Supplementary material for "Hierarchical spatio-temporal change-point detection"

Mehdi Moradi

Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

E-mail: mehdi.moradi@umu.se

Ottmar Cronie

Department of Mathematical Sciences, Chalmers University of Technology & University of Gothenburg, Gothenburg, Sweden.

Unai Pérez-Goya

Department of Statistics, Computer Science, and Mathematics, Public University of Navarre, Pamplona, Spain

Jorge Mateu

Department of Mathematics, University Jaume I, Castellón, Spain.

Abstract

The following supplementary material document contains additional background to the manuscript, including some preliminaries on multivariate change-point detection, and hierarchical clustering with details regarding linkage functions, and methods to estimate optimal clusters for spatially dependent data in Section S1. Additional simulation studies can be found in Section S2 which covers various scenarios concerning changes in the mean, the variance, multiple changes, and scalability of HSTCPD. Further descriptions and results about our real datasets (i.e. LST in Spain from February 2000 to November 2021, and Afghan War Diary (AWD) data which runs monthly from January 2004 to December 2009) are also presented in Section S3.

S1 Extra background

This section is devoted to providing some technical preliminaries (Section S1.1), examples of image time series over which it could be of interest to perform change-point analysis (Section S1.2), and hierarchical cluster analysis (Section S1.3) together with a detailed background on different linkage functions (Section S1.3.1), ways to find an optimal number of clusters (Section S1.3.2), and spatially dependent functional clustering (Section S1.3.3).

S1.1 Preliminaries

Throughout, let T, with $T_0 = \inf T$ and $T_\infty = \sup T$, be a time domain, given either by an interval $T = [T_0, T_\infty]$ in the Euclidean space $\mathbb{R}$ or a discrete collection of times, i.e. $T = \{T_0, \ldots, T_\infty\}$. We further denote the Euclidean norm on $\mathbb{R}^d$, $d \geq 1$, by $|\cdot|$.

We here consider a set of locations, $z_1, \ldots, z_n \in W \subset \mathbb{R}^d$, $n \geq 1$, $d \geq 1$, which typically represent a set of monitoring places/sites. Next, given a time interval T, we attach a multivariate random function/process $X(t) = (X_1(t), \ldots, X_n(t))$, $t \in T$, to these spatial locations. More specifically, we assume that X is a random element in the product space $L^2(T)^n$, where the Hilbert space

$$L^2(T) = \{f : T \to \mathbb{R} \text{ such that } \|f\| = \left(\int_T f^2(t)\mathrm{d}t\right)^{1/2} < \infty\},$$

has inner product $\langle f, g\rangle = \int_T f(t)g(t)\mathrm{d}t$, and metric

$$d_{L^2}(f, g) = \|f - g\| = \langle f - g, f - g\rangle^{1/2} = \sqrt{\int_T (f(t) - g(t))^2 \mathrm{d}t} \tag{S1}$$

for $f, g \in L^2(T)$ (Ramsey and Silverman, 2005).

S1.1.1 Multivariate change-point detection

Given an n-dimensional stochastic process $X(t) = (X_1(t), \ldots, X_n(t))$, $t \in T$, either in continuous or discrete time, let $\Theta_i(t) = (\theta_{i,1}(t), \ldots, \theta_{i,q}(t))$, $t \in T$, $1 \leq i \leq n$, correspond to $q \geq 1$ characteristics/parameters of X_i such as mean, variance, etc. In multivariate change-point detection the main interest lies in alerting when, for some i, $\Theta_i(t)$ undergoes changes (Aston

and Kirch, 2012; Matteson and James, 2014; Liu et al., 2020; Grundy et al., 2020). Formally speaking, to discover change-points $\{\tau_1, \ldots, \tau_k\} \subset T$, for each $i = 1, \ldots, n$ one deals with the hypotheses

$$\begin{aligned} \mathcal{H}_0 \quad &: \quad \Theta_i(t) = \Theta(t) = \Theta_0, \quad \text{for any} \quad t \in T,\, i = 1, \ldots, n, \\ \mathcal{H}_1 \quad &: \quad \text{We can find } i \in \{1, \ldots, n\} \text{ s.t. } \exists \{\tau_1', \ldots, \tau_{k'}'\} \subset \{\tau_1, \ldots, \tau_k\} \text{ where } \Theta_i(t) = \Theta_i^j(t), \\ &\qquad \tau_{j-1}' < t \leq \tau_j', \Theta_i^j \neq \Theta_i^{j+1}, \text{ for all } j = 1, \ldots, k'+1, \end{aligned} \tag{S2}$$

where $\tau_0' = T_0$, $\tau_{k'+1}' = T_\infty$, and Θ_i^j is a q-dimensional function on T. In other words, the alternative hypothesis $\mathcal{H}_1$ claims different behaviours for $\Theta_i(t)$ for any two consecutive time periods, separated by τ_j', $1 \leq j \leq k'$. If $k = 1$ the problem reduces to the detection of at most one change (AMOC). Note that, in contrast to the classical setting, we do not assume that the changes occur jointly across all marginals X_i. When dealing with multivariate data, it is rarely the case that all X_i's face changes of the same type or undergo changes at the same time. Moreover, nearby locations usually tend to experience similar changes in terms of time but possibly of different magnitudes, which might be a sign of spatial dependence.

A frequently used non-parametric change-point detection approach for multivariate time-series $X(t) = (X_1(t), \ldots, X_n(t))$, $t \in T$, i.e. multivariate discrete time stochastic processes, is that of Matteson and James (2014). They proposed to detect the number of changes and their positions through a multivariate divergence measure and then hierarchically estimate change-point time indices. In the case of AMOC, the divergence measure used by Matteson and James (2014) is of the form

$$\mathcal{E}_\alpha(X, s) \propto \sum_{T_0 \leq i \leq s} \sum_{s < j \leq T_\infty} |X(i) - X(j)|^\alpha - \sum_{T_0 \leq i < k \leq s} |X(i) - X(k)|^\alpha - \sum_{s < j < k \leq T_\infty} |X(j) - X(k)|^\alpha,$$

where $\alpha \in (0, 2)$. They label the time $\widehat{\tau} \in T$, which maximises the above divergence measure, a change-point, i.e.,

$$\widehat{\tau} = \arg\max_{s \in T} \; \mathcal{E}_\alpha(X, s).$$

Thus, if the parameter under consideration was e.g. the mean function $\mu(\cdot)$, under the

assumption of AMOC, one may rewrite an estimator of a discretely sampled version of (1) as

$$X_i(t) = \mu_i^1(t)\mathbf{1}\{t \in [T_0, \widehat{\tau}]\} + \mu_i^2(t)\mathbf{1}\{t \in (\widehat{\tau}, T_\infty]\} + \epsilon_i(t), \qquad i = 1, \ldots, n,$$

where $\mathbf{1}\{\cdot\}$ is an indicator function and $\mu_i^1(\cdot) \neq \mu_i^2(\cdot)$. Matteson and James (2014) further proposed to iteratively apply the above technique to search for multiple change-points. The statistical significance of such change-points is, at a later step, checked through a permutation test which gives rise to approximate p-values.

S1.2 Examples of image time series

Below we give two examples where one deals with image time series.

Example 1 (Time series of satellite images) *Such a time series constitutes a set of time-ordered satellite images (taken from a fixed region) of some quantity of interest, e.g. Land Surface Temperature (LST) or Normalised Difference Vegetation Index (NDVI), where the spatial locations $z_1, \ldots, z_n$ are (regularly dispersed) pixels. One may here want to monitor the temporal distributional behaviour and detect significant changes in the quantity of interest (Militino et al., 2020; Moradi et al., 2022).*

Example 2 (Time series of point patterns) *One may be interested in monitoring time-ordered point patterns, i.e. realisations of point processes, that correspond to locations of events or objects of interest, which are distributed over a fixed region (Baddeley et al., 2015). A typical example of a quantity of interest, which we may want to monitor here, is the estimated intensity surface; when and where was there a significant change in the intensity of e.g. crimes? Note that $z_1, \ldots, z_n$ represent a discretisation of the spatial domain (Chaudhuri et al., 2021).*

S1.3 Hierarchical cluster analysis

Cluster analysis is usually carried out to discover groups of observations as homogeneous as possible, based on a pre-selected similarity measure (distance metric). The output of such an analysis is a collection of groups of observations such that pairwise dissimilarities between

observations within the same group are smaller than dissimilarities between different groups. Among clustering approaches, hierarchical clustering, which is quite frequently used in practice, refers to labelling observations by cluster IDs through hierarchical representations. In order to perform such clustering, one can follow: i) a bottom-up strategy (agglomerative), which considers each observation as a single cluster and then starts to merge them recursively at each level of the hierarchy until the highest level that contains a single cluster with all observations, or ii) a top-down procedure (divisive) which puts all observations in a single cluster and then, at each level of the hierarchy, starts to split one of the existing clusters into two clusters. In agglomerative approaches, which is what we will consider in this paper, the idea is to merge (based on some rule) the two most similar clusters at each level of the hierarchy. We next go through some commonly used linkage functions, which govern how the merging of clusters occurs.

S1.3.1 Linkage functions

In the lowest level of the hierarchy, merging begins with putting the two closest (most similar) observations in a single cluster. Therefore, and since in the next level one deals with merging clusters of one observation and a single cluster of two observations, a measure of dissimilarity should be specified to quantify the dissimilarity between any of the clusters of size one and that of size two. In the literature, various *linkage* functions to measure such dissimilarities have been proposed, and we next briefly look at a few such proposals.

Let $\mathcal{G}$ and $\mathcal{H}$ be two arbitrary clusters, at any level of the hierarchy. We use the notation $\mathfrak{d}(i, j)$ to denote the dissimilarity between two observations i and j. Three frequently encountered linkage functions are:

- *Single*: this linkage function is also known as the nearest-neighbour technique, and considers the minimum distance between the members of $\mathcal{G}$ and $\mathcal{H}$ when merging clusters. More specifically, the dissimilarity between clusters $\mathcal{G}$ and $\mathcal{H}$ is given by

$$\mathfrak{d}^S(\mathcal{G}, \mathcal{H}) = \min\{\mathfrak{d}(i, j) : \ i \in \mathcal{G} \text{ and } j \in \mathcal{H}\},$$

  which means that the only requirement to merge two clusters is to have at least two very close/similar members.

- *Complete*: this linkage function is also called furthest-neighbour technique, and considers the maximum distance between the members of $\mathcal{G}$ and $\mathcal{H}$ when merging clusters. Hence, the dissimilarity between clusters $\mathcal{G}$ and $\mathcal{H}$ is given by

$$\mathfrak{d}^C(\mathcal{G},\mathcal{H}) = \max\{\mathfrak{d}(i,j) : \ i \in \mathcal{G} \text{ and } j \in \mathcal{H}\},$$

  which means that two clusters can be merged if all pairwise dissimilarities between their members are small. The clusters generated by complete linkage would typically be more dense than those obtained by single linkage.

- *Ward*: this linkage function aims to minimise the within-cluster dispersion at each level of the hierarchy, based on a sum-of-squares criterion that, in turn, deals with a more complicated merging process. Since the first definition by Ward (1963), there have been a few modifications for this linkage function; see Murtagh and Legendre (2014) for details. Throughout we consider the version known as *Ward.D2*.

S1.3.2 Optimal number of clusters

At this stage, the remaining task is to decide what level of the hierarchy should be considered as the optimal classification. This can be achieved through a dendrogram, which provides a way to display the clusters at all levels of the hierarchy. Letting $\mathcal{C}_{h,r}$, $1 < h \leq r < n$, denote the obtained clusters at the level r of the dendrogram hierarchy, which forms an r-sized partition of $\{1,\ldots,n\}$, the task is to find an optimal value for r, which is done by maximising an index $\alpha(r)$, $1 < h \leq r < n$. Below we briefly review three commonly used indices to find the optimal (number of) clusters:

- *Calinski and Harabasz (Ch)* : Caliński and Harabasz (1974) proposed to select the optimal number of clusters by minimising the within-cluster dispersion and maximising the between-clusters dispersion simultaneously. According to their proposal, the level of hierarchy that maximises

$$\alpha(r) = \mathrm{ch}(r) = \frac{n-r}{r-1}\left(\frac{1/n\sum_{i,j=1}^{n}\mathfrak{d}(i,j)^2}{\sum_{h=1}^{r}(1/\#\mathcal{C}_{h,r})\sum_{i,j\in\mathcal{C}_{h,r}}\mathfrak{d}(i,j)^2} - 1\right),$$

where $\#\mathcal{C}_{h,r}$ denotes the cardinality of $\mathcal{C}_{h,r}$, leads to an optimal number of clusters (Hennig and Liao, 2013).

- *Dunn*: Dunn (1974) introduced the index

$$\alpha(r) = \mathrm{d}(r) = \min_{i=1,\dots,r-1} \left\{ \min_{j=i+1,\dots,r} \left\{ \frac{\mathfrak{d}^S(\mathcal{C}_{i,r}, \mathcal{C}_{j,r})}{\max_{h=1,\dots,r} \mathrm{diam}(\mathcal{C}_{h,r})} \right\} \right\},$$

where $\mathrm{diam}(\mathcal{C}_{h,r}) = \max_{i,j \in \mathcal{C}_{h,r}} \mathfrak{d}(i,j)$. Since the interest is to maximise the similarities (dissimilarities) within (between) clusters, $\arg\max_r \mathrm{d}(r)$ provides an estimate of the optimal number of clusters (Halkidi et al., 2001).

- *Dunn2*: this is another version of the *Dunn* index, which instead suggests to select a level of hierarchy that maximises

$$\alpha(r) = \mathrm{d}^2(r) = \frac{\min_{1 \le i < j \le r} \left\{ (\#\mathcal{C}_{i,r} \#\mathcal{C}_{j,r})^{-1} \sum_{i' \in \mathcal{C}_{i,r}} \sum_{j' \in \mathcal{C}_{j,r}} \mathfrak{d}(i', j') \right\}}{\max_{1 \le h \le r} \left\{ 2(\#\mathcal{C}_{h,r}(\#\mathcal{C}_{h,r} - 1))^{-1} \sum_{i \in \mathcal{C}_{h,r}} \sum_{j \in \mathcal{C}_{h,r} \setminus \{i\}} \mathfrak{d}(i,j) \right\}}.$$

This is a ratio of minimum average dissimilarities between pairs of clusters and maximum average dissimilarities within clusters.

S1.3.3 Spatially dependent functional clustering

In order to find similar groups within a set of functional data, $\{f_i\}_{i=1}^n$, we need a metric $d(\cdot,\cdot)$ which allows us to measure distances $\mathfrak{d}(i,j) = d(f_i, f_j)$ between two functions f_i, f_j.

In the classical functional data analysis setting, where $\{f_i\}_{i=1}^n$ is treated as a realisation of an independent and identically distributed (iid) sample of random functions, a natural choice for $d(\cdot,\cdot)$ is the metric $d_{L^2}(\cdot,\cdot)$ found in (S1). However, when the sample of random functions exhibits spatial dependence, (S1) could become problematic for the purpose of describing similarities between the underlying random functions since closeness between two functions may be the result of i) similarity between e.g. their means and/or ii) spatial dependence between them, which forces nearby functions to resemble each other.

The literature offers only a few proposals for measuring distances between functions while incorporating potential underlying dependencies. In particular, Giraldo et al. (2012)

proposed to use the weighted distances

$$d(f_i, f_j) = d_{L^2}(f_i, f_j)\widehat{\gamma}(h),$$

where $\gamma(h)$ is an estimate of the so-called trace-variogram. More specifically, we assume that f_i, $i = 1, \ldots, n$, are realisations of second-order stationary and isotropic processes $X_i(t)$, $t \in T$, $i = 1, \ldots, n$, with trace-variogram function

$$\gamma(h) = \gamma(\|z_i - z_j\|) = \frac{1}{2}\mathbb{E}\left[\int_T (X_i(t) - X_j(t))^2 \mathrm{d}t\right], \tag{S3}$$

where z_i, z_j denote the locations corresponding to $X_i(\cdot)$ and $X_j(\cdot)$, respectively. The trace-variogram is estimated by means of

$$\widehat{\gamma}(h) = \frac{1}{2\#N(h)} \sum_{(i,j)\in N(h)} \int_T (X_i(t) - X_j(t))^2 \mathrm{d}t, \tag{S4}$$

where $N(h)$ consists of all distinct pairs (i, j) with a spatial distance of exactly h units. In situations where $N(h)$ is empty or does not include enough observations, the exact distance h would be replaced by an interval $h \pm \epsilon$, $\epsilon > 0$. Although we here assume second-order stationarity, we have seen that, in comparison to $d_{L^2}(\cdot, \cdot)$, the proposed metric works well also when this assumption may not hold. In practice, when we sample $X_i(t)$, $t \in T$, $i = 1, \ldots, n$, over a set of discrete times, thus yielding a multivariate time series, the integrals in (S1) and (S4) are approximated by sums.

S2 Numerical Evaluation

This section presents additional parts of the results for scenario I together with the entire results of scenarios II, III, IV, V, and VI. Before turning to further details we emphasise that, throughout the paper, the considered spatial random field is Gaussian with mean 0 and covariance function

$$((x_1, y_1), (x_2, y_2)) \mapsto 10e^{-(||(x_1,y_1)-(x_2,y_2)||/5)^2}, \qquad (x_1, y_1), (x_2, y_2) \in \mathbb{R}^2. \tag{S5}$$

Concerning scenarios I, II, and III, the results of the approach of Matteson and James (2014) (ecp) are given in Section S2.1, S2.2, and S2.3, respectively, while all the results for the adaptive method of Liu et al. (2020) (AdaptiveCpt) and geometrical mapping approach of Grundy et al. (2020) (geomcp) are presented in Section S2.4 and S2.5, respectively. The results of scenario IV, which focuses on multiple changes in the mean per each cloud, are presented in Section S2.6, and our findings corresponding to scenario V, which deals with the case where there are changes in both the mean and the variance in each cloud, are given in Section S2.7. We further present our findings within scenario VI, which concerns the scalability of HSTCPD, in Section S2.8.

S2.1 Scenario I

S2.1.1 Spatially correlated time series of images

Figure S1 and S2 show the proportion of the estimated change-points when the actual change-point for cl_2 occurs at time 90 and 120, respectively. One can easily see that, in all cases regardless of the clustering approach, HSTCPD clearly outperforms the Classic approach while providing results quite similar to the Conditional. Additionally, we can see that increasing the time difference between the two change-points in cl_1 and cl_2 improves the performance of both Classic and HSTCPD; compare Figure S1 and S2 with Figure 3 in the main paper. We add that such an improvement for HSTCPD is quite minor, but having large time difference between the two actual change-points can slightly facilitate the detection of optimal clusters which in turn may lead to a better detection of actual change-points. However, increasing the time difference between the two change-points has a big effect on the performance of the Classic approach by enhancing its detection rate. Note that, the Classic approach can still not discover the marginals which undergo changes.

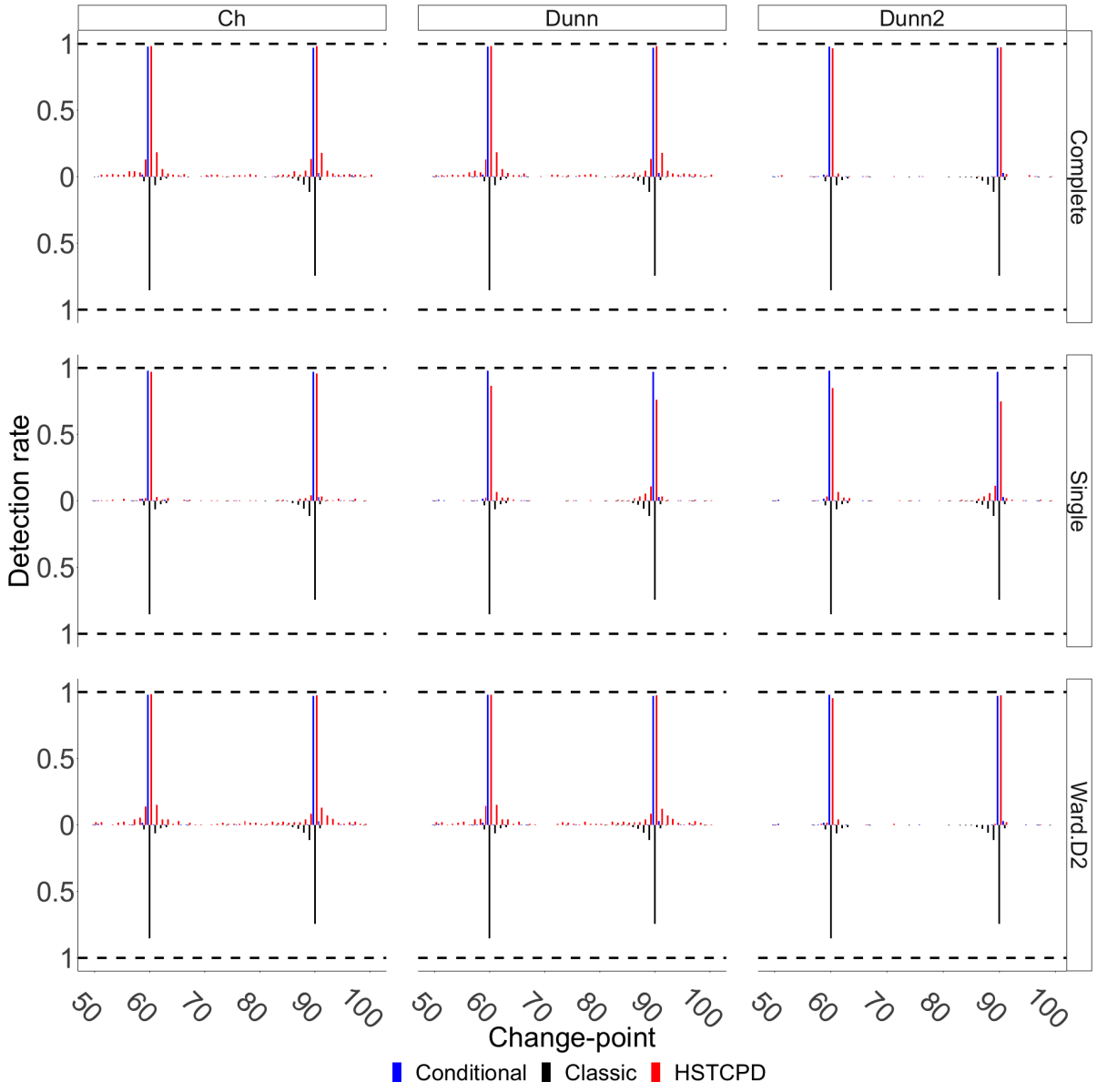

Figure S1: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

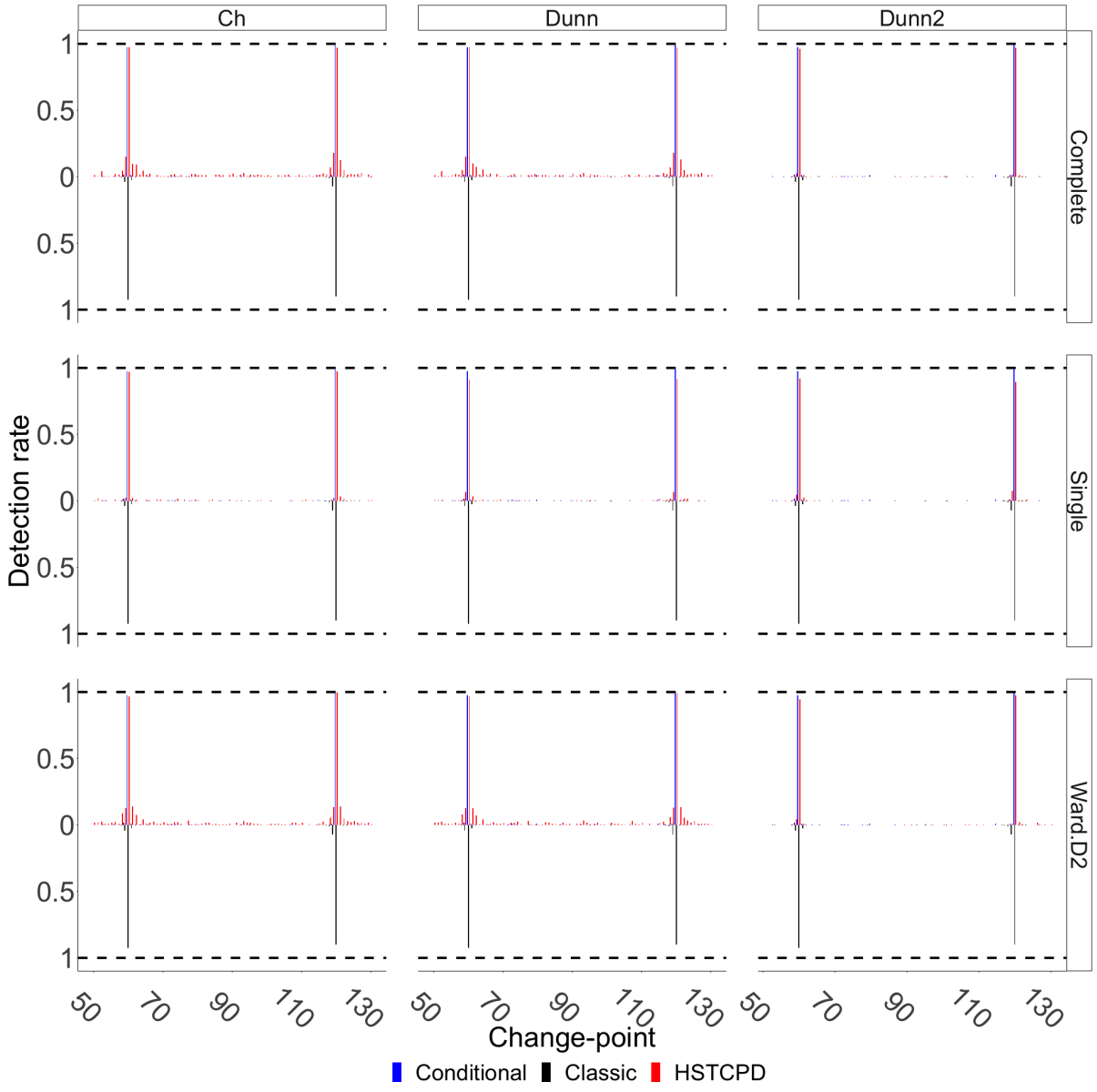

Figure S2: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.1.2 Independent pixel time series coming from the standard normal distribution

From Table S1 and Figure S3 one can see that, in this scenario, the joint performance of Dunn approach and any of the linkage functions seems to be the best by leading to properly disclosing the borders of two clouds as both **RI** and **SI** are nearly one. Other combinations of linkage functions and methods to discover optimal clusters could not manage to disclose the borders of the two clouds properly; one cloud is lost within the part of image with no change.

Table S1: Scenario I with independent pixel time series. An upward change with magnitude one happens at time 60 in cl_1, whereas the time index of a downward change of the same magnitude, happening in cl_2, varies between 70 and 120. Average Rand index **RI** (separation index **SI**) are reported.

Linkage	**Optimal**	**2nd change-point**					
		70	**80**	**90**	**100**	**110**	**120**
Single	**Ch**	0.97 (1.00)	0.93 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
	Dunn	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.99 (1.00)	0.98 (1.00)	0.96 (1.00)
	Dunn2	0.93 (0.99)	0.91 (0.98)	0.90 (0.97)	0.90 (0.97)	0.90 (0.96)	0.91 (1.00)
Complete	**Ch**	0.97 (1.00)	0.93 (1.00)	0.92 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
	Dunn	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.98 (1.00)	0.96 (1.00)
	Dunn2	0.92 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
Ward.D2	**Ch**	0.96 (1.00)	0.92 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
	Dunn	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	0.98 (1.00)	0.96 (1.00)
	Dunn2	0.92 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)

Figure S4-S6 show the detection rate when the change for cl_2 happens at times 70, 90, and 120, respectively. Similar to other cases, one can easily see how the pre-clustering component of HSTCPD can enhance the detection/precision rate, especially when the time difference between the two change-points in cl_1 and cl_2 is short.

