The Annals of Applied Statistics
2012, Vol. 6, No. 3, 1162–1184
DOI: 10.1214/12-AOAS548

TREE MODELS FOR DIFFERENCE AND CHANGE DETECTION IN A COMPLEX ENVIRONMENT

BY YONG WANG, ILZE ZIEDINS, MARK HOLMES AND NEIL CHALLANDS

University of Auckland, University of Auckland, University of Auckland and New Zealand Fire Service

A new family of tree models is proposed, which we call "differential trees." A differential tree model is constructed from multiple data sets and aims to detect distributional differences between them. The new methodology differs from the existing difference and change detection techniques in its nonparametric nature, model construction from multiple data sets, and applicability to high-dimensional data. Through a detailed study of an arson case in New Zealand, where an individual is known to have been laying vegetation fires within a certain time period, we illustrate how these models can help detect changes in the frequencies of event occurrences and uncover unusual clusters of events in a complex environment.

1. Introduction. We propose a new family of tree models that can be used to uncover distributional differences between multiple data sets. These models, which we call "differential trees," are suitable for solving sophisticated, multivariate problems. They can be applied, for instance, to change detection and work effectively in an online surveillance fashion.

The research was motivated by a real-world problem. Fire service departments are often interested in detecting changes in the frequencies of different types of fire incidents, automatically from large amounts of data and informatively to shed light on potential causes. This change detection problem is certainly not unique to fire incidents. Similar problems can easily be found in many fields such as climatology, epidemiology and economics.

To investigate the problem in depth, one particular scenario has been chosen as a case study, and is used exclusively in this paper to illustrate and investigate the new methodology. It was known that an individual had been laying vegetation fires between October 2006 and January 2007 in the urban area of Blenheim, New Zealand. The New Zealand Fire Service wishes to be able to automatically detect such a sequence of events as early as possible and isolate them from the rest. At first glance, there seems to be a lack of information to relate the scenario to frequency change detection, since no fire maliciously set by an individual could be definitely known as such in reality. However, a surrogate variable can be used. All fire incidents are categorized by on-the-spot fire fighters as either suspicious or not.

Received March 2011; revised December 2011.

Key words and phrases. Tree models, change detection, event data, p-value adjustment, arson case study.

Since the maliciously-set fires should be highly correlated to those labeled suspicious, we turn the vaguely-defined practical problem into one of detecting changes in the frequencies of: (a) suspicious and other fires, (b) suspicious fires only, as a more direct approach, or (c) fire incidents of a different categorization, as a less direct approach. We consider the frequency changes as distributional differences.

The problem above poses a number of challenges for traditional change detection methods that rely on parametric assumptions [Basseville and Nikiforov (1993), Gustafsson (2000), Poor and Hadjiliadis (2009)]. For this and similar problems, there may exist a number of potentially relevant variables, which can be either numerical or categorical and may contain missing values. The distribution of fire incidents may depend on many factors, such as geographical, seasonal, time-of-day and day-of-week effects, and is simply impossible to model parametrically. Moreover, an arsonist may operate in certain time periods and in certain neighborhoods, and light fires of certain types.

By contrast, the proposed methodology is particularly suitable for solving such problems. Though belonging to the family of tree models [Breiman et al. (1984), Morgan and Sonquist (1963), Quinlan (1993)], a differential tree is constructed from multiple data sets, as opposed to from a single data set by a conventional method, and purpose-built for difference detection. Intuitively, the method stacks the data sets on top of one another (imagine a two-dimensional case) and then, via recursive space partitioning of tree-structured models, looks for the local areas with heterogeneity. By ignoring variations in individual data sets that are common to all and thus irrelevant to changes, such as geographical and seasonal effects in the arson case, it makes more efficient use of data information than an approach that builds one model from each data set. Hence, it achieves a gain in power which is similar in spirit to that of the paired t-test or blocking in experimental design.

The arson data used throughout the paper contains information for all fire incidents that occurred within and around Blenheim, a moderately sized town (population 30,200), between 1/Jan/2004 and 31/Dec/2007, as stored in 11 variables, with names, meanings and possible values given in Table 1. During the quadrennial period, there were a total of 704 fire incidents, 171 of which were labeled suspicious. Two variables, `heatsource` and `objignited`, contain, respectively, 342 and 275 missing values. Pairs of disjoint subsets of the data will be produced in various ways below, and will be used to construct differential trees. Our main focus will be on contrasting the subsets in two biennial periods, 2004–2005 and 2006–2007, to uncover the unusual cluster(s) of fire incidents in the latter period that are likely related to the arson case. We shall also apply the methodology in a sequential detection fashion and compare two consecutive annual periods by shifting time periods progressively. Random subsets will also be produced by permutation or bootstrapping for assessing and enhancing performance.

The rest of the paper is organized as follows. Section 2 briefly reviews the problem of change detection and tree models and gives an overview of the proposed methodology. Section 3 describes in detail the differential tree models and their

TABLE 1
Variables

Name	Meaning	Values
`x`	Map grid east	Real
`y`	Map grid north	Real
`Urban`	Whether an urban or rural area	{1 = urban, 0 = rural}
`Alarm`	Alarm method code	{1 = 111 emergency call, 2 = exchange phone call, 3 = running call, 4 = police/ambulance, 5 = private fire alarm, 6 = other}
`Firetype`	Type of fire incident	{1 = structure, 2 = mobile property, 3 = vegetation, 4 = chemical, 5 = rubbish, 6 = other}
`Heatsource`	Heat source	{1 = outside fire lit for lawful purpose, 2 = gas/liquid fuelled equipment, 3 = solid fuelled equipment, 4 = electrical equipment, 5 = hot object, 6 = fireworks, 7 = cigarette/smoking materials, 8 = act of nature, 9 = exposure fire}
`Objignited`	Object ignited	{1 = structure component, 2 = furniture/appliances, 3 = soft goods/bedding, 4 = decoration/recreational materials, 5 = storage containers and materials, 6 = electrical equipments/tyres/insulators, 7 = outdoor items, 8 = hazardous substances and fuels, 9 = other}
`Time`	Time of day	[0, 24)
`day`	Day of the quadrennial period	{1, 2, . . . , 1461}
`Dayweek`	Day of the week	{1 = Monday, . . . , 7 = Sunday}
`label`	Category labeled by fire fighters	{`suspicious` = suspicious fire, `other` = other type}

construction. A primary study of the arson case is presented in Section 4. The performance of the method will be assessed and enhanced in Section 5, with an application in a sequential detection fashion given in Section 6. Section 7 investigates building differential trees using other responses, and Section 8 gives some concluding remarks.

The data and computer code for carrying out the analysis presented in the paper are given in the supplementary material [Wang et al. (2012)].

2. An overview.

2.1. *Change detection*. Change detection has a long history of research in statistics, with a focus on detecting change points [Lai (1995), MacEachern, Rao and Wu (2007), Page (1954), Shewhart (1931)]. These methods, however, rely on

parametric assumptions and are applied to situations, such as industrial process control, where such assumptions can be safely made.

Another very useful technique for detecting changes is scan statistics [Glaz, Naus and Wallenstein (2001), Naus (1965)]. This technique looks for unusual clusters of temporal or spatial events in a single data set by using a scanning window to locate clusters of observations that differ in distribution from the rest. Because of the high computational cost, it is only applicable to low-dimensional problems.

2.2. *Tree models*. Tree models are often used to solve difficult, high-dimensional problems. There are two major families, classification and regression trees, for a categorical and a continuous response variable, respectively [Breiman et al. (1984)]. Other families also exist but are less used, for example, Poisson regression trees for a count response [Chaudhuri et al. (1995), Therneau and Atkinson (1997)] and survival trees for a failure time response with censoring [Davis and Anderson (1989), Ishwaran et al. (2008)].

