# Change-point Detection and Image Segmentation for Time Series of Astrophysical Images

Cong Xu[1] , Hans Moritz Günther[2] , Vinay L. Kashyap[3] , Thomas C. M. Lee[1] , and Andreas Zezas[4]
[1] Department of Statistics, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA; cngxu@ucdavis.edu
[2] MIT, Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[3] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[4] Department of Physics, University of Crete, 710 03 Heraklion, Crete, Greece

## Abstract

Many astrophysical phenomena are time-varying, in the sense that their intensity, energy spectrum, and/or the spatial distribution of the emission suddenly change. This paper develops a method for modeling a time series of images. Under the assumption that the arrival times of the photons follow a Poisson process, the data are binned into 4D grids of voxels (time, energy band, and $x$-$y$ coordinates), and viewed as a time series of non-homogeneous Poisson images. The method assumes that at each time point, the corresponding multiband image stack is an unknown 3D piecewise constant function including Poisson noise. It also assumes that all image stacks between any two adjacent change points (in time domain) share the same unknown piecewise constant function. The proposed method is designed to estimate the number and the locations of all of the change points (in time domain), as well as all of the unknown piecewise constant functions between any pairs of the change points. The method applies the minimum description length principle to perform this task. A practical algorithm is also developed to solve the corresponding complicated optimization problem. Simulation experiments and applications to real data sets show that the proposed method enjoys very promising empirical properties. Applications to two real data sets, the XMM observation of a flaring star and an emerging solar coronal loop, illustrate the usage of the proposed method and the scientific insight gained from it.

Unified Astronomy Thesaurus concepts: Time series analysis (1916); Poisson distribution (1898); Maximum likelihood estimation (1901); Detection (1911); Spatial point processes (1915); Astronomy data analysis (1858)

## 1. Introduction

Many phenomena in the high-energy universe are time-variable, from coronal flares on the smallest stars to accretion events in the most massive black holes. Often, this variability can just be seen "by-eye" but at other times, we need to use robust methods founded in statistics to distinguish random noise from significant variability. Realizing where the change has occurred is critical for subsequent scientific analyses, e.g., spectral fitting and light-curve modeling. Such analyses must focus on those intervals in data space that are properly tied to the changes in the physical processes that generate the observed photons. Therefore, it is of importance to identify sources as well as to locate their spatial boundaries. Our goal is to detect change points in the time direction; that is, the times at which sudden changes happened during the underlying astrophysical process.

Change-point detection in time series is well studied, and several algorithms employing different philosophies have been developed. For example, Aue & Horváth (2013) employed hypothesis testing to study structural break detection, using both nonparametric approaches like cumulative sum (CUSUM) and parametric methods like likelihood ratio statistic to deal with different kinds of structural breaks. Another likelihood-based approach commonly used to analyze astronomical time series is Bayesian Blocks (Scargle et al. 2013), which finds change points by fitting piecewise constant models between change points. A good example of the model driven approach is the Auto-PARM procedure developed by Davis et al. (2006). By modeling the piecewise-stationary time series, the procedure is able to simultaneously estimate the number of change

points, their locations, and the parametric model for each piece. Here the minimum description length (MDL) principle by Rissanen (1989, 2007) is applied in the model selection procedure. Davis & Yau (2013) proved the strong consistency of the MDL-based change-point detection procedure. Another example is Automark by Wong et al. (2016), who developed an MDL-based methodology that detects the changes in observed emission from astronomical sources in 2D time–wavelength space.

Dey et al. (2010) loosely classified image segmentation techniques into two categories: (i) image driven approaches and (ii) model driven approaches. Image driven segmentation techniques are mainly based on the discrete pixel values of the image. For example, the graph-based algorithm by Felzenszwalb & Huttenlocher (2004) treats pixels as vertices, and the weights of the edges are based on the similarity between the features of pixels. The evidence for a boundary between two regions can be measured based on the graph. Such methods work in many complicated cases as there is no underlying model for images. Model driven approaches rely upon the information of the structure of the image. These methods are based on the assumption that the pixels in the same region have similar characteristics. The Blobworld framework by Carson et al. (1999) assumes that the features of pixels are from a underlying multivariate Gaussian mixture model. Neighboring pixels whose features are from the same Gaussian distribution are grouped into the same region.

Here we follow the model driven approach. In order to develop a change-point detection method for image time series data, we begin by specifying an underlying statistical model for the images between any two consecutive change points. In

doing so we also study the statistical properties of the change-point detection method. We assume that the underlying Poisson rate for each of the images follows a piecewise constant function. Therefore, the seeded region growing algorithm developed by Adams & Bischof (1994) for gray-value images can be naturally applied.

Given the previous successes of applying the MDL principle (Rissanen 1989, 2007) to other time series change-point detection and image segmentation problems (e.g., Lee 2000; Davis et al. 2006; Wong et al. 2016), here we also use MDL to tackle our problem of joint change-point detection and image segmentation for time series of astronomical images. Briefly, MDL defines the best model as the one that produces the best lossless compression of the data. There are different versions of MDL, and the one we use is the so-called two-part code; a gentle introduction can be found in Lee (2001). When comparing with other versions of MDL such as normalized maximum likelihood, one advantage of the two-part version is that it tends to be more computationally tractable for complex problems such as the one this paper considers. It has also been shown to enjoy excellent theoretical and empirical properties in other model selection tasks (e.g., Lee 2000; Davis et al. 2006; Aue & Lee 2011; Davis & Yau 2013). Based on MDL, we develop a practical algorithm that can be applied to simultaneously estimate the number and locations of the change points, as well as to perform image segmentation on each of the images.

