Change-point analysis as a tool to detect abrupt climate variations

BY CLAUDIE BEAULIEU1,*, JIE CHEN2 AND JORGE L. SARMIENTO1

1Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ 08540, USA
2Department of Mathematics and Statistics, University of Missouri–Kansas City, Kansas City, MO 64110, USA

Recently, there have been an increasing number of studies using change-point methods to detect artificial or natural discontinuities and regime shifts in climate. However, a major drawback with most of the currently used change-point methods is the lack of flexibility (able to detect one specific type of shift under the assumption that the residuals are independent). As temporal variations in climate are complex, it may be difficult to identify change points with very simple models. Moreover, climate time series are known to exhibit autocorrelation, which corresponds to a model misspecification if not taken into account and can lead to the detection of non-existent shifts. In this study, we extend a method known as the informational approach for change-point detection to take into account the presence of autocorrelation in the model. The usefulness and flexibility of this approach are demonstrated through applications. Furthermore, it is highly desirable to develop techniques that can detect shifts soon after they occur for climate monitoring. To address this, we also carried out a simulation study in order to investigate the number of years after which an abrupt shift is detectable. We use two decision rules in order to decide whether a shift is detected or not, which represents a trade-off between increasing our chances of detecting a shift and reducing the risk of detecting a shift while in reality there is none. We show that, as of now, we have good chances to detect an abrupt shift with a magnitude that is larger than that of the standard deviation in the series of observations. For shifts with a very large magnitude (three times the standard deviation), our simulation study shows that after only 4 years the probabilities of shift detection reach nearly 100 per cent. This reveals that the approach has potential for climate monitoring.

Keywords: change-point detection; autocorrelation; regime shift; abrupt climate change

1. Introduction

It is widely recognized that long-term changes in global temperature or other climatic variables exhibit abrupt shifts and nonlinearities in their behaviour. Several types of abrupt shifts can be encountered. Regime shifts, such as that

*Author for correspondence (beaulieu@princeton.edu).

One contribution of 13 to a Theme Issue ‘Climate predictions: the influence of nonlinearity and randomness’.
Change-point detection techniques have been developed to detect abrupt shifts in the parameters of a distribution or in the coefficients of a regression model. These techniques make it possible to detect the timing of the shift under the model containing a shift and to determine whether there is a change or not using a decision rule. The usual problem is that it is difficult to obtain the distribution of the statistic under the null hypothesis of no change. Different approaches can be applied to approximate this distribution, such as the Bonferroni inequality [20], the asymptotic theory [21] and Monte Carlo methods [22]. These approaches are not presented in detail here, as this is outside the scope of this paper. For details about the development of change-point techniques, the reader can refer to...
general literature reviews of change-point detection techniques presented by Bai & Perron [23] and Chen & Gupta [24]. In this section, we present the different models that have been used for climate-related applications of change-point detection. We refer to a few studies that have looked at climate variations using these models, but we do not intend to provide an exhaustive list owing to the scope of this study. The list of notation used in this paper is presented in Table 1.

Table 1. List of notation.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>trend of the linear regression model</td>
</tr>
<tr>
<td>$\hat{\beta}$</td>
<td>estimate of the trend of the linear regression model</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>trend before the shift</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>estimate of the trend before the shift</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>trend after the shift</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>estimate of the trend after the shift</td>
</tr>
<tr>
<td>$\Delta$</td>
<td>magnitude of the shift</td>
</tr>
<tr>
<td>$\epsilon_t$</td>
<td>random errors of the model at time $t$</td>
</tr>
<tr>
<td>$\eta$</td>
<td>quadratic trend of the regression model</td>
</tr>
<tr>
<td>$\eta_1$</td>
<td>quadratic trend of the regression model before the shift</td>
</tr>
<tr>
<td>$\eta_2$</td>
<td>quadratic trend of the regression model after the shift</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>intercept of the linear regression model</td>
</tr>
<tr>
<td>$\hat{\lambda}$</td>
<td>estimate of the intercept</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>intercept before the shift</td>
</tr>
<tr>
<td>$\hat{\lambda}_1$</td>
<td>estimate of the intercept before the shift</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>intercept after the shift</td>
</tr>
<tr>
<td>$\hat{\lambda}_2$</td>
<td>estimate of the intercept after the shift</td>
</tr>
<tr>
<td>$\mu$</td>
<td>overall mean</td>
</tr>
<tr>
<td>$\hat{\mu}$</td>
<td>estimate of the mean</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>mean before the shift</td>
</tr>
<tr>
<td>$\hat{\mu}_1$</td>
<td>estimate of the mean before the shift</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>mean after the shift</td>
</tr>
<tr>
<td>$\hat{\mu}_2$</td>
<td>estimate of the mean after the shift</td>
</tr>
<tr>
<td>$\rho_i$</td>
<td>lag-$i$ autocorrelation coefficient</td>
</tr>
<tr>
<td>$\hat{\rho}_i$</td>
<td>estimate of the lag-$i$ autocorrelation coefficient</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>overall variance</td>
</tr>
<tr>
<td>$\sigma^2_1$</td>
<td>variance before the shift</td>
</tr>
<tr>
<td>$\sigma^2_2$</td>
<td>variance after the shift</td>
</tr>
<tr>
<td>$k_j$</td>
<td>number of parameters to estimate in model $j$</td>
</tr>
<tr>
<td>$L(\hat{\Theta}_j)$</td>
<td>maximum likelihood of model $j$</td>
</tr>
<tr>
<td>$m$</td>
<td>autocorrelation order</td>
</tr>
<tr>
<td>$N$</td>
<td>normal distribution</td>
</tr>
<tr>
<td>$n$</td>
<td>number of observations</td>
</tr>
<tr>
<td>$p$</td>
<td>time of the shift</td>
</tr>
<tr>
<td>RSS</td>
<td>residual sum of squares</td>
</tr>
<tr>
<td>SIC$_j$</td>
<td>Schwarz information criterion for model $j$</td>
</tr>
<tr>
<td>$t$</td>
<td>time (years)</td>
</tr>
<tr>
<td>$y_t$</td>
<td>response variable at time $t$</td>
</tr>
<tr>
<td>$\bar{y}$</td>
<td>sample mean of the response variable</td>
</tr>
<tr>
<td>$\bar{y}_1$</td>
<td>sample mean of the response variable before the shift</td>
</tr>
<tr>
<td>$\bar{y}_2$</td>
<td>sample mean of the response variable after the shift</td>
</tr>
</tbody>
</table>
Change-point detection in climate