As in the references above, the basic idea of tree modeling is to partition the space of explanatory variables recursively into increasingly smaller regions so that a simple model fits well to the data in a minimal region. We call such a simple model an *atomic model*, which can be, for example, the constant function or the normal distribution for a continuous response, or a multinomial distribution for a categorical response, as for regression and classification trees, respectively. A tree model is the composite of the atomic models in the minimal regions and is best represented by a rooted tree, in which a node corresponds to a region, a terminal node a minimal region, and the branching under an internal node a space partitioning. Each internal node thus also has a *subtree model*.

Building a tree model typically consists of splitting and pruning stages. Splitting proceeds in a top-down fashion, by selecting at each node a split in the form of a logical condition from a large number of candidates, which aims to maximize the homogeneity in subregions. Univariate binary splits are commonly used, for example, $x \leq 3.5$ for a continuous variable, or $\texttt{season} = \{\texttt{spring}, \texttt{autumn}\}$ for a categorical variable. Splitting continues until homogeneity is reached in a region. An exhaustive splitting is generally beneficial, and allows for uncovering relations hidden deep under the surface. However, a tree grown only by splitting is likely to overfit the data. Hence, it is often followed by pruning, which replaces spurious subtrees with their root nodes in a bottom-up fashion.

Terminal nodes are important for a tree model and the features of interest at those nodes are described by the atomic models. We shall often use the word "pattern" to specifically indicate a terminal node, including its associated region and observations, atomic model, and assessment results.

2.3. *Differential trees*. In this paper we relate the methodology of tree modeling to difference and change detection. By following the general methodology

described in Section 2.2, a differential tree is built from *multiple* data sets to discover distinguishing patterns between them. Its atomic model is for observations from all data sets, but only the parameters that account for differences in distribution are of direct interest and examined by a homogeneity test. In particular, we will use the Poisson distribution for the count of observations at each level of the response in each data set to form the atomic model, and contrast the event rates in all data sets with a likelihood ratio test of homogeneity.

This new change detection method differs from those described in Section 2.1, in its nonparametric nature and applicability to high-dimensional data. As for other families of tree models, it has the advantages of fast training (relative to most other data mining models), easy handling of different types of variables and dealing nicely with missing values. The resulting models are easily comprehensible, which can be important for change detection, since it helps suggest possible causes behind complicated phenomena.

3. Building differential trees.

3.1. *A likelihood-based framework.* We adopt the likelihood-based approach for building a differential tree and for subsequent analysis. Using the likelihood for tree construction is not rare in the literature, but it sometimes takes an implicit or approximate form. For example, for classification trees the information gain splitting criterion of Quinlan (1993) is equivalent to using the likelihood ratio test, whereas the χ^2 criterion [Kass (1980)] and the Gini splitting criterion [Breiman et al. (1984)] are approximations. Su, Wang and Fan (2004) use the likelihood method, in place of the least squares criterion, for building regression trees and obtain simpler yet more accurate tree models in general. Using the likelihood method for tree construction gives results a statistical interpretation, deals with splitting and pruning in one framework, and permits the handling of many families of atomic models in a coherent way. For building differential trees we take one further step, by making use of the p-values of the likelihood ratio tests. In general, this helps resolve several difficult issues: (a) splitting in multiple ways; (b) adjusting in the presence of missing values; (c) assessing patterns by their statistical significance; and (d) adjusting for multiple hypothesis testing.

Within this framework, the likelihood ratio test or its statistic can also be conveniently used to assess and compare models, even if there exist nuisance parameters, as in the case of differential trees. We will make extensive use of the fact that the log-likelihood ratio statistic W is asymptotically χ^2_ν, with degrees of freedom ν equal to the number of free parameters for a simple hypothesis or the difference in the number of free parameters for a composite one.

3.2. *Likelihood ratio test.* The Poisson distribution with probability mass function

$$f(n;\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}, \qquad \lambda > 0, n = 0, 1, 2, \ldots,$$

is widely used to model the number of occurrences of an event over time or in space. Let Y_i $(i = 1, 2)$ have the Poisson distribution with rate λ_i. For testing homogeneity

$$H_0 : \lambda_1 = \lambda_2,$$

the log-likelihood ratio statistic is given by

$$W = 2\{\log f(Y_1; Y_1) + \log f(Y_2; Y_2) - \log f(Y_1; \bar{Y}) - \log f(Y_2; \bar{Y})\},$$

where $\bar{Y} = (Y_1 + Y_2)/2$. Under H_0, W is asymptotically χ_1^2.

There are two parameters here, (λ_1, λ_2), or, with reparametrization, $(\lambda_1, \lambda_2 - \lambda_1)$. The focus is on whether $\lambda_2 - \lambda_1 = 0$, while λ_1 is a nuisance parameter.

3.3. *Atomic models.* Suppose there are d data sets and the response variable has c levels. For node τ, let D^τ denote the data in its associated subregion, and assume Y_{ij}^τ $(i = 1, \ldots, c, j = 1, \ldots, d)$, the number of observations of level i in data set j, is Poisson distributed with mean λ_{ij}^τ in that subregion. The atomic model thus has $c \times d$ unknown parameters, λ_{ij}^τ, or, equivalently,

$$\begin{pmatrix} \lambda_{11}^\tau & \lambda_{12}^\tau - \lambda_{11}^\tau & \cdots & \lambda_{1d}^\tau - \lambda_{11}^\tau \\ \lambda_{21}^\tau & \lambda_{22}^\tau - \lambda_{21}^\tau & \cdots & \lambda_{2d}^\tau - \lambda_{21}^\tau \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{c1}^\tau & \lambda_{c2}^\tau - \lambda_{c1}^\tau & \cdots & \lambda_{cd}^\tau - \lambda_{c1}^\tau \end{pmatrix}.$$

Any nonzero difference in the matrix implies a *distributional difference* between the data sets. Of direct interest to us is whether all the differences are exactly zero, while those in the first column are nuisance parameters.

We can hence perform a homogeneity test under the null hypothesis

$$H_0 : \lambda_{i1}^\tau = \cdots = \lambda_{id}^\tau \qquad \text{for } i = 1, \ldots, c. \tag{3.1}$$

Letting $\bar{Y}_i^\tau = d^{-1} \sum_{j=1}^d Y_{ij}^\tau$ $(i = 1, \ldots, c)$, the log-likelihood ratio statistic becomes

$$W(\tau; D^\tau) = 2\left\{ \sum_{i=1}^c \sum_{j=1}^d \log f(Y_{ij}^\tau; Y_{ij}^\tau) - \sum_{i=1}^c \sum_{j=1}^d \log f(Y_{ij}^\tau; \bar{Y}_i^\tau) \right\}, \tag{3.2}$$

which is asymptotically $\chi^2_{(d-1)c}$ under (3.1). The test provides evidence for preference between two settings of the atomic model.

As an example, consider the most significant pattern produced by the differential tree shown later in Figure 2. This pattern covers 22 other and 0 suspicious fires in the first data set, and 43 other and 41 suspicious fires in the second. The test statistic value is

$$\begin{aligned} W &= 2\{\log f(22; 22) + \log f(43; 43) - \log f(22; 32.5) - \log f(43; 32.5)\} \\ &\quad + 2\{\log f(0; 0) + \log f(41; 41) - \log f(0; 20.5) - \log f(41; 20.5)\} \\ &\approx 63.75, \end{aligned}$$

which yields a p-value of 1.4×10^{-14} under χ^2_2.