## 2. Methodology

Our method is applied to 4D data cubes where 2D spatial slices in several energy passbands are stacked in time. Such data cubes are commonly available, though in high-energy astrophysics, data are usually obtained in the form of a list of photons. The list contains the 2D spatial coordinates where the photons were recorded on the detector, the times they were recorded, and their energies or wavelengths. To facilitate our analysis, we bin these data into a 4D rectangular grid of boxes. After the binning of the original data, we obtain a 4D table of photon counts indexed by the 2D coordinates $(x, y)$, time index $t$, and energy band $w$. The data set is thus a series of multiband images with counts of photons as the values of the pixels. Since the emission times of photons can be considered a non-homogeneous Poisson process, and the grids do not overlap one another, the counts in each pixel are independent, and the image slices are also independent.

We first partition these images into a set of nonoverlapping region segments using a seeded region growing (SRG) method, and then merging adjacent segments to minimize MDL (see Section 2.1). The counts in each segment are modeled as Poisson counts (see Section 2.2; the implementation details of the algorithm are described in Section 2.3). We minimize the MDL criterion across the images by iteratively removing change points along the time axis and applying the SRG segmentation onto the images in each of the time intervals. Key pixels that are influential in how the segmentations and change points are determined are then identified through searching for changes in the fitted intensities (see Section 2.4). Such regions are the focus of follow-up analyses.

We list the variables, parameters, and notation used here in Table 1.

### 2.1. Region Growing and Merging

As a first step in the analysis, a suitable segmentation method must be applied to the images to delineate regions of interest (ROIs). For this, we use the SRG method of Adams & Bischof (1994) to obtain a segmentation of the image. We choose SRG over other image segmentation algorithms for its speed and reliability (Fan & Lee 2015). Also, it can be straightforwardly incorporated to the Poisson setting.

At the beginning of SRG, we select a set of seeds, manually or automatically, from the image. Each seed can be a single pixel or a set of connected pixels. A seed comprises an initial region. Then each region starts to grow outward until the whole image is covered. (Section 2.3.2 offers some suggestions on the selection of initial seeds.)

At each step, the unlabeled pixels, which are neighbors to at least one of the current regions, comprise the set of candidates for growing the region. One of these candidates is selected to merge into the region, based on the Poisson likelihood that measures the similarity between a candidate pixel and the corresponding region. We repeat this process until all of the pixels are labeled, thus producing an initial segmentation by SRG.

At the end of the SRG process, we are left with an over-segmentation, i.e., with the image split into a larger-than-optimal number of segments. We then merge these segments based on the largest reduction or smallest increase in the MDL criterion (see below). From this sequence of segmentations, we select the one that gives the smallest value of the MDL criterion as the final ROIs.

### 2.2. Modeling a Poisson Image Series

#### 2.2.1. Input Data Type

We require that the data are binned into photon counts in an $N_I \times N_T \times N_W$ tensor $\{y_{i,t,w}\}$, $i = 1,\ldots,N_I$, $t = 1,\ldots,N_T$, $w = 1,\ldots,N_W$, where $y_{i,t,w}$ is the photon counts within the $i$th spatial rectangular region, the $t$th time interval $[T_{t-1}, T_t)$, and the $w$th energy range $[W_{w-1}, W_w)$. After binning, the data can be viewed as a time series of images in different energy bands. The values of each pixel are the photon counts in the different bands in the corresponding spatial region.

Notice that compared with Automark (Wong et al. 2016), we incorporate 2D spatial information into the model, thus extending the analysis from two (wavelength/energy and time) to four dimensions (wavelength/energy, time, and projected sky location). We also relax the restriction that the bin sizes along any of the axes are held fixed. Thus, sharp changes are more easily detected.

As the data in high-energy astrophysics are photon counts, we use a Poisson process to model the data,

$$y_{i,t,w} \overset{i.i.d.}{\sim} \text{Poisson}(\lambda_{i,t,w}\Delta T_t), \qquad (1)$$

where $\Delta T_t = (T_t - T_{t-1})$.

Our goal is to infer model intensities $\lambda_{i,t,w}$ from the observed count data $\{y_{i,t,w}\}$. We are especially interested in detecting significant changes of $\lambda_{i,t,w}$ over time. If there are changes, we also want to estimate the number and locations of the change points.