Figure 1. Examples of time series with a change point in (a) the mean, (b) the variance, (c) both the mean and the variance, (d) the intercept of a linear regression model and (e) both the intercept and the trend of a linear regression model, and (f) no change point, but a strong positive autocorrelation.

(a) Mean and/or variance

Techniques for the detection of shifts in the mean and variance have received significant attention in the statistical literature [20,25–31]. Examples of synthetic series with a shift in the mean, in the variance and in both the mean and the variance are presented in figure 1a–c. Typically, the most likely time for a shift is identified, and then the model with this shift is compared to a model without any shift. To test for a shift in the mean only, a model with a constant mean and variance, and a model with a shift in the mean only are fitted. These models can be expressed, respectively, as

\[ y_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, \ldots, n), \]  
\[ y_t = \begin{cases} 
\mu_1 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, \ldots, p), \\
\mu_2 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = p + 1, \ldots, n),
\end{cases} \]

where \( y_t \) is the response variable, \( \mu \) represents the overall mean, \( \varepsilon_t \) are the normally distributed \( (N) \) random errors with mean 0 and overall variance \( \sigma^2 \), \( t \) represents the time, \( n \) is the number of observations, and \( \mu_1 \) and \( \mu_2 \) are the means before and after the unknown change point at time \( p \). These models have been used widely to detect artificial shifts in temperature, pressure or precipitation series [13,32–36]. Undocumented changes in the measurement procedures can be detected by applying change-point techniques to series of ratios or differences between the observations at several neighbouring sites. The detection and correction of these artificial shifts are very important in producing reliable time series suitable for the analysis of climate trends and climate variability and for the detection of anthropogenic climate change [37].

Phil. Trans. R. Soc. A (2012)
Models representing a shift in the variance, and in both the mean and the variance, respectively, can be expressed as

\[
\begin{align*}
\forall t \in \{1, \ldots, p\}: & \quad \mu + \varepsilon_t \sim N(0, \sigma^2) \\
\forall t \in \{p + 1, \ldots, n\}: & \quad \mu + \varepsilon_t \sim N(0, \sigma^2_2)
\end{align*}
\]

where \(\sigma^2_1\) and \(\sigma^2_2\) are the variances before and after the unknown change point at time \(p\). These models can also be compared with the model with a constant mean and variance (equation (2.1)). Change-point detection in the variance has been applied mostly in finance to study volatility in stock market prices \([14,27,29]\). Recently, Killick et al. \([38]\) used change-point detection to study changes in the variance of hindcast time series of significant wave height during 1900–2005 in the Gulf of Mexico. This analysis revealed abrupt changes in the variance of significant wave height during the period 1900–1933, which are thought to be due to the underestimation of some storms in the early twentieth century in hindcast time series.

\((b)\) Linear regression

A linear regression model with a shift in the intercept and/or trend is another case that has been extensively studied in the statistical literature \([39–44]\). We represent a simple linear regression model by

\[
y_t = \lambda + \beta t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, \ldots, n),
\]

where \(\lambda\) represents the intercept, \(\beta\) represents the trend and \(\varepsilon_t\) are the random errors. A model with a shift in the intercept at the unknown time \(p\) is expressed by

\[
y_t = \begin{cases} 
\lambda_1 + \beta t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, \ldots, p), \\
\lambda_2 + \beta t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = p + 1, \ldots, n),
\end{cases}
\]

where \(\lambda_1\) and \(\lambda_2\) are the intercept before and after the shift at time \(p\). Finally, a model with a shift in both the intercept and trend at time \(p\) can be fitted:

\[
y_t = \begin{cases} 
\lambda_1 + \beta_1 t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = 1, \ldots, p), \\
\lambda_2 + \beta_2 t + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2) \quad (t = p + 1, \ldots, n),
\end{cases}
\]

where \(\beta_1\) and \(\beta_2\) represent the trend before and after the shift at time \(p\). Examples of synthetic series with a shift in the intercept only and a shift in the intercept and trend are presented in figure 1d\textit{d},e. Solow \([45]\) used these models to search for shifts in trends of temperature time series. In hydrology, techniques allowing a choice between these models, as well as models with a shift in the mean only, have been developed to study shifts in river streamflows \([46,47]\). Streamflows are likely to exhibit these types of changes if there are changes in the precipitation, changes in deforestation or the construction of hydraulic structures in the surrounding environment \([47]\). Cermak et al. \([11]\) used these models to detect change points...
in aerosol optical depth and cloud datasets and to investigate whether these changes are consistent with ‘global brightening’. These models have also been used to detect artificial shifts in climate time series [48–50].