The appropriate atomic model depends on the problem under study. By assuming equal rates, null hypothesis (3.1) implies that all data sets were obtained under the same exposure, for example, over time periods of equal length. While this applies to our analysis presented below due to our special partitioning of the data set on an annual basis, one could also consider the case where exposures are different. If the exposures are known, say, e_j for data set j, one needs to modify H_0 to

$$H_0' : e_1\lambda_{i1}^\tau = \cdots = e_d\lambda_{id}^\tau \qquad \text{for } i = 1, \ldots, c, \tag{3.3}$$

where λ_{ij}^τ is a rate per unit exposure. Reassigning $\bar{Y}_i^\tau = \sum_{i=1}^d e_j Y_{ij}^\tau / \sum_{i=1}^d e_j$, we have

$$W'(\tau; D^\tau) = 2\left\{\sum_{i=1}^{c}\sum_{j=1}^{d} \log f(Y_{ij}^\tau; Y_{ij}^\tau) - \sum_{i=1}^{c}\sum_{j=1}^{d} \log f(Y_{ij}^\tau; e_j\bar{Y}_i^\tau)\right\},$$

which is also asymptotically $\chi^2_{(d-1)c}$.

If, however, the exposures are unknown, it is impossible to test a null hypothesis of type (3.1) or (3.3). Instead, one can investigate if every data set has the same distribution for the proportions of all response levels, namely,

$$H_0'' : p_{i1}^\tau = \cdots = p_{id}^\tau \qquad \text{for } i = 1, \ldots, c, \tag{3.4}$$

where p_{ij}^τ is the probability an observation in data set j is of level i. Therefore, one can assume that $(Y_{1j}^\tau, \ldots, Y_{cj}^\tau)^\top$ has a multinomial distribution with probabilities $(p_{1j}^\tau, \ldots, p_{cj}^\tau)^\top$. The log-likelihood ratio statistic is

$$W''(\tau; D^\tau) = 2\left\{\sum_{i=1}^{c}\sum_{j=1}^{d} n_j^\tau \log \hat{p}_{ij}^\tau - \sum_{i=1}^{c}\sum_{j=1}^{d} n_j^\tau \log \hat{p}_i^\tau\right\},$$

where $n_j^\tau = \sum_{i=1}^c Y_{ij}^\tau$, $\hat{p}_{ij}^\tau = Y_{ij}^\tau / n_j^\tau$ and $\hat{p}_i^\tau = \sum_{j=1}^d Y_{ij}^\tau / \sum_{j=1}^d n_j^\tau$. Under H_0'', W'' is asymptotically $\chi^2_{(c-1)(d-1)}$.

Throughout our study, we assume that the underlying distribution of counts is Poisson distributed, and only the null hypothesis (3.1) and the resulting statistic (3.2) are used. In general, altering the atomic model alters the family of differential trees being considered, but the framework for analysis remains the same.

3.4. *Subtree models.* Denote by T^τ the subtree rooted at node τ and by $\widetilde{T}^\tau$ the set of its terminal nodes. The log-likelihood ratio statistic for T^τ is given by

$$W(T^\tau; D^\tau) = \sum_{t \in \widetilde{T}^\tau} W(t; D^t). \tag{3.5}$$

The statistic $W(T^\tau; D^\tau)$ is approximately χ^2, with degrees of freedom given by the sum of the degrees of freedom of the individual terms.

3.5. *Splitting*. We only consider univariate binary splits, which use data information most efficiently, allow surrogate splitting in the presence of missing values, and treat numerical variables no differently from ordinal ones. We further turn categorical variables into ordinal ones by using their pre-given order of levels, instead of considering all possible combinations. This avoids the overfitting introduced by level grouping, which can be severe when a categorical variable has many levels. It is also helpful when the pre-given levels are partially ordinal.

Without missing values, the primary split at node τ is determined by

$$s_*^\tau = \arg\max_{s \in S^\tau} W(T_s^\tau; D^\tau), \tag{3.6}$$

where T_s^τ is the two-child-node tree defined by a univariate binary split s and S^τ is the set of all candidate splits at node τ. For S^τ, we consider every explanatory variable and every midpoint between two consecutive distinct values of the variable from all data sets, but we exclude the cases where small subsets (having less than a total of $5c$ observations, by default) are produced. With s_*^τ, all data sets are split accordingly and the tree is grown with two new child nodes. The splitting process starts with a single node for all data and proceeds in a top-down, recursive style, until a stop-splitting criterion is met, for example, too few observations left.

To find the primary split in the presence of missing values, a slight adjustment is made using p-values, which takes account of different sample sizes caused by missing values. For the kth variable at node τ, denote by n_k^τ the number of observations without missing values and by p_{k*}^τ the smallest p-value of all likelihood ratio tests for the n_k^τ observations. The adjusted p-value is given by

$$\tilde{p}_{k*}^\tau = p_{k*}^\tau + \gamma \sqrt{p_{k*}^\tau (1 - p_{k*}^\tau)/n_k^\tau}, \tag{3.7}$$

where $\gamma > 0$ is a constant, which is defaulted to 2 in our implementation. Similar in spirit to the 1-SE rule of Breiman et al. (1984), the adjustment tends to favor variables with fewer missing values.

To determine the correct branch for an observation when the primary splitting variable at a node has a missing value, a surrogate split can be used, as described in Breiman et al. (1984), Section 5.3. A surrogate split is made using a different variable, chosen so that the surrogate split is as similar to the primary split as possible for the observations without missing values. We measure the similarity of two splits by the number of common observations in their resulting subsets. An ordered list of surrogate splits can be constructed according to their similarities to the primary split.

3.6. *Pruning*. The pruning of an initially grown tree is necessary for removing spurious subtrees. It works in a bottom-up style, by choosing either the atomic model at an internal node or its subtree model. To do this, one could use the cost-complexity measure, which here is just the log-likelihood penalized by the degrees of freedom. For a subtree, it is defined as

$$W_\alpha(T^\tau; D^\tau) = W(T^\tau; D^\tau) + \alpha \, \mathrm{DF}(\widetilde{T}^\tau),$$

where $\mathrm{DF}(\widetilde{T}^\tau)$ is the number of degrees of freedom for all the atomic models in $\widetilde{T}^\tau$ and α the complexity parameter. Note that the atomic model at a node is just a tree with a single node, so its cost-complexity measure is

$$W_\alpha(\tau; D^\tau) = W(\tau; D^\tau) + \alpha\,\mathrm{DF}(\tau).$$

The pruning criterion is as follows:

(3.8) Choose the model with the larger value of W_α.

The value of α can be determined by a model selection criterion, such as AIC or BIC, or cross-validation. In principle, replacing a subtree with its root node implies that the event frequencies cannot be *further* differentiated between the data sets in all subregions.

If the main goal for building a differential tree is to find the most significant differences between data sets, we can simply preserve the most significant patterns in a constructed tree. Let $p_{\min}(\tau)$ be the p-value of the hypothesis test performed at the node τ, for example, the likelihood ratio test based on the statistic (3.2); and $p_{\min}(T^\tau)$ be the smallest p-value of all hypothesis tests performed at the atomic nodes of the subtree T^τ. The new pruning criterion is as follows:

(3.9) Choose the model with the smaller value of $p_{\min}$.

By doing so, each subtree preserves the node with the most significant pattern and keeps it as a terminal node. This also facilitates the p-value adjustments, as described in Section 5.1.

In addition, one may set up a threshold p-value, say, p_{cut}, such that a subtree is cut off directly if its $p_{\min}(T^\tau) \geq p_{\mathrm{cut}}$. It helps remove the less significant patterns, while keeping the most significant ones. The tree model may thus be greatly simplified and can be interpreted more easily. In our implementation, we set $p_{\mathrm{cut}} = 10^{-6}$ as default. For the arson case study, this roughly corresponds to $p'' = 0.25$; see Section 5.1 for the definition of p''.

3.7. *Pseudo code*. To serve as a summary, Algorithm 1 gives the pseudo code of the recursive function that we implemented for building a differential tree from two data sets.