To simplify the presentation, we first develop a time-homogeneous model, i.e., one where there are no change points and $\lambda_{i,t,w}$ is unchanging with $t$ (Section 2.2.2). We will then

**Table 1**
Statistical Notations

| Notation | Definition |
|---|---|
| $N_I$ | number of pixels in each 2D spatial image |
| $N_T$ | number of time bins |
| $N_W$ | number of energy bands |
| $\Delta T_t$ | duration of the $t$th time bin |
| $y_{i,t,w}$ | photon counts within the $i$th spatial pixel, the $t$th time interval, and the $w$th energy range |
| $\lambda_{i,t,w}$ | Poisson rate for the $i$th spatial pixel, the $t$th time interval, and the $w$th energy range |
| $K$ | number of change points |
| $\tau_k$ | location of the $k$th change point |
| $m^{(k)}$ | number of region segments for the $k$th interval between two consecutive change points |
| $a_h^{(k)}$ | the area (number of pixels) of the $h$th region segment of the $k$th interval between two consecutive change points |
| $b_h^{(k)}$ | the "perimeter" (number of pixel edges between this and neighboring regions) of the $h$th region of the $k$th interval |
| $\mu_{h,w}^{(k)}$ | Poisson rate for the $h$th region segment and the $w$th energy range of the $k$th interval |
| $\hat{\mu}_{h,w}^{(k)}$ | fitted Poisson rate for the $h$th region segment and the $w$th energy range of the $k$th interval |

consider more complex cases, where change points are added to the model so that $\lambda_{i,t,w}$ is allowed to change over time (Section 2.2.3).

### 2.2.2. Piecewise Constant Model

First consider a temporally homogeneous Poisson model without any change points. Then each image can be treated as an independent Poisson realization of the same, unknown, true image.

We model the image as a 3D piecewise constant function. That is, the 2D space of $x$-$y$ coordinates is partitioned into $m$ nonoverlapping regions such that all of the pixels in a given region have the same Poisson intensity. Different energy bands share the same spatial partitioning. Rigorously, the Poisson parameter $\lambda_{i,t,w}$ can be written as a summation of region-specific Poisson rates $\mu_{h,w}$ times the corresponding indicator functions of regions ($I_{\{i \in R_h\}}$ is 1 for pixel $i$ in region $R_h$ and 0 otherwise) in the following format:

$$\lambda_{i,t,w} = \sum_{h=1}^{m} \mu_{h,w} I_{\{i \in R_h\}}. \tag{2}$$

Here $i \in R_h$ means "the $i$th pixel in the $h$th region," and $I$ is the indicator function. $R_h$ is the index set of the pixels within the $h$th region, with $R_h \subseteq \{1,\ldots,N_I\}$. Also, $\mu_{h,w}$ is the Poisson rate for the $w$th band of the $h$th region. The partition of the image is specified by $\boldsymbol{R} = \{R_h | h = 1,\ldots,m\}$.

### 2.2.3. Adding Change Points to the Model

Now we allow the underlying Poisson parameter $\lambda_{i,t,w}$ to change over time $t$. We model $\lambda_{i,t,w}$ as a piecewise constant function of $t$.

Suppose that these $N_T$ images can be partitioned into $K + 1$ homogeneous intervals by $K$ change points

$$\boldsymbol{\tau} = \{\tau_0 = 0,\ \tau_1,\ \tau_2,\ldots,\tau_K,\ \tau_{K+1} = N_T\}.$$

For the $t$th image, suppose that it belongs to the $k$th time interval; i.e., $t \in (\tau_{k-1}, \tau_k]$. For each given $t$, let $\lambda$ be a 2D piecewise constant function with $m^{(k)}$ constant regions. Then $\lambda$ can be represented by

$$\lambda_{i,t,w} = \sum_{k=1}^{K+1} I_{\{t \in (\tau_{k-1}, \tau_k]\}} \sum_{h=1}^{m^{(k)}} \mu_{h,w}^{(k)} I_{\{i \in R_h^{(k)}\}}, \tag{3}$$

where $m^{(k)}$ is the number of regions within the $k$th interval. Let $\mathcal{M} = \{m^{(k)} | k = 1, 2,\ldots,K + 1\}$. The partition of the images within interval $k$ is specified by $\boldsymbol{R}^{(k)} = \{R_h^{(k)} | h = 1,\ldots,m^{(k)}\}$. And the overall partition is $\mathcal{R} = \{\boldsymbol{R}^{(k)} | k = 1, 2,\ldots,K + 1\}$. The Poisson rate $\mu_{h,w}^{(k)}$ is the value for the $w$th band in the $h$th region of the $k$th interval. Let $\boldsymbol{\mu}^{(k)} = \{\mu_{h,w}^{(k)} | h = 1,\ldots,m^{(k)}, w = 1,\ldots,N_W\}$. And let $\boldsymbol{\mu} = \{\boldsymbol{\mu}^{(k)} | k = 1,\ldots,K+1\}$, and $i \in R_h^{(k)}$ means "the $i$th pixel is in the $h$th region of the $k$th interval."

### 2.2.4. Model Selection Using MDL

Given the observed images $\{y_{i,t,w}\}$, we aim to obtain an estimate of $\lambda_{i,t,w}$. In other words, we want an estimate of the image partitions and the Poisson rates of the regions for each band. It is straightforward to estimate the Poisson intensities given the region partitioning, but the partitioning is a much more complicated model selection problem.

We will apply MDL to select the best-fitting model. Loosely speaking, the idea behind MDL for model selection is to first obtain an MDL criterion for each possible model, and then define the best-fitting model as the minimizer of this criterion. MDL defines the best model as the one that produces the best compression of the data. The criterion can be treated as the code length, or amount of hardware memory required to store the data.

First we present the MDL criterion for the homogeneous Poisson model, then follow it by the MDL criterion for the general case (i.e., with change points).

