(-m) m^{3/2} (dm/-2k^{5/2})$$

$$\propto 1/\sqrt{C} k^{-(n-1)/4-5/2} \int_0^\infty m^{(n+1)/4+3/2} exp(-m) dm$$

The integral is a proper Gamma integral and hence a constant. So, our end result after replacing k

$$p(\tau|w) \propto \frac{1}{\sqrt{C}}((A-B^2)/C)^{-(n+9)/4}$$

References

[1] Antoniadis, A., and Fan, J. (2001), "Regularization of wavelets approximations" (with discussion), *J. Amer. Statist. Assoc.*, 96, 939-967.

[2] Bilen, C. and Huzurbazar, S.(2002), "Wavelet-based detection of Outliers in Time Series," *Journal Computational and Graphical Statistics*, 11(2), 311-327.

[3] Cai, T. Tony, and Brown, D. L. (1998), "Wavelet shrinkage for nonequispaced samples", *The Annals of Statistics*, 26, 1783-1799.

[4] Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian analysis of changepoint problems", *Appl. Stat.*, 41, 389-405.

[5] Carter, R. L., and Blight, B. J. N. (1981), "A Bayesian Change Point Problem with an Application to the Prediction and Detection of Ovulation in Women", *Biometrics*, 37, 743-751.

[6] Fearnhead, P. (2005), "Exact Bayesian curve fitting and signal segmentation",*IEEE Transactions on Signal Processing*, 53, 21602166.

[7] Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990), "Illustration of Bayesian inference in normal data models using Gibbs sampling", *Journal of the American Statistical Association*, 85, 972-85.

[8] Heaton, T. J., and Silverman, B. W. (2008), "A wavelet- or lifting-scheme-based imputation method," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 70, 567-587.

[9] Jansen, M., Nason, G. P., and Silverman, B. W. (2001), "Scattered data smoothing by empirical Bayesian shrinkage of second generation wavelet coefficients", *Proc. SPIE*, 4478, 87-97.

[10] Jansen, Mark J. (2004), "Semiparametric Bayesian inference of long-memory stochastic volatility models", *Journal of Time Series Analysis*, 25, 895-922.

[11] Kovac, A. and Silverman, B.W. (2000), "Extending the Scope of Wavelet Regression Methods by Coefficient-dependent Thresholding", *Journal of the American Statistical Association*, 95, 172-183.

[12] Nason, G.P. (2002), "Choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage", *Statistics and Computing*, 12, 219-227.

[13] Nunes, M., Knight, M., and Nason, G. P. (2006), "Adaptive lifting for nonparametric regression", *Statist. Comput.* , 16, 143-159.

[14] Ogden, R.Todd, and Lynch, James D. (1998), Bayesian analysis of Change-point models, *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidokovic eds., Springer-Verlag, New York.

[15] Rasmussen, P. (2001), "Bayesian estimation of change points using the general linear model", *Water Resour. Res.*, 37, 2723-2731.

[16] Richwine, J. E. (1996), "Bayesian estimation of change-points using Haar wavelets," Master's Thesis at the University of South Carolina.

[17] Smith, A.F.M., (1975), "A Bayesian approach to inference about change-point in sequence of random variables," *Biometrika*, 62, 407416.
[18] Stephens, D. A. (1994), "Bayesian retrospective multiple-changepoint identification", *Appl. Stat.*, 43, 159 178.
[19] Sweldens, W. (1996), "The lifting scheme: a custom-design construction of biorthogonal wavelets", *Applied and Computational Harmonic Analysis*, 3, No. 2, 186-200.
[20] Tanner, M. (1993), *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions*, Springer Verlag, New York.
[21] Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385-397.

Arunendu Chatterjee, PhD, is an associate professor at the department of Mathematics, University of Wisconsin River Falls. He is also a member of American Statistical Association (ASA) and International Society for Bayesian Analysis (ISBA). Dr. Chatterjee's work has been published in different statistical journals including Statistical Methodology and Advances and Applications in Statistics. He is an author/reviewer of Journal of Statistics Applications and Probability Letters and a reviewer of American Journal of Mathematical and Management Sciences.