4. A primary study of the arson case.

4.1. *Setup*. In this section we compare two approaches to solving the arson problem. Both use `label` as the response; one utilizes traditional classification trees, and the other builds a differential tree directly. For our case study the latter is more efficient at discovering differential patterns.

We divide the arson data set into two subsets, covering two time periods, 2004–2005 and 2006–2007, respectively (and reset 1/Jan/2006 to `day` $= 1$ and similarly

Algorithm 1 Differential Tree Construction (from Two Data Sets)

function difftree($D1$, $D2$)

Require: Data sets $D1$ and $D2$

Create a terminal node τ
Compute $W(\tau; (D1, D2))$ and p-value, using (3.2)
if too few observations in $D1$ and $D2$ **then**
    return τ
end if
Label τ as an internal node
for each predictor variable **do**
    Find all potential splits from its distinct values in $D1$ and $D2$
    Compute $W(T_s^\tau; (D1, D2))$ for each potential split s, using (3.5)
end for
Find the primary split s_*^τ, using (3.6) [or (3.7) in the presence of missing values]
Find all surrogate splits of s_*^τ
Use s_*^τ (and possibly surrogate splits) to partition $D1$ into $(D1_{\text{left}}, D1_{\text{right}})$ and $D2$ into $(D2_{\text{left}}, D2_{\text{right}})$
$\tau\$\text{left} = \text{difftree}(D1_{\text{left}}, D2_{\text{left}})$
$\tau\$\text{right} = \text{difftree}(D1_{\text{right}}, D2_{\text{right}})$
if τ is preferred over T^τ by (3.9) **then**
    Discard $\tau\$\text{left}$ and $\tau\$\text{right}$ and label τ as a terminal node
end if
return τ

the days after). The two subsets contain, respectively, 318 and 386 fire incidents, of which 80 and 91 are suspicious. Both suspicious and other fires are included in the study, because a maliciously-set fire is not necessarily labeled suspicious or vice versa, and because it illustrates the application of the method to a multiple category problem. In Section 7 we apply the proposed method to detect changes in the frequencies of suspicious fires only, and of fire incidents with a different categorization.

4.2. *Two classification trees.* To discover differential patterns, let us first consider building two classification trees, one from each subset, and then testing all the patterns induced from a classification tree against the other subset. The rationale is that classification trees, if constructed properly, are consistent estimators of the underlying distributions [Breiman et al. (1984), Chapter 12] and thus their differences are also consistent for estimating the true distributional differences. Note that each individual classification tree is only built to model the underlying relation between the response and explanatory variables for a *single* data set and thus

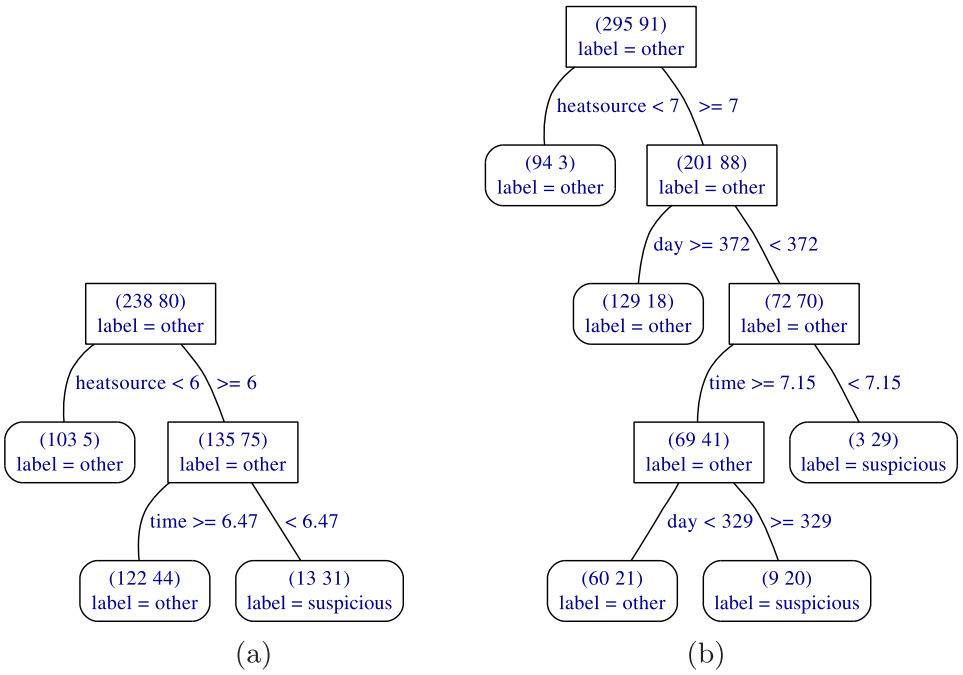

FIG. 1. *Classification trees built from fire incidents in Blenheim during*: (a) 1*/Jan/*2004–31*/Dec/*2005 *and* (b) 1*/Jan/*2006–31*/Dec/*2007. *Inside the parentheses at a node are the numbers of observations for each response level, here "other" and "suspicious."*

inevitably may include patterns that are common with the other, for example, for seasonal effects.

We use the R package "`rpart`" [Therneau and Atkinson (1997)] for classification tree construction. The classification tree built from the first subset is shown in Figure 1(a). The tree identifies three situations or patterns, as also listed in the upper part of Table 2, in ascending order of their estimated proportions of suspicious fires. To eliminate the patterns that are irrelevant to differences, we test them against the second subset, using (3.2). Hence, the remaining significant patterns can only be attributed to the distributional differences between the two subsets. After this screening, only pattern 2 remains significant, with a p-value of 1.2×10^{-3}. Nonetheless, its significance is mainly due to an increase of "other" in 2006–2007, rather than a change in the frequency of suspicious fires.

Analogously, we can find patterns from the second subset and test them against the first subset. The classification tree built from the second subset is shown in Figure 1(b). It contains five patterns, as listed in the lower part of Table 2. Pattern 7 is the most significant, with a p-value of 4.5×10^{-7}, which corresponds to a remarkable increase of 20 suspicious fires and appears to be related to the arson case. Specifically, it indicates a significant increase in the proportion of suspicious fires between day 329 (25/Nov/2006) and day 371 (6/Jan/2007), for time after

TABLE 2
Patterns obtained from each of the two classification trees and tested by their covered observations in both subsets

	2004–2005			2006–2007			
Pattern	**Other**	**Suspicious**	**Proportion**	**Other**	**Suspicious**	**Proportion**	***p*-value**
(a)		*Training set*			*Test set*		
1	103	5	0.046	90	2	0.022	3.3×10^{-1}
2	122	44	0.265	183	55	0.231	1.2×10^{-3}
3	13	31	0.705	22	34	0.607	2.9×10^{-1}
(b)		*Test set*			*Training set*		
4	105	7	0.062	94	3	0.031	3.2×10^{-1}
5	77	38	0.330	129	18	0.122	3.4×10^{-5}
6	36	17	0.321	60	21	0.259	3.9×10^{-2}
7	15	0	0.000	9	20	0.690	4.5×10^{-7}
8	5	18	0.783	3	29	0.906	2.1×10^{-1}

7:12 am, with heat source that includes cigarettes/matches/candles. It possesses a very different characteristic from pattern 8, which classifies fires as highly suspicious that occur between 0:00 am and 7:12 am, due to heat source ≥ 7. However, with a p-value of 0.21, pattern 8 does not suggest a change, although it merits further investigation by itself. Pattern 5 is also highly significant but corresponds to a decrease of $38 - 18 = 20$ suspicious fires, as well as a substantial increase of $129 - 77 = 52$ other fires; this change occurred after day 371 (6/Jan/2007).