Following similar arguments as in Lee (2000; see their Appendix B), the MDL criterion for segmenting $N_T$ homogeneous images is

$$\mathrm{MDL}(m, R, \hat{\mu}) = m\log(N_I) + \frac{\log(3)}{2} \sum_{h=1}^{m} b_h$$
$$+ \frac{N_W}{2} \sum_{h=1}^{m} \log(N_T a_h) - \sum_{w=1}^{N_W} \sum_{t=1}^{N_T} \tag{4}$$
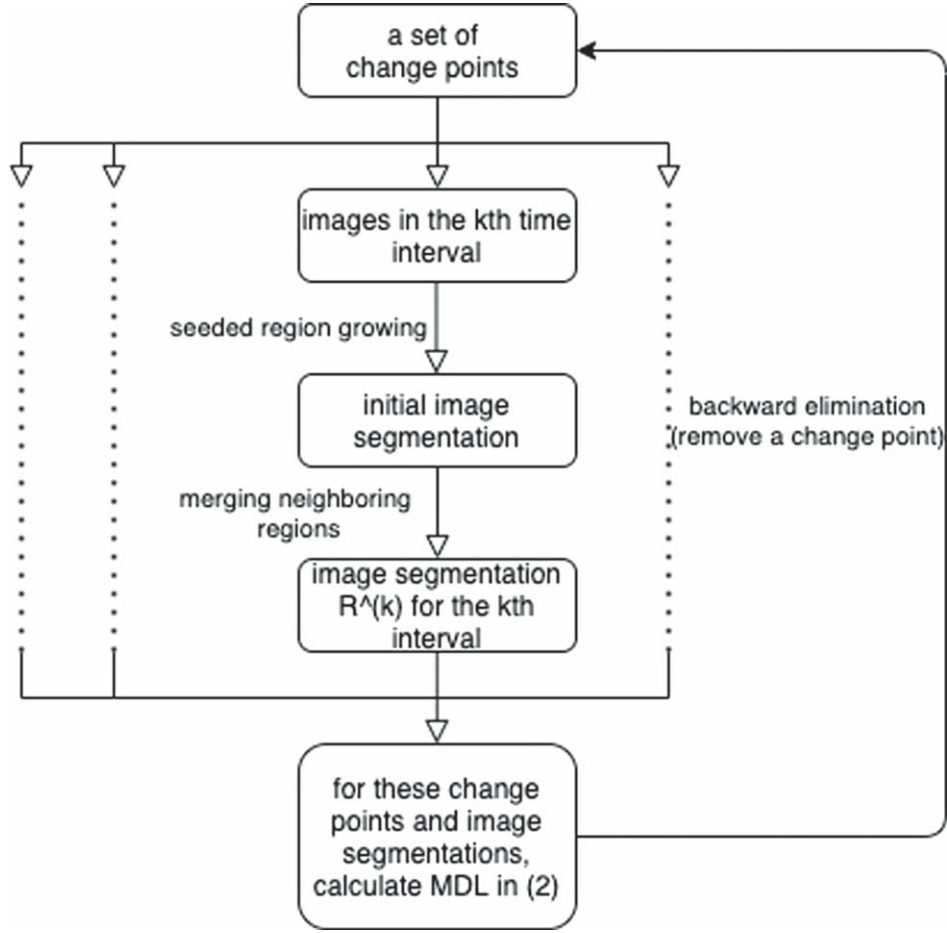$$\sum_{h=1}^{m} \sum_{i \in R_h} y_{i,t,w} \log(\hat{\mu}_{h,w}),$$

**Figure 1.** Schematic illustration of the minimization algorithm.

where $a_h$ and $b_h$ are, respectively, the "area" (number of pixels) and "perimeter" (number of pixel edges) of region $R_h$, and

$$\hat{\mu}_{h,w} = \frac{1}{\sum_{t=1}^{N_T} \Delta T_t a_h} \sum_{t=1}^{N_T} \sum_{i \in R_h} y_{i,t,w} \qquad (5)$$

is the maximum likelihood estimate of the Poisson rate in the corresponding region. Note that the indices of $\hat{\mu} = \{\hat{\mu}_{hw}\}$ run over the region segments $h = 1..m$ and the passbands $w = 1..N_W$.

For the Poisson model with change points, once the number of change points $K$ and the locations $\boldsymbol{\tau} = \{\tau_1,\ldots,\tau_K\}$ are specified, for each $k \in (1, 2,\ldots,K+1)$, $m^{(k)}$ and $\boldsymbol{R}^{(k)}$ can be estimated independently. Using the previous argument, the MDL criterion for images within the same homogeneous interval is

$$
\begin{aligned}
&\text{MDL}(\tau_{k-1}, \tau_k, m^{(k)}, \boldsymbol{R}^{(k)}, \hat{\boldsymbol{\mu}}^{(k)}) \\
&= m^{(k)} \log(N_I) + \frac{\log(3)}{2} \sum_{h=1}^{m^{(k)}} b_h^{(k)} \\
&\quad + \frac{N_W}{2} \sum_{h=1}^{m^{(k)}} \log((\tau_k - \tau_{k-1}) a_h^{(k)}) \\
&\quad - \sum_{w=1}^{N_W} \sum_{t=\tau_{k-1}+1}^{\tau_k} \sum_{h=1}^{m^{(k)}} \sum_{i \in R_h^{(k)}} y_{i,t,w} \log(\hat{\mu}_{h,w}^{(k)}).
\end{aligned} \qquad (6)
$$