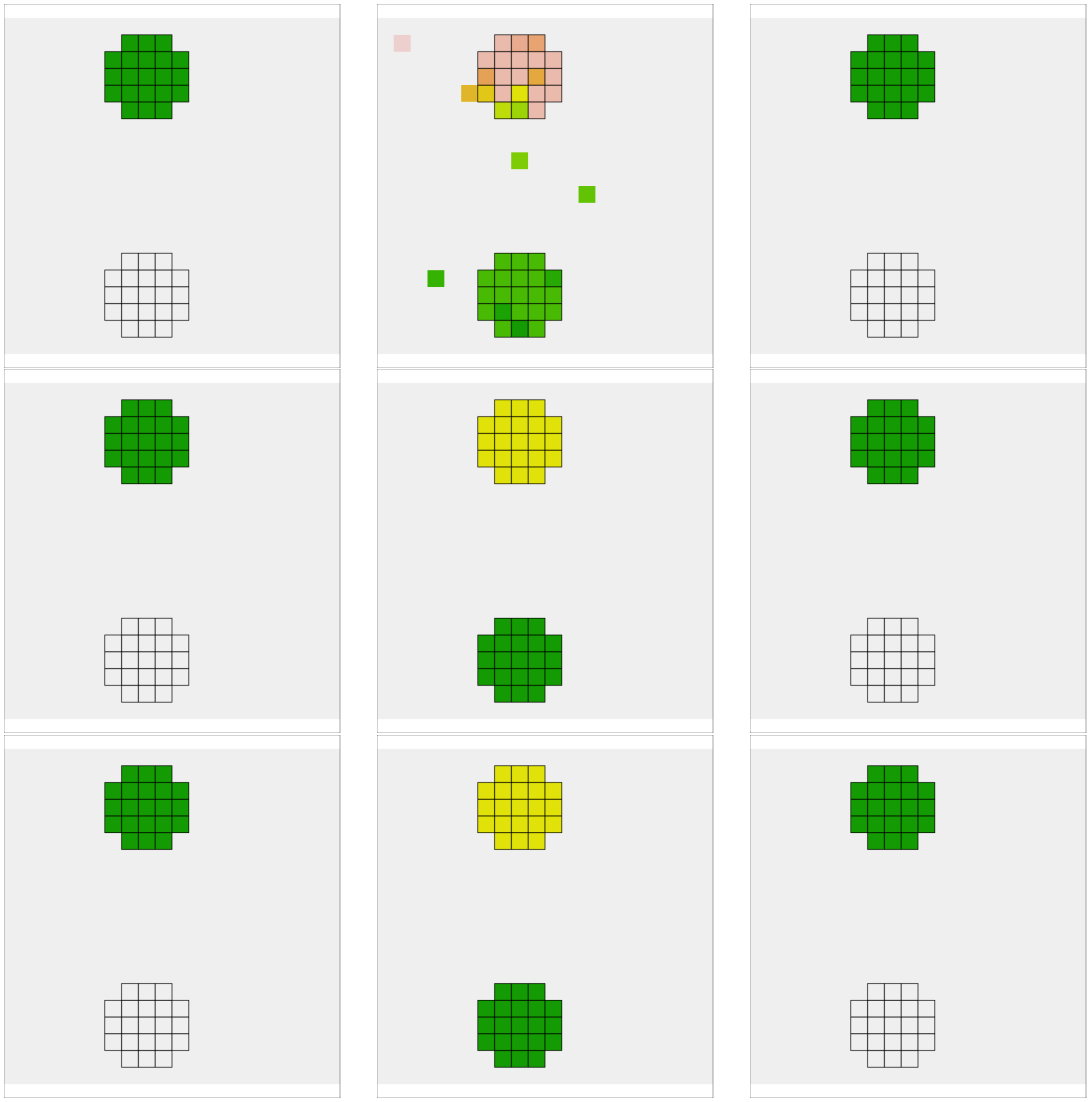

Figure S3: Individual examples (one out of 200 simulations) of the clustering results for Scenario I, with independent pixel time series, when the change-point for cl_2 happens at time 90. Rows from top to bottom: Single, Complete, and Ward.D2 linkage functions. Columns from left to right: Ch, Dunn, and Dunn2. Each colour represents a cluster, and the two clouds are displayed as grids.

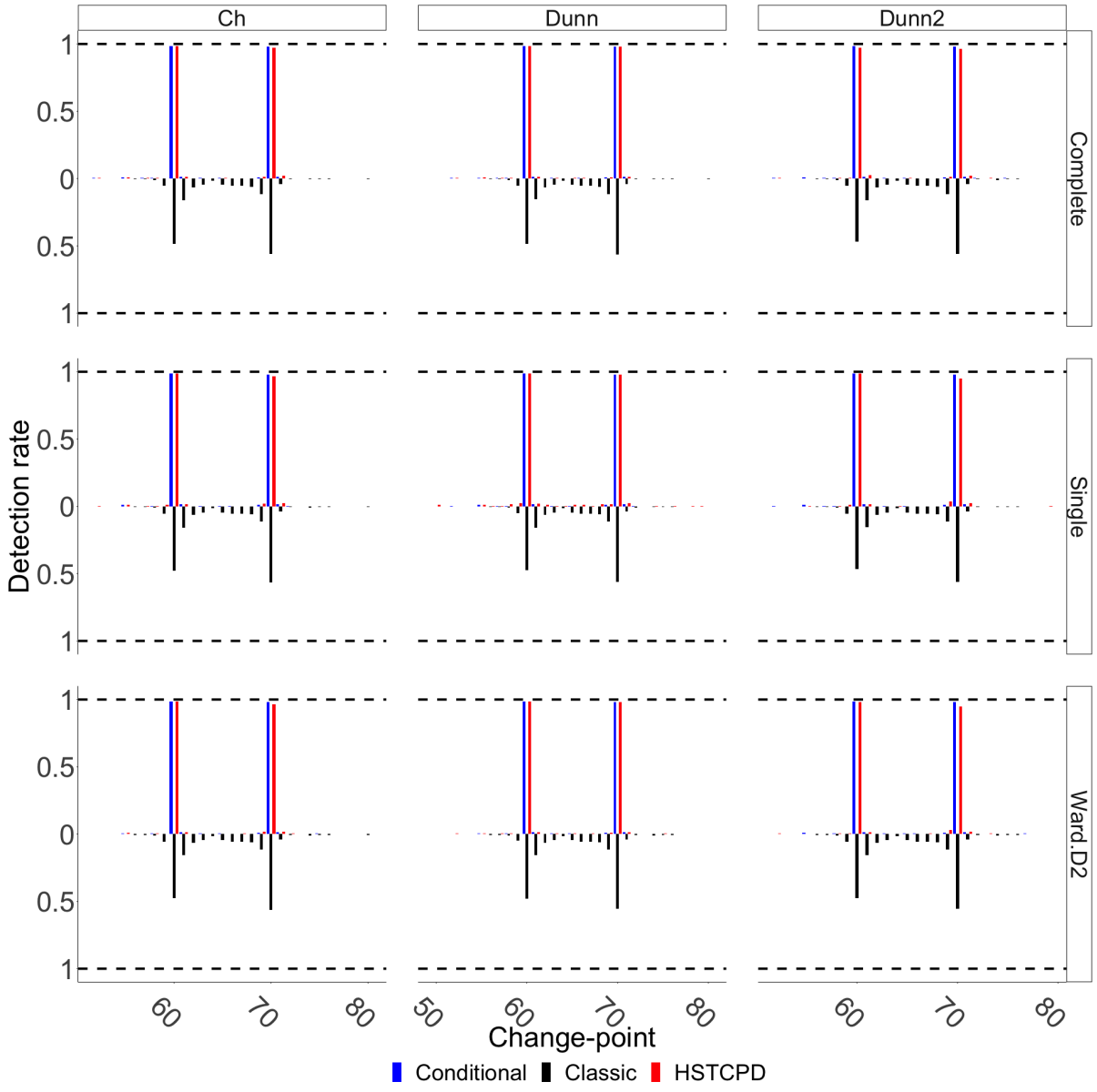

Figure S4: Detection rate based on 200 simulations from Scenario I, with independent pixel time series, when the change-point for cl_2 occurs at time 70.

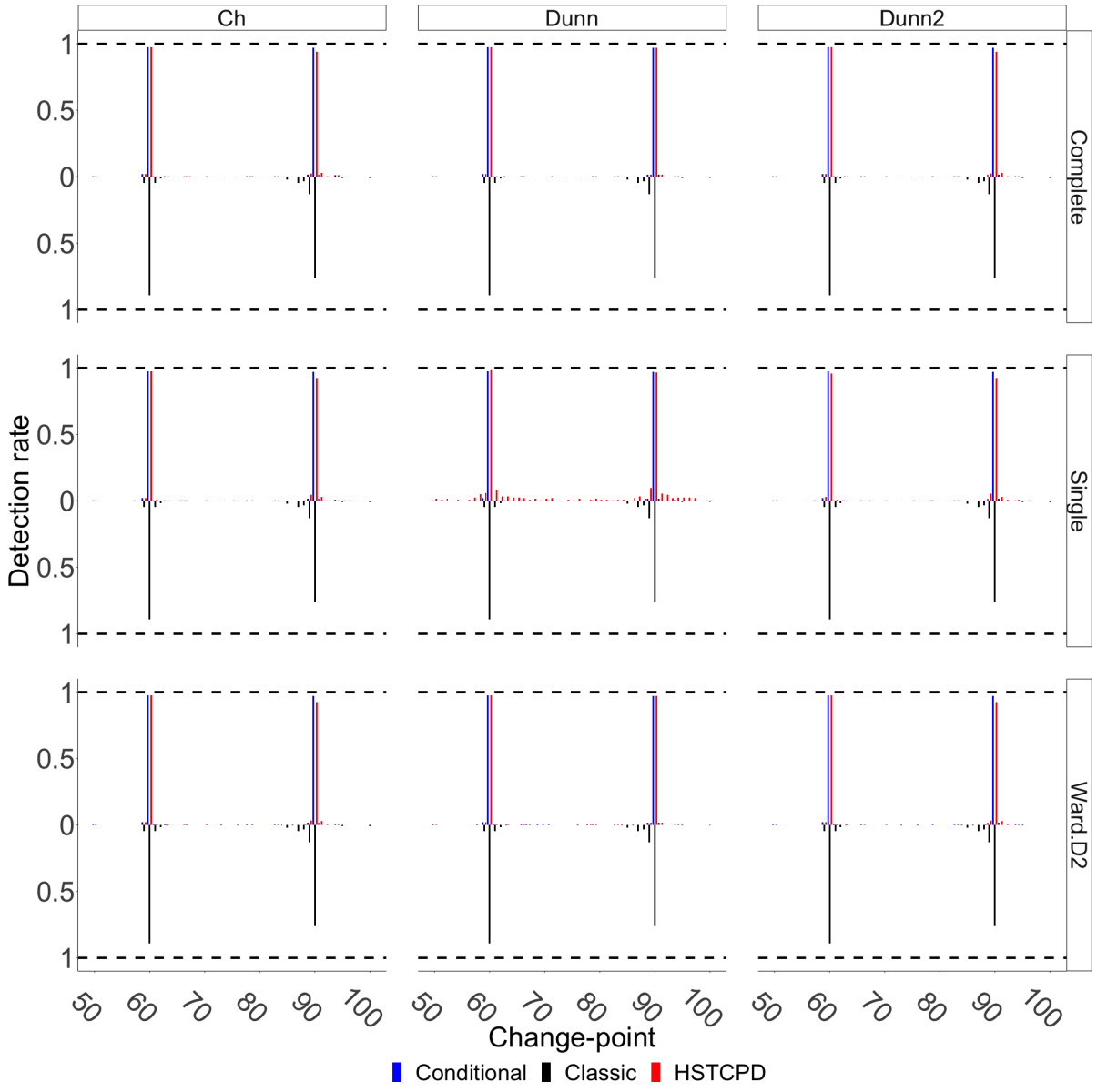

Figure S5: Detection rate based on 200 simulations from Scenario I, with independent pixel time series, when the change-point for cl_2 occurs at time 90.

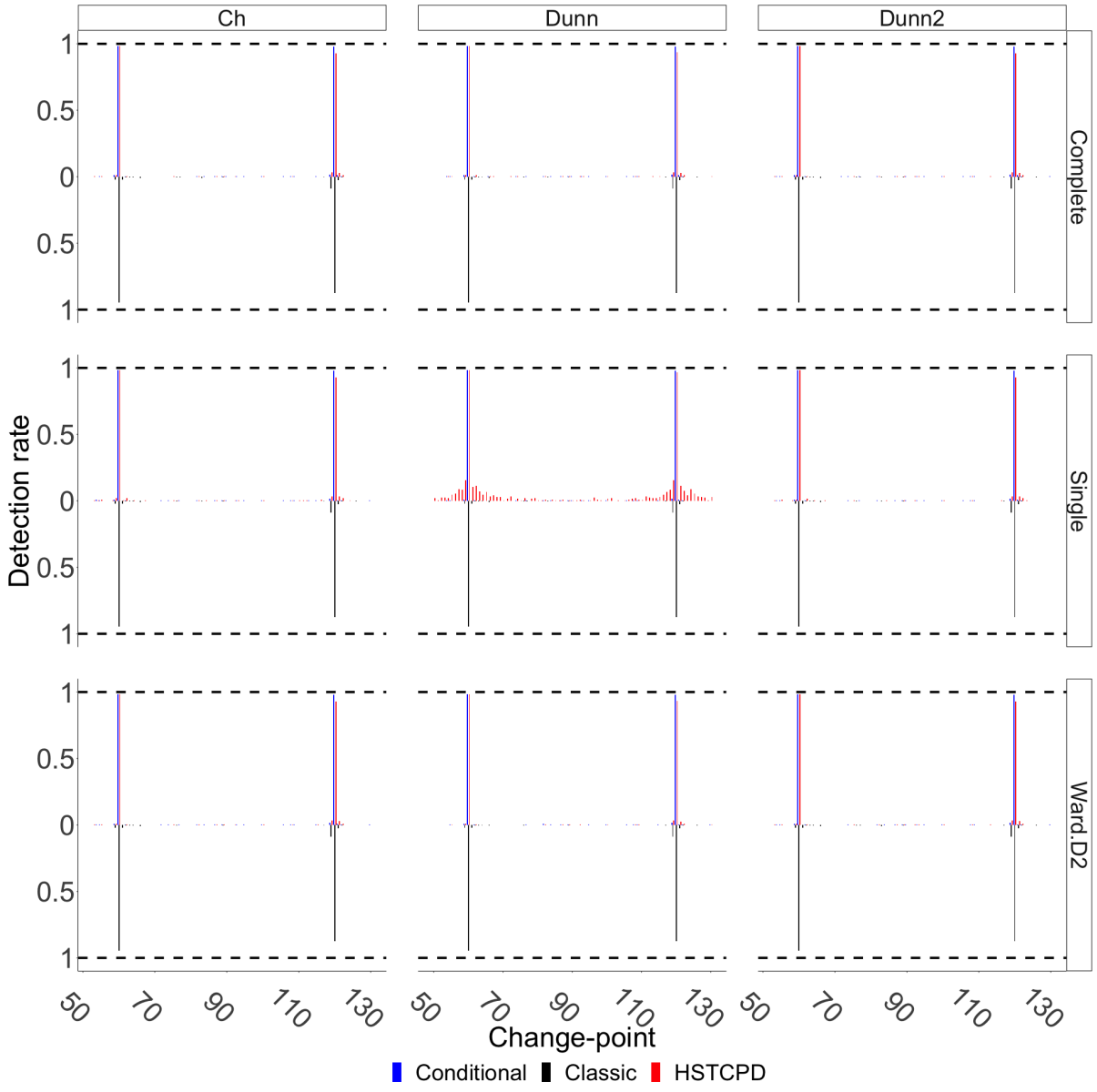

Figure S6: Detection rate based on 200 simulations from Scenario I, with independent pixel time series, when the change-point for cl_2 occurs at time 120.

S2.2 Scenario II

In this scenario we consider the case in which the two clouds, cl_1 and cl_2, undergo upward changes but of different magnitudes. We are interested to see if a change of larger magnitude inhibits the Classic approach's ability to detect both of the changes. In addition, we want to see if our HSTCPD approach manages to properly detect both of the change-points.

S2.2.1 Spatially correlated time series of images

Table S2 shows the obtained average **RI** and **SI** from 200 simulations. The main difference with respect to Table 1 is that here, in Scenario II, the Single linkage function generally performs better in terms of **SI**. We can further see that apart from Single-Dunn2, for which **SI** varies between 0.41 and 0.48, the rest of the combinations of linkage functions and methods to estimate the optimal number of clusters properly separate the two clouds. Similarly, we can see that, regardless of the method to estimate the number of clusters, Single linkage favours few clusters. The combination of Ward.D2/Complete with Ch/Dunn leads to substantially more clusters compared to their combination with Dunn2. Moreover, Ward.D2/Complete with Ch/Dunn estimate the borders of the clouds better than Ward.D2/Complete with Dunn2. We further see that **RI** and **SI** do not vary by increasing the time distance between the two change-points.

In Figure S7 we show the clustering outcomes for an individual simulation, when the change-point for cl_2 happens at time 90, with the intention of clarifying the meaning of the obtained values for **RI** and **SI** in Table S2. The combination of Single and either of Ch or Dunn properly detects the borders of one of the clouds, while the other cloud is included/lost within a large detected second cluster, which essentially covers the rest of the image. Moreover, Single and Dunn2 perform comparably less well, with an **SI** of approximately 0.45, and in Figure S7 we see one of 200 simulations where separation was not successful. The joining of Ward.D2/Complete and Dunn2 separates the clouds cl_1 and cl_2 by dividing the image time series into two big clusters, so that the estimated clouds become bigger than their actual sizes. However, according to **RI**, Complete-Dunn2 performs slightly better than Ward.D2-Dunn2. The combination of Ward.D2/Complete with Ch/Dunn leads to many clusters but the borders of cl_1 and cl_2 are essentially well estimated.

Table S2: Scenario II with spatially correlated pixel time series. An upward change with magnitude one happens at time 60 in cl_1, whereas the time index of the upward change of magnitude two, happening in cl_2, varies between 70 and 120. Average Rand index **RI** (separation index **SI**) are reported.

Linkage	**Optimal**	**2nd change-point**					
		70	**80**	**90**	**100**	**110**	**120**
Single	**Ch**	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)
	Dunn	0.90 (1.00)	0.90 (1.00)	0.90 (1.00)	0.90 (0.99)	0.90 (1.00)	0.90 (0.99)
	Dunn2	0.85 (0.44)	0.85 (0.46)	0.85 (0.48)	0.85 (0.44)	0.84 (0.42)	0.84 (0.41)
Complete	**Ch**	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)
	Dunn	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)
	Dunn2	0.54 (1.00)	0.52 (1.00)	0.51 (1.00)	0.51 (1.00)	0.51 (1.00)	0.49 (1.00)
Ward.D2	**Ch**	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)
	Dunn	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)	0.24 (1.00)
	Dunn2	0.49 (1.00)	0.49 (1.00)	0.48 (1.00)	0.48 (1.00)	0.47 (1.00)	0.48 (1.00)

For all 200 simulated image time series, Figure S8 shows the proportion of all detected change-points within all estimated clusters obtained by the 9 considered combinations. The Classic approach always detects the change-point in cl_2, which is of magnitude two, but in only 69% of times it manages to detect the change-point in cl_1, which is of magnitude one. Thus, one can see how changes of bigger magnitudes can dominate/hide changes of smaller magnitudes.

The results for the Conditional approach confirm that the change-point in cl_1 (cl_2) is detected 96.5% (100%) of the time. A few changes are also detected at times 59 and 61, which correspond to the change-point in cl_1. Such differences in the change-point detection rate of the Conditional approach and the Classic approach reveal the importance of carrying out classification prior to running multivariate change-point detection analyses.

Turning to the performance of HSTCPD, in accordance with what we saw in Table S2, the poorest results are those for Single-Dunn2 which did not manage to separate the clouds sufficiently well. This leads to 77% and 100% precise detection of the changes in cl_1 and cl_2. Note, however, that this is in contrast with the Classic approach, which has 69% detection rate for cl_1. Furthermore, other combinations of linkage functions and methods to find optimal clusters give rise to a high rate of simultaneous detection of both changes. For instance, Ward.D2-Dunn (Complete-Dunn) detects both changes in cl_1 and cl_2 with rates

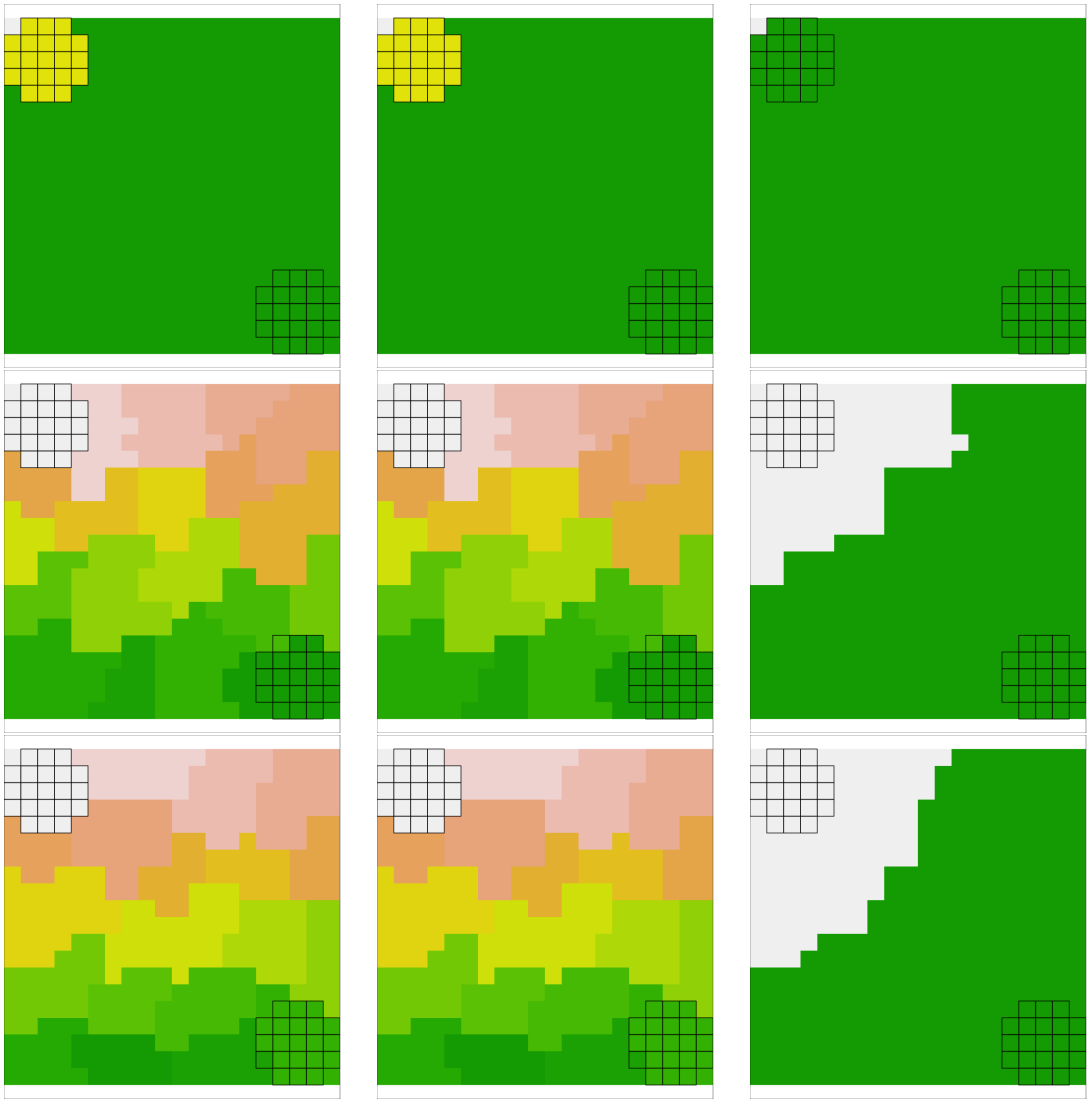

Figure S7: Individual examples (one out of 200 simulations) of the clustering results for Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 happens at time 90. Rows from top to bottom: Single, Complete, and Ward.D2 linkage functions. Columns from left to right: Ch, Dunn, and Dunn2. Each colour represents a cluster, and the two clouds are displayed as grids.

96% and 100% (97% and 100%), the estimated clusters are compact and the border of the detected clouds are closer to the original cloud borders. The detection rates for Ward.D2-/Complete-Dunn2 are 95% and 100%, and the clouds are estimated to be larger than their actual sizes.

The results concerning the cases where the change in cl_2 happens at times 90 or 120 can be found in Figure S9 and S10.

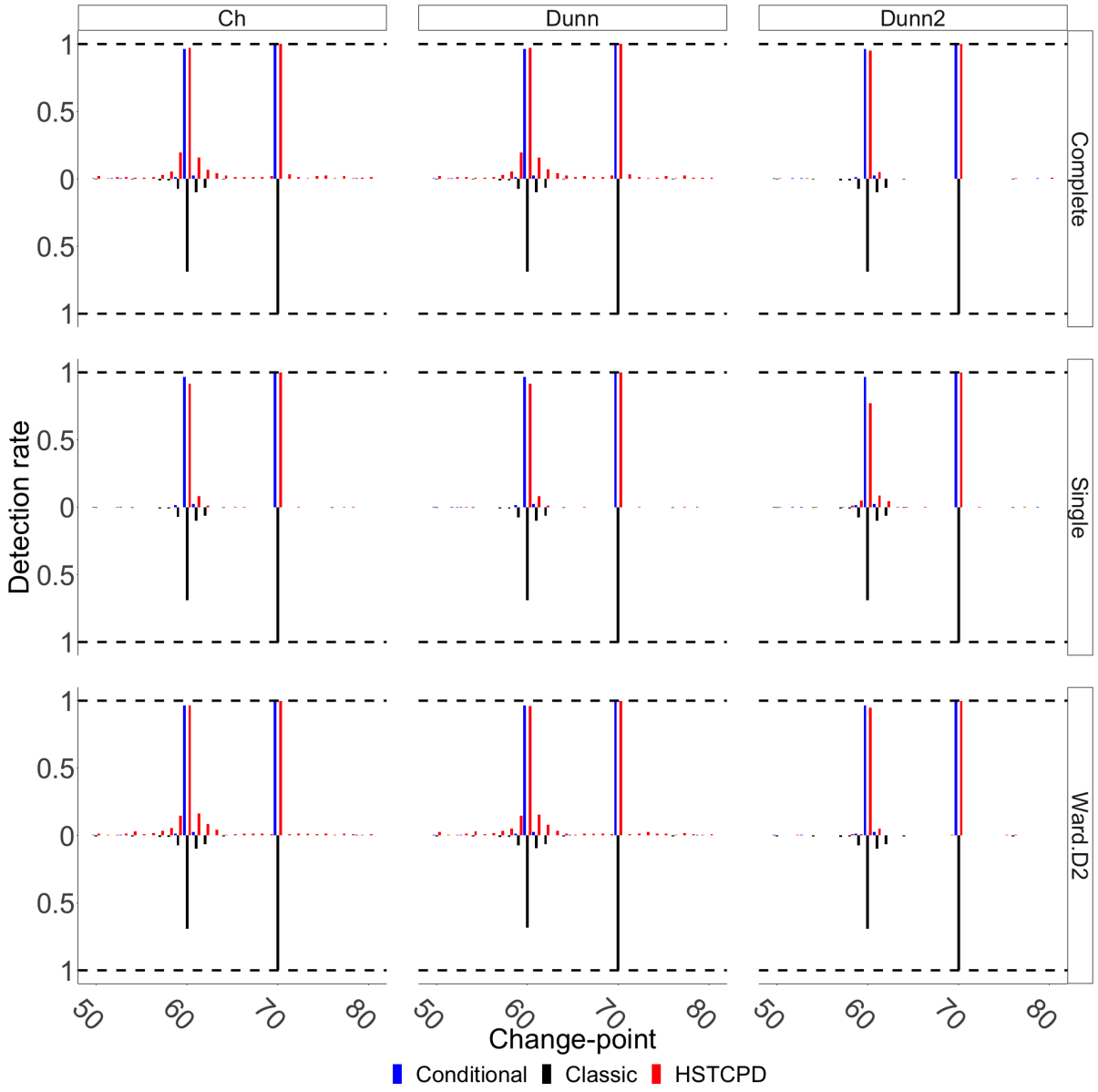

Figure S8: Detection rate for 200 simulations of Scenario II, with spatially correlated pixel time series, when the change-point in cl_2 occurs at time 70.