(c) Several shifts

The assumption that there is at most one shift in the time series can be unrealistic for the study of climate time series. For example, although attention for Pacific regime shifts originally focused on the 1976–1977 shift, it later became clear that this was but one of a sequence of regime shifts that are now thought to be associated with the PDO [2,4–6]. Several authors have developed techniques facilitating the detection of several shifts in the mean or in a linear regression model for climate-specific applications. Some authors have proposed using tests developed to detect at most one shift and to apply them iteratively [1,51]. Model selection approaches have also been used to detect the number of change points in a model [52]. Lu et al. [53] combine a model selection approach with an algorithm allowing the determination of the number and the positions of the change points. Bayesian approaches have been developed for climate or hydrological applications by Seidou & Ouarda [54] and Hannart & Naveau [55]. Techniques able to detect several shifts in the mean or in a linear regression model have been useful to study shifts in temperature, precipitation and climate indices such as the PDO [1,10] and also useful to detect several shifts in streamflows [54,56].

(d) Non-normal distributions

Not all climatic time series are normally distributed. Examples of non-normal distributions of climatic variables can be found in the study of Wilks [57], e.g. the intensity of tornado counts (Poisson), wind speed (gamma), cloudiness (beta) or climatic extremes (Weibull or generalized extreme value). Some authors have developed techniques that can detect shifts in the parameter of different distributions. Zhao & Chu [58] developed an approach to detect shifts in hurricane counts by modelling the counts by a Poisson distribution, where the intensity is represented by a gamma distribution. Jarušková [59], Jarušková & Rencová [60], Zhao & Chu [61] and Dierckx & Teugels [62] presented approaches to detect shifts in extreme events (e.g. typhoons, heavy rainfall, heat waves, temperature extrema).

More generally, non-parametric approaches based on ranks, such as the change-point method developed by Pettitt [63], are useful in detecting shifts in time series without having parametric specifications. This method has been applied to detect shifts in time series of precipitation and streamflows [64]. Lanzante [65] also applied a non-parametric change-point method to detect artificial shifts in radiosonde temperature.

(e) Autocorrelation

A common feature of climate time series is the time dependence in the observations (autocorrelation), especially at the monthly or smaller time scale [19]. The presence of strong positive autocorrelation creates patterns in time series that can be easily confused with change points, especially if the magnitude of the change point is small [17]. For example, figure 1f presents a series that is strongly and positively autocorrelated. One can easily misinterpret the variations in this
time series and identify apparent shifts, even though there are none. This is a challenging problem in change-point detection, as most techniques were developed for independent observations. In the presence of autocorrelation, the risk of false detection tends to increase and the power of detection diminishes [18,66].

Several authors have proposed approaches that can take into account autocorrelation when applying a test for a shift in the mean [21,67,68]. A detailed example of the application of these methods is presented in Jandhyala et al. [69]. Seidel & Lanzante [70] integrated the autocorrelation in the SIC formulation by using the concept of effective sample size. This approach allows a first-order autocorrelation model (AR(1)) or a second-order autocorrelation model (AR(2)) to be taken into account in the analysis. More recently, Lund et al. [18] presented an approach for detecting a shift in the intercept of a linear regression model with periodicity and autocorrelation. Wang [19] extended the penalized maximal t-test [33] to detect a shift in the mean that takes first-order autocorrelation into account. Robbins et al. [71] proposed a test based on cumulative sums adjusted for autocorrelation. Finally, Kirch [72] proposed an approach to approximate the critical values in a change-point problem with dependent observations.

3. The informational approach

The previous section surveyed a number of change-point models that have proved useful to study past changes in climate fields through the analysis of time series. There are several approaches that have been presented in the statistical literature to discriminate between these models, such as the likelihood-ratio test [20], Bayesian approach [25], cumulative sums tests [29], wavelets [73] and the informational approach [15]. A review of these approaches is presented in Chen & Gupta [24]. In this paper, we focus on the informational approach, as it is a general model selection technique that can be adapted to a diverse set of situations. This approach was used by Karl et al. [74] and by Seidel & Lanzante [70] to select among a hierarchy of piecewise regression models to explain changes in global temperature. It was also used by Killick et al. [38] to detect shifts in the variance of wave heights and by Beaulieu et al. [12] to discriminate between models representing a constant mean, an abrupt shift in the mean, a simple linear regression, a linear regression with abrupt shifts in the intercept and/or trend, and a multiple linear regression with covariate effects and a shift in the intercept. The models of Beaulieu et al. [12] were fitted to the land uptake of carbon, which seems to have abruptly shifted after 1988 [75]. The informational approach is also useful for climate time series, as their variability can be driven, for instance, by volcanic eruptions, the El Niño Southern Oscillation, the PDO or the North Atlantic Oscillation. Integrating these covariate effects allows one to explain a part of the variability in the time series and to attribute the shifts detected to other factors.

(a) General formulation of the informational approach

The proposed approach consists of the use of an information criterion to identify the unknown position of the shift under a change-point model and to discriminate, among a collection of models, the one that is the most likely to fit...
the data. The SIC, developed by Schwarz [76], and also commonly known as the Bayesian information criterion, can be used to this end [14–16,24]. The general formulation of the SIC to select among $M$ models can be expressed by

$$SIC_j = -2 \log L(\hat{\Theta}_j) + c_j \log n, \quad j = 1, 2, \ldots, M,$$

(3.1)

where $SIC_j$ is the SIC for model $j$, $L(\hat{\Theta}_j)$ represents the maximum-likelihood function for model $j$ and $c_j$ is the number of parameters to be estimated for model $j$. The SIC is based on the maximum-likelihood function of a given model penalized by the number of parameters that are estimated in the model. The model that minimizes the SIC is considered to be the most appropriate model. It represents the best compromise between parsimony (few parameters) and good fit (small residuals). The penalty term ensures that the model chosen does not over-fit the data. For example, the SIC of the model with a constant mean and variance (equation (2.1)) can be expressed as

$$SIC_1 = n \log(RSS_1) + n(1 + \log(2\pi)) + (2 - n) \log(n),$$

(3.2)

where $RSS_1 = \sum_{t=1}^{n} (y_t - \bar{y})^2$ is the residual sum of squares, $y_t$ ($t = 1, \ldots, n$) are the observations, $\bar{y}$ is the observation mean and $n$ is the number of observations. The SIC associated with the model representing a shift in the mean (equation (2.2)) can be expressed as

$$SIC_2(k) = n \log(RSS_2) + n(1 + \log(2\pi)) + (3 - n) \log(n), \quad k = 2, \ldots, n - 2,$$