Then the overall MDL criterion for the model with change points is

$$
\begin{aligned}
&\text{MDL}_{\text{overall}}(K, \boldsymbol{\tau}, \mathcal{M}, \mathcal{R}, \hat{\boldsymbol{\mu}}) \\
&= K \log(N_T) + \sum_{k=1}^{K+1} \text{MDL}(\tau_{k-1}, \tau_k, m^{(k)}, \boldsymbol{R}^{(k)}, \hat{\boldsymbol{\mu}}^{(k)}). \qquad (7)
\end{aligned}
$$

To sum up, using the MDL principle, the best-fit model is defined as the minimizer of the criterion (7). The next subsection presents a practical algorithm for carrying out this minimization.

### 2.2.5. Statistical Consistency

An important step to demonstrating the efficacy of our method is to establish its statistical consistency. That is, if it is shown that as the size of the data increases, the differences between the estimated model parameters and the true values decrease to zero, then the method can be said to be free of asymptotic bias, can be applied in the general case, and is elevated above a heuristic. We prove in the Appendix that the MDL-based model selection to choose the region partitioning, as well as the corresponding Poisson intensity parameters, is indeed strongly statistically consistent under mild assumptions of maintaining the temporal variability structure of $\lambda_{i,t,w}$.

## 2.3. Practical Minimization

### 2.3.1. An Iterative Algorithm

Given its complicated structure, global minimization of $\mathrm{MDL_{overall}}(K, \tau, \mathcal{M}, \mathcal{R}, \hat{\mu})$ (Equation (7)) is virtually infeasible when the number of images $N_T$ and the number of pixels $N_I$ are not small, because the time complexity of the exhaustive search is of the order of $2^{N_T N_I}$.

We iterate the following two steps to (approximately) minimize the MDL criterion (7).

1. Given a set of change points, apply the image segmentation method to all of the images belonging to the first homogeneous time interval and obtain the MDL best-fitting image for this interval. Repeat this for all remaining intervals. Calculate the MDL criterion (7).
2. Modify the set of change points by, for example, adding or removing one change point. In terms of what modification should be made, we use the greedy strategy to select the one that achieves the largest reduction of the overall MDL value in (7).

For Step 1 we begin with a large set of change points (i.e., an over-fitted model). In Step 2 we remove one change point (i.e., merge two neighboring intervals) to maximize the reduction of the MDL value. The procedure stops and declares that the optimization is done if no further MDL reduction can be achieved by removing change points. This is similar to backward elimination for statistical model selection problems. See Figure 1 for a flowchart of the whole procedure.

### 2.3.2. Practical Considerations

Here we list some practical issues that are crucial to the success of the above minimization algorithm.

*Initial seed allocation for SRG:* The selection of the initial seeds in SRG plays an important role in the performance of the algorithm. To obtain a good initial over-segmentation, there must be at least one seed within each true region. Currently we use all of the local maxima as well as a subset of the square lattice as the initial seeds. Based on simulation results (see Section 3.2), when the number of initial seeds is inadequate, the SRG will underfit the images, which will in turn lead to an over-fitting of the change points and will lead to an increased false-positive rate. On the other hand, it could be time consuming when the number of initial seeds is very large, especially for high-resolution images. We developed an algorithm that allocates initial seeds automatically based on locating local maxima. However, an optimal selection of initial seeds almost certainly requires expert intervention, as it depends on the type of data that we work with. To reduce the chance of obtaining a poor over-segmentation with SRG, in practice one could try using different sets of initial seeds and select the over-segmentation that gives the smallest MDL value.

*Counts per bin:* The photon counts in the image pixels cannot be too small, otherwise the algorithm could fail to produce meaningful output; see Section 3.2. In some sense, small photon counts can be seen as a low signal level, which means that the proposed method requires a minimum level of signal to operate with. Therefore, care must be exercised when deciding the size for the bin. As a rule of thumb, it should be enough to have around 100 counts for each pixel belonging to
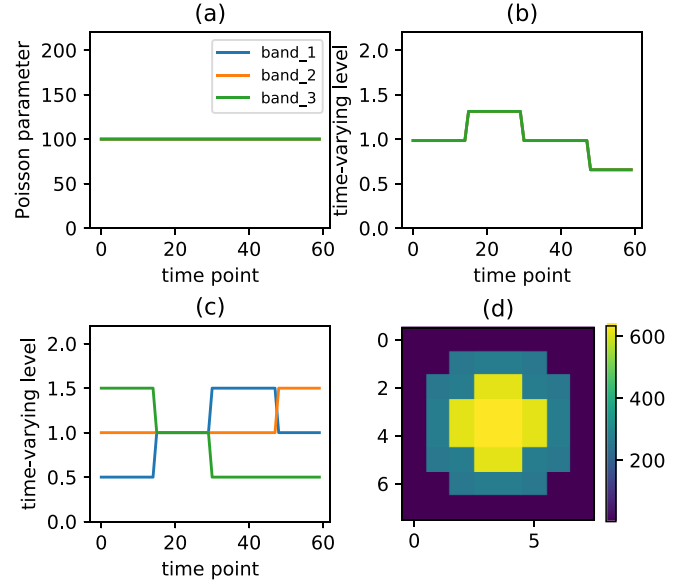