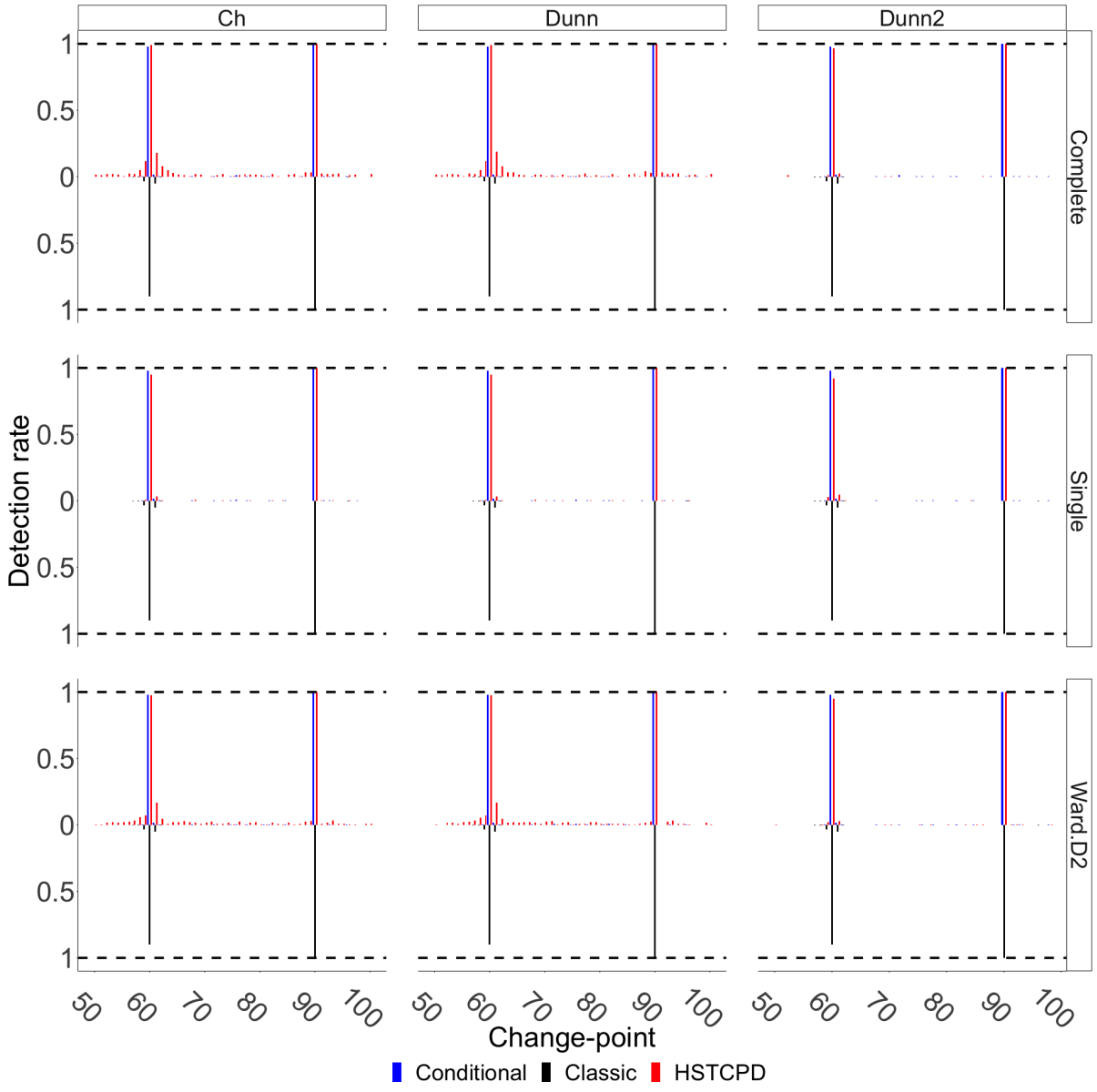

Figure S9: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

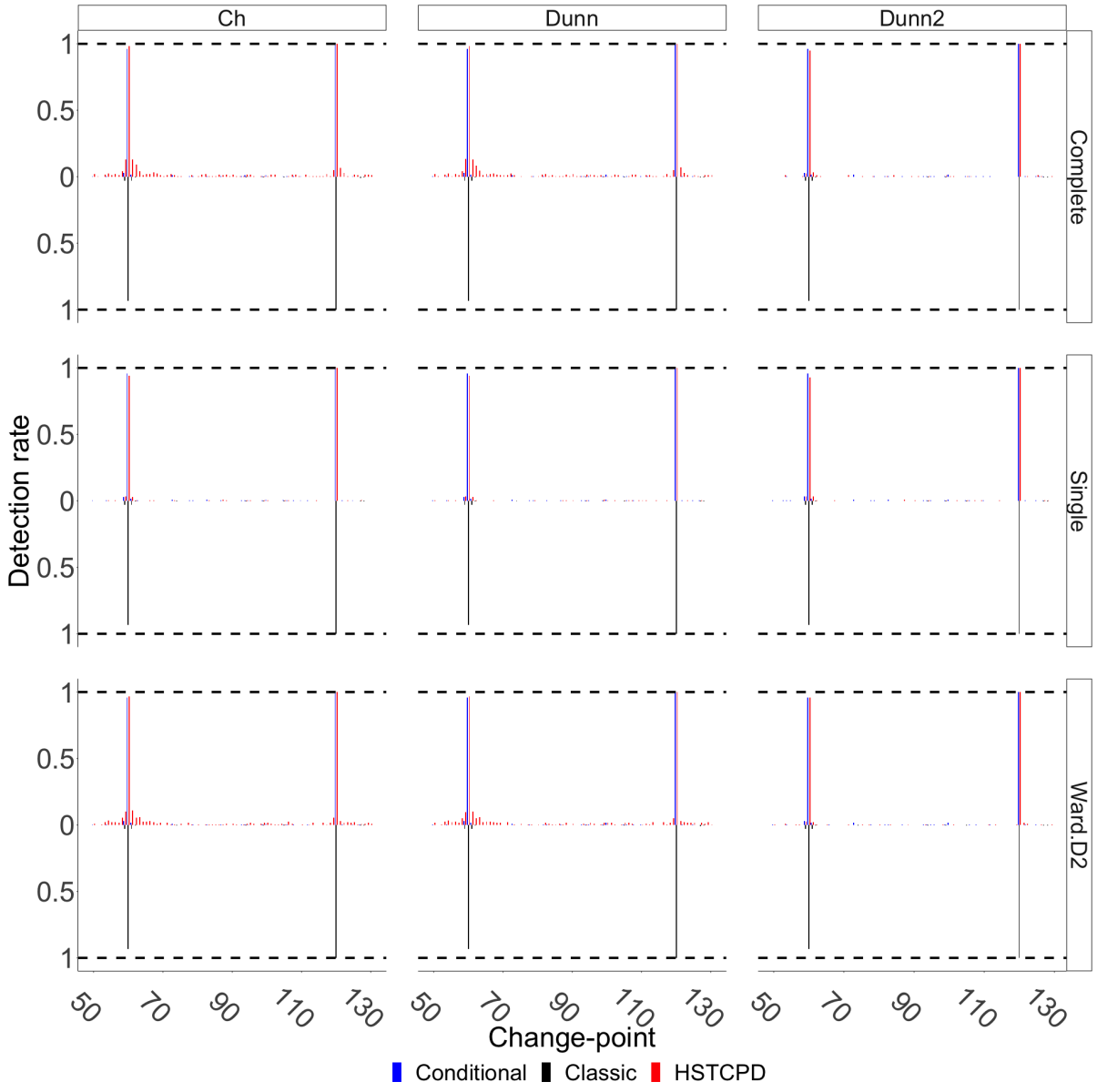

Figure S10: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.2.2 Independent pixel time series coming from the standard normal distribution

Table S3 shows the obtained average **RI** and **SI** based on 200 simulations in this case with similar results to those in Table S1; the combination of Dunn and any of the linkage functions seems to perform the best. The joint performance of Single and Ch/Dunn2 could only detect one of the clouds and the other one is lost within the part of image with no change. The combinations of Complete/Ward.D2 and Ch perform quite poorly in the sense that they could not properly separate the two clouds so that Complete-Ch approximately has an average value of 0.14 for **SI** while average **SI** for Ward.D2-Ch is zero. The performance of Complete/Ward.D2 and Dunn2 clearly depends on the time difference between the two change-points; it clearly improves by increasing such time difference.

Table S3: Scenario II with independent pixel time series. An upward change with magnitude one happens at time 60 in cl_1, whereas the time index of the upward change of magnitude two, happening in cl_2, varies between 70 and 120. Average Rand index **RI** (separation index **SI**) are reported.

Linkage	**Optimal**	**2nd change-point**					
		70	**80**	**90**	**100**	**110**	**120**
Single	**Ch**	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
	Dunn	0.94 (1.00)	0.95 (1.00)	0.96 (1.00)	0.98 (1.00)	0.98 (1.00)	0.99 (1.00)
	Dunn2	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)	0.91 (1.00)
Complete	**Ch**	0.98 (0.20)	0.98 (0.14)	0.98 (0.14)	0.98 (0.11)	0.98 (0.12)	0.99 (0.07)
	Dunn	0.98 (1.00)	0.99 (1.00)	0.99 (1.00)	0.99 (1.00)	1.00 (1.00)	1.00 (1.00)
	Dunn2	0.98 (0.20)	0.98 (0.14)	0.98 (0.18)	0.99 (0.50)	0.99 (0.84)	0.99 (0.98)
Ward.D2	**Ch**	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
	Dunn	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
	Dunn2	0.99 (0.00)	0.99 (0.00)	0.99 (0.05)	1.00 (0.44)	1.00 (0.84)	1.00 (0.98)

Figure S12-S14 show the proportion of the detected change-points when the change in cl_2 happens at times 70, 90, and 120, which, in accordance with the results in Table S3, confirm a better performance for HSTCPD compared to the Classic approach in all cases.

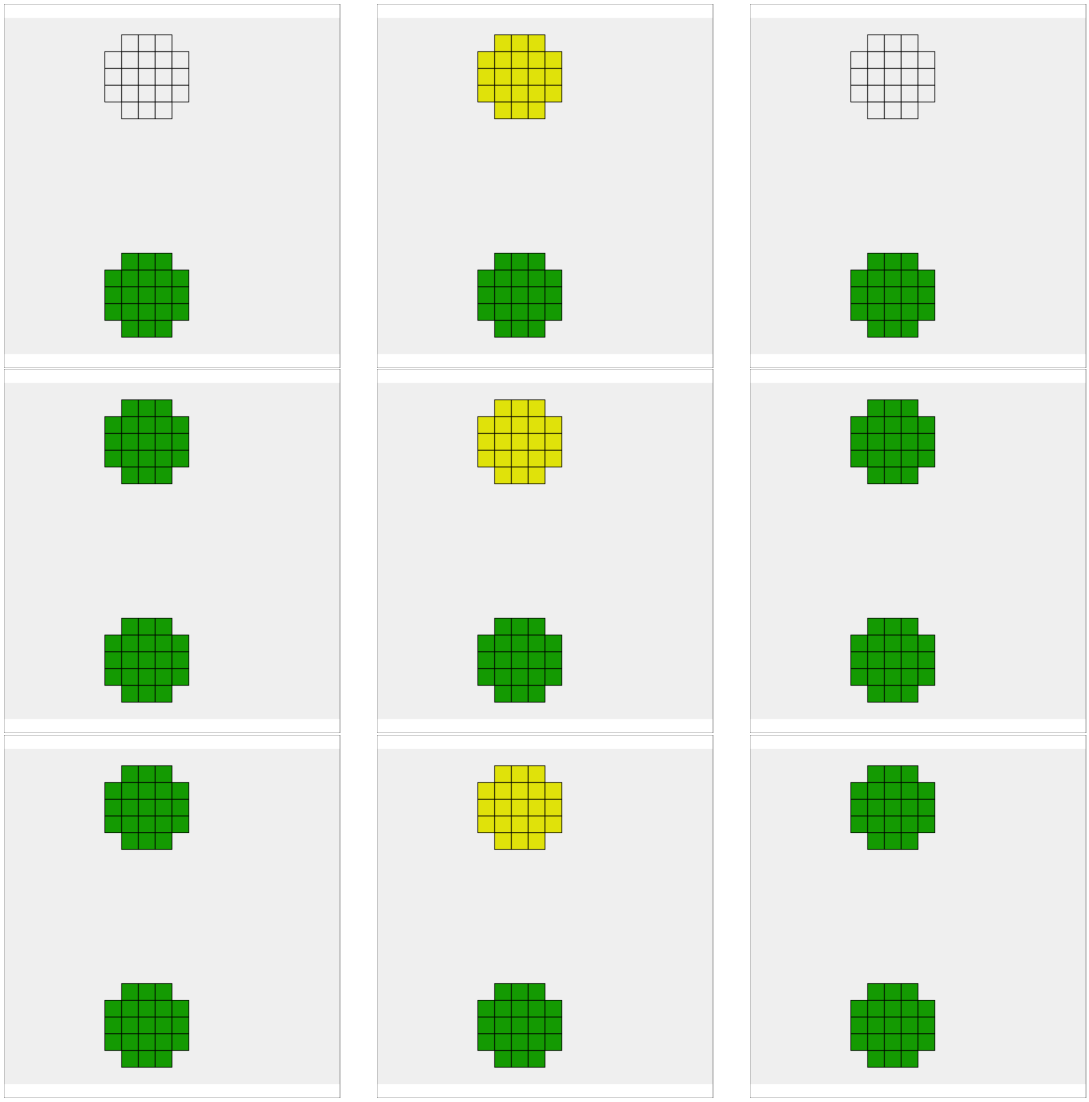

Figure S11: Individual examples (one out of 200 simulations) of the clustering results for Scenario II, with independent pixel time series, when the change-point for cl_2 happens at time 90. Rows from top to bottom: Single, Complete, and Ward.D2 linkage functions. Columns from left to right: Ch, Dunn, and Dunn2. Each colour represents a cluster, and the two clouds are displayed as grids.

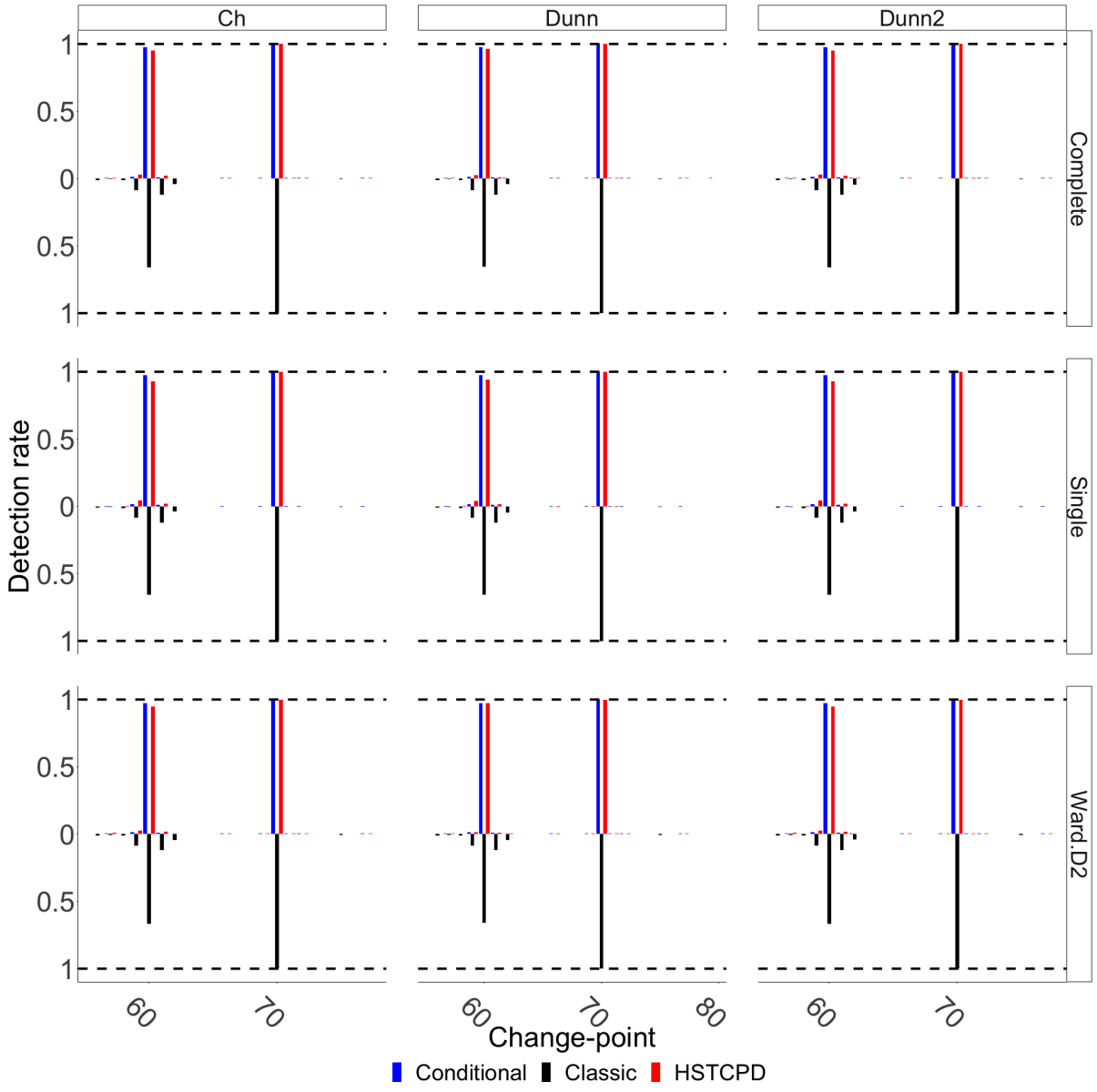

Figure S12: Detection rate based on 200 simulations from Scenario II, with independent pixel time series, when the change-point for cl_2 occurs at time 70.

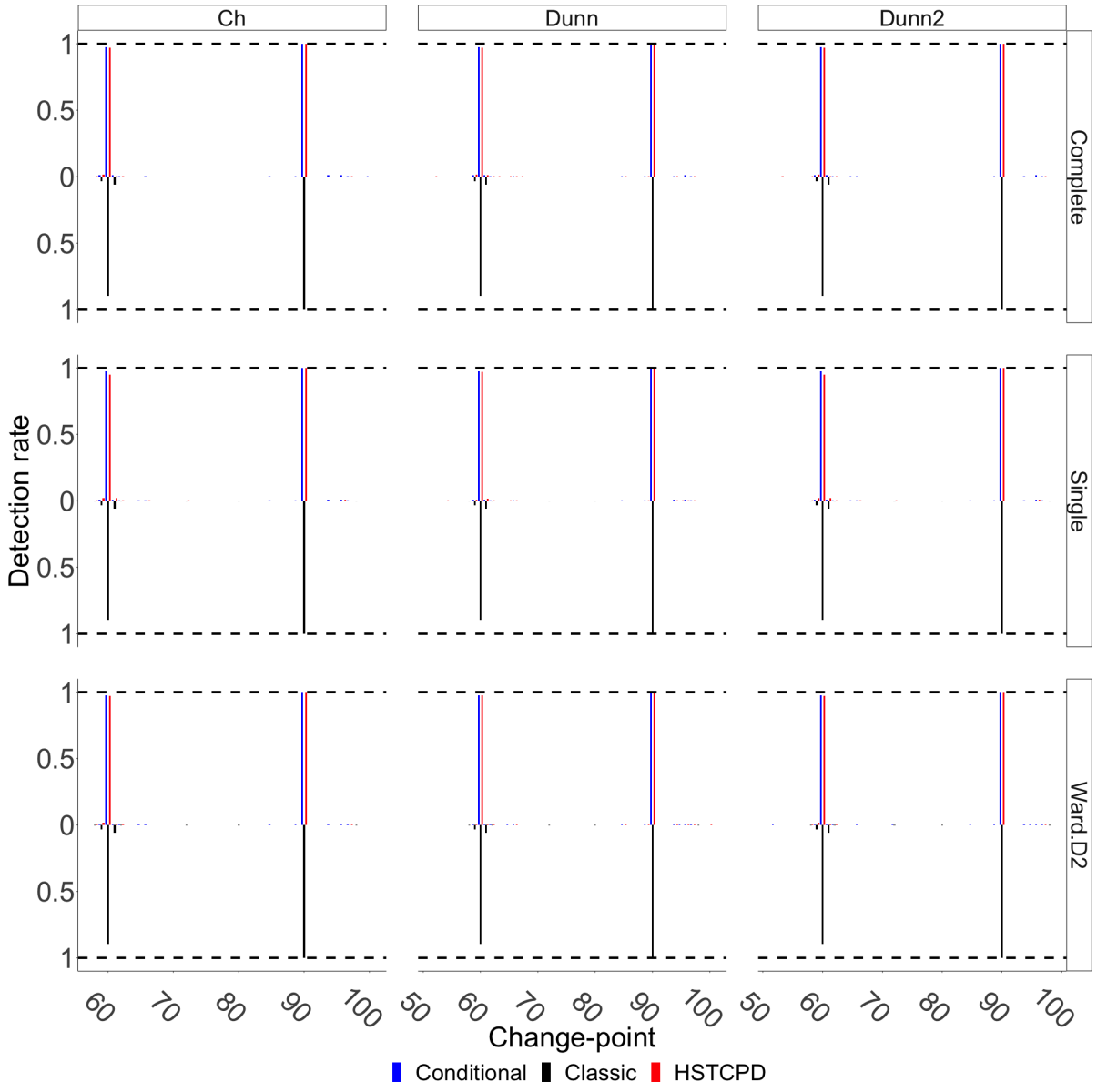

Figure S13: Detection rate based on 200 simulations from Scenario II, with independent pixel time series, when the change-point for cl_2 occurs at time 90.

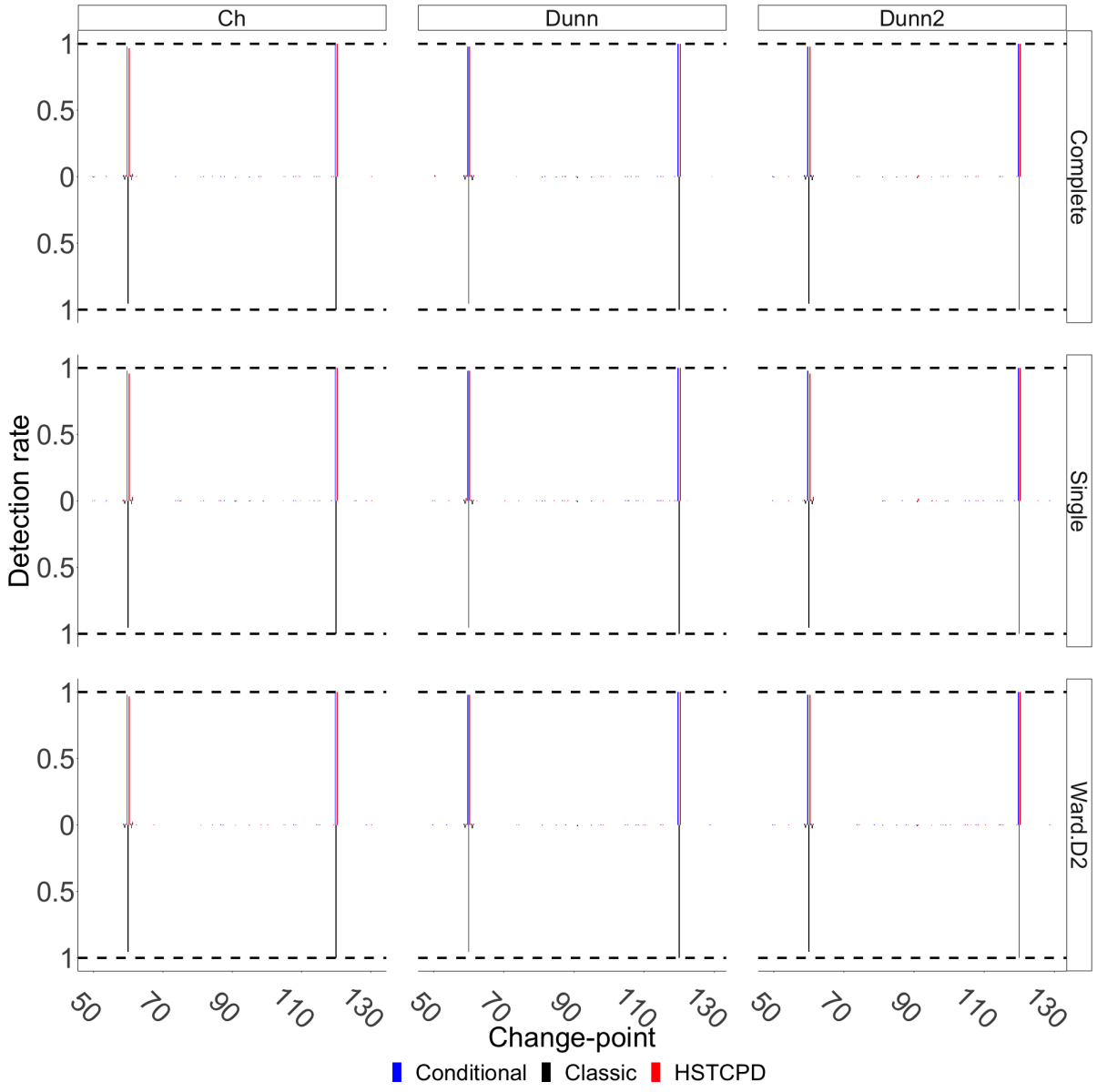

Figure S14: Detection rate based on 200 simulations from Scenario II, with independent pixel time series, when the change-point for cl_2 occurs at time 120.

S2.3 Scenario III

This scenario focuses on a situation where we have two changes of similar magnitudes but in different regions and at distinct times. The sizes of the regions exhibiting changes are also different. Hence, we study whether the sizes of regions undergoing changes tend to mislead change-point detection methods, and consequently yielding erroneous outcomes.

S2.3.1 Spatially correlated time series of images

Table S4 shows the average Rand index **RI** and the separation index **SI**. Similarly to the other two scenarios, we can see that the combination of Complete/Ward.D2 and Ch/Dunn yields several compact clusters, which in turn pushes down **RI** while the two clouds, cl_1 and cl_2, do not share any mutual pixels/functions. The combinations of Complete and Ward.D2 with Dunn2 perform similarly in terms of **RI** while Complete-Dunn2 performs slightly better than Ward.D2-Dunn2 in terms of **SI**. The Single linkage performs best when it is combined with Ch and it has high values for both **RI** and **SI**. The performance of Single-Dunn is slightly poorer than Single-Ch, but slightly better than Single-Dunn2.

Table S4: Scenario III with spatially correlated pixel time series. An upward change of magnitude one happens at time 60 in cl_1, while the time index of an upward change of similar magnitude, which happens in cl_2, varies between 70 and 120. The sizes of cl_1 and cl_2 correspond to 5.25% and 17.25% of the full image. Average Rand index **RI** (separation index **SI**) are reported.