(3.3)

where $RSS_2 = \sum_{t=1}^{k} (y_t - \bar{y}_1)^2 + \sum_{t=k+1}^{n} (y_t - \bar{y}_2)^2$ and where $\bar{y}_1$ and $\bar{y}_2$ are, respectively, the sample means before and after the shift at time $k$. The formulation has to be modified according to the number of parameters in the model and to the residual sum of squares of each respective model. The most likely position for a shift is selected as the one that minimizes the SIC, $SIC_2(p) = \min\{SIC_2(k), \quad k = 2, \ldots, n - 2\}$. The model with a change after time $p$ is selected if

$$SIC_2(p) < SIC_1,$$

(3.4)

otherwise, it seems more likely that there is no shift in the model [16]. The advantage of using information criteria such as the SIC is that it provides a very simple approach for exploring the presence of a change point in the data with no need to resort to the significance level. However, when the SICs are very close, one may question whether the small difference between the SICs might be due to fluctuation in the data instead of a change. If one wants to conclude that there is a change point with a significance level, then a critical value can be added to the decision rule [15]. In this case, a model with a change after time $p$ is selected (with $1 - \alpha$ confidence level) if

$$SIC_2(p) + c_\alpha < SIC_1.$$

(3.5)

The critical values $c_\alpha$ for different series lengths obtained through the asymptotic distribution are presented in Chen & Gupta [15] and can also be obtained by Monte Carlo simulations [12].
(b) Integrating the autocorrelation

Change-point techniques based on the informational approach as presented in Chen & Gupta [14] are based on the independence assumption. Here, we make an extension for the presence of autocorrelation when detecting a shift in the mean. We present the equations for an autoregressive model of order $m$ (AR($m$)). We can rewrite the models with a constant mean (equation (2.1)) and a shift in the mean (equation (2.2)) in the presence of autocorrelation as

$$y_t = \mu + \sum_{i=1}^{m} \rho_i y_{t-i} + \varepsilon_t \quad (t = 1, \ldots, n),$$  \hspace{1cm} (3.6)

$$y_t = \begin{cases} 
\mu_1 + \sum_{i=1}^{m} \rho_i y_{t-i} + \varepsilon_t \quad (t = 1, \ldots, p), \\
\mu_2 + \sum_{i=1}^{m} \rho_i y_{t-i} + \varepsilon_t \quad (t = p+1, \ldots, n),
\end{cases}$$  \hspace{1cm} (3.7)

where $\rho_i \ (i = 1, \ldots, m)$ represents the autocorrelation coefficients and the errors are independent and normally distributed ($\varepsilon_t \sim N(0, \sigma^2)$). The SIC formulation for the model with a constant mean and $m$th-order autocorrelation (equation (3.6)) would become

$$\text{SIC}_3 = n \log(\text{RSS}_3) + n(1 + \log(2\pi)) + (m + 2 - n) \log(n),$$  \hspace{1cm} (3.8)

where $\text{RSS}_3 = \sum_{t=1}^{n} (y_t - \hat{\mu} - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i})^2$ represents the sum of squares of the residuals after fitting a model with a constant mean in the presence of $m$th-order autocorrelation, and $\hat{\mu}, \hat{\rho}_1, \ldots, \hat{\rho}_m$ are the maximum-likelihood estimators of the model parameters (equation (3.6)). The last term of the equation shows that there are $m$ additional parameters to estimate (the autocorrelation coefficients), as opposed to the SIC formulation of the corresponding model without autocorrelation (equation (2.9)). The SIC for the model with a mean shift in the presence of $m$th-order autocorrelation (equation (3.7)) would be

$$\text{SIC}_4(k) = n \log(\text{RSS}_4) + n(1 + \log(2\pi)) + (m + 3 - n) \log(n),$$  
$$k = m + 2, \ldots, n - (m + 2),$$  \hspace{1cm} (3.9)

where $\text{RSS}_4 = \sum_{t=1}^{k} (y_t - \hat{\mu}_1 - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i})^2 + \sum_{t=k+1}^{n} (y_t - \hat{\mu}_2 - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i})^2$ and $\hat{\mu}_1, \hat{\mu}_2, \hat{\rho}_1, \ldots, \hat{\rho}_m$ are the maximum-likelihood estimators of the model parameters (equation (3.7)).

Following the same logic, it is also possible to integrate the autocorrelation in simple linear regression models, and in multiple linear regression models with covariate effects, and to allow the autocorrelation coefficients to change after the shift. Regression models with autocorrelated errors can be estimated using generalized least squares. We present the SIC for the linear regression models in the presence of autocorrelation and for shifts in the autocorrelation in table 2.
Table 2. List of models having \( m \)-th order autocorrelation in the errors and their associated SIC formulation. All these models rely on the assumption that the random errors are independent and identically normally distributed \((\varepsilon_t \sim N(0, \sigma^2))\).