**Figure 2.** The Poisson rate functions $\lambda_{i,t,w}$ used in the simulation experiments. (a): $\lambda_{i,t,w}$ used in the single pixel experiment without change point (Section 3.1.1). The $x$-axis denotes the time points, while the $y$-axis shows the values of $\lambda_{i,t,w}$ for different band $w$. (b): the $\lambda_{i,t,w}$ relative to the no-change case (top left), used in the single pixel simulation with changing intensity (Section 3.1.2). (c): the $\lambda_{i,t,w}$ for different passbands (marked in blue, orange, and green) relative to the no-change case (top left), used in the single pixel simulation with changing intensity and spectra (Section 3.1.3). (d): the spatial structure used for the second group of experiments. The size of the image is $m = n = 8$.

an astronomical object, while pixels from the background can have a very low or even zero count.

*Initial change-point selection:* Although the stepwise greedy algorithm is capable of saving a significant amount of computation time, it could still be time consuming if the initial set of change points is too large, as it might need many iterations to reach a local minimum. It is recommended to select the initial change points based on prior knowledge, if available, in order to accelerate the algorithm.

*Computation Time:* In each iteration, the main time-consuming part is to apply the SRG. When the total number of pixels and the number of seeds are large, during the process, the number of candidates is large. Therefore, the comparison among all of the candidates and the following updating manner lead to most of the computation burden. As an example, we found that it takes about 40 minutes to apply SRG and for merging once on $64 \times 64$ images with about 200 seeds on a Linux machine with an octa-core 2.90 GHz Intel Xeon processor.

## 2.4. Highlighting the Key Pixels

After the change points are located, it is necessary to locate the pixels or regions that contribute to the estimation of the change points. The manner by which such *key pixels* are identified depends on the scientific context. Below we present two methods that are applicable to the real-world examples we discuss in Section 4.

We focus here on images in a single passband, i.e., gray-valued images. For multiband images, one can first transform a multiband image into a single-band image by, for example, summing the pixel values in different bands, or by using only the principal component image of the multiband image.
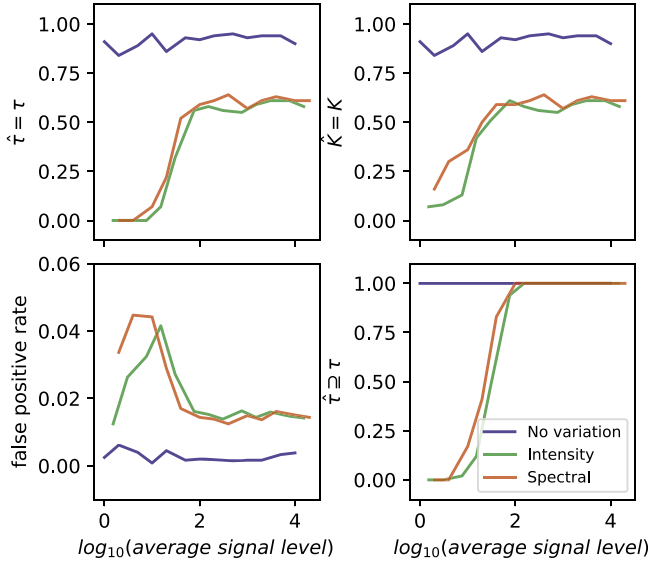
**Figure 3.** Simulation results for Group 1 experiments (single pixel). For all four plots, the *x*-axes denote the logarithm of the average counts of photons over all time points and all bands. *Top left:* fraction of the fitted change points $\hat{\tau}$ that are identical to the true change points $\tau$. *Top right:* fraction of the fitted number of change point $\hat{K}$ that equals the true number of change point $K$. *Bottom left:* false-positive rate. *Bottom right:* fraction of fitted change points contains true change points; i.e., $\tau \subseteq \hat{\tau}$. Note that the legend in the bottom right plot holds for all four plots.
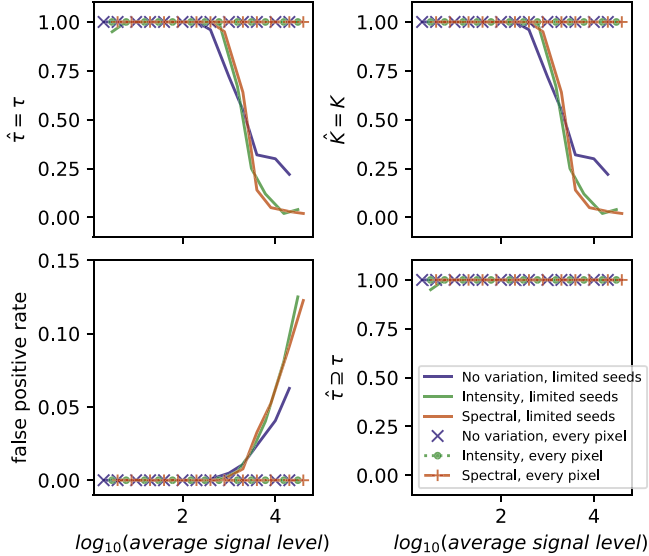