4.3. *One differential tree*. A differential tree between the two subsets is constructed, as shown in Figure 2, which contains six terminal nodes. The most significant, with a p-value of 1.4×10^{-14}, appears to relate directly to the arson case. Specifically, it suggests that a change has occurred between day 284 (11/Oct/2006) and day 383 (18/Jan/2007), with all types but property fires. The change is due to a substantial increase of $41 - 0 = 41$ suspicious fires, as well as an increase of $43 - 22 = 21$ other fires. To gain more information about these 41 suspicious fires, the histograms/barplots for all predictor variables are shown in Figure 3. These fires are exclusively due to `heatsource` $= 7$ (=cigarettes/matches/candles), mainly of `firetype` $= 3$ (=Vegetation), largely distributed along a horizontal strip (variable `y`), and having an increasing trend over time (variable `day`).

The second most significant pattern has a p-value of 4.5×10^{-10} and specifies a situation where there is a decrease in the number of suspicious fires and yet an increase in the number of other fires. This change took place between day 420 (24/Feb/2007) and day 586 (9/Aug/2007).

Note that the general conclusions drawn here are similar to those in Section 4.2. This is not really surprising, since both methods provide consistent estimators

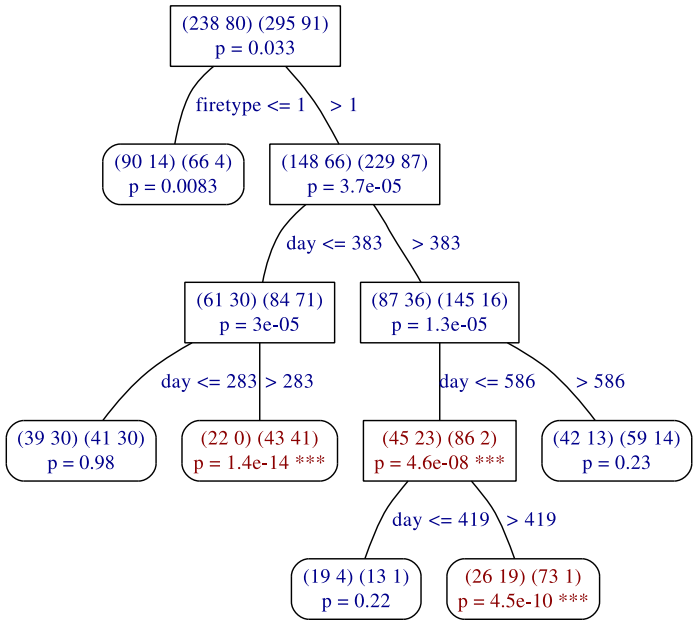

FIG. 2. *Differential tree built directly by contrasting the fire incidents from* 1*/Jan/*2004–31*/Dec/*2005 *with those from* 1*/Jan/*2006–31*/Dec/*2007. *Each pair of parentheses at a node contains the numbers of observations for all response levels in a data set, and for the arson data here* (`#other`, `#suspicious`). *Any p-value less than* 10^{-5} *is marked "***."*

for detecting differences between the two underlying distributions. However, we should also notice that the patterns found by the differential tree, that searches for changes directly by ignoring irrelevant patterns, are statistically much more signif-

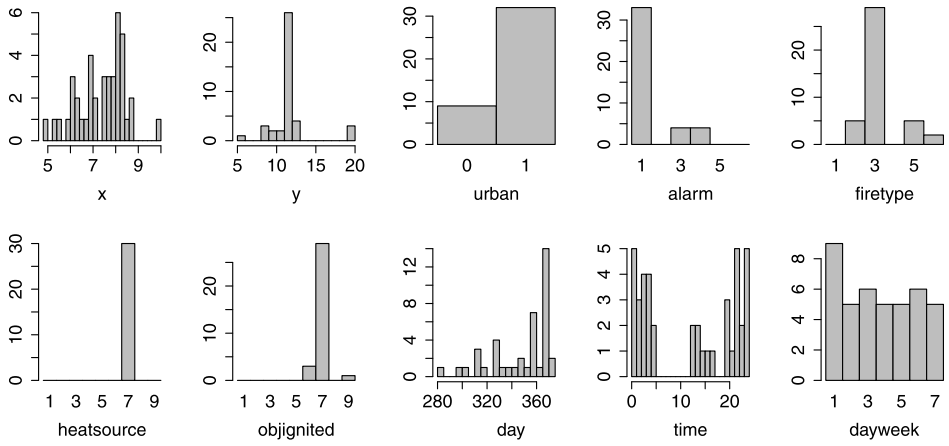

FIG. 3. *Histograms/barplots for the* 41 *suspicious fires covered by the most significant pattern.*

icant [even using the properly adjusted p-value (5.4) or (5.5)]. This suggests that differential trees are the more efficient approach to change detection. There must therefore be situations where real changes can be detected by the differential tree approach, but not by the other, and especially so when a data set contains many significant patterns not attributable to changes.

5. Performance assessment and enhancement.

5.1. *Significance adjustment*. Since "significant patterns" can always be found with an exhaustive search, one should consider their possible spuriousness. In the following, we consider adjusting p-values using the Bonferroni and the permutation method.

The Bonferroni method is the simplest and most conservative. For m tests performed, it adjusts their smallest p-value, say, p, by

$$p' = \min\{mp, 1\}. \tag{5.1}$$

For building the differential tree shown in Figure 2, there are 13,414 tests performed in total. This includes all candidate splits examined at the splitting stage, including those at the nodes that are cut off later, but not any comparisons at the pruning stage due to their irrelevance in determining the minimum p-value. Its adjusted p-value for the most significant pattern is thus

$$p' = 13{,}414 \times 1.4 \times 10^{-14} \approx 1.9 \times 10^{-10}, \tag{5.2}$$

which remains highly significant, despite the conservativeness of the method.

The permutation method adjusts a p-value by using it as a statistic and is based on the fact that, under the null hypothesis, the adjusted p-value has the uniform distribution on $[0, 1]$. The p-value to be adjusted can be either p or p' in (5.1), and, for the arson data, it does not appear to make much difference. In general, we are inclined to use p' since it guards against the situation where an extremely small p-value is produced through an exhaustive search. The empirical null distribution can be obtained by permuting either the entire data under investigation, which may nonetheless contain irregular changes and hence reduce the power of detection, or, better, some comparable, "clean" historical data. For the arson case, we choose to permute the entire data here, and later in Section 6 some historical data.

Specifically, our adjustment proceeds as follows. Each observation in the two subsets created in Section 4.1 is randomly reallocated to either the first or second biennial period by tossing a fair coin (without changing its date within a biennial period), thus ensuring the null hypothesis (3.1) is satisfied. This shuffling destroys all distributional differences between the two periods, but preserves all the relations among the variables such as geographical clusters and seasonal effects. For each pair of random subsets, a differential tree is constructed, and a minimum p-value obtained and adjusted by (5.1). With R (=1000

throughout the paper) random replications, R copies of the p'-value are obtained and ordered into $p'_{(1)} \le \cdots \le p'_{(R)}$, whose self-adjusted values are, respectively, $1/(R+1), \ldots, R/(R+1)$, namely, their expectations under the null hypothesis. Letting $p'_{(0)} = 0$ and $p'_{(R+1)} = 1$, a new p' can be adjusted by interpolation:

$$p'' = \frac{j+r}{R+1} \qquad \text{if } p'_{(j)} \le p' \le p'_{(j+1)}, \qquad j = 0, \ldots, R, \tag{5.3}$$

where $r = (p' - p'_{(j)})/(p'_{(j+1)} - p'_{(j)})$.

From the 1000 differential trees constructed, we obtained $p'_{(1)} = 8.4 \times 10^{-6}$, and, therefore, the permutation adjusted p-value for the most significant pattern in the differential tree shown in Figure 2 is

$$p'' = \frac{1.9 \times 10^{-10}/8.4 \times 10^{-6}}{1001} \approx 2.3 \times 10^{-8}. \tag{5.4}$$

This is still an extremely small p-value, indicating that it is highly unlikely that this discovered pattern occurred purely by chance.