<table>
<thead>
<tr>
<th>Model Description and Notation</th>
<th>Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant mean and ( m )-th order autocorrelation</td>
<td>[ y_t = \mu + \sum_{i=1}^m \rho_i y_{t-i} + \varepsilon_t \quad (t = 1, \ldots, n), ] (3.10)</td>
</tr>
<tr>
<td>[ \text{SIC} = n \log(\text{RSS}) + n(1 + \log(2\pi)) + (m + 2 - n) \log(n), ]</td>
<td></td>
</tr>
<tr>
<td>[ \text{RSS} = \sum_{t=1}^n \left( y_t - \hat{\mu} - \sum_{i=1}^m \hat{\rho}<em>i y</em>{t-i} \right)^2, ]</td>
<td></td>
</tr>
<tr>
<td>Where ( \hat{\mu} ) and ( \hat{\rho}_i ) ((i = 1, \ldots, m)) are the maximum-likelihood estimates of ( \mu ) and ( \rho_i ) ((i = 1, \ldots, m)).</td>
<td></td>
</tr>
<tr>
<td>Shift in the mean and ( m )-th order autocorrelation</td>
<td>[ y_t = \begin{cases} \mu_1 + \sum_{i=1}^m \rho_i y_{t-i} + \varepsilon_t &amp; (t = 1, \ldots, p), \ \mu_2 + \sum_{i=1}^m \rho_i y_{t-i} + \varepsilon_t &amp; (t = p + 1, \ldots, n), \end{cases} ] (3.11)</td>
</tr>
<tr>
<td>[ \text{SIC}(p) = n \log(\text{RSS}) + n(1 + \log(2\pi)) + (m + 3 - n) \log(n), ]</td>
<td></td>
</tr>
<tr>
<td>[ \text{RSS} = \sum_{t=1}^p \left( y_t - \hat{\mu}<em>1 - \sum</em>{i=1}^m \hat{\rho}<em>i y</em>{t-i} \right)^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ + \sum_{t=p+1}^n \left( y_t - \hat{\mu}<em>2 - \sum</em>{i=1}^m \hat{\rho}<em>i y</em>{t-i} \right)^2, ]</td>
<td></td>
</tr>
<tr>
<td>Where ( \hat{\mu}_1, \hat{\mu}_2 ) and ( \hat{\rho}_i ) ((i = 1, \ldots, m)) are the maximum-likelihood estimates of ( \mu_1, \mu_2 ) and ( \rho_i ) ((i = 1, \ldots, m)).</td>
<td></td>
</tr>
<tr>
<td>Shift in the mean and shift in the ( m )-th order autocorrelation</td>
<td>[ y_t = \begin{cases} \mu_1 + \sum_{i=1}^m \rho_{1i} y_{t-i} + \varepsilon_t &amp; (t = 1, \ldots, p), \ \mu_2 + \sum_{i=1}^m \rho_{2i} y_{t-i} + \varepsilon_t &amp; (t = p + 1, \ldots, n), \end{cases} ] (3.12)</td>
</tr>
<tr>
<td>[ \text{SIC}(p) = n \log(\text{RSS}) + n(1 + \log(2\pi)) + (2m + 3 - n) \log(n), ]</td>
<td></td>
</tr>
<tr>
<td>[ \text{RSS} = \sum_{t=1}^p \left( y_t - \hat{\mu}<em>1 - \sum</em>{i=1}^m \hat{\rho}<em>{1i} y</em>{t-i} \right)^2 ]</td>
<td></td>
</tr>
<tr>
<td>[ + \sum_{t=p+1}^n \left( y_t - \hat{\mu}<em>2 - \sum</em>{i=1}^m \hat{\rho}<em>{2i} y</em>{t-i} \right)^2, ]</td>
<td></td>
</tr>
<tr>
<td>Where ( \hat{\mu}<em>1, \hat{\mu}<em>2, \hat{\rho}</em>{1i} ) ((i = 1, \ldots, m)) and ( \hat{\rho}</em>{2i} ) ((i = 1, \ldots, m)) are the maximum-likelihood estimates of ( \mu_1, \mu_2, \rho_{1i} ) ((i = 1, \ldots, m)) and ( \rho_{2i} ) ((i = 1, \ldots, m)).</td>
<td></td>
</tr>
<tr>
<td>Intercept and linear trend with ( m )-th order autocorrelation</td>
<td>[ y_t = \lambda + \beta t + \sum_{i=1}^m \rho_i y_{t-i} + \varepsilon_t \quad (t = 1, \ldots, n), ] (3.13)</td>
</tr>
<tr>
<td>[ \text{SIC} = n \log(\text{RSS}) + n(1 + \log(2\pi)) + (m + 3 - n) \log(n), ]</td>
<td></td>
</tr>
<tr>
<td>[ \text{RSS} = \sum_{t=1}^n \left( y_t - \hat{\lambda} - \hat{\beta} t - \sum_{i=1}^m \hat{\rho}<em>i y</em>{t-i} \right)^2, ]</td>
<td></td>
</tr>
<tr>
<td>Where ( \hat{\lambda}, \hat{\beta} ) and ( \hat{\rho}_i ) ((i = 1, \ldots, m)) are the maximum-likelihood estimates of ( \lambda, \beta ) and ( \rho_i ) ((i = 1, \ldots, m)).</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. (Continued.)