**Figure 4.** Simulation results for Group 2 experiments (with spatial structure). For all four plots, the *x*-axes denote the logarithm of the average count of photons over all time points and all bands. The solid curves denote the results when the number of initial seeds was inadequate, while the dotted curves show the results when every pixel was designated as an initial seed. *Top left:* fraction of the fitted change points $\hat{\tau}$ that are identical to the true change points $\tau$. *Top right:* fraction of the fitted number of change points $\hat{K}$ that is identical to the true number of change points $K$. *Bottom left:* false-positive rates. *Bottom right:* fraction of fitted change points that contains the true change points; i.e., $\tau \subseteq \hat{\tau}$. The legend in the bottom right plot holds for all four plots.

Alternatively, one can also apply the method to each band individually, and merge the results from each band.

### 2.4.1. Based on Pixel Differences

The first method is to highlight key pixels based on the distribution of pixel differences before and after change points.

The rationale is that a pixel with different fitted values before and after a change point is very likely a key pixel.

Suppose that the fitted values for pixel $i$ in time intervals $k$ and $(k+1)$ are $\hat{\lambda}_i^{(k)}$ and $\hat{\lambda}_i^{(k+1)}$, respectively. Given the Poisson nature of the data, we first apply a square-root transformation to normalize the fitted values. Define the difference $d_i$ for pixel $i$ as

$$d_i = \sqrt{\hat{\lambda}_i^{(k+1)}} - \sqrt{\hat{\lambda}_i^{(k)}}. \qquad (8)$$

A pixel is labeled as a key pixel if its $d_i$ is far away from the mean of all of the differences. To be specific, pixel $i$ is labeled as a key pixel if

$$\left| \frac{d_i - \hat{\mu}}{\hat{\sigma}} \right| > \Phi^{-1}\left(1 - \frac{1}{2}p\right), \qquad (9)$$

where $\hat{\mu} = \frac{1}{N_I}\sum_1^{N_I} d_i$ and $\hat{\sigma} = \frac{1}{\Phi^{-1}(3/4)}\text{MAD}$. Here $\text{MAD} = \text{median}(|d_i - \tilde{d}|)$ is the median absolute deviation with $\tilde{d} = \text{median}(d_i)$, and is used to obtain a robust estimate (i.e., a measure that minimizes the effect of outliers) of the standard deviation of the $d_i$'s (see, e.g., Rousseeuw & Croux 1993). $\Phi^{-1}(\cdot)$ is the quantile of the standard normal distribution, and $p$ is the pre-specified significance level.[5] Notice that by checking the sign of $d_i - \hat{\mu}$, we can deduce if pixel $i$ has increased or decreased after this change point.

### 2.4.2. Based on Region Differences

Another method to locate key pixels is to compare pairs of regions. For any region in the time interval $k$, there must exist at least one region in time interval $(k+1)$ such that these two regions have overlapping pixels. We then test if the difference between the means of the pixels from these two regions is significant or not.

As before, we apply the square-root transformation to the pixels within each of the regions. Then we calculate the sample means $\hat{\mu}_1$ and $\hat{\mu}_2$ and sample variances $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ of these two groups of square-rooted values. Then we can for example test whether the difference between $\hat{\mu}_1$ and $\hat{\mu}_2$ is large enough with

$$\left| \frac{\hat{\mu}_2 - \hat{\mu}_1}{\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \right| > \Phi^{-1}\left(1 - \frac{1}{2}p\right). \qquad (10)$$

See Section 4 for the applications of these two methods on some real data sets.

Lastly we note that the selection of $p$, the significance level, deserves a more careful consideration. As in reality one may need to do comparisons for many change points and energy bands, this becomes a multiple-testing problem where the number of tests is large. Therefore, one should adjust the value of $p$ in order to control for false positives.

### 3. Simulations

Two groups of simulations were conducted to evaluate the empirical performance of the proposed method. A specially designed $\lambda_{i,t,w}$ was used for each of the experiments. For each

---

[5] $\Phi^{-1}$ is related to the standard Normal error function, with exemplar values $\Phi^{-1}(\{0.75, 0.841, 0.977, 0.9986, 1 - \frac{10^{-5}}{2}, 1 - \frac{10^{-10}}{2}, 1 - \frac{10^{-15}}{2}\})=\{0.6745, 1, 2, 3, 4.417, 6.467, 8.014\}$. We typically choose $p = 1 - \frac{10^{-15}}{2}$ as our threshold.

experiment, we tested 13 signal levels, defined as the average number of photon counts per pixel. For each signal level, 100 data sets were generated according to Equation (1), with $\Delta T = 1$. The number of spectral bands $N_W = 3$, so $w = 1, 2, 3$, and the number of time points $N_T = 60$.

### 3.1. Group 1: Single Pixel

The first group of experiments were designed to evaluate the ability of the proposed method for detecting change points, under the condition that there are no spatial variations. To be more specific, the size of the images is $1 \times 1$; i.e., only one pixel. In other words, $N_I = 1$ and the $i$ in $\lambda_{i,t,w}$ is a dummy index. Three $\lambda_{i,t,w}$'s of increasing complexity were used:

1. $\lambda_{i,t,w}$ is constant; i.e., no change point (as depicted in Figure 2(a)).
2. $\lambda_{i,t,w}$ shows intensity changes, but all three bands are identical at any given $t$ (see Figure 2(b)).
3. $\lambda_{i,t,w}$ shows spectral changes (see Figure 2(c)).