Figure 4 shows the minimum p-values from the first 100 permutations, along with their adjustments. Despite the pure randomness of the permutations, the minimum p-values produced by differential trees are remarkably small, indicating the necessity of adjustment. What surprises us most, as can also be seen in results given later, is that p'' is almost always larger than p', because the Bonferroni adjustment is theoretically the most conservative. How could this happen? We think that the reason may lie in the fact that, conditional on the data, patterns with the smallest p-values are sought in a deterministic manner, and this has some similarity to a deterministic optimization process, which violates the underlying assumption of

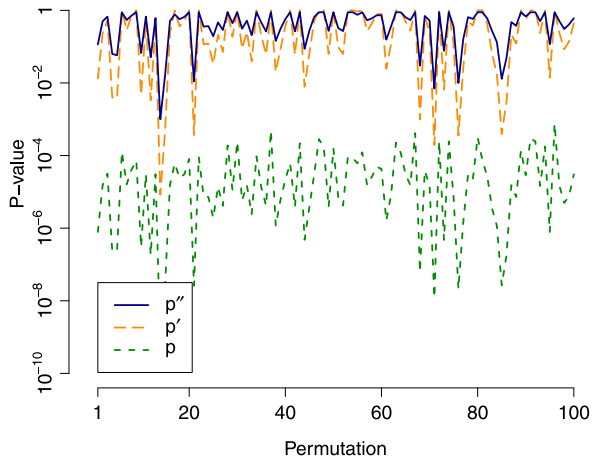

FIG. 4. *Minimum p-values and adjustments in the differential trees constructed from the first* 100, *out of* 1000, *random permutations of the* 4*-year data.*

randomness for multiple hypothesis testing. If this is true, it has profound implications for many modern statistical methods of modeling and hypothesis testing that involve extensive data manipulation to find the "best" solutions. The bias introduced by such data manipulation may be very high, so high that even the most conservative method can fail to bound it.

5.2. *Bootstrap aggregating.* One problem with tree models is instability [Breiman (1996b)], which means that a small perturbation in the data may result in a tree with a substantially different structure. In general, an unstable estimator tends to exhibit high variation and low predictive power. For differential trees, this is relevant for discovered differential patterns and their significance levels. Instability, however, can be reduced, often considerably, by using meta-learning techniques, such as boosting [Freund and Schapire (1997)], bagging (bootstrap aggregating) [Breiman (1996a)] or random forests [Breiman (2001)], which resort to building a number of models by perturbing the data. In the following we consider the bagging technique to stabilize the estimation of the minimum p-value in a differential tree.

To use bagging on the two subsets described in Section 4.1, we draw a bootstrap sample from each subset and build a differential tree from the pair of bootstrap samples, which gives a minimum p-value and its Bonferroni adjustment p'. This is repeated B (=50 throughout the paper) times. The median of the B resulting p'-values is then taken as the bagging estimate of the Bonferroni-adjusted minimum p-value. From a random run, we obtained an estimate $p' = 1.8 \times 10^{-11}$.

To find the empirical null distribution of the bagging estimator for permutation adjustment, 1000 random replications of the 4-year data were produced with random allocations to the two biennial periods, in a similar fashion to Section 5.1. The above bagging estimator is then applied to each replication. The five-number summary of the resulting p'-values is $(6.4 \times 10^{-7}, 2.7 \times 10^{-5}, 5.5 \times 10^{-5}, 1.2 \times 10^{-4}, 5.9 \times 10^{-4})$. Thus, the permutation adjusted p-value is

$$p'' = \frac{1.8 \times 10^{-11}/6.4 \times 10^{-7}}{1001} \approx 2.8 \times 10^{-8}. \tag{5.5}$$

As we shall see in Section 5.3, the bagging-based adjusted p-values are less variable and, when there exist true differences, tend to be smaller than those that are produced without using bagging.

One problem with bagging is that it does not produce one but many trees, which loses the interpretability of a single differential tree. A possible remedy is to associate a p-value with each observation, for example, using the median p''-value of all patterns that cover the observation. Then we know which observations are associated with changes, and how significantly. Areas containing observations with small p-values can perhaps be derived subsequently.

5.3. *A simulation study.* In order to gauge the efficiency and stability of the differential tree method, we conducted a simulation study and made use of the arson data in a way that mimicked the arson case. To produce random data for two biennial periods, all 318 (238 other and 80 suspicious) fire incidents in 2004–2005 are duplicated once and then randomly reallocated to either the first or second biennial period by coin tossing (as in Section 5.1). We did not include the data in 2006–2007 to avoid contamination. Then we added $n_\Delta \in \{0, 10, \ldots, 50\}$ distinctive fire incidents to the second biennial period, randomly drawn from those 2006–2007 incidents covered by the most significant pattern discovered in Section 4.3, in the proportions of 30% other and 70% suspicious fires. For each $n_\Delta \in \{0, 10, \ldots, 50\}$, 100 such data sets were generated, and thus 100 (without bagging) and 100×50 (with bagging) differential trees were built. To adjust p-values, 1000 permutations were carried out both with and without bagging, thus producing $1000 + 1000 \times 50$ differential trees. A total of 81,600 differential trees were built in the simulation study.

Figure 5 shows summaries of the p-values (p or p'') produced by three methods: a direct evaluation using (3.2) without building any differential tree (which gives the same p-value as the root node of a differential tree), building one differential tree, and building differential trees with bagging. The central 50% interval of the empirical distribution of the p-value is plotted for each case. It can be seen that, when $n_\Delta = 0$, all three p-values appear to conform well with the uniform distribution on $[0, 1]$. We can use the medians of these p-values to gauge efficiency and the widths of the central 50% intervals to gauge stability. As n_Δ increases,

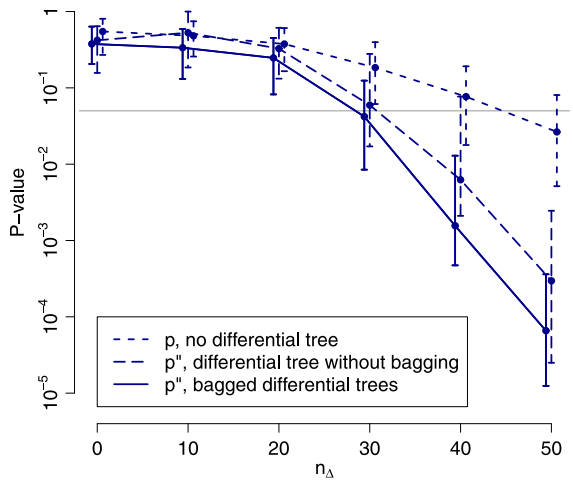

FIG. 5. *Each vertical line segment represents the central* 50% *interval of an empirical distribution obtained from* 100 p- *or* p''*-values, and a solid point the median. Some line segments are slightly shifted horizontally for distinguishing purposes. The horizontal line is where* p*-value* $= 0.05$.

each p-value decreases, and at an increasing rate. The directly evaluated p, however, decreases slowly and this approach, on average, is unable to detect the change at the 5% significance level until $n_\Delta \approx 44$. An intuitive explanation is that a dramatic change deep under the surface may only manifest as ripples on the surface, that is, at the root node. Also, multiple changes may even cancel out the effects of one another and leave no trace on the surface, as is the case of the differential tree shown in Figure 2. By contrast, with differential trees it "dives" down and seeks differences between the data sets in increasingly smaller areas. The method can thus uncover local differences more efficiently and, for the arson problem, is able to start detecting the change for $n_\Delta \approx 31$ (without bagging) and ≈ 27 (with bagging). As n_Δ increases, there are also clearly widening gaps between the p-values produced by the direct evaluation method and the differential tree methods. For $n_\Delta = 50$, the median p is only 2.6×10^{-2}, being barely significant, while the median p'' is 3.0×10^{-4} (without bagging) or 6.6×10^{-5} (with bagging). It is also clear that the bagging technique helped reduce instability and increase efficiency. The arson case has n_Δ close to 60, which we could not include in the simulation study since it requires 42 suspicious fires but the most significant pattern has only 41. However, with a visual extrapolation of the curves in Figure 5 to where $n_\Delta = 60$, it should be clear that the proposed method is quite effective for discovering the changes in the arson case.