Linkage	**Optimal**	**2nd change-point**					
		70	**80**	**90**	**100**	**110**	**120**
Single	**Ch**	0.92 (0.98)	0.92 (0.98)	0.90 (0.95)	0.87 (0.93)	0.86 (0.97)	0.81 (0.92)
	Dunn	0.85 (0.90)	0.85 (0.93)	0.82 (0.90)	0.81 (0.84)	0.81 (0.90)	0.77 (0.85)
	Dunn2	0.76 (0.72)	0.75 (0.75)	0.72 (0.74)	0.70 (0.63)	0.71 (0.74)	0.70 (0.70)
Complete	**Ch**	0.41 (1.00)	0.41 (0.99)	0.41 (1.00)	0.41 (1.00)	0.41 (0.99)	0.41 (1.00)
	Dunn	0.41 (1.00)	0.41 (0.99)	0.41 (1.00)	0.41 (1.00)	0.41 (0.99)	0.41 (1.00)
	Dunn2	0.55 (0.83)	0.55 (0.79)	0.54 (0.78)	0.54 (0.85)	0.54 (0.80)	0.53 (0.80)
Ward.D2	**Ch**	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)
	Dunn	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)	0.41 (1.00)
	Dunn2	0.55 (0.74)	0.55 (0.71)	0.55 (0.73)	0.54 (0.74)	0.54 (0.80)	0.54 (0.73)

Figure S15 shows the clustering outcomes for one of 200 simulations when the change

for cl_2 occurs at time 90. We can see that the Single linkage, regardless of the choice of method to estimate the number of clusters, precisely detects the borders of cl_1 while the rest of the pixels/functions are classified as belonging to a sole cluster. Similar to the previous scenarios, the combination of Complete/Ward.D2 with Ch/Dunn yields compact clusters which are relatively small. The two clouds cl_1 and cl_2 are clearly separated here. Note that the borders of the actual clouds are not well detected and, in particular, the bigger cloud belongs to several different clusters. Since, at a later step, we independently carry out individual change-point detection analyses for all detected clusters, the clusters which intersect the clouds and also have the majority of the pixels/functions not facing any change, might influence the quality of the detected changes. The performance of Complete/Ward.D2 when combined with Dunn2 results in few clusters, with the sizes of the detected clouds/clusters being bigger than the actual cloud sizes. It further seems that Ward.D2-Dunn2 gives rise to more clusters in which the borders of the bigger cloud are better detected. We also see that the time difference between the two changes apparently has no effect on the performance of the clustering approaches.

Figure S16 shows the proportion of all detected changes for all considered approaches when the change-point for cl_2 happens at time 70. The performance of the Classic approach is influenced by the existence of two changes in regions of different sizes, and it only in approximately half of times detect the change in the smaller cloud.

The Conditional approach is able to detect the actual change in cl_1 97% of the time while the other change, in cl_2, is always detected. This stresses the importance of pre-classification, just as in the other two scenarios.

Turning to the performance of HSTCPD, for the Single linkage the best separation between the two clouds happens when it is combined with Ch, as noted in Table S4. Here we obtain 95.5% and 100% precise detection rate for the changes in cl_1 and cl_2, respectively. The rate of precise detection for the change in cl_1 is 93% and 88.5% for Dunn and Dunn2, respectively, in combination with the Single linkage function. The proportion of times that changes in cl_1 and cl_2 are precisely detected when we make use of Complete/ward.d2 together with ch are 97% and 100%, respectively. The same proportions of precise detection are obtained when Dunn is used together with Complete/Ward.d2. Concerning the performance of Dunn2 in this scenario, for the placed change in cl_1 we obtain a precise detection rate of

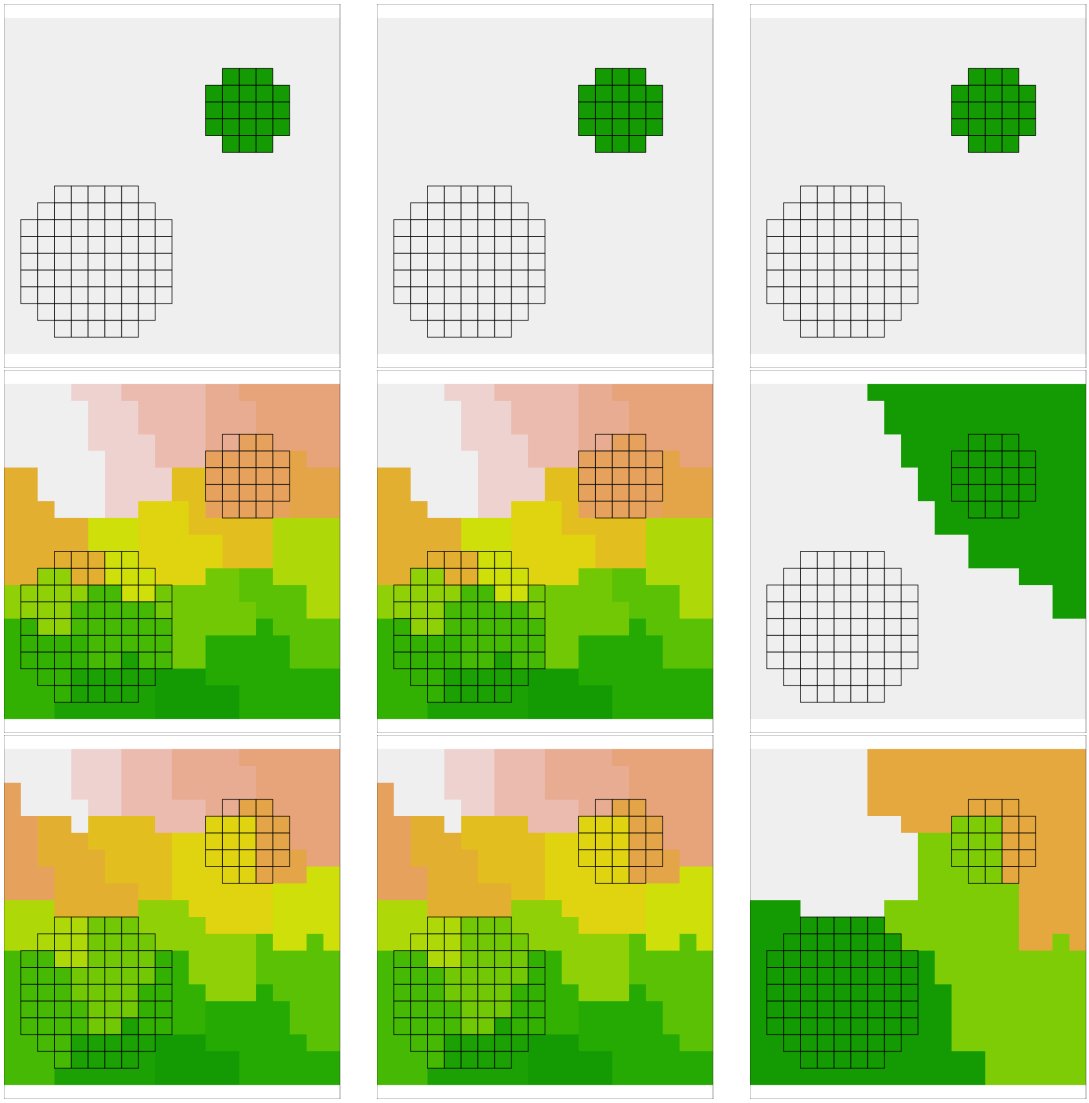

Figure S15: Individual examples (one out of 200 simulations) of the clustering results for Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 happens at time 90. Rows from top to bottom: Single, Complete, and Ward.D2 linkage functions. Columns from left to right: Ch, Dunn, and Dunn2. Each colour represents a cluster, and the two clouds are displayed as grids.

92.5% if we construct the dendrogram by means of the Complete linkage, while this rate is 94.5% if we use the Ward.D2 linkage. Not surprisingly, the quality of the change-point detection strongly depends on how well one manages to put different sub-regions with distinct temporal behaviours in disjoint clusters. The cases for which pieces of the clouds belong to different estimated clusters may decrease the performance of the employed change-point detection method. If this is the case, since we would consider clusters for which not all the

included functions actually undergo changes, the changes in these clusters may be detected with small/negligible errors, which can be seen in Figure S16. This is more visible when Ch/Dunn is combined with Complete/Ward.D2 to estimate the optimal clusters; note the small/compact clusters.

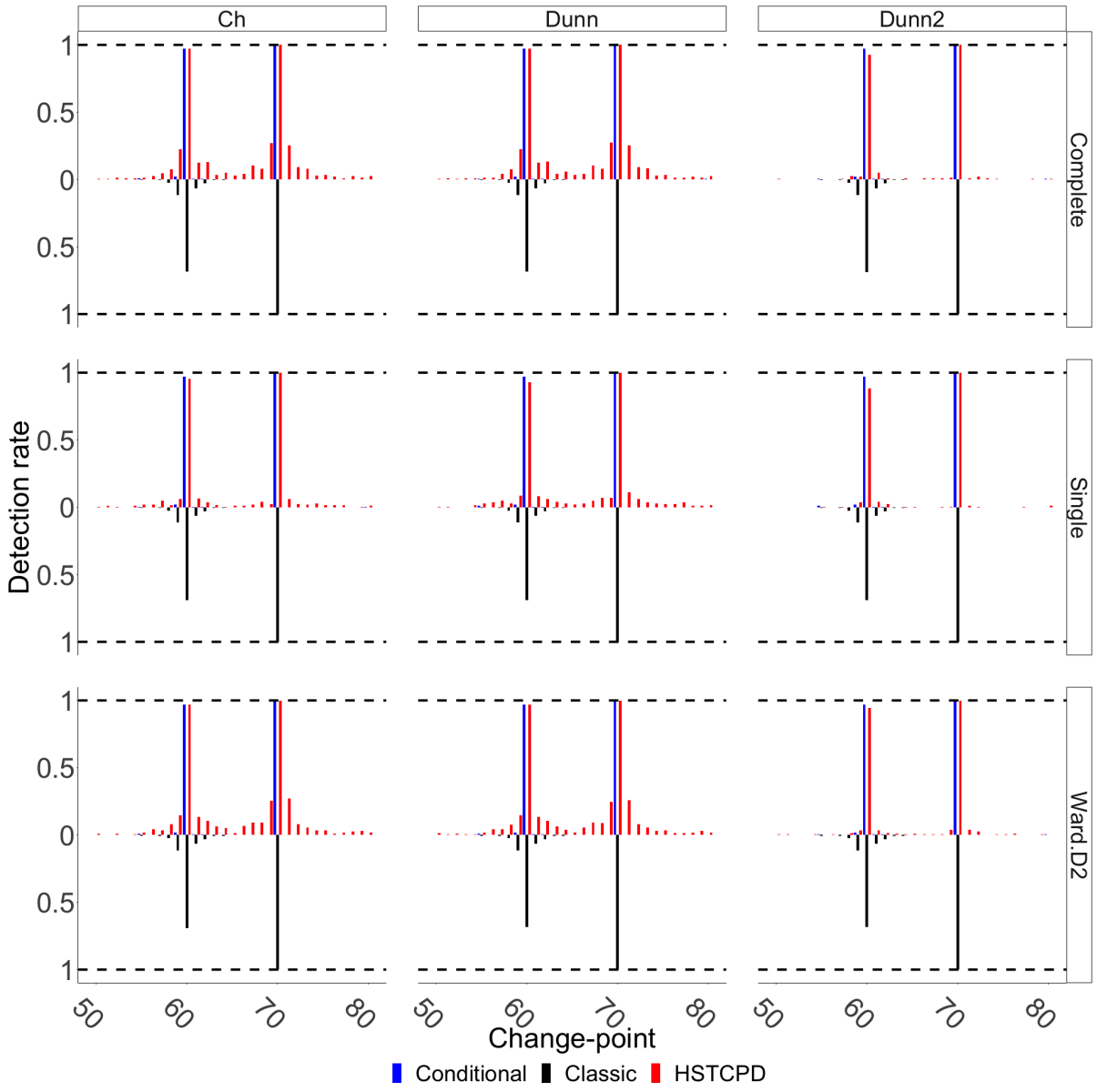

Figure S16: Detection rate based on 200 simulations of Scenario III, with spatially correlated pixel time series, when the change-point in cl_2 occurs at time 70.

The results corresponding to the case where the change-point for cl_2 happens at time 90 and 120 can be seen in Figure S17 and S18.

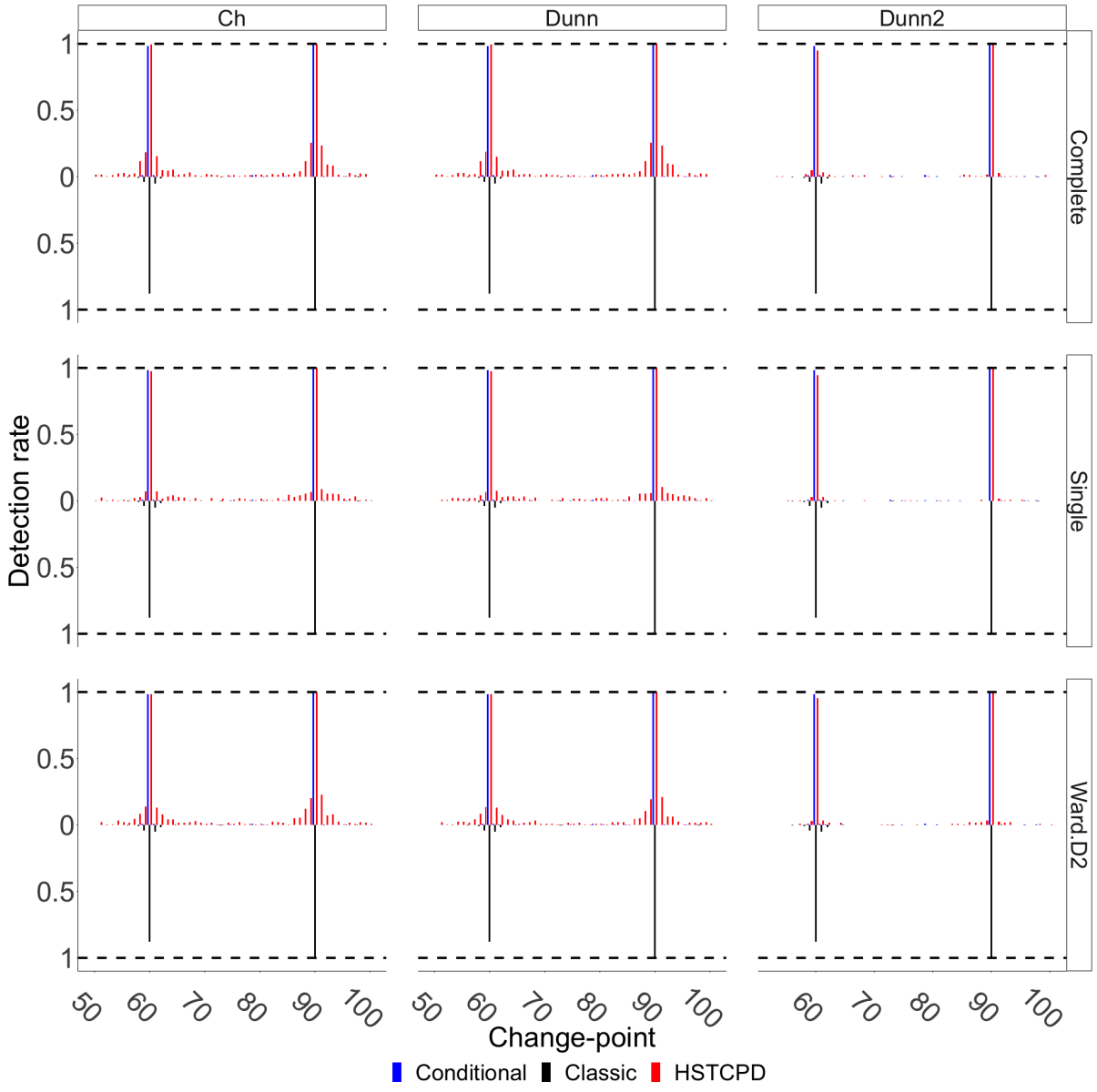

Figure S17: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

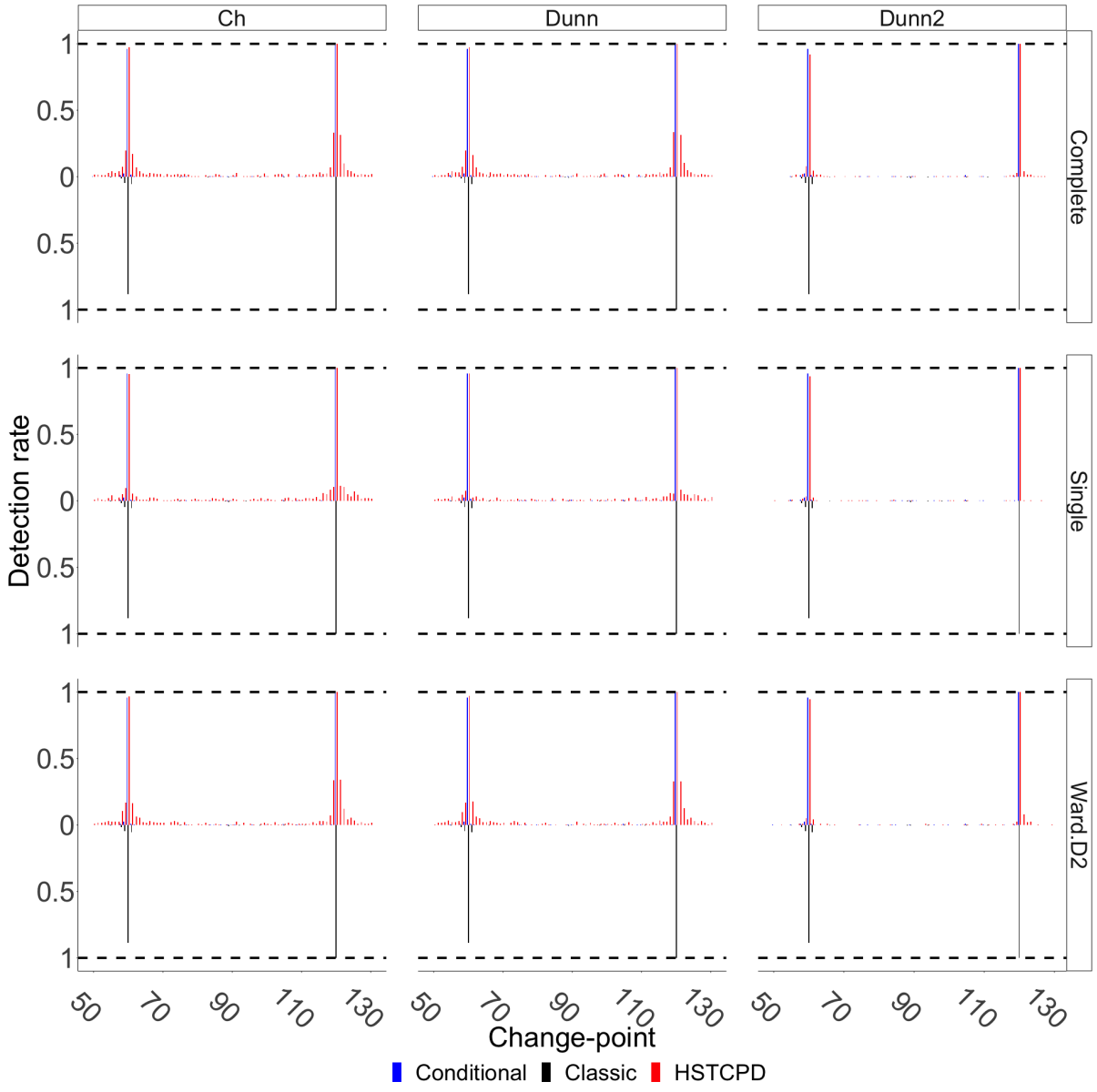

Figure S18: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.3.2 Independent pixel time series coming from the standard normal distribution

Table S5 shows the obtained average values for both **RI** and **SI** based on 200 simulations for this scenario. Surprisingly, none of the combinations of the linkage functions and methods to estimate optimal clusters could properly separate the two clusters; **SI** for the combinations of Complete/Ward.D2 with Ch/Dunn2 is zero, i.e. mistakenly put the two clouds in a unique cluster, apart from when the change-point for cl_2 happens at time 120 which is 0.14. However, in all other case, we can see that **SI** increases as the time difference between the two change-points increases. This is clearly reflected in the performance of HSTCPD which even in such cases performs better than the Classic approach (see Figure S20-S22).

Table S5: Scenario III with independent pixel time series. An upward change of magnitude one happens at time 60 in cl_1, while the time index of an upward change of similar magnitude, which happens in cl_2, varies between 70 and 120. The sizes of cl_1 and cl_2 correspond to 5.25% and 17.25% of the full image. Average Rand index **RI** (separation index **SI**) are reported.

Linkage	**Optimal**	**2nd change-point**					
		70	**80**	**90**	**100**	**110**	**120**
Single	**Ch**	0.97 (0.02)	0.98 (0.09)	0.96 (0.22)	0.94 (0.44)	0.90 (0.68)	0.81 (0.92)
	Dunn	0.97 (0.04)	0.98 (0.20)	0.96 (0.37)	0.95 (0.75)	0.95 (0.93)	0.89 (0.97)
	Dunn2	0.94 (0.01)	0.91 (0.07)	0.86 (0.15)	0.80 (0.32)	0.76 (0.55)	0.72 (0.69)
Complete	**Ch**	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)
	Dunn	0.97 (0.01)	0.98 (0.05)	0.97 (0.26)	0.99 (0.64)	0.98 (0.93)	0.97 (0.99)
	Dunn2	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.14)
Ward.D2	**Ch**	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)
	Dunn	0.97 (0.03)	0.98 (0.12)	0.98 (0.32)	0.99 (0.67)	0.99 (0.91)	0.99 (1.00)
	Dunn2	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.00)	0.98 (0.14)

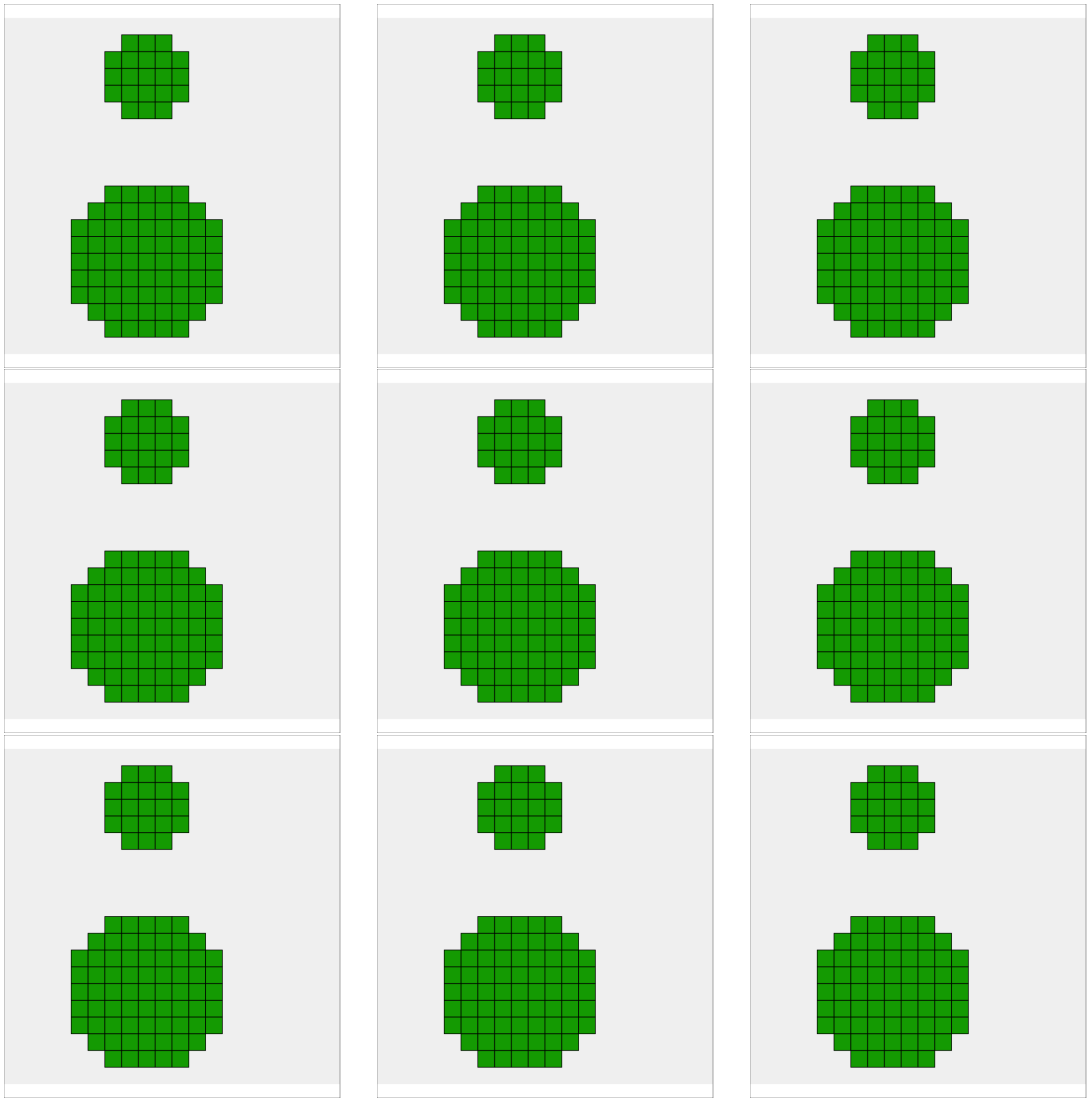

Figure S19: Individual examples (one out of 200 simulations) of the clustering results for Scenario III, with independent pixel time series, when the change-point for cl_2 happens at time 90. Rows from top to bottom: Single, Complete, and Ward.D2 linkage functions. Columns from left to right: Ch, Dunn, and Dunn2. Each colour represents a cluster, and the two clouds are displayed as grids.

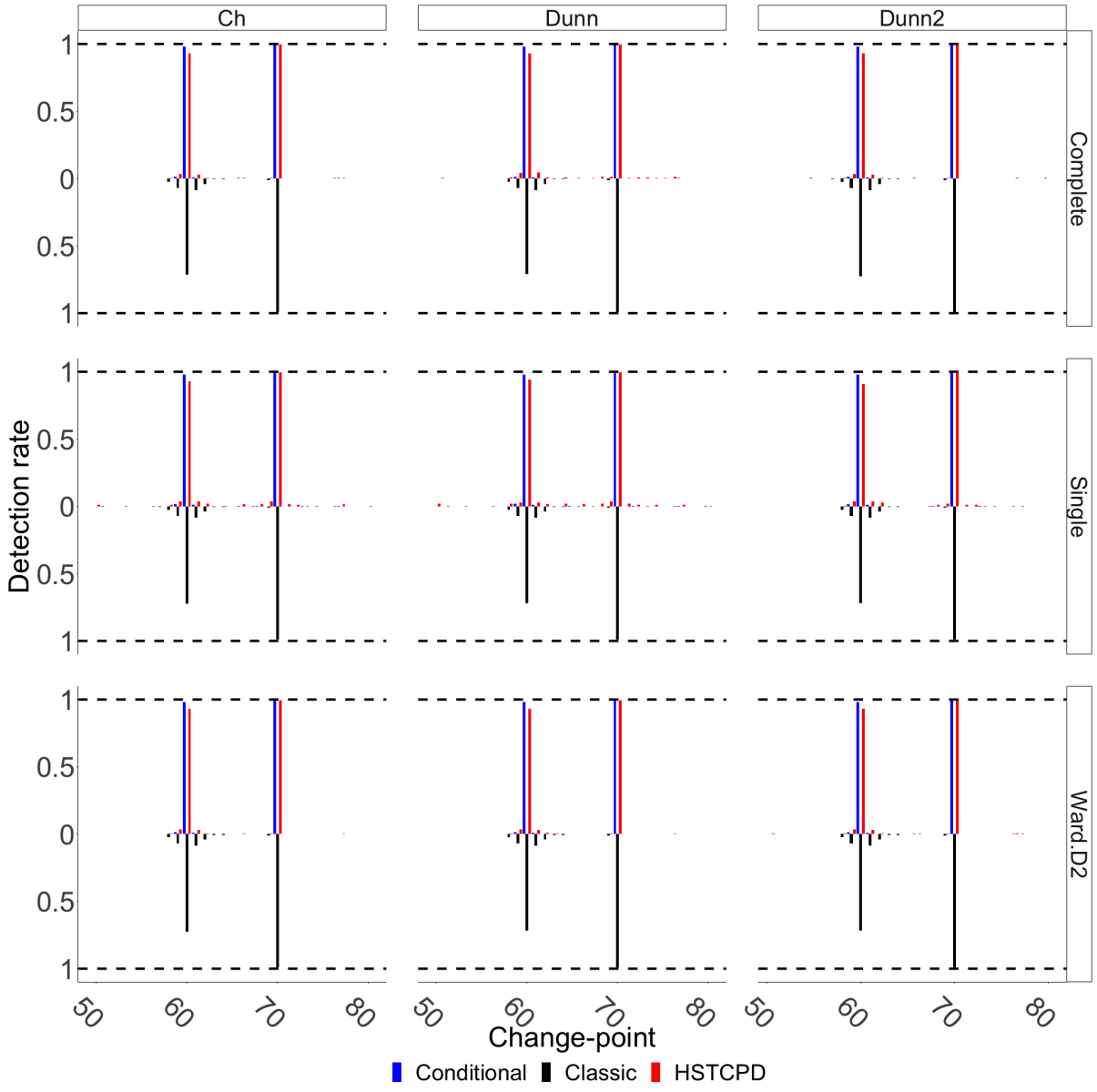

Figure S20: Detection rate based on 200 simulations from Scenario III, with independent pixel time series, when the change-point for cl_2 occurs at time 70.