<table>
<thead>
<tr>
<th>model description (and notation)</th>
<th>equations</th>
</tr>
</thead>
</table>
| shift in the intercept and same linear trend with mth-order autocorrelation | \[ y_t = \begin{cases} 
\lambda_1 + \beta t + \sum_{i=1}^{m} \rho_i y_{t-i} + \epsilon_t & (t = 1, \ldots, p), \\
\lambda_2 + \beta t + \sum_{i=1}^{m} \rho_i y_{t-i} + \epsilon_t & (t = p + 1, \ldots, n), 
\end{cases} \quad (3.14) \]
| \[ \mathrm{SIC}(p) = n \log(\mathrm{RSS}) + n(1 + \log(2\pi)) + (m + 4 - n) \log(n), \]
| \[ \mathrm{RSS} = \sum_{t=1}^{p} \left( y_t - \hat{\lambda}_1 - \hat{\beta} t - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i} \right)^2 + \sum_{t=p+1}^{n} \left( y_t - \hat{\lambda}_2 - \hat{\beta} t - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i} \right)^2, \] |
| where \( \hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta}, \hat{\beta}1, \hat{\beta}2 \) and \( \hat{\rho}_i (i = 1, \ldots, m) \) are the maximum-likelihood estimates of \( \lambda_1, \lambda_2, \beta, \) and \( \rho_i (i = 1, \ldots, m) \) |

shift in both the intercept and linear trend with mth-order autocorrelation

\[ y_t = \begin{cases} 
\lambda_1 + \beta_1 t + \sum_{i=1}^{m} \rho_i y_{t-i} + \epsilon_t & (t = 1, \ldots, p), \\
\lambda_2 + \beta_2 t + \sum_{i=1}^{m} \rho_i y_{t-i} + \epsilon_t & (t = p + 1, \ldots, n), 
\end{cases} \quad (3.15) \]
| \[ \mathrm{SIC}(p) = n \log(\mathrm{RSS}) + n(1 + \log(2\pi)) + (m + 5 - n) \log(n), \]
| \[ \mathrm{RSS} = \sum_{t=1}^{p} \left( y_t - \hat{\lambda}_1 - \hat{\beta}_1 t - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i} \right)^2 + \sum_{t=p+1}^{n} \left( y_t - \hat{\lambda}_2 - \hat{\beta}_2 t - \sum_{i=1}^{m} \hat{\rho}_i y_{t-i} \right)^2, \] |
| where \( \hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta}_1, \hat{\beta}_2 \) and \( \hat{\rho}_i (i = 1, \ldots, m) \) are the maximum-likelihood estimates of \( \lambda_1, \lambda_2, \beta_1, \beta_2 \) and \( \rho_i (i = 1, \ldots, m) \) |

4. Applications

(a) Atmospheric carbon dioxide concentrations at Mauna Loa

The atmospheric CO₂ concentration at Mauna Loa has been expressed as a superposition of a linear and quadratic trend [77]. Lund & Reeves [49] applied a change-point detection approach to annual CO₂ concentrations at Mauna Loa by fitting a regression model with a linear and a quadratic trend and searching for a shift in the regression coefficients. A shift in 1989 was detected in the three regression coefficients (intercept, linear trend and quadratic trend) with a 95% confidence level. To compare the two models, Lund & Reeves [49] used a maximum Fisher statistic. Here, we refine this change-point analysis using the observations until 2010 by fitting not only the same models as Lund & Reeves [49], but also models with a change point only in the intercept, in the linear trend, in the quadratic trend or a combination of these, in order to investigate whether the shift occurs in all the regression coefficients and
whether we could simplify the model. We apply these models to annual CO₂ concentrations at Mauna Loa for the period 1959–2010 from the National Oceanic and Atmospheric Administration, Earth System Research Laboratory [78] (available at http://www.esrl.noaa.gov/gmd/ccgg/trends/; see figure 2). The models fitted to the data and their associated SIC are presented in table 3. The model leading to the smallest SIC is also represented in figure 2. The most likely model has a shift in 1991 in both the intercept and the quadratic trend coefficient, but the linear trend coefficient remains the same after the shift. This model has an SIC of 64, which is very close to the model with a shift in the three parameters, with an SIC of 68. However, this model should be chosen as it is a simpler model with fewer parameters. It is also very close to the SIC of the model with a shift in both the intercept and the linear trend coefficient (SIC of 65). As these two models are very similar and have the same number of parameters, they might be equally good for representing annual CO₂ concentrations at Mauna Loa. Both these models are simpler than the one fitted in Lund & Reeves [49], and the change-point timing coincides with the Mount Pinatubo eruption that occurred in 1991.

\( (b) \) Interhemispheric gradient of \( \Delta ^{14}C \)

The \( \Delta ^{14}C \) of CO₂ (\( \Delta ^{14}C \) for short here) is a widely used tracer of past climate changes for both the ocean and the atmosphere. We analyse the interhemispheric gradient of \( \Delta ^{14}C \) over the period 850–1830 covering the Medieval Climate Anomaly (approx. 950–1250) and the Little Ice Age (approx. 1500–1800) using INTCAL04 [79] and SHCAL04 [80] tree-ring data. The \( \Delta ^{14}C \) seems to have shifted abruptly in the transition between the Medieval Climate Anomaly and the Little
We fitted models for which the parameters of the second-order autoregressive model follow a second-order autoregressive model \[82\]. Thus, we fit a model without and with a shift in the mean and find that the most likely change and the associated SIC.