6. Sequential detection. The method developed above can also be used in a sequential detection manner. Let us consider comparing the data of the two consecutive annual periods immediately before a "detection" day (the first day after the two year period). With the quadrennial data available, we start the detection from 1/Jan/2006, by building a differential tree that compares the two time periods, 1/Jan/2004–31/Dec/2004 and 1/Jan/2005–31/Dec/2005, and build new differential trees by shifting the detection day at intervals of seven days, until all data have been examined. From every tree constructed the smallest p-value is extracted and adjusted by the Bonferroni and permutation methods, using (5.1) and (5.3). The empirical null distribution of the minimum p-value in a differential tree that is needed by the permutation adjustment is obtained by permuting 1000 times the historical fire incidents that occurred during 1/Jan/2004–31/Dec/2005. We have also produced an empirical null distribution by permuting random halves of all the quadrennial data and found that the resulting adjusted p-values are only slightly larger, due to the contamination of the irregular changes in the latter two years. The conclusions, however, remain largely the same. To use bagging, one only needs to replace each single differential tree described above with 50 trees obtained under bootstrap sampling (Section 5.2).

The results are shown in Figure 6. The sequential detection results with bagging shown in Figure 6(b) are clearly more stable than those without bagging in Figure 6(a). From Figure 6(a), after the initial 45 weeks with basically no significant change discovered and a smallest p''-value of 0.025, a sudden decrease

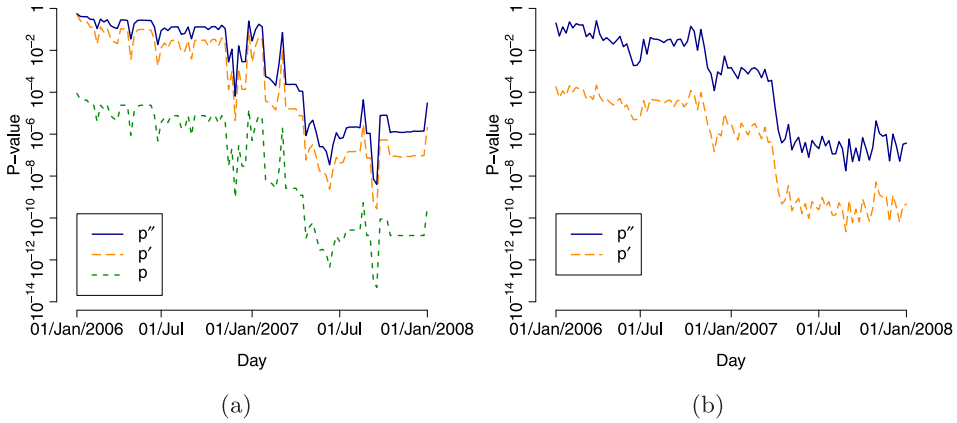

FIG. 6. *Minimum p-values and their adjustments in sequential detection*: (a) *without bagging*; (b) *with bagging*. *In particular, p is the minimum p-value in a differential tree, p' the Bonferroni adjustment of p, and p'' the permutation adjustment of p'.*

of the p''-value occurred on detection day 316 (12/Nov/2006) with $p'' = 0.0040$. This is clearly a sign that some significant change(s) have occurred in the underlying data-generating mechanism. Similar conclusions can be drawn from the more stable estimates in Figure 6(b).

By monitoring the change of (adjusted) p-values, it is straightforward for an online system to set up different levels of warning in an easily comprehensible sense.

7. Using different responses.

7.1. *Using only suspicious fires*. Instead of using both suspicious and other fires as done in the study so far, we can use suspicious fires only. Figure 7 displays the differential tree, built analogously to that in Figure 2, from the two biennial subsets. Interestingly, the two most significant patterns are comparable in both trees: one concerning a substantial increase of suspicious fires during almost the same time period and the other a decrease of suspicious fires after it. Note that one cannot use the classification tree approach here, since the response variable has only one level.

The minimum p-values and their adjustments for sequential detection are plotted in Figure 8. The sudden change has also been successfully detected, but at a delayed date, as compared with that in Figure 6. This is because most suspicious fires occurred in the latter part of the biennial period (see the histogram of `day` in Figure 3), and because in the earlier part of the time period there is an increase of fire incidents that are not labeled "suspicious," which are thus excluded from the study here. In this case, including all fire incidents is preferable—it gives an earlier warning!

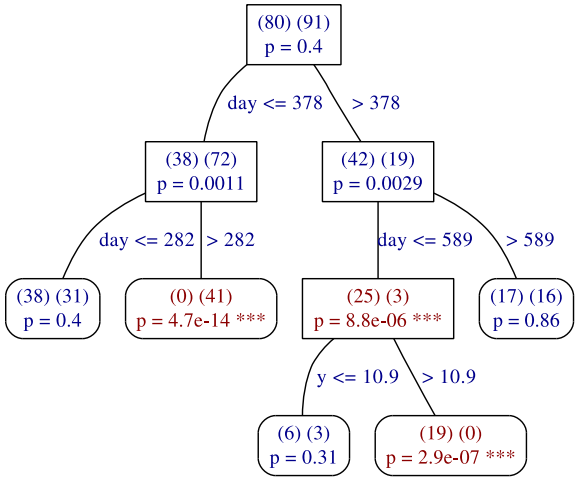

FIG. 7. *Differential tree built from using suspicious fires only.*

7.2. *Using an alternative response variable.* One can also use a different response variable, as if for a general surveillance, in total ignorance of what has happened. Let us this time treat the variable `firetype` as the response. The differential tree built from the two biennial subsets is shown in Figure 9. This tree appears to be less informative and its most significant pattern is also less significant, as compared with the trees shown in Figures 2 and 7. However, this discovered pattern is still remarkably significant, showing that the difference is mainly due to an increase of vegetation fires, jumping from 45 cases to 121 for `day` > 86 and $x > 5.9$.

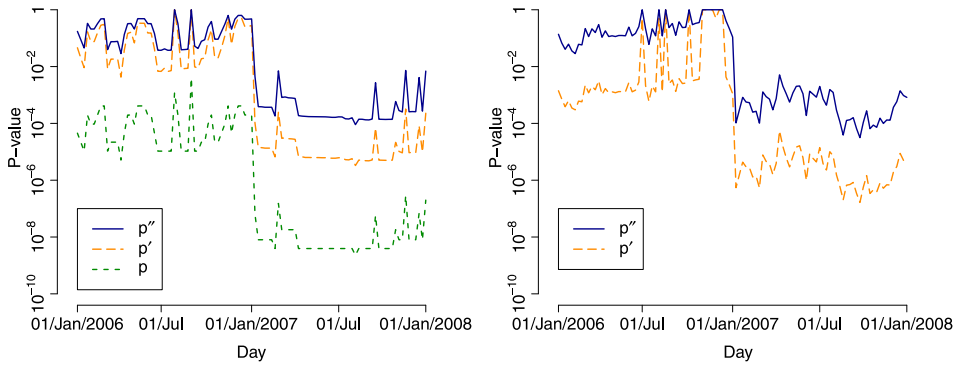

FIG. 8. *p-values in sequential detection using suspicious fires only*: (a) *without bagging*; (b) *with bagging*.