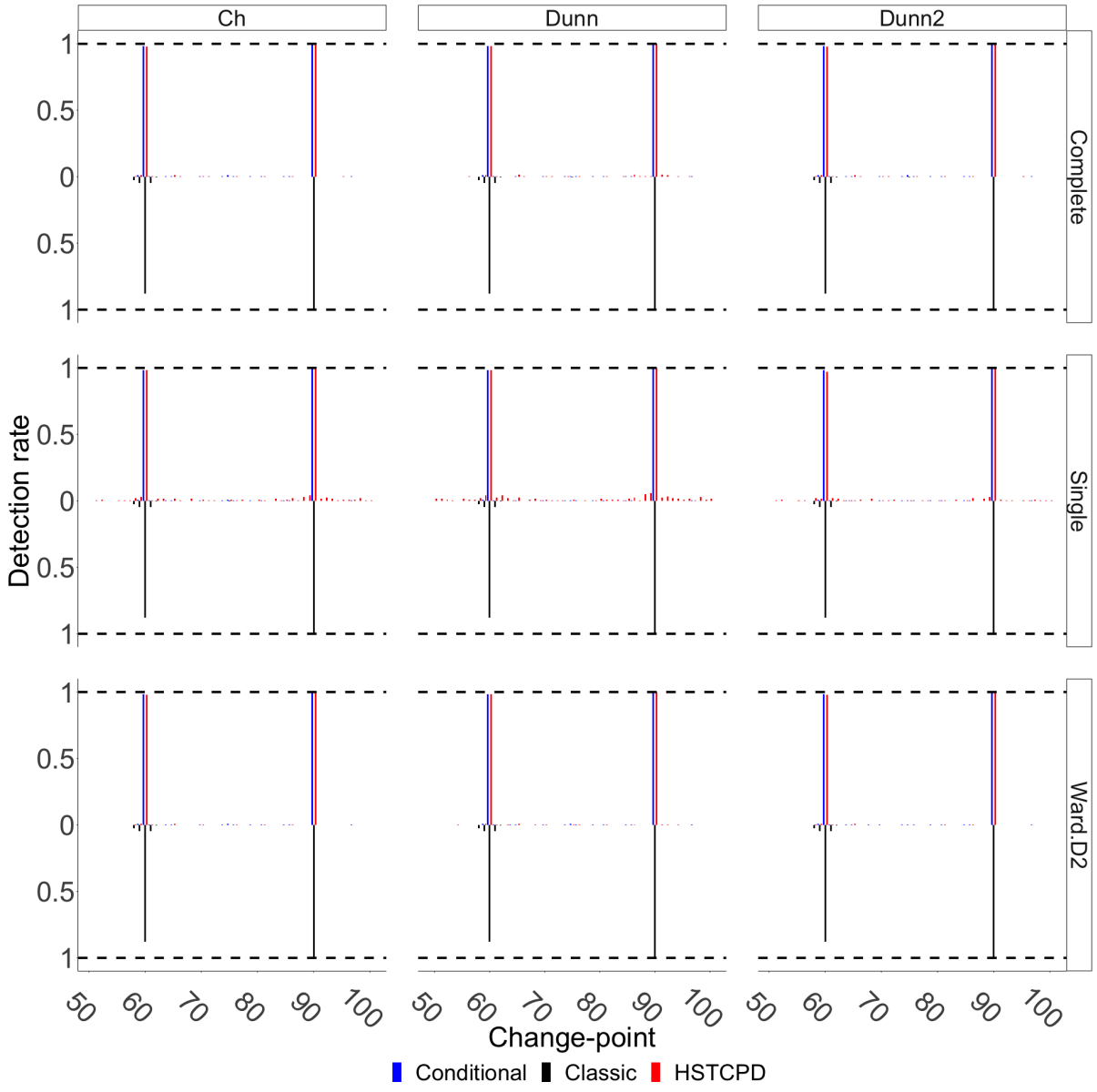

Figure S21: Detection rate based on 200 simulations from Scenario III, with independent pixel time series, when the change-point for cl_2 occurs at time 90.

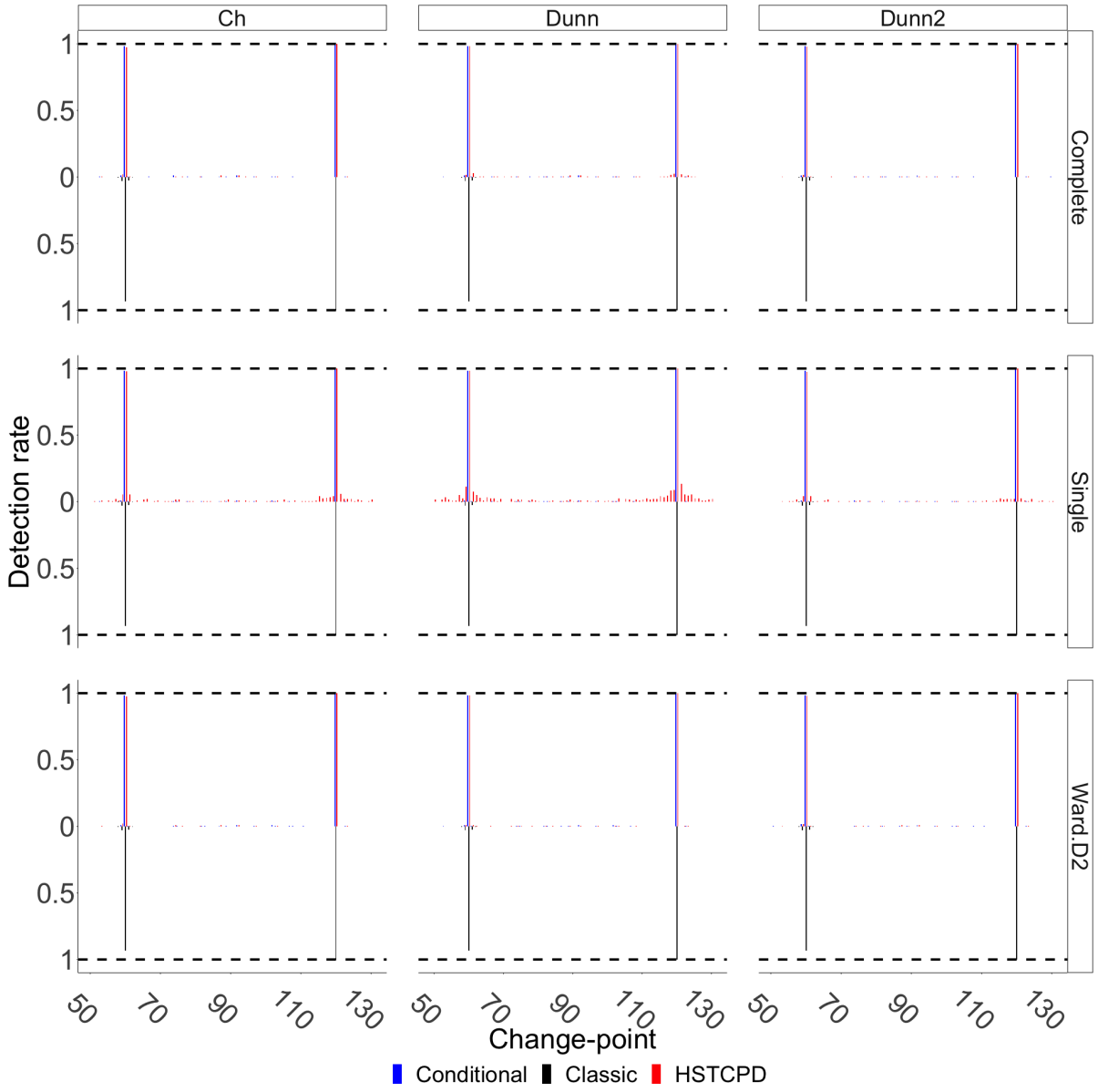

Figure S22: Detection rate based on 200 simulations from Scenario III, with independent pixel time series, when the change-point for cl_2 occurs at time 120.

S2.4 AdaptiveCpt

S2.4.1 Scenario I

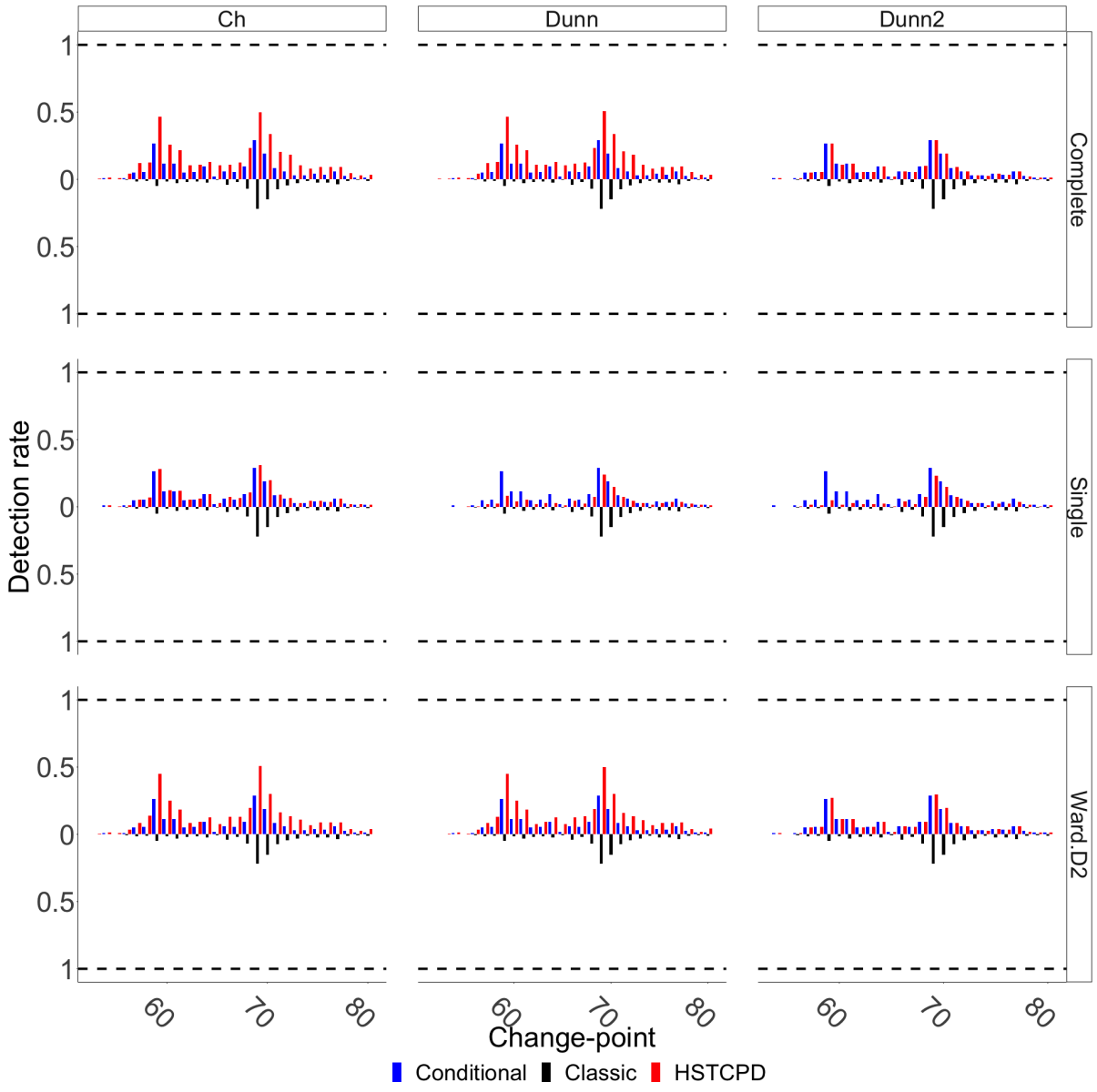

Figure S23: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

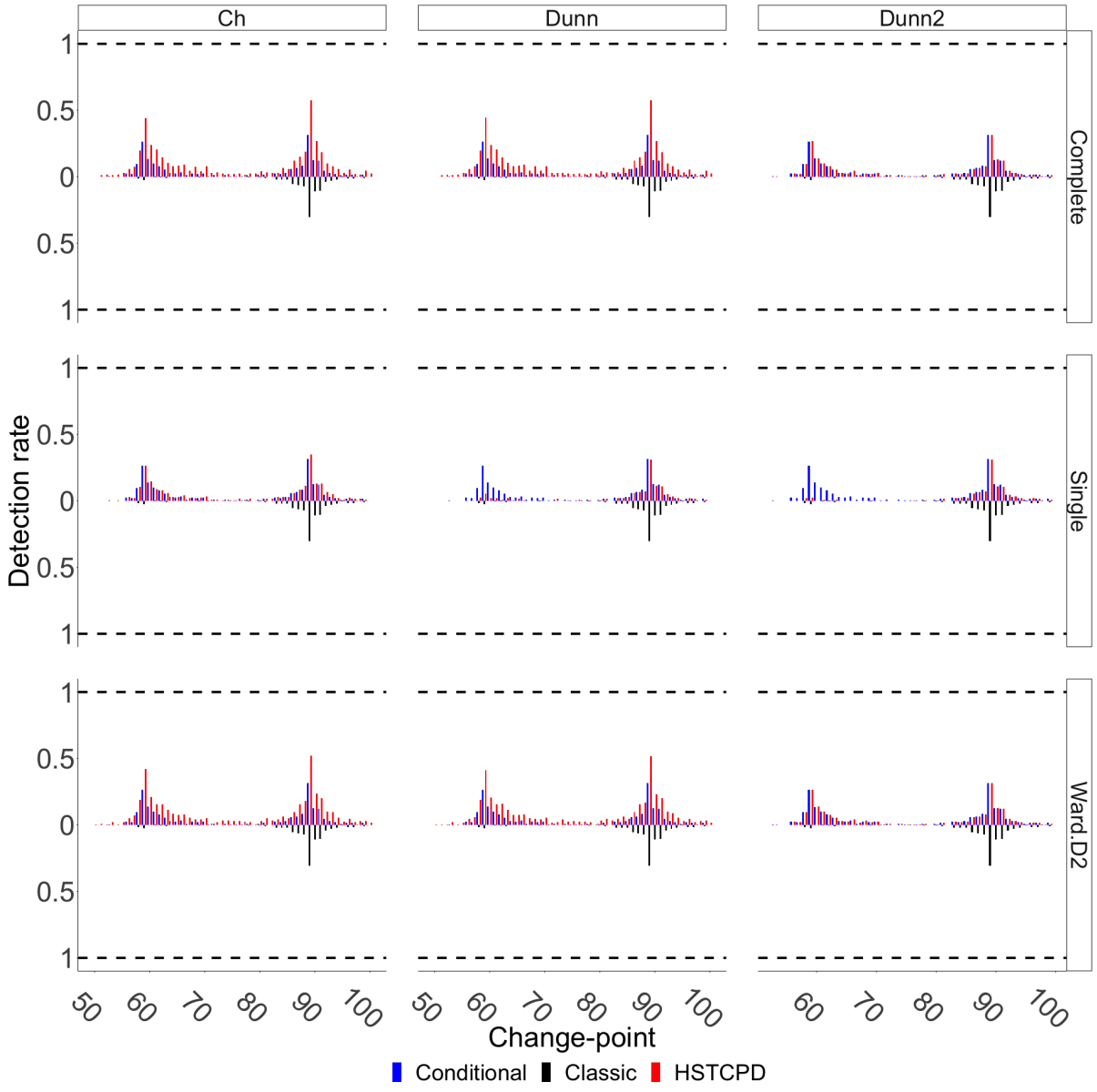

Figure S24: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

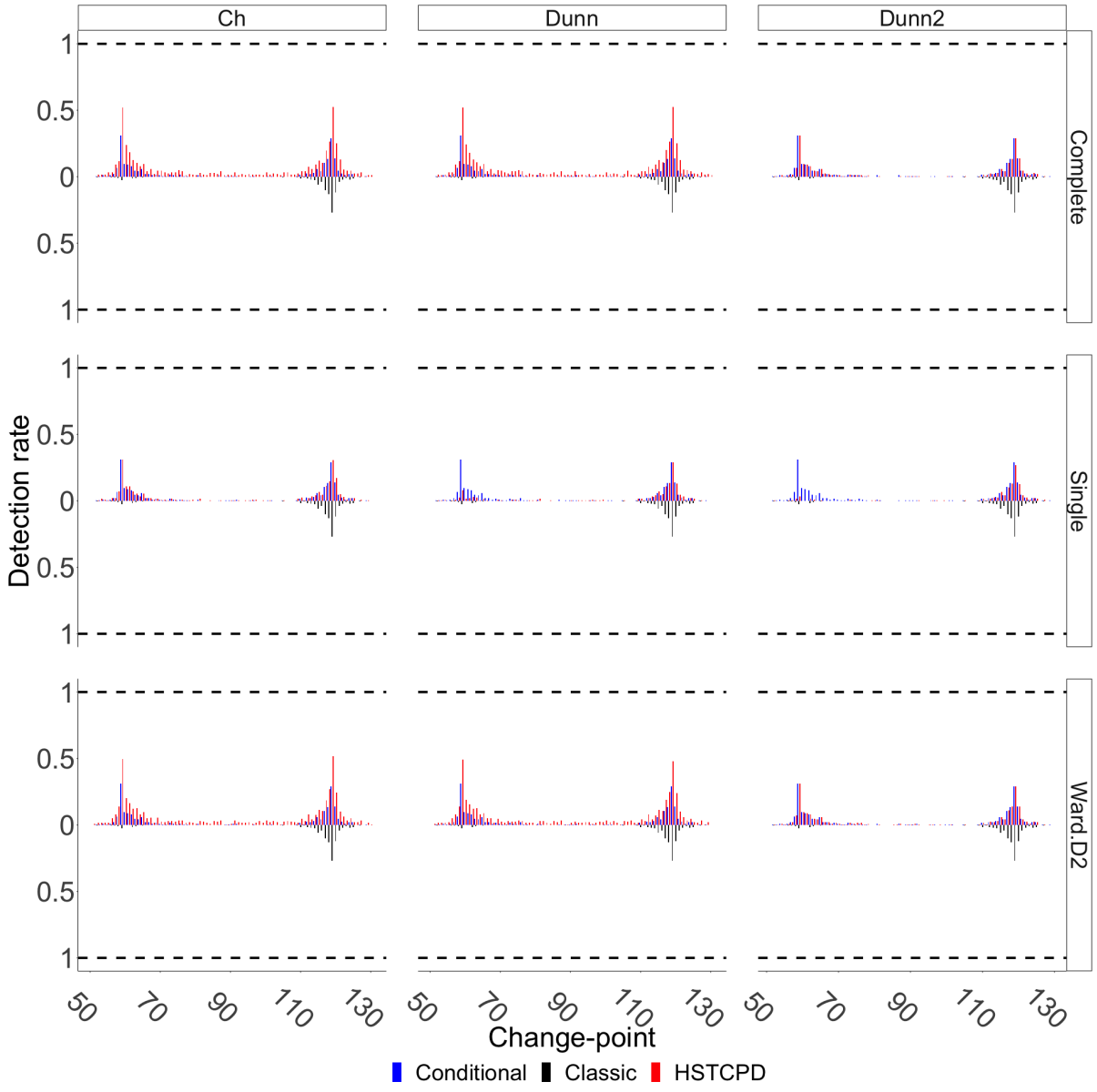

Figure S25: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.4.2 Scenario II

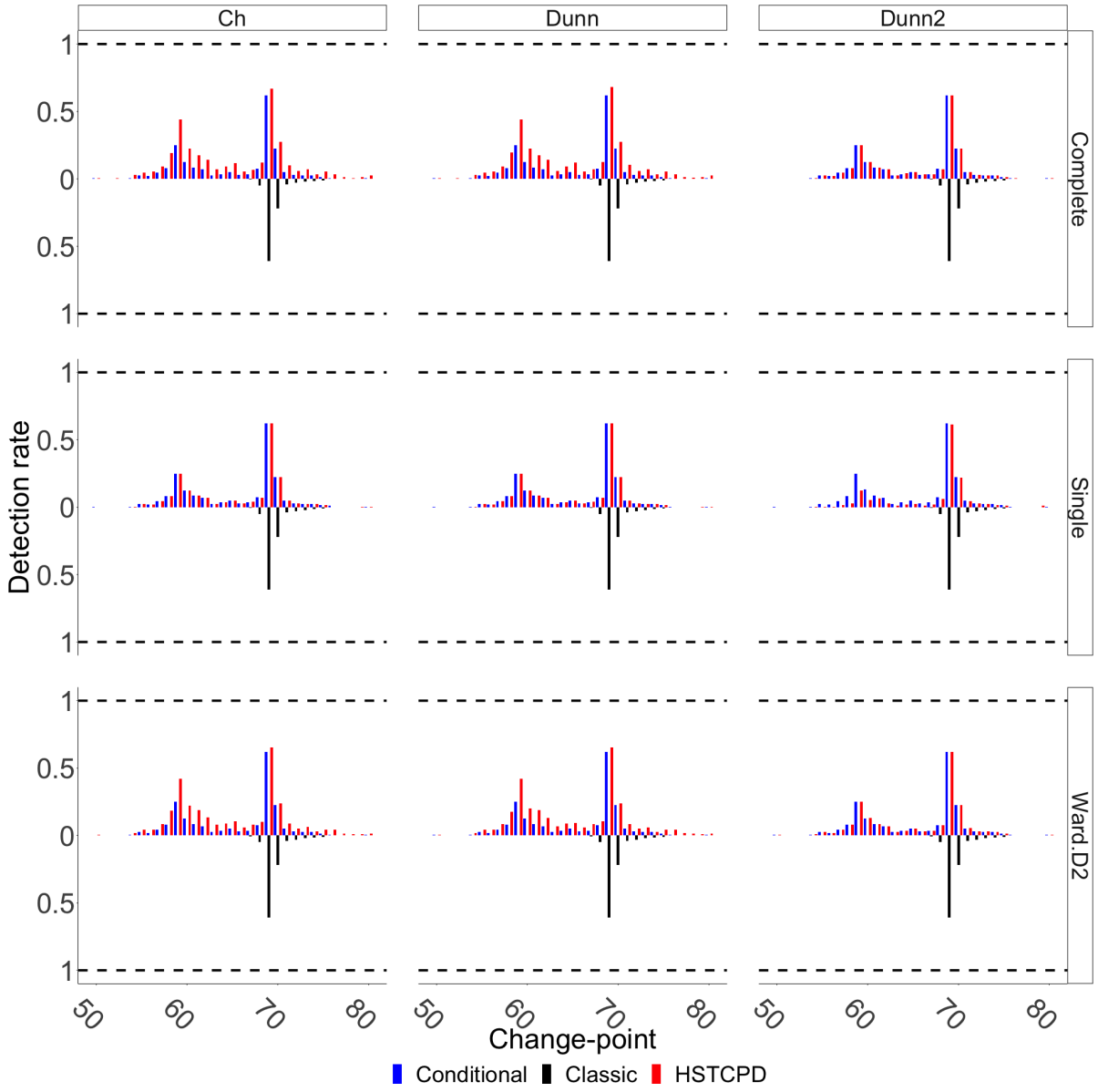

Figure S26: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

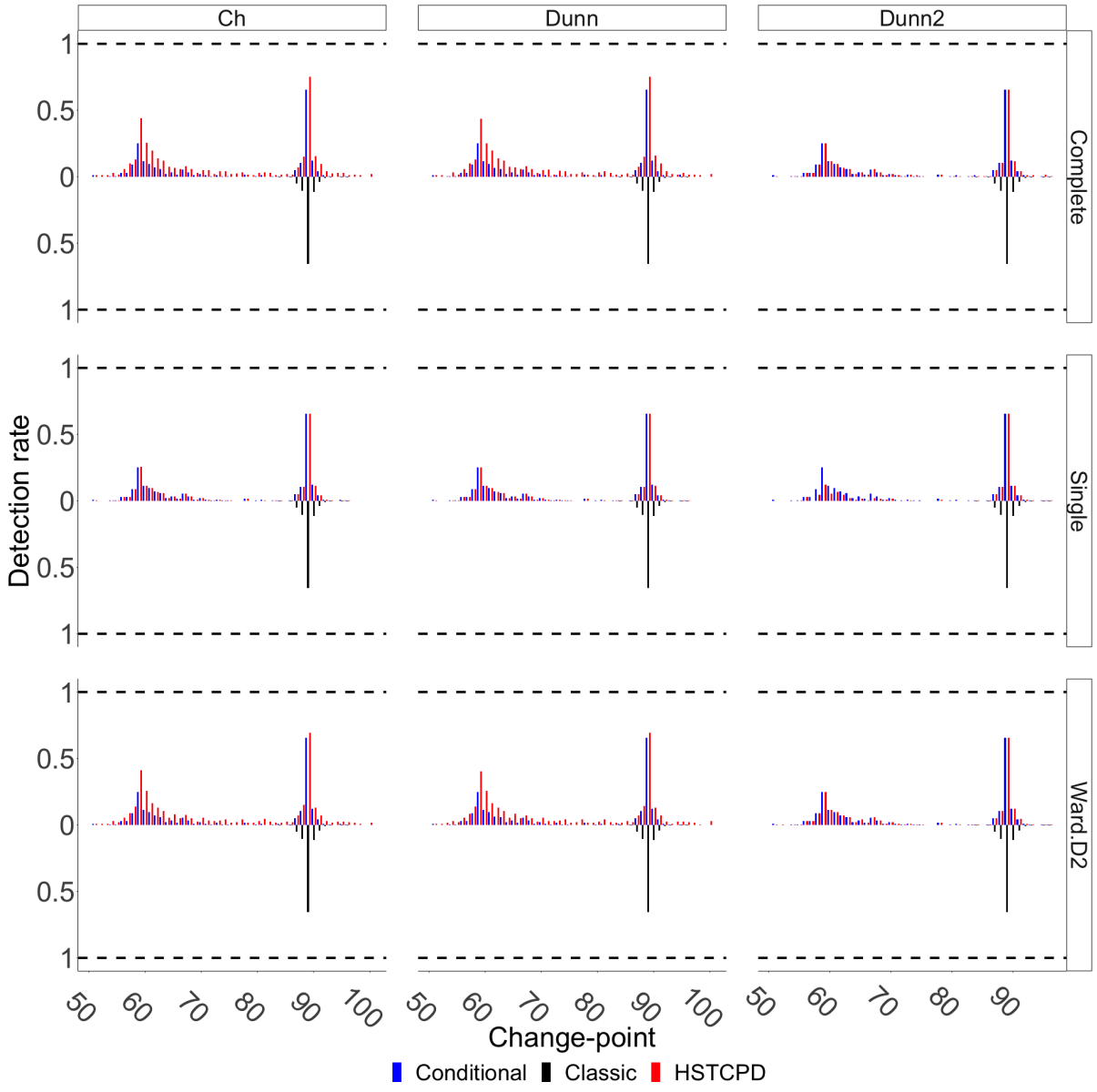

Figure S27: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

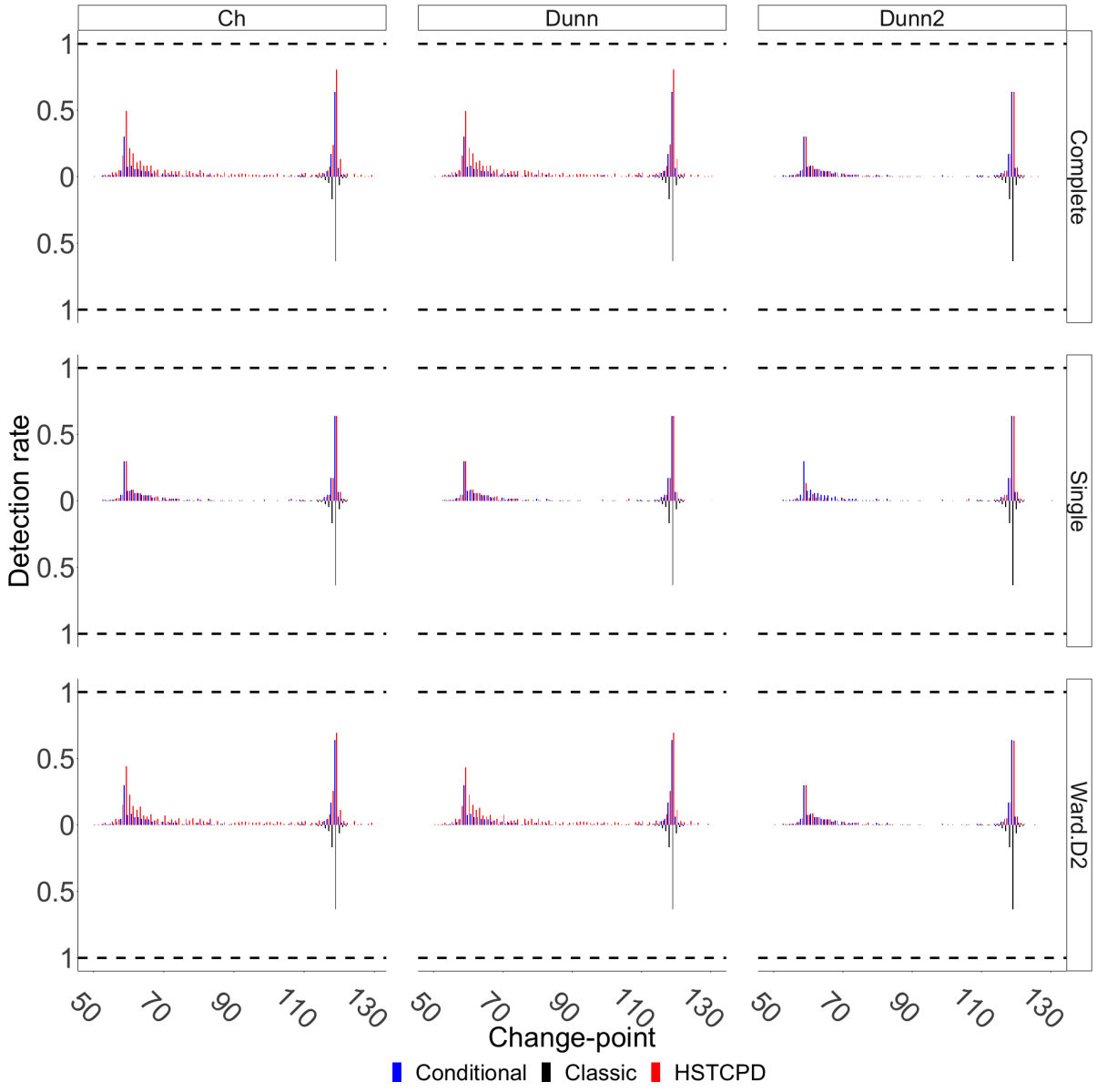

Figure S28: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.4.3 Scenario III