The first $\lambda_{i,t,w}$ was used to study the level of false positives, while the remaining two were used to study false negatives.
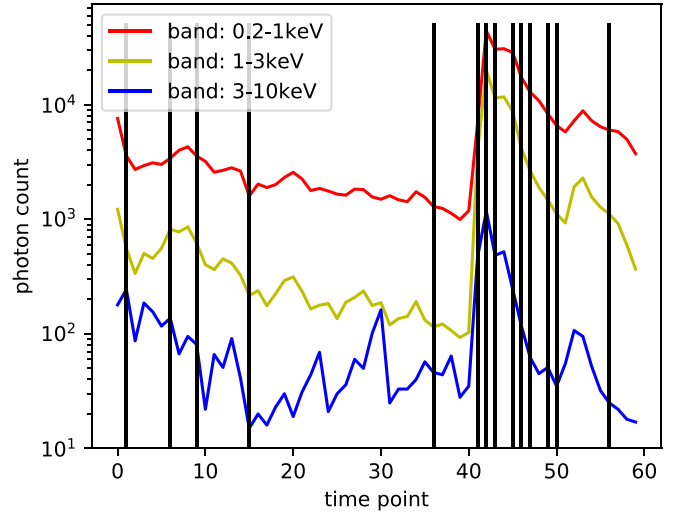
#### 3.1.1. No Change Point

As there is no change point in $\lambda_{i,t,w}$, this experiment is ideal for studying the relationship between the false-positive rate and the signal level; recall that the latter is defined as the average number of photon counts per pixel.

The results of this experiment (together with the next two experiments) are summarized as the blue curves in Figure 3. The figure captures how well the simulation recovers the location of the change points (top left), the number of change points (top right), the excess number of change points (false positives; bottom left), and the deficit in change points (false negatives; bottom right). The top left plot reports the fraction of simulated data sets for which the set of the fitted change points $\hat{\tau}$ is identical to the set of true change points $\tau$. The top right plot presents the fraction of simulated data sets for which the fitted number of change points $\hat{K}$ equals the true number of change points $K$. The bottom left plot shows the average false-positive rate, which is defined as the average number of falsely detected change points per possible location. The bottom right plot presents the fraction of simulated data sets for which $\hat{\tau}$ contains $\tau$, i.e., $\tau \subseteq \hat{\tau}$. One can see that the false discovery rate seems to be quite stable across different signal levels. Notice that the last curve is always 1 because $\tau$ is empty for this experiment.

#### 3.1.2. Varying Intensity

In this experiment, we introduced variation in $\lambda_{i,t,w}$ by multiplying (a) and (b) in Figure 2 together. The results are reported as the green curve in Figure 3. One can see that when the signal level is small ($\log_{10}$(average signal level) $< 1.0$), increasing the signal level leads to more false positives: this range of signals levels is too small to provide enough information for the detection of the true change points.

When $\log_{10}$(average signal level) is between 1.0 and 2.0, the proposed method starts to be able to detect the true change points as the signal level increases. Also, the false-positive rate begins to drop. One possible explanation is that for any two consecutive homogeneous time intervals, the location of the



**Figure 5.** Light curves of Proxima Centauri in different bands. Each curve denotes the number of photons within the corresponding band at a given time point index. The vertical black lines denote the locations of the detected change points.

change point might not be clear when the signal level is relatively small.

As the signal level continues to increase, the proposed method becomes more successful in detecting the true change-point locations. For example, when the average number of counts in each bin is greater than 100 (i.e., $\log_{10}$(average signal level) $> 2.0$), the signal is strong enough so that all of the true change points can be detected successfully. We note that there are always some false positives due to the Poisson randomness, and it seems that the false-positive rate stabilizes as the signal level increases.

#### 3.1.3. Varying Spectrum

Here we allow different bands $w$ to vary at the change points. The rate $\lambda_{i,t,w}$ was obtained by multiplying (a) and (c) in Figure 2 together. The results, which are about the same as the previous experiment (varying intensity), are reported as the red curve in Figure 3. When the signal level is small ($\log_{10}$(average signal level) $< 1.0$), the proposed method fails to detect the true change points. When $\log_{10}$(average signal level) is between 1.0 and 2.0, as the signal level increases, the false-positive rate begins to decrease while the true positive rate increases. When $\log_{10}$(average signal level) $> 2.0$, all of the true change points can be detected successfully, while the false-positive rate stays at the same level as the signal level increases.

### 3.2. Group 2: Spatial Structure

Instead of having a constant spatial signal (i.e., single pixel), in this second group of experiments, a spatial varying structure is introduced to study the empirical performance of the proposed method. As before, three Poisson rate functions $\lambda_{i,t,w}$ are considered. The size of the image is set to $N_I = 8 \times 8$. To illustrate the importance of initial seed placement, we tested two allocation strategies: (i) we deliberately placed an inadequate number of initial seeds, and (ii) we used every pixel as an initial seed.