Table 3 presents the models fitted and the associated SIC. First, we fit a model without and with a shift in the mean and simultaneously integrate the autocorrelation in the residuals. Here, we apply the informational approach to search for a shift in the mean and simultaneously integrate the autocorrelation in the analysis. We also consider the possibility of a shift in the autocorrelation parameters. Table 3 presents the models fitted and the associated SIC. First, we fit a model without and with a shift in the mean and find that the most likely time for a shift would be in 1375, as presented in Rodgers et al. \[82\]. However, the residuals follow a second-order autoregressive model \[82\]. Thus, we fit a model with and without a shift in the mean and a second-order autoregressive model (table 3). We fitted models for which the parameters of the second-order

<table>
<thead>
<tr>
<th>application</th>
<th>model</th>
<th>year</th>
<th>SIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO₂ concentrations at Mauna Loa</td>
<td>[ y_t = \lambda + \beta t + \eta t^2 + \varepsilon_t \quad (t = 1, \ldots, n) ]</td>
<td></td>
<td>118.16</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \lambda_1 + \beta t + \eta t^2 + \varepsilon_t &amp; (t = 1, \ldots, p) \ \lambda_2 + \beta t + \eta t^2 + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1992</td>
<td>92.60</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \lambda_1 + \beta_1 t + \eta_1 t^2 + \varepsilon_t &amp; (t = 1, \ldots, p) \ \lambda_2 + \beta_2 t + \eta_2 t^2 + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1991</td>
<td>65.31</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \lambda_1 + \beta_1 t + \eta_1 t^2 + \varepsilon_t &amp; (t = 1, \ldots, p) \ \lambda_2 + \beta_2 t + \eta_2 t^2 + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1991</td>
<td>64.06a</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \lambda_1 + \beta_1 t + \eta_1 t^2 + \varepsilon_t &amp; (t = 1, \ldots, p) \ \lambda_2 + \beta_2 t + \eta_2 t^2 + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1991</td>
<td>68.02</td>
</tr>
<tr>
<td>interhemispheric gradient of ¹⁴CO₂</td>
<td>[ y_t = \mu + \varepsilon_t \quad (t = 1, \ldots, n) ]</td>
<td></td>
<td>747.73</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \mu_1 + \varepsilon_t &amp; (t = 1, \ldots, p) \ \mu_2 + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1375</td>
<td>690.16</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \mu + \sum_{i=1}^{2} \rho_i y_{t-i} + \varepsilon_t \quad (t = 1, \ldots, n) ]</td>
<td></td>
<td>475.72</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \mu_1 + \sum_{i=1}^{2} \rho_i y_{t-i} + \varepsilon_t &amp; (t = 1, \ldots, p) \ \mu_2 + \sum_{i=1}^{2} \rho_i y_{t-i} + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1600</td>
<td>468.68</td>
</tr>
<tr>
<td></td>
<td>[ y_t = \begin{cases} \mu_1 + \sum_{i=1}^{2} \rho_i y_{t-i} + \varepsilon_t &amp; (t = 1, \ldots, p) \ \mu_2 + \sum_{i=1}^{2} \rho_i y_{t-i} + \varepsilon_t &amp; (t = p + 1, \ldots, n) \end{cases} ]</td>
<td>1455</td>
<td>430.95a</td>
</tr>
</tbody>
</table>

\[ Model \text{ leading to the smallest SIC.}\]
autoregressive model remain the same before and after the shift and a model for which the parameters also change along with the shift in the mean. According to the SIC, it is most likely that the mean of the $\Delta^{14}C$ and its autocorrelation parameters abruptly shifted in 1455. This result is slightly different from the result of Rodgers et al. [82], but is still in agreement, as the shift is detected during the transition period between the Medieval Warm Period and the Little Ice Age. The magnitude of the shift remains the same (2‰) and is approximately one standard deviation (1 s.d.). The advantage of the present methodology is the simultaneous estimation of the change-point position and the autocorrelation, which allows one to avoid interference while estimating them.

(c) Length of time series and shift detection

In order to determine how many years of observations are required to detect an abrupt shift, we carry out a simulation study using the change-point methodology described in the previous section. This example is motivated by time series of atmospheric CO$_2$ growth rates, measured since 1958 at Mauna Loa, as well as the uptake of carbon by the land, and thought to have experienced an abrupt shift around 1990 [12,75]. Cermak et al. [11] also detected an abrupt shift in the early 1990s in trends of aerosols and clouds. Thus, we generate synthetic series of annual ‘observations’ corresponding to the years 1961–2100. We set an abrupt shift in the mean in the year 1990. We vary the magnitude relative to the s.d. ($\Delta = |\mu_1 - \mu_2|/\sigma$) from 0.5 to 3. We use this formulation to ease comparison with other time series. The synthetic series generated are normally and independently distributed with a mean of zero and a s.d. of one. For each combination of magnitude and number of years after the shift,
we generate 2000 synthetic series. We also generated 2000 synthetic series without introducing any shifts. Figure 4a presents examples of the synthetic series generated.

We apply the change-point method to detect a shift in the mean of independent observations (equations (3.2) and (3.3)) to the synthetic series generated for the 1961–1992 period only, 1961–1994 only and so on until 1961–2100. We use the two decision rules based on the SIC values of the two models only (equation (3.4)) and based on a 95% confidence level (equation (3.5)). In the rest of the paper, we denote these decision rules as decision rule 1 and decision rule 2, respectively. We use this approach in order to determine how soon the shifts are detectable. We compute the hit rates, which we define as: the percentage of cases in the 2000 synthetic series for which the model with a shift in the mean is selected and the shift detected is located at most 2 years away from the real shift in 1990. The 2 years rule was used in several studies comparing techniques for the detection of artificial shifts in temperature and precipitation [35, 66, 83] and provides a measure of the power of detection of the technique.

Figure 4b presents the results of this simulation study. It shows that shifts of any magnitude are detectable soon after their occurrence. For very large shifts (with a magnitude that is three times the s.d.), the hit rate reaches 99 per cent
only 4 years after the shift occurs if decision rule 1 is used, and 6 years after
the shift occurs with decision rule 2. For shifts having a magnitude of 2 s.d., the
hit rate reaches 90 per cent after 4 years (decision rule 1) and after 10 years
(decision rule 2). For shifts of a magnitude the same as the s.d., the hit rate is
approximately 50 per cent after 6 years and keeps increasing until 60 per cent
after 100 years (decision rule 1). If decision rule 2 is used, the hit rate is much
smaller (50% after 40 years). For shifts with a magnitude smaller than the s.d.
in the series, the hit rate remains very small: 40 per cent for a magnitude of 0.75
and 20 per cent for a magnitude of 0.5 (decision rule 1). The hit rates are even
smaller when decision rule 2 is used. In general, if the shift is not detected after 30
years of observations, the probability of detecting it will not increase much with
more observations. Hence, the probability of detecting a shift depends mainly on
the magnitude of the shift and less on the length of the time series.