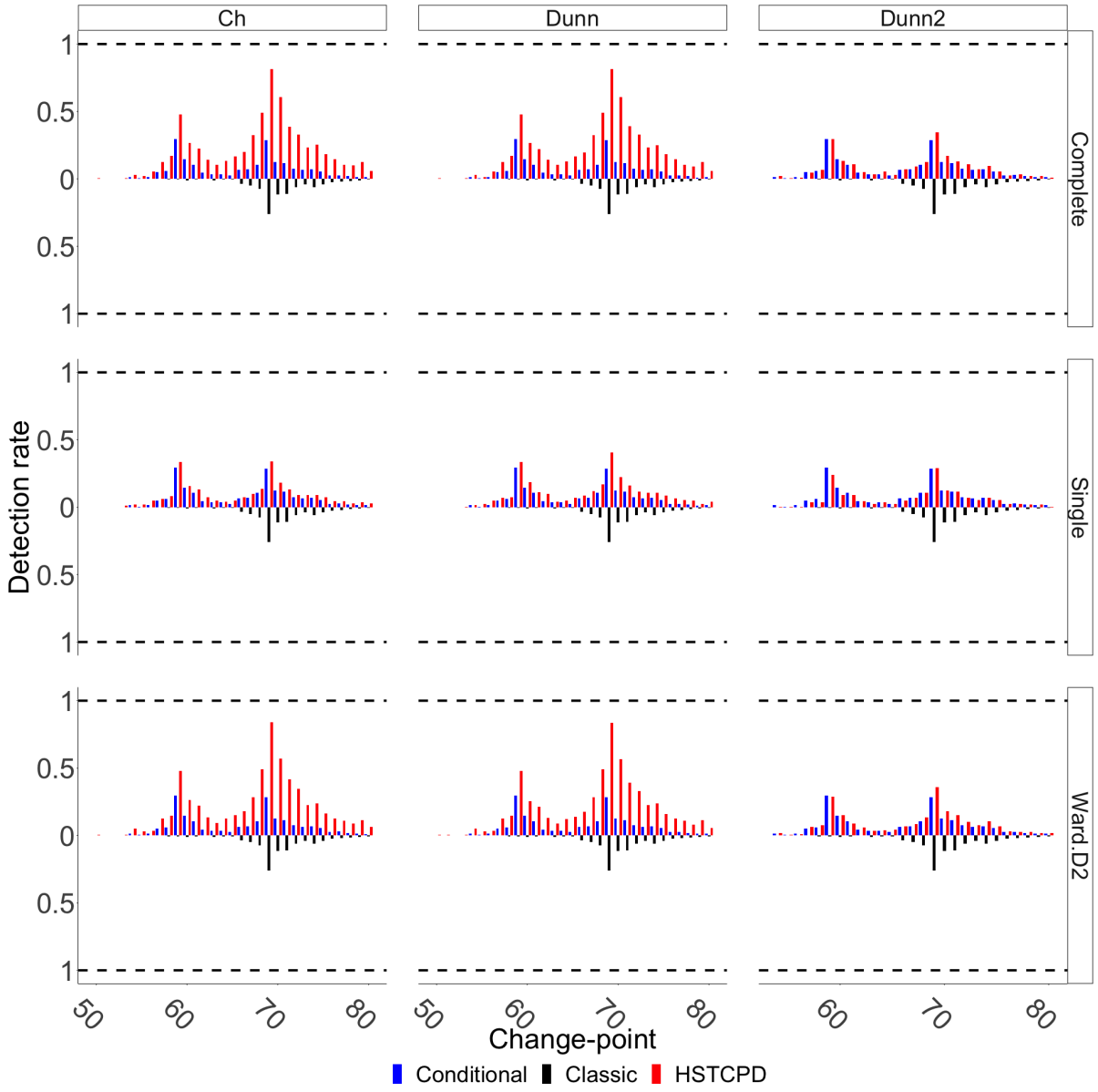

Figure S29: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

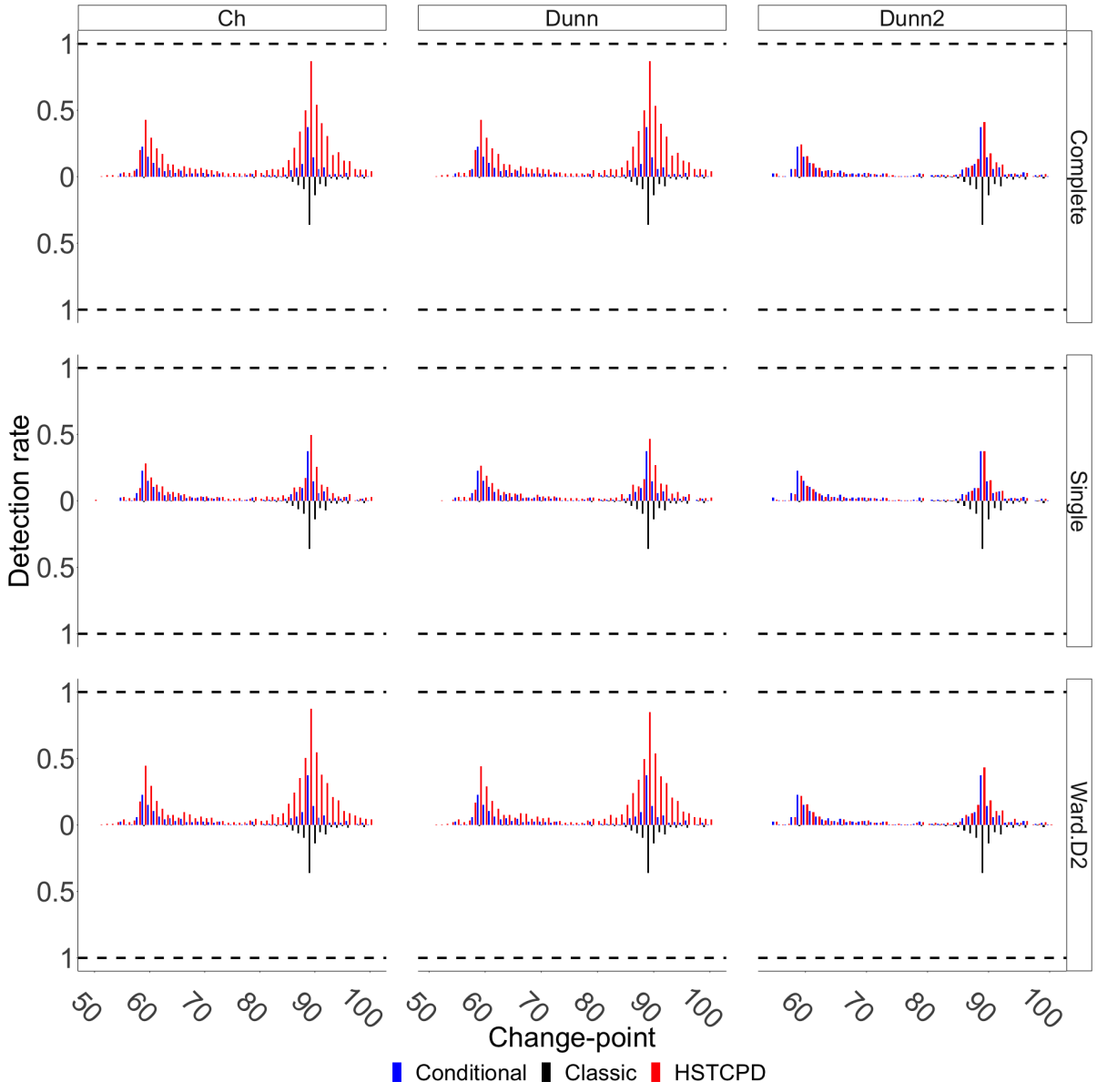

Figure S30: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

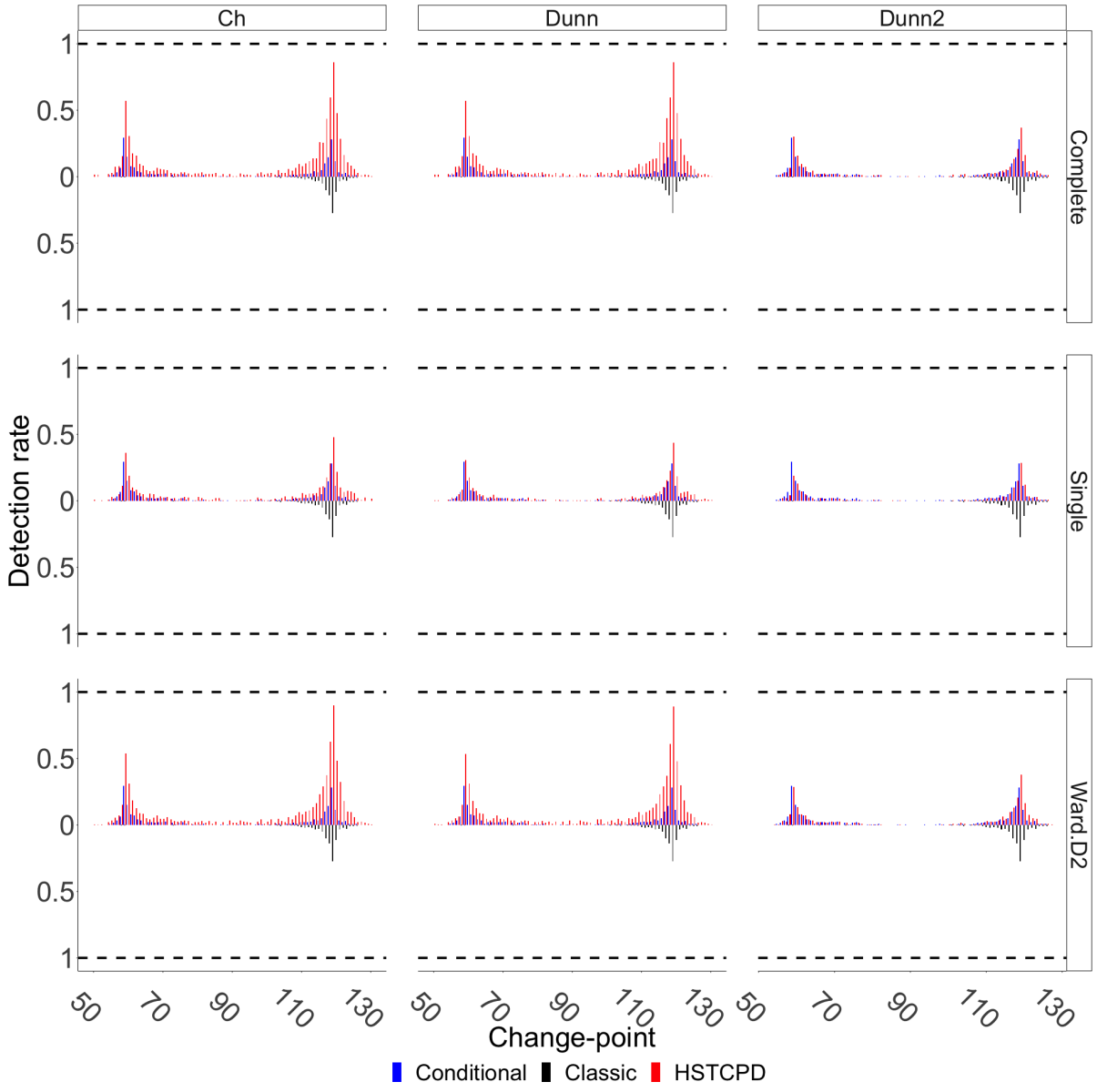

Figure S31: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.5 Geompcp

S2.5.1 Scenario I

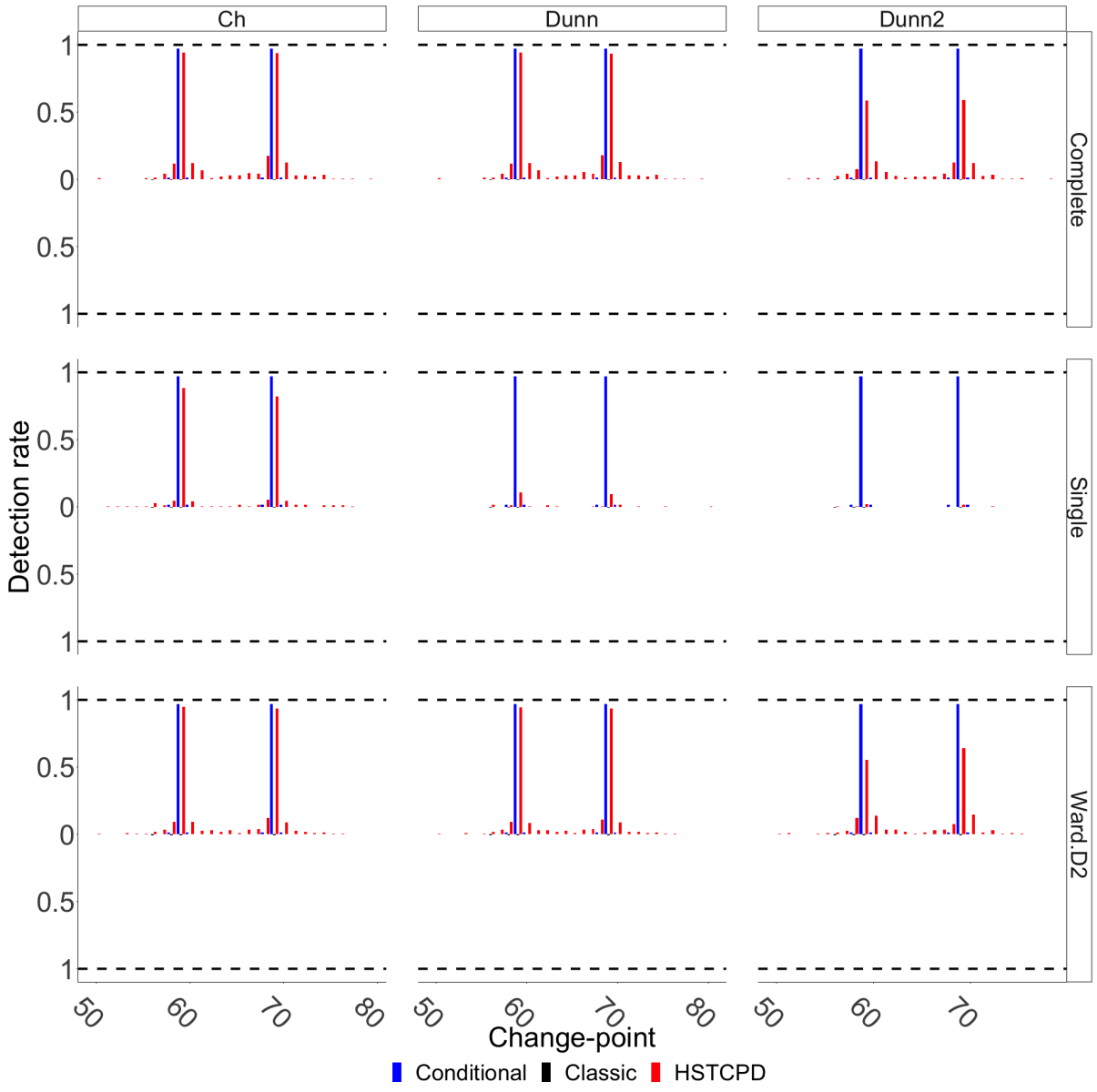

Figure S32: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

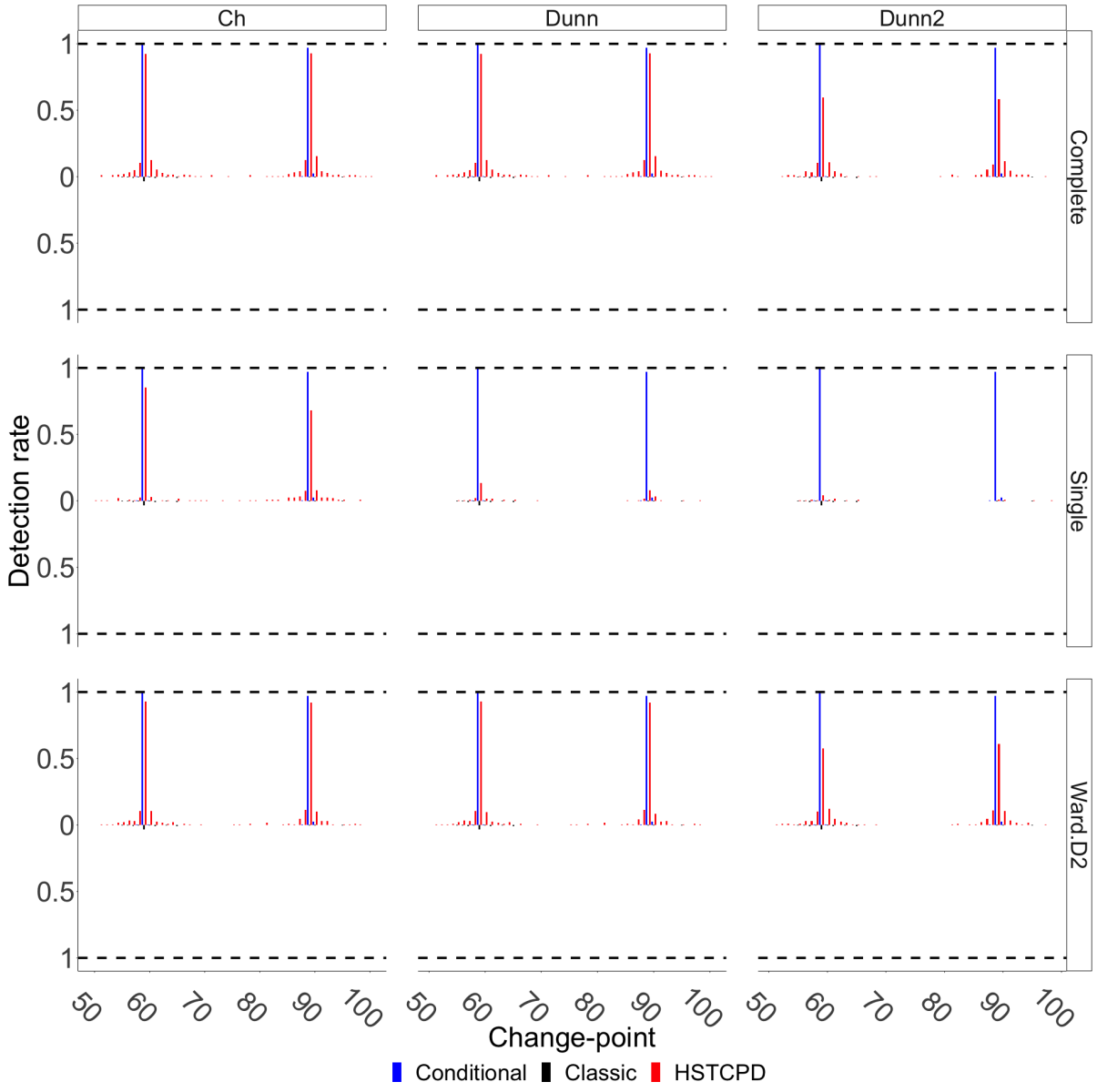

Figure S33: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

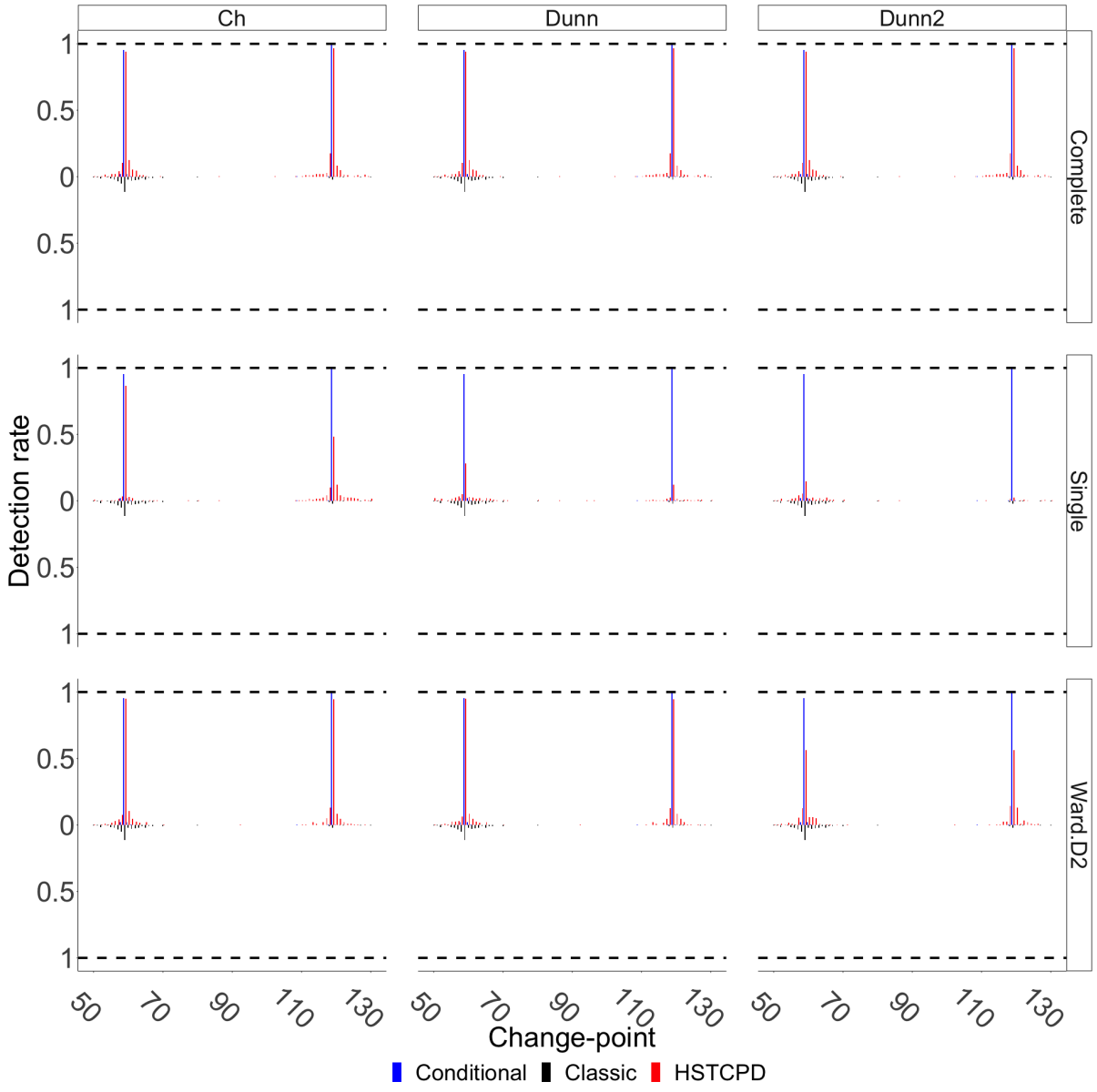

Figure S34: Detection rate based on 200 simulations from Scenario I, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.5.2 Scenario II