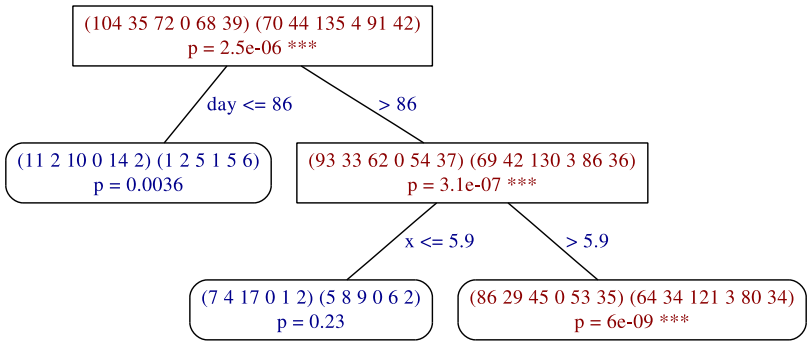

FIG. 9. *Differential tree built with a different response variable.*

The minimum p-values and their adjustments obtained via sequential detection are shown in Figure 10. It is clear that the change has also been detected, at a later date and less dramatically than that in Section 6.

These results are perhaps the most one could hope for when conducting a general surveillance.

8. Concluding remarks. There are two main new ideas in our proposed method for change or difference detection. One is to contrast data sets and model their distributional differences and the other to use tree models to uncover local, irregular changes and provide interpretable results. We followed the general methodology for tree construction. Variants with improved performance likely exist, as in the literature for other families of the tree model. Extensions to other types of difference detection seem fairly straightforward.

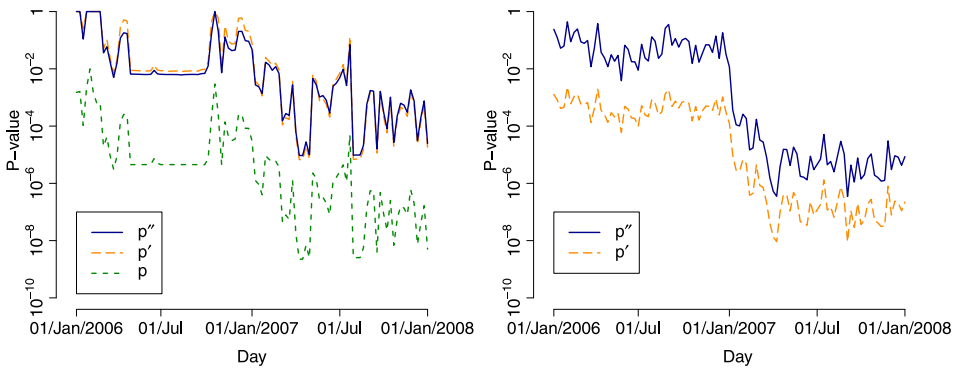

FIG. 10. *p-values in a sequential detection using* `firetype` *as response*: (a) *without bagging*; (b) *with bagging*.

Building a differential tree is reasonably fast. With our implementation in R [R Development Core Team (2011)], it took, respectively, 5.0, 1.4 and 6.8 seconds to build the trees shown in Figures 2, 7 and 9, on a workstation with a 2.93 GHz CPU. This made possible the demanding numerical studies reported earlier. If implemented in FORTRAN or C, it is likely much faster.

Finally, we give a rationale for using differential trees in a complex environment. A general alternative is to compare the data with a reference model that can be either exactly known, which is virtually impossible in a complex environment, or estimated from a reference data set, just as we did in Section 4.2. Since building a model from one data set and testing it against the other can waste data information on discovering patterns irrelevant to differences and we are essentially comparing two data sets, why do not we just build one model that directly describes their differences? This is exactly what a differential tree does.

Acknowledgments. We are grateful to the Editor, Associate Editor and two reviewers for their helpful and constructive comments that led to many improvements in the manuscript.

SUPPLEMENTARY MATERIAL

Supplement A: Data (DOI: 10.1214/12-AOAS548SUPPA; .csv). The file `arson.csv` contains the Arson data that is described in Section 1 and used in the analysis throughout the paper. The meanings of the variables and the values they take on are available in Table 1.

Supplement B: Software (DOI: 10.1214/12-AOAS548SUPPB; .R). The file `difftree.R` contains the R code for carrying out the analysis in the paper.

REFERENCES

BASSEVILLE, M. and NIKIFOROV, I. V. (1993). *Detection of Abrupt Changes*: *Theory and Application*. Prentice Hall, Englewood Cliffs, NJ. MR1210954

BREIMAN, L. (1996a). Bagging predictors. *Machine Learning* **24** 123–140.

BREIMAN, L. (1996b). Heuristics of instability and stabilization in model selection. *Ann. Statist.* **24** 2350–2383. MR1425957

BREIMAN, L. (2001). Random forests. *Machine Learning* **45** 5–32.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, CA. MR0726392

CHAUDHURI, P., LO, W. D., LOH, W.-Y. and YANG, C. C. (1995). Generalized regression trees. *Statist. Sinica* **5** 641–666. MR1347613

DAVIS, R. B. and ANDERSON, J. R. (1989). Exponential survival trees. *Stat. Med.* **8** 947–961.

FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. MR1473055

GLAZ, J., NAUS, J. and WALLENSTEIN, S. (2001). *Scan Statistics*. Springer, New York. MR1869112

GUSTAFSSON, F. (2000). *Adaptive Filtering and Change Detection*. Wiley, Chichester, UK.

ISHWARAN, H., KOGALUR, U. B., BLACKSTONE, E. H. and LAUER, M. S. (2008). Random survival forests. *Ann. Appl. Stat.* **2** 841–860. MR2516796

KASS, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *J. Appl. Stat.* **29** 119–127.

LAI, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B* **57** 613–658. MR1354072

MACEACHERN, S. N., RAO, Y. and WU, C. (2007). A robust-likelihood cumulative sum chart. *J. Amer. Statist. Assoc.* **102** 1440–1447. MR2412559

MORGAN, J. N. and SONQUIST, J. A. (1963). Problems in the analysis of survey data, and a proposal. *J. Amer. Statist. Assoc.* **58** 415–434.

NAUS, J. I. (1965). The distribution of the size of the maximum cluster of points on a line. *J. Amer. Statist. Assoc.* **60** 532–538. MR0183041

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850

POOR, H. V. and HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge Univ. Press, Cambridge. MR2482527

QUINLAN, J. R. (1993). *C*4.5: *Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.

R DEVELOPMENT CORE TEAM (2011). *R*: *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

SHEWHART, W. A. (1931). *Economic Control of Manufactured Products*. Van Nostrand-Reinhold, New York.

SU, X., WANG, M. and FAN, J. (2004). Maximum likelihood regression trees. *J. Comput. Graph. Statist.* **13** 586–598. MR2087716

THERNEAU, T. M. and ATKINSON, E. J. (1997). An introduction to recursive partitioning using the `rpart` routine. Technical Report 61, Section of Biostatistics, Mayo Clinic, Rochester, NY.

WANG, Y., ZIEDINS, I., HOLMES, M. and CHALLANDS, N. (2012). Supplement to "Tree models for difference and change detection in a complex environment". DOI:10.1214/12-AOAS548SUPPA, DOI:10.1214/12-AOAS548SUPPB.

Y. WANG
I. ZIEDINS
M. HOLMES
DEPARTMENT OF STATISTICS
UNIVERSITY OF AUCKLAND
PRIVATE BAG 92019, AUCKLAND
NEW ZEALAND
E-MAIL: yongwang@auckland.ac.nz
i.ziedins@auckland.ac.nz
m.holmes@auckland.ac.nz

N. CHALLANDS
NEW ZEALAND FIRE SERVICE
P.O. BOX 2133, WELLINGTON
NEW ZEALAND
E-MAIL: Neil.Challands@fire.org.nz