We repeat the exercise with the synthetic series in which we did not introduce
any shift. We compute the percentage of false detections: the cases for which a
model with a shift is selected. Figure 4c presents the percentage of false detections
according to the length of the synthetic series. If decision rule 1 is used, the
overall risk of false detection (integral under each curve) is very high for small
samples (50% for the 1961–1992 sample) and decreases with larger samples (38%
for the 1961–2100 sample). However, this high percentage of false detection is
inflated by the higher risk of false detection at the beginning or end of the time
series. When analysing a shift that could have occurred in 1990, we can see that
the risk of false detection is very high if we perform the analysis for shorter
periods (1961–1992 and 1961–2000). However, if the time series lasts until 2010
or longer, then the risk of false detection becomes approximately 2 per cent or
even smaller. If decision rule 2 is used with a 95% confidence level, there is
a very small overall risk of false detection (2% for the 1961–2100 sample and
less for smaller samples). One can expect a higher risk of false detection and a
higher probability of detecting a shift when performing the analysis with a smaller
confidence level. Similarly, one can expect a smaller risk of false detection and
a smaller probability of detecting a shift when performing the analysis with a
higher confidence level.

5. Discussion and conclusion

In this paper, we reviewed commonly used change-point models to study past
changes in climate-related variables and demonstrated their use in a number
of applications. We gave a detailed explanation of change-point methodology
based on the informational approach and provided an extension to take into
account nth-order autocorrelation in the detection of a shift in the mean and of
a shift in the coefficients of a simple linear regression, and to detect a shift in the
mean and in the autocorrelation. In order to show the flexibility of the proposed
approach, we presented an example of an application to the CO2 concentrations at
Mauna Loa, which we modelled by the sum of a linear and a quadratic trend and
which showed a shift in the regression coefficients in 1991. A second application
to the interhemispheric gradient of Δ\textsuperscript{14}C shows an abrupt shift in the mean
and in the autocorrelation parameters occurring during the transition from the
Medieval Climate Anomaly and the Little Ice Age, which agrees with the results
of Turney & Palmer [81] and Rodgers et al. [82]. These examples showed that the proposed methodology can be easily adapted to several cases that might be encountered in climatic time series.

It is of critical importance to be able to detect abrupt shifts very soon after they have occurred (e.g. climate monitoring). Thus, we have verified by simulation how long it would take to detect an abrupt shift in the mean having different magnitudes. To make a realistic case study, we have generated synthetic time series covering the period 1961–2100 and introduced a shift in 1990, because a few studies show evidence for abrupt shifts around this time. This method shows the potential to start analysing shifts that might have occurred around 1990 if the observations started in 1961 or earlier. Hence, we have good chances to start detecting the shifts with a magnitude of at least 1 s.d. with both decision rules (with and without a confidence level associated). Shifts having a small magnitude (less than 1 s.d.) will be more difficult to detect, especially if decision rule 2 is used. Thus, for monitoring small shifts, decision rule 1 might be better. However, one should be cautious when using decision rule 1 because of the high risk of false detection. This is especially true if a shift is detected close to the beginning or end of the series, as the risk of false alarm is higher. To obtain a significance level associated with the change-point analysis, we have shown that this can be included through the decision rule. The critical values for the SIC test to search for a shift in the mean and/or in the variance are presented by Chen & Gupta [14]. The critical values for different change-point models can be obtained easily through Monte Carlo experiments. The choice between the two different decision rules should depend on the context and need of different applications.

Other model selection approaches could be used, such as reversible jump Markov chain Monte Carlo [84], birth–death Markov chain Monte Carlo [85], the deviance information criterion [86] or the minimum description length [87]. However, there is no study comparing these model selection approaches for discriminating between several change-point models. Comparative simulation studies should be performed in order to identify the advantages and disadvantages of several model selection approaches for change-point detection. A number of efforts have been dedicated to identifying the best-performing homogenization techniques for temperature and/or precipitation [34,35,66,83]. These types of comparison studies would also benefit the development of optimal methods for change-point detection in the climate system.

Change-point analysis is a tool that can be used to describe past climate variations. Change-point techniques are not intended to have predictive skill with respect to future shifts. More significant development would be necessary in order to develop a framework for monitoring shifts and predicting them in advance. Techniques able to provide early warning signals for tipping points in the climate and ecosystems are presented by Livina & Lenton [88], Dakos et al. [89], Scheffer et al. [90], Lenton et al. [91] and Sieber & Thompson [92]. The shifts discussed in this paper cannot be predicted in advance with these early warning signal methods, as they are not accompanied by a critical slowing down.

A part of this work was completed while C. Beaulieu was a visiting fellow at the Isaac Newton Institute for Mathematical Sciences, Cambridge University, for the ‘Mathematical and Statistical Approaches to Climate Modelling and Prediction’ programme. The authors acknowledge financial support from the Fonds Québécois de la Recherche sur la Nature et les Technologies, the
Department of Energy (grant number DE-FG02-07ER64467), the Cooperative Institute for Climate Science, and BP and Ford Motor Company through the Carbon Mitigation Initiative at Princeton University. The authors would also like to thank the editors, an anonymous reviewer, D. Jaruskova, R. Lund and K. Rodgers for constructive and helpful comments that led to improvements in the manuscript.

References