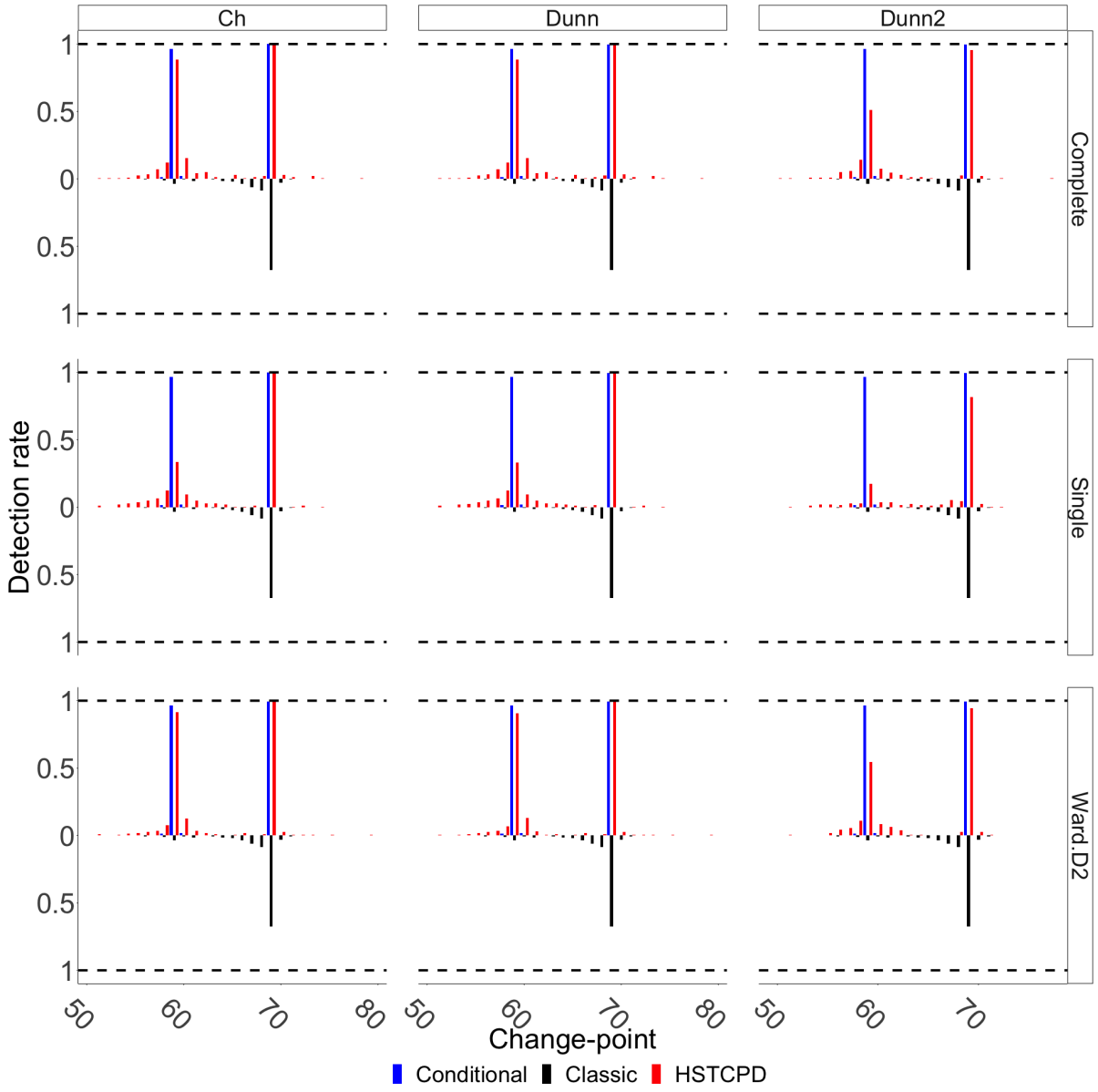

Figure S35: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

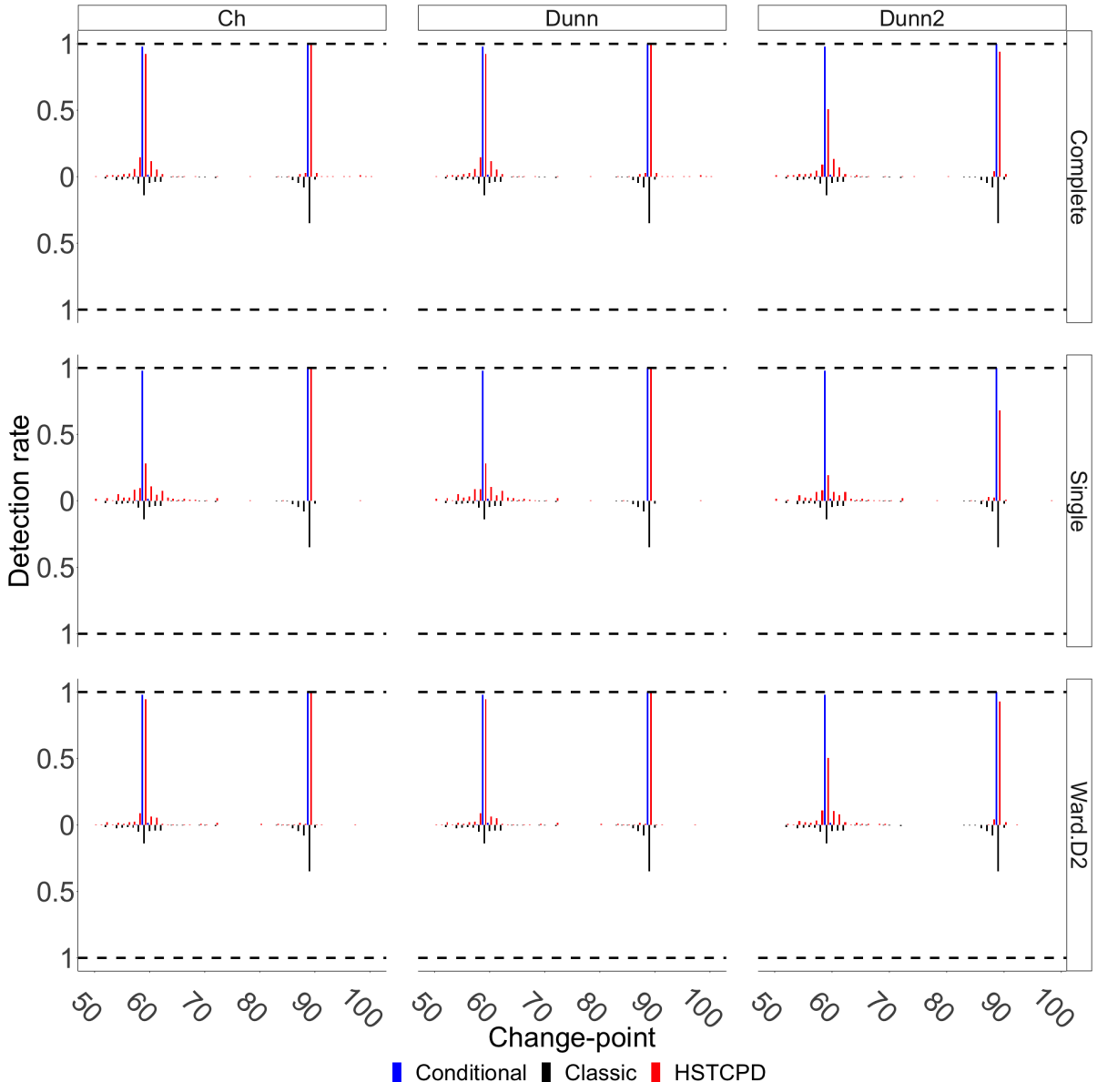

Figure S36: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

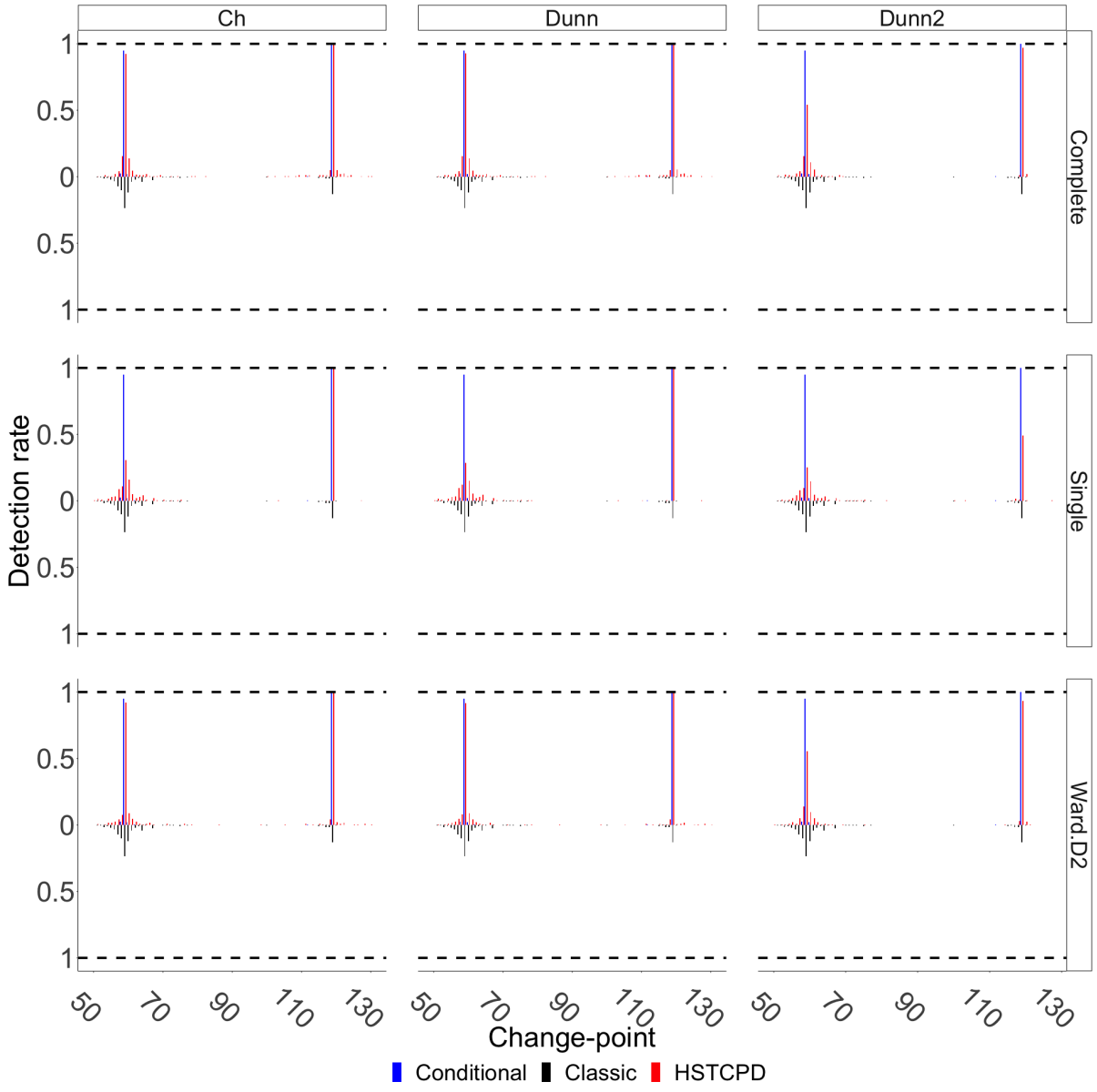

Figure S37: Detection rate based on 200 simulations from Scenario II, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.5.3 Scenario III

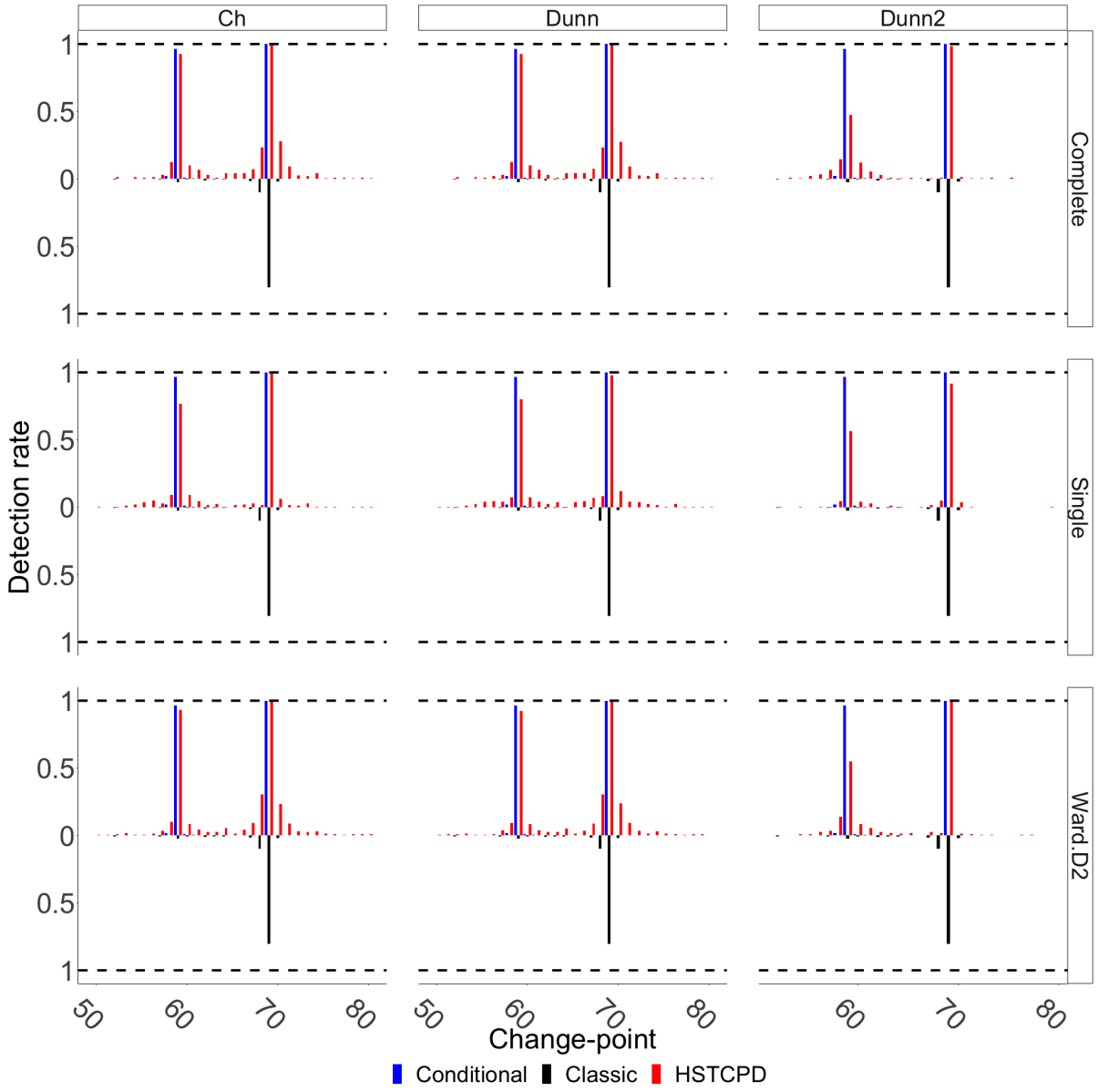

Figure S38: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 70.

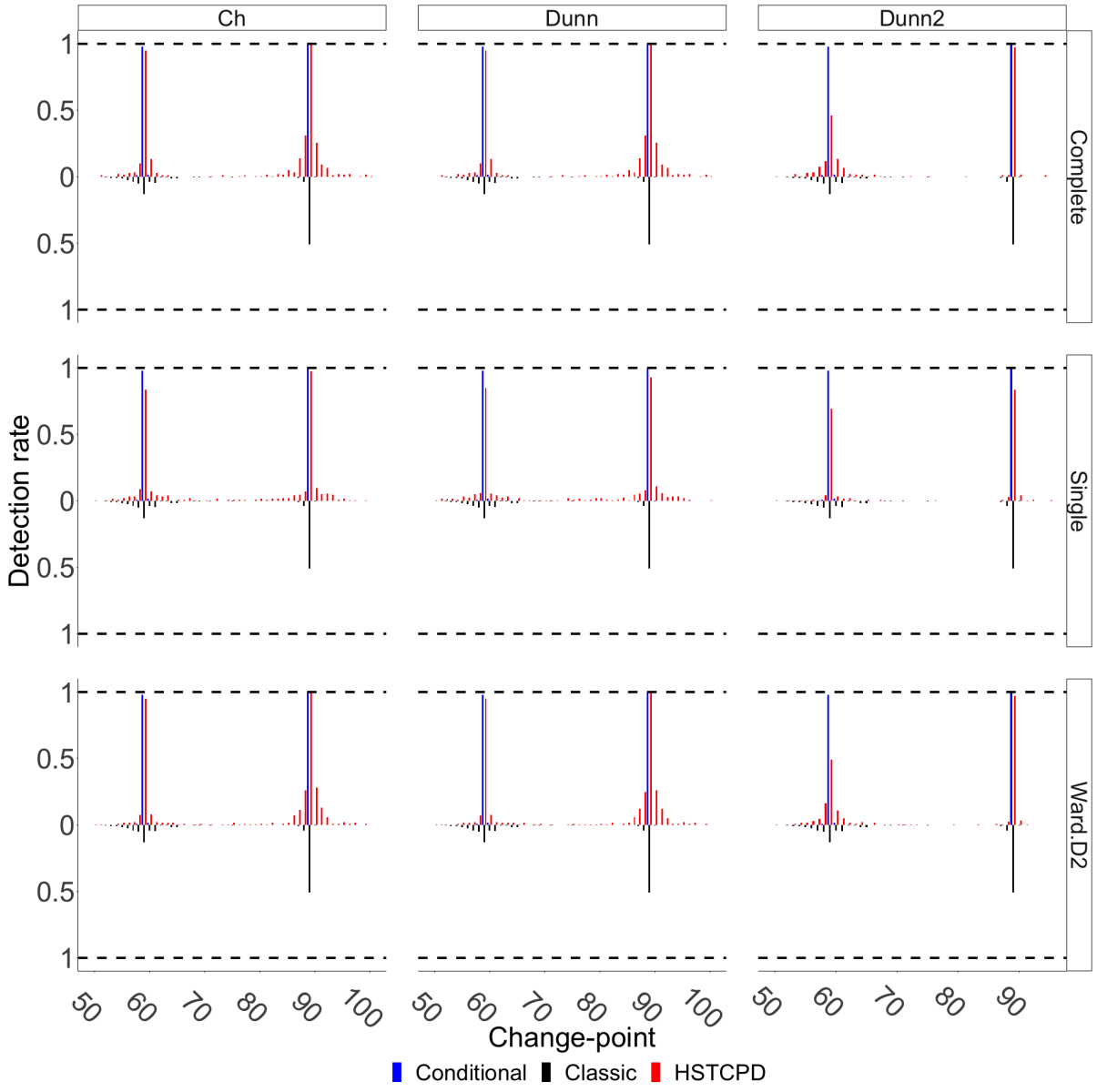

Figure S39: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 90.

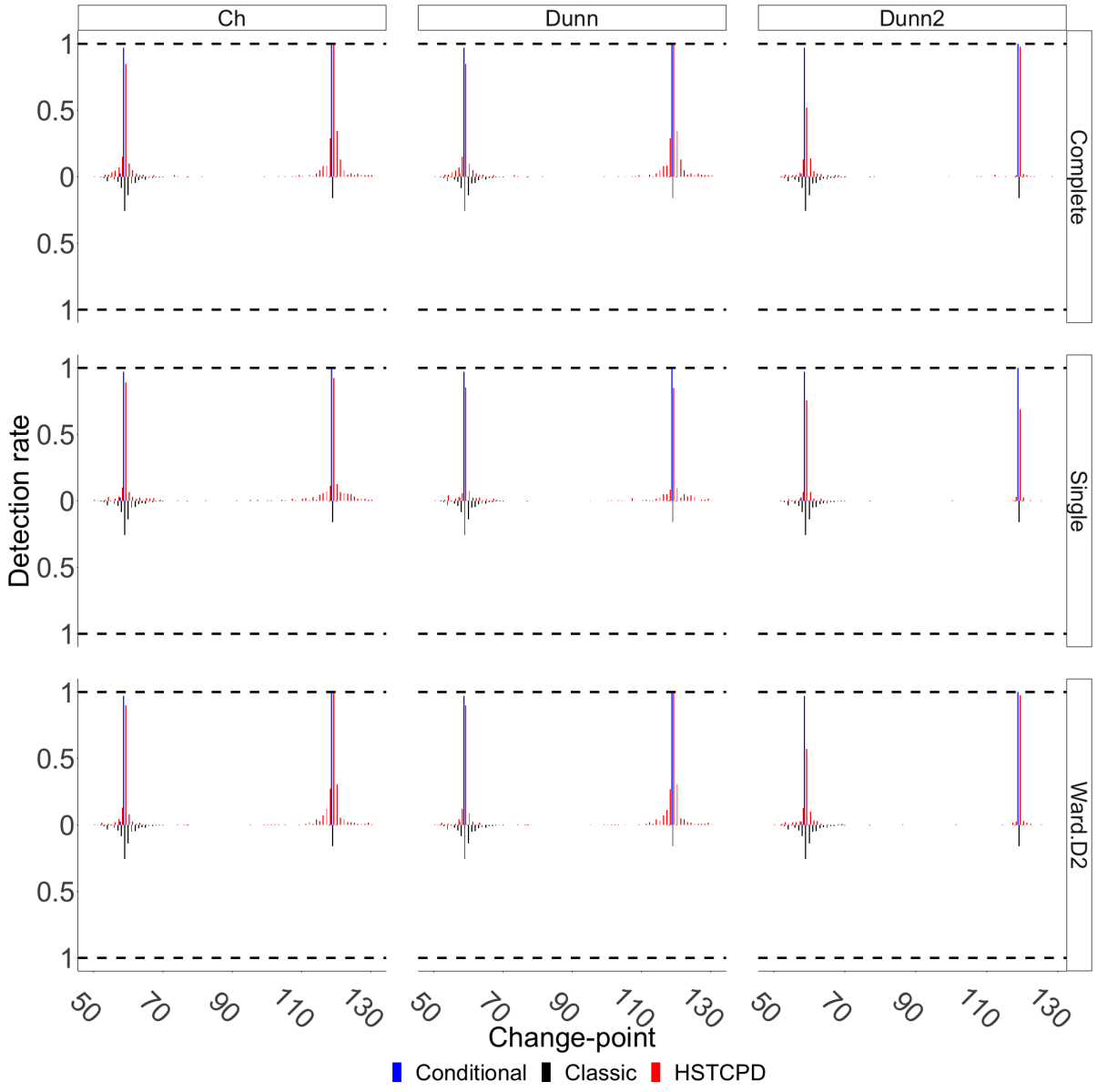

Figure S40: Detection rate based on 200 simulations from Scenario III, with spatially correlated pixel time series, when the change-point for cl_2 occurs at time 120.

S2.6 Multiple mean change-point detection

In this Section, under the existence of spatial correlation between pixel time series and using clouds of the same sizes as scenarios I and II, we consider a situation where we have two mean change-points per each cloud. The considered change-points in different clouds act in different directions and they are all of magnitude one (similar to scenario I). There is an upward change at time 60 and a downward change at time 120 in cl_1, whereas in cl_2 there is a downward change at time 70 and an upward change at time 130. In other words, after the second change-points time series return to their past state (i.e. the state before their first change-point). In order to perform clustering, we make use of the Ward.D2 linkage functions and Dunn criteria to estimate optimal number of clusters, and concerning the change-point method, we employ the ecp approach (Matteson and James, 2014). Figure S41 shows the detection rate of change-points where one can easily see that the Classic approach can hardly detect the actual change-points and its detection rate for all four change-points is around 0.5. However, the HSTCPD approach has a detection rate of 0.97 for all actual change-points which is generally the same as that of Conditional.

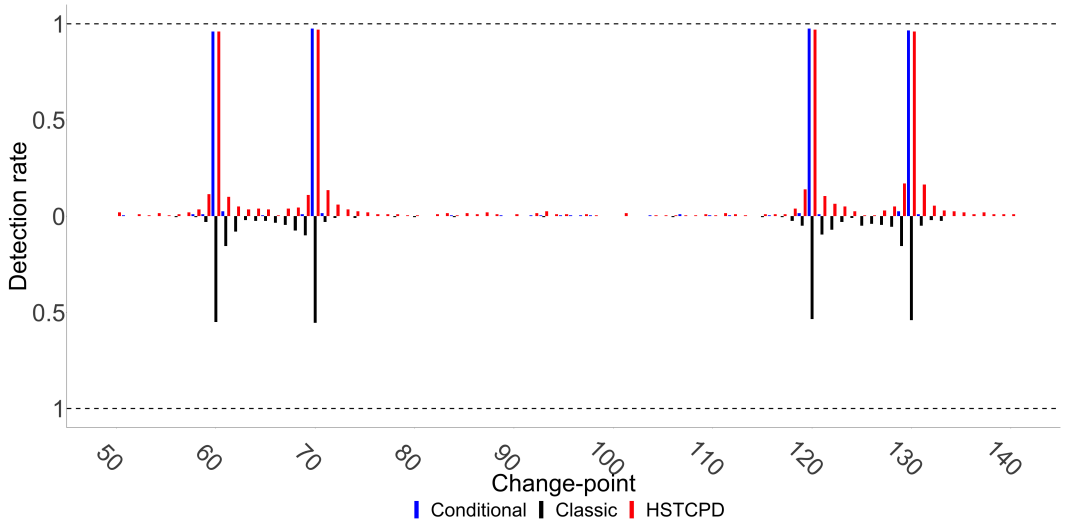

Figure S41: Detection rate when there are change-points 60(upward)/120(downward) in cl_1 and 70(downward)/130(upward) in cl_2. All changes are of magnitude one, and spatially correlated pixel time series are considered for this example.

S2.7 Simultaneous mean and variance change-point detection

This section is devoted to the case where in each cloud there exist changes in both mean and variance, pixel time series are completely independent, and clouds are of the same sizes as scenarios I and II. Regarding the changes in mean parameter, there is an upward change at time 60 in cl_1 and a downward change at time 70 in cl_2; both changes are of magnitude one similar to scenario I. The changes corresponding to variance occur at times 120 and 130 in cl_1 and cl_2, respectively, in which variance in cl_1 (cl_2) changes from one to 0.3 (3). In order to perform clustering before turning to the change-point detection part, similar to Section S2.6, we make use of Ward.D2 in combination with Dunn.

Figure S42 shows the detection rate of the ecp approach (Matteson and James, 2014) where one can see that the Classic approach could hardly detect the actual changes; the mean change at time 60 (70) is detected only 19.5% (43.5%) of the time and the variance change at time 120 (130) is detected 15% (0.5%)of times. The performance of HSTCPD in detecting the actual mean changes is generally as good as the Conditional one; the mean change at time 60 (70) is detected 97% (94%) of times. In terms of changes in variance, the change at time 120 is detected 98.5% (98%) by HSTCPD (Conditional) while the change at time 130 is detected 38.5% (99.5%). It is also visible that the not well detecting the borders of clouds has led HSTCPD to detect some change-points with an error margin of few units.

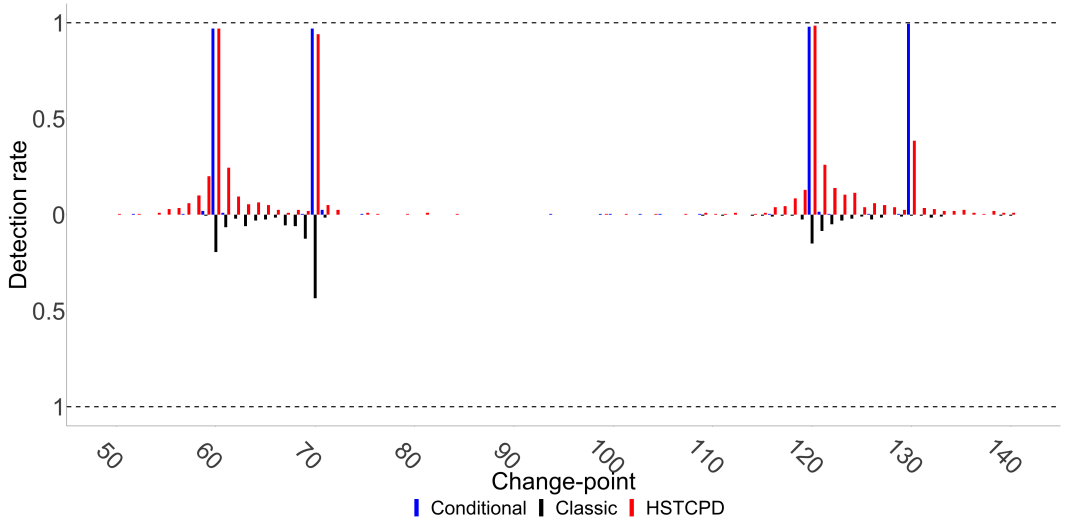

Figure S42: Detection rate of ecp approach for the change-points 60(mean)/120(variance) in cl_1 and 70(mean)/130(variance) in cl_2. No spatial correlation between pixel time series is considered in this scenario.

Since the method of Grundy et al. (2020) is inherently built to simultaneously detect change-points in both mean and variance when dealing with multivariate data, we next employ the geomcp approach (Grundy, 2020) to further study its performance in this particular case. Figure S43 shows its detection rate in which apart from poor performance of Classic approach compared to HSTCPD/Conditional, one can see that geomcp sometimes mistakenly detects changes in mean as if they were changes in variance and vice versa. For the case of Conditional approach, the change-points 120 and 130, which correspond to variance, are 53% and 61% of times detected as if they were changes in mean. Similarly, the change-points 60 and 70, which correspond to mean, are 20% and 47.5% of times detected as if they were changes in variance. This highlights that not only geomcp can suffer from masking, but it may mistakenly introduce mean changes as changes in variance and vice versa.

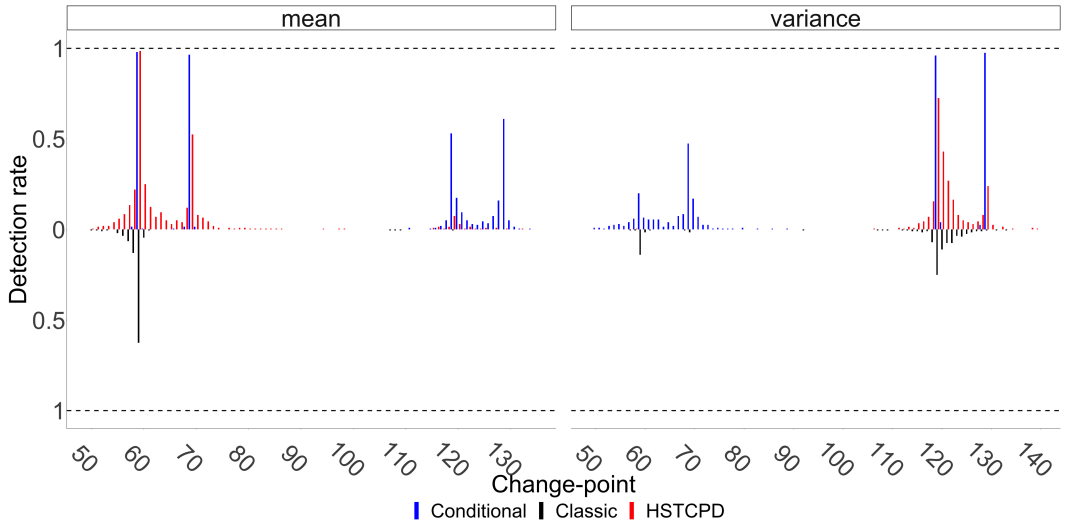

Figure S43: Detection rate of geomcp for the change-points 60(mean)/120(variance) in cl_1 and 70(mean)/130(variance) in cl_2. No spatial correlation between pixel time series is considered in this scenario.

S2.8 Simulation on large images

This section replicates scenario I, as explained in Section 4.1, for large images (i.e. images of the same size as the Spanish LST data and the AWD data) to further study the performance of our proposal, i.e. HSTCPD, when the number of pixel time series is large. To do so, as a showcase, we here only consider the case where there is complete independence between

pixel time series, and concerning the clustering part, similarly to Section S2.6 and S2.7, we make use of Ward.D2 in combination with Dunn. For the case where we simulate time series on images with the same size as the LST data, the two actual change-points happen at times 80 (upward in cl_1) and 90 (downward in cl_2), whereas for the case of AWD the two change-points occur at times 20 (upward in cl_1) and 30 (downward in cl_2). The corresponding time of these changes are chosen to be proportionally similar to the case of the change-points 60 and 70 in scenario I. Regarding the choice of multivariate change-point detection method, we have here employed ecp (Matteson and James, 2014) and geomcp (Grundy et al., 2020) since they have previously showed a better performance than the data-adaptive approach of Liu et al. (2020). Figure S44 shows the estimated clusters for two individual cases together with the locations of cl_1 and cl_2.

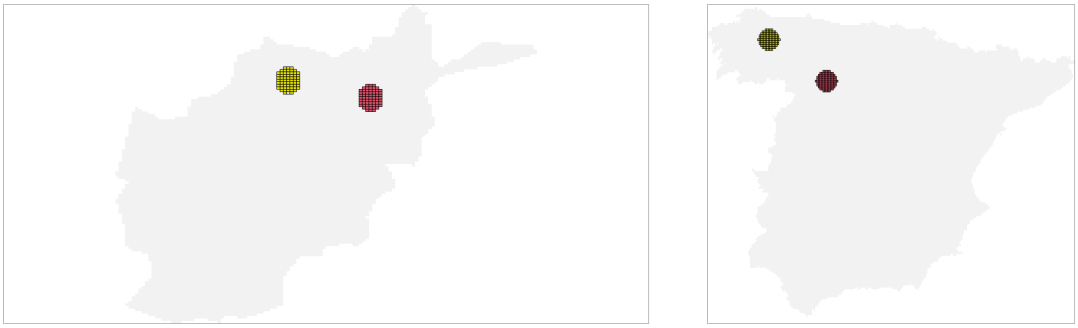

Figure S44: Individual examples (one out of 200 simulations) of the clustering results for Scenario VI. Left: Afghanistan, Right: Spain. Colours show the estimated clusters and grids display the clouds cl_1 and cl_2.

Figure S45 and S46 show the detection rates for both cases. We can see that, similarly to other scenarios, HSTCPD outperforms the Classic approach. More specifically, one can see that HSTCPD has its biggest impact on the geometrical mapping approach of Grundy et al. (2020). In both cases, the Classic approach, represented by geomcp, does not detect any of the actual change-points. However, for the image time series simulated over Spain, the detection rates of HSTCPD, for the actual change-points 80 and 90, are equal to 100%. Concerning the case of the image time series simulated over Afghanistan, the detection rates of HSTCPD, for the actual change-points 20 and 30, are 100% and 78.5%.

With respect to the detection rates of the ecp approach (Matteson and James, 2014), for the data simulated over Spain, the actual change-points 80 and 90 are detected at a rate of

44% and 43%, whereas HSTCPD detects these changes 100% of the time. For the image time series simulated over Afghanistan, the change-points 20 and 30 are detected 33% and 57% of the time, while HSTCPD detects these changes 75% and 81.5% of the time. This slightly poorer performance is due to the fact that the actual change-points are near the tails of time series, see Moradi et al. (2022) for details.

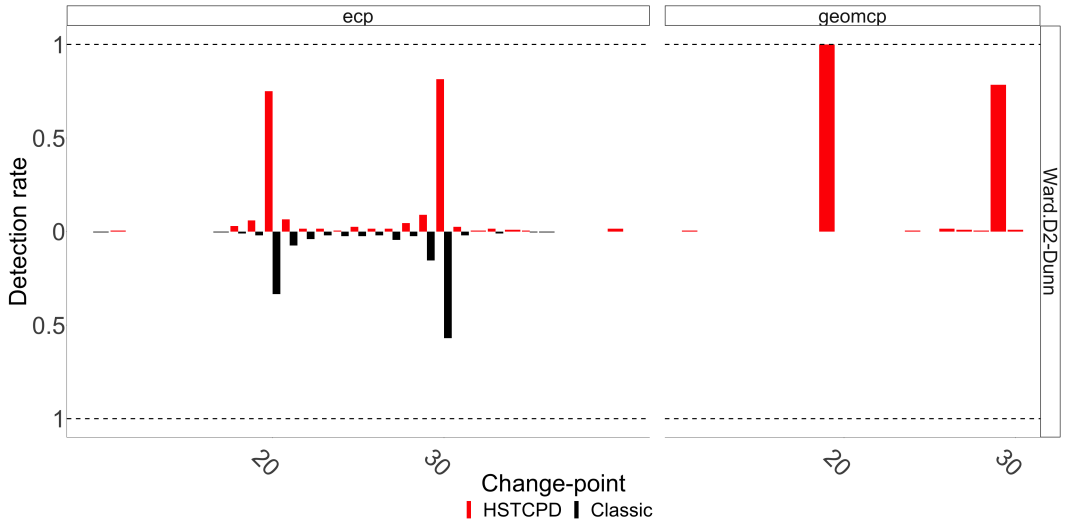

Figure S45: Detection rate for scenario VI when time series are simulated over the images of the same size as AWD in Afghanistan.

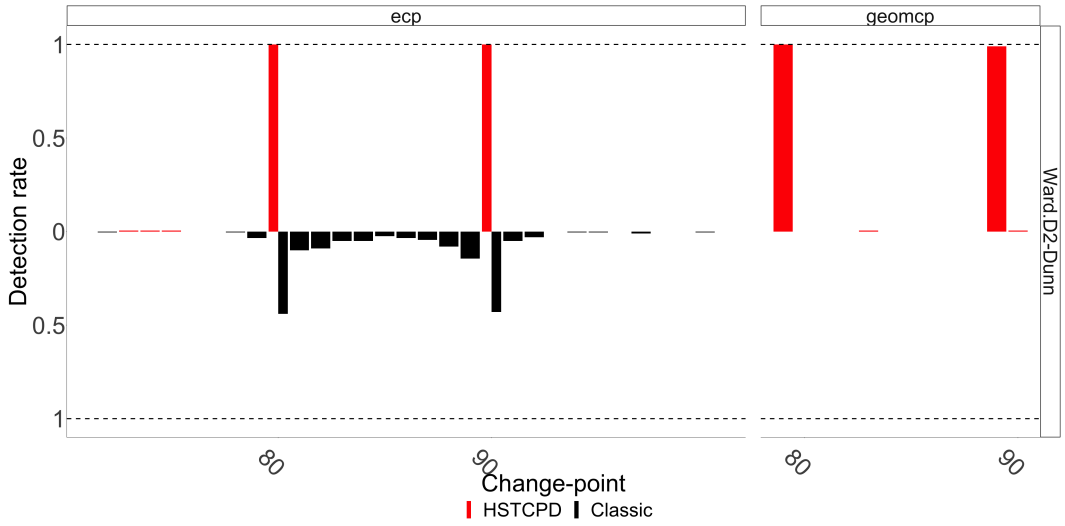

Figure S46: Detection rate for the scenario VI when time series are simulated over the images of the same size as LST in Spain.

S3 Real data

In this Section we present additional information regarding our real datasets.

S3.1 Land Surface Temperature (LST) in Spain from 2000 to 2021

The Modis images are provided in a sinusoidal projection, which represents pixels of different sizes, i.e. larger in the north when approaching the Poles and smaller in the south when approaching the Equator. Therefore, the images are re-projected into the UTM coordinate system to ensure that all the pixels represent the same size.

Figure S47 shows the LST data for the study period. All the detected change-points per cluster are presented in Figure S48 and for additional intuition we also present the average elevation within the estimated clusters in Figure S49. We can see that the general distribution of change-points is spatially dependent, such that there are clusters with different numbers of change-points from zero to 12. Moreover, there are few changes that are detected within several clusters, but generally there is no agreement between the clusters regarding the number and position of the change-points. Note that following a no pre-clustering approach allows the change-points which are visible in the majority of clusters to dominate/conceal other actual change-points.

S3.2 Afghan War

Figure S50 shows the monthly spatial point patterns corresponding to the AWD dataset. The most dense pattern corresponds to August 2009 with 3619 events, and the most sparse pattern is that of April 2004 with 81 events. In Figure S51 we present the monthly estimated intensity surfaces for such point patterns. One can observe an increasing tendency in the east of Afghanistan from the middle of 2006 which apparently reaches its peak around May/June 2007. Thereafter, it experiences a decreasing trend up to July 2008. Then, the estimated intensity in the east and south-east seems to go through some spatially dependent variations including an increasing tendency which starts in June 2009 and reaches its maximum around August 2009, and then slightly decreases. One can further see that the temporal behaviour of the estimated intensity is clearly spatially dependent, i.e. the spatial distribution tends to locally change with time. Figure S52 presents the detected change-points within the

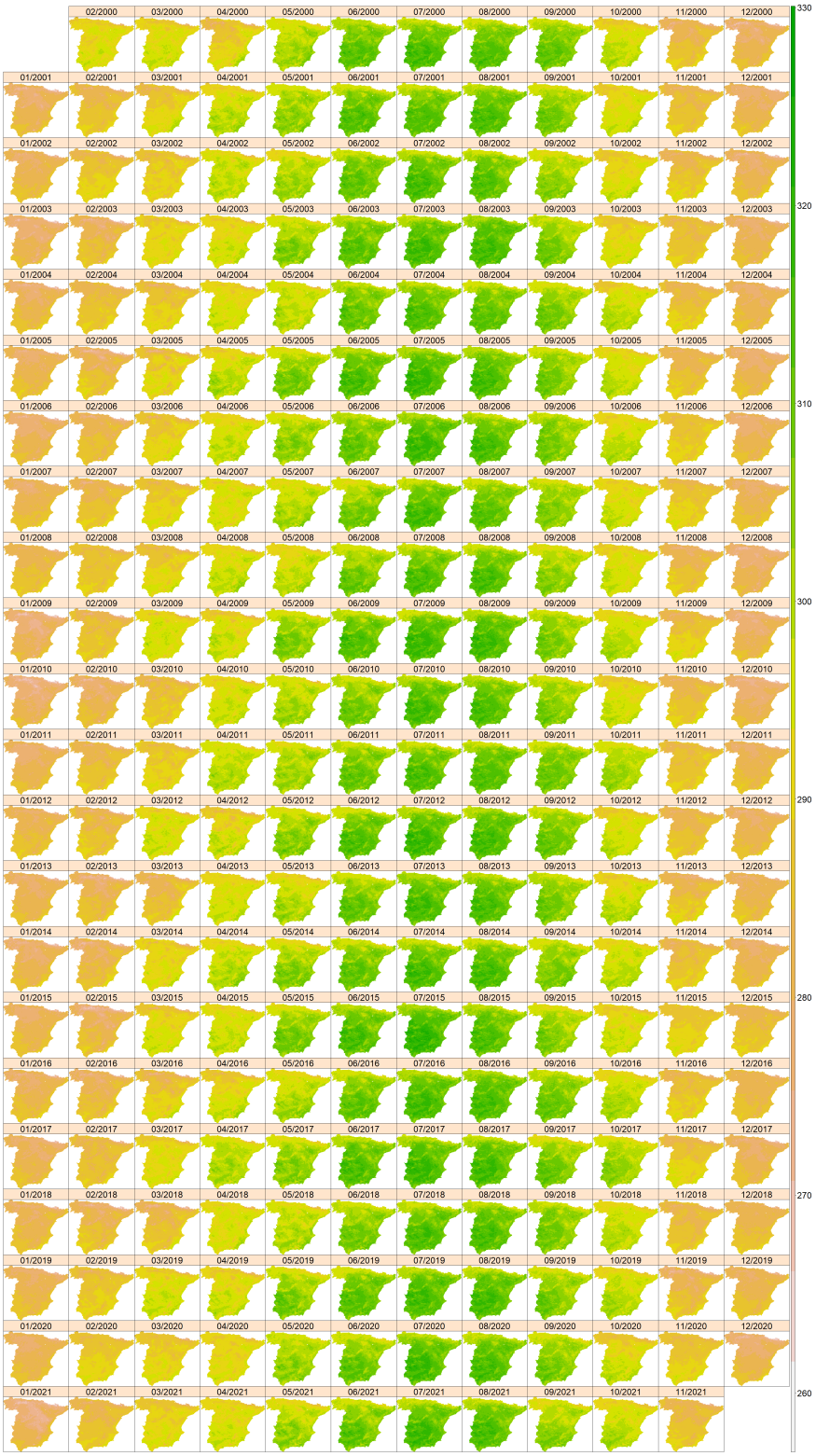

Figure S47: LST data in Spain from the year 2000 to 2021. Values are LST based on Kelvin unit.

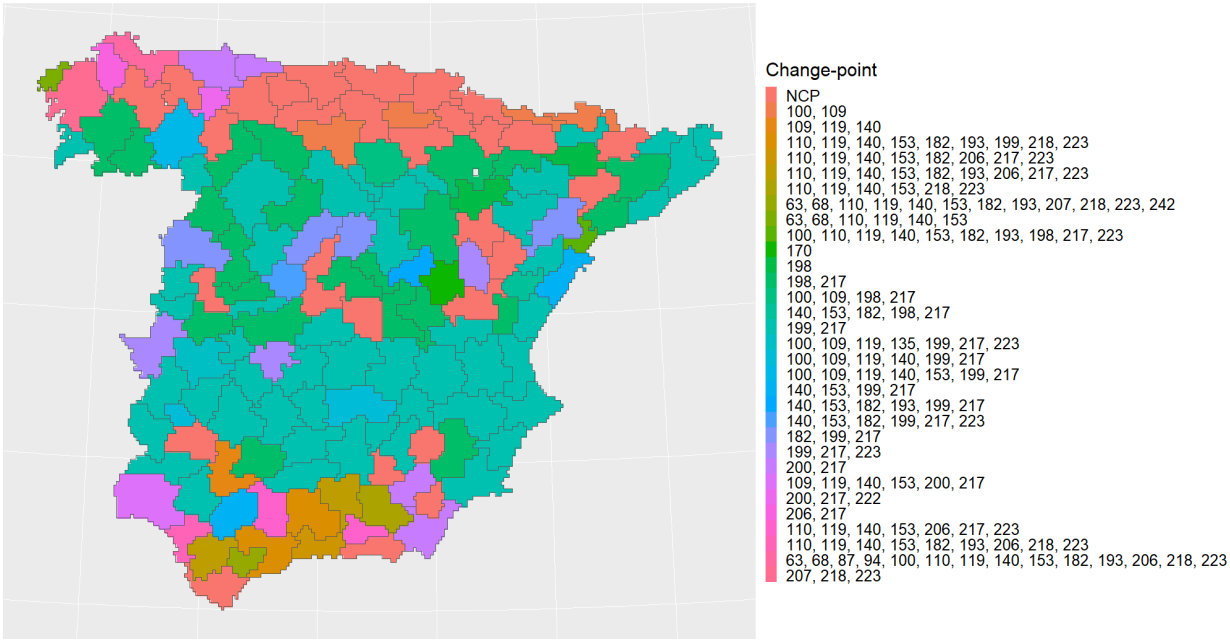

Figure S48: Detected change-points for the LST dataset in Spain within the estimated clusters. Values are change-points.

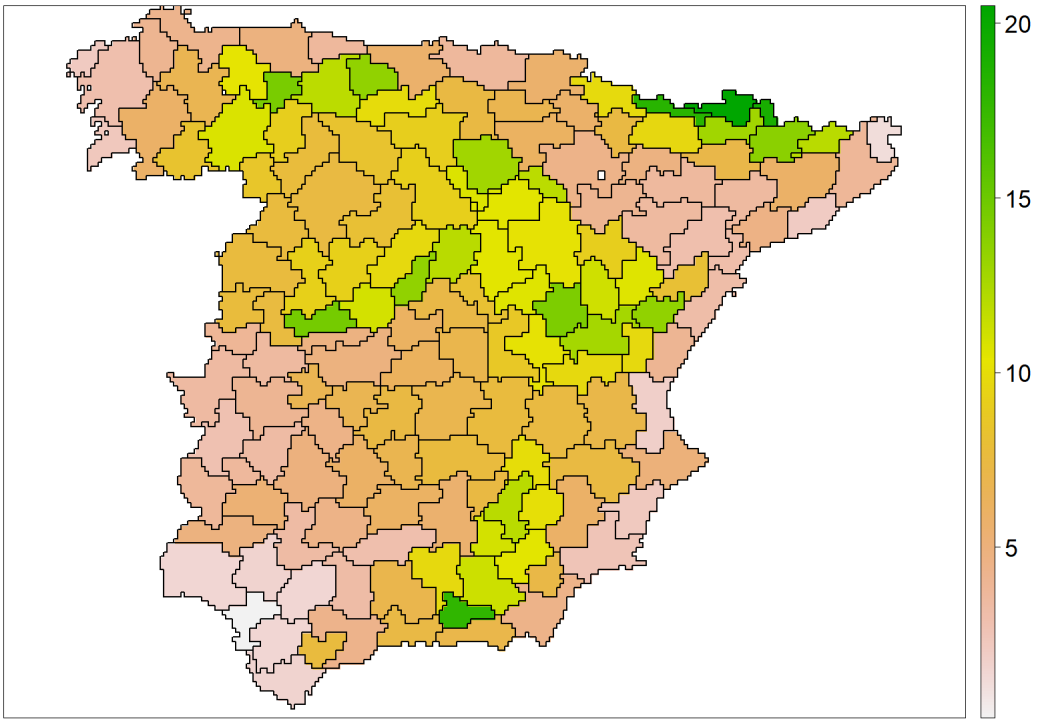

Figure S49: Average elevation in the estimated clusters. Values are elevation divided by 100.

estimated clusters in which one can see that neighbour clusters experience similar changes. Although there are some changes that are detected in several clusters, we can see that there is no general agreement concerning the detected change-points within the estimated clusters which is not visible if one does not perform pre-clustering/classification prior to change-point detection.

Figure S50: The AWD data. Rows from top to bottom represent the years 2004 to 2009, respectively. Columns from left to right represent the months January to December, respectively.

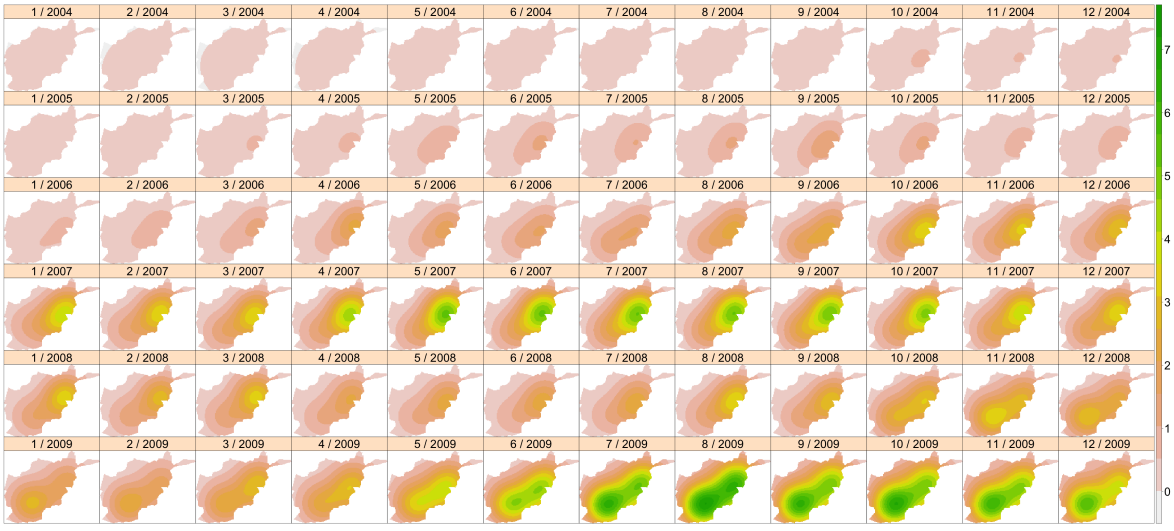

Figure S51: Monthly estimated intensity images for AWD data in Figure S50. Intensity values are points per 10^9 meters.

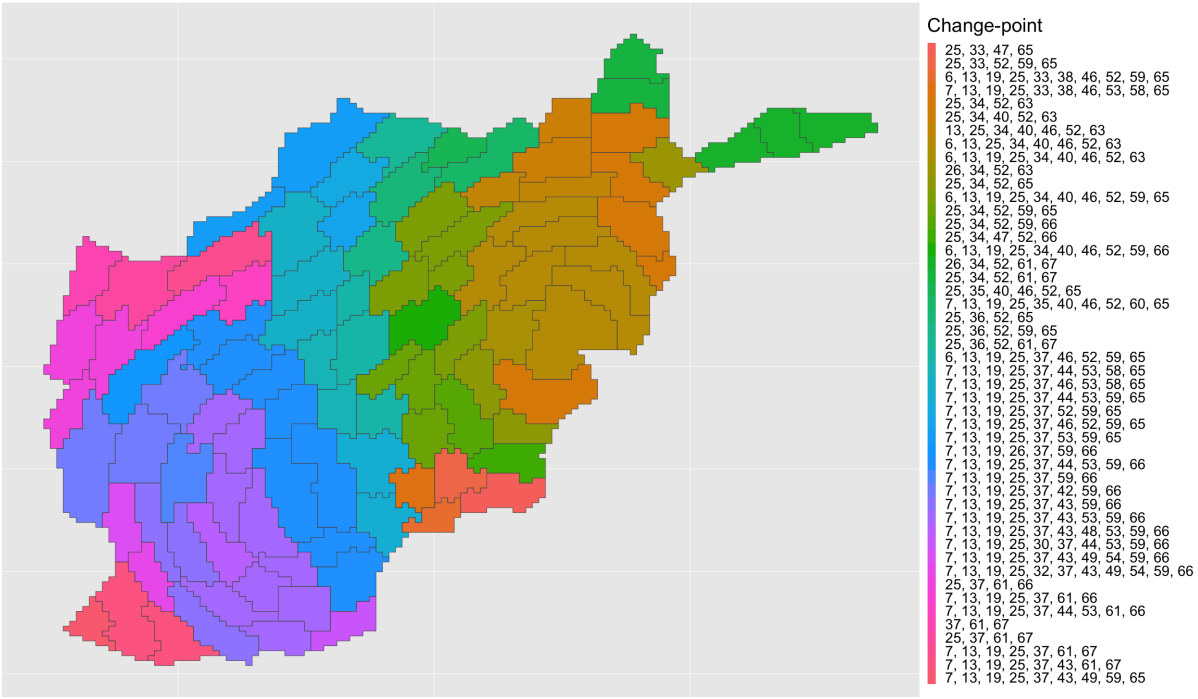

Figure S52: Detected change-points for the AWD data within the estimated clusters. Values are change-points.

References

Aston, J. A. and Kirch, C. (2012), 'Detecting and estimating changes in dependent functional data', *J Multivariate Anal* **109**, 204–220.

Baddeley, A., Rubak, E. and Turner, R. (2015), *Spatial Point Patterns: Methodology and Applications with R*, CRC Press.

Caliński, T. and Harabasz, J. (1974), 'A dendrite method for cluster analysis', *Communications in Statistics-Theory and Methods* **3**(1), 1–27.

Chaudhuri, S., Moradi, M. and Mateu, J. (2021), 'On the trend detection of time-ordered intensity images of point processes on linear networks', *Communication in Statistics- Simulation and Computation* .

Dunn, J. C. (1974), 'Well-separated clusters and optimal fuzzy partitions', *Journal of Cybernetics* **4**(1), 95–104.

Giraldo, R., Delicado, P. and Mateu, J. (2012), 'Hierarchical clustering of spatially correlated functional data', *Statistica Neerlandica* **66**(4), 403–421.

Grundy, T. (2020), *changepoint.geo: Geometrically Inspired Multivariate Changepoint Detection*. R package version 1.0.1.

Grundy, T., Killick, R. and Mihaylov, G. (2020), 'High-dimensional changepoint detection via a geometrically inspired mapping', *Stat. Comput.* **30**(4), 1155–1166.

Halkidi, M., Batistakis, Y. and Vazirgiannis, M. (2001), 'On clustering validation techniques', *Journal of Intelligent Information systems* **17**(2), 107–145.

Hennig, C. and Liao, T. F. (2013), 'How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **62**(3), 309–369.

Liu, B., Zhou, C., Zhang, X. and Liu, Y. (2020), 'A unified data-adaptive framework for high dimensional change point detection', *J. R. Statist. Soc. B* **82**(4), 933–963.

Matteson, D. S. and James, N. A. (2014), 'A nonparametric approach for multiple change point analysis of multivariate data', *JASA* **109**(505), 334–345.

Militino, A. F., Moradi, M. and Ugarte, M. D. (2020), 'On the performances of trend and change-point detection methods for remote sensing data', *Remote Sensing* **12**(6), 1008.

Moradi, M., Montesino-SanMartin, M., Ugarte, M. D. and Militino, A. F. (2022), 'Locally adaptive change-point detection (LACPD) with applications to environmental changes', *Stoch Env Res Risk A* **36**(1), 251–269.

Murtagh, F. and Legendre, P. (2014), 'Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion?', *Journal of Classification* **31**(3), 274–295.

Ramsey, J. and Silverman, B. W. (2005), 'Functional Data Analysis', *Springer, New York* .

Ward, J. H. (1963), 'Hierarchical grouping to optimize an objective function', *JASA* **58**(301), 236–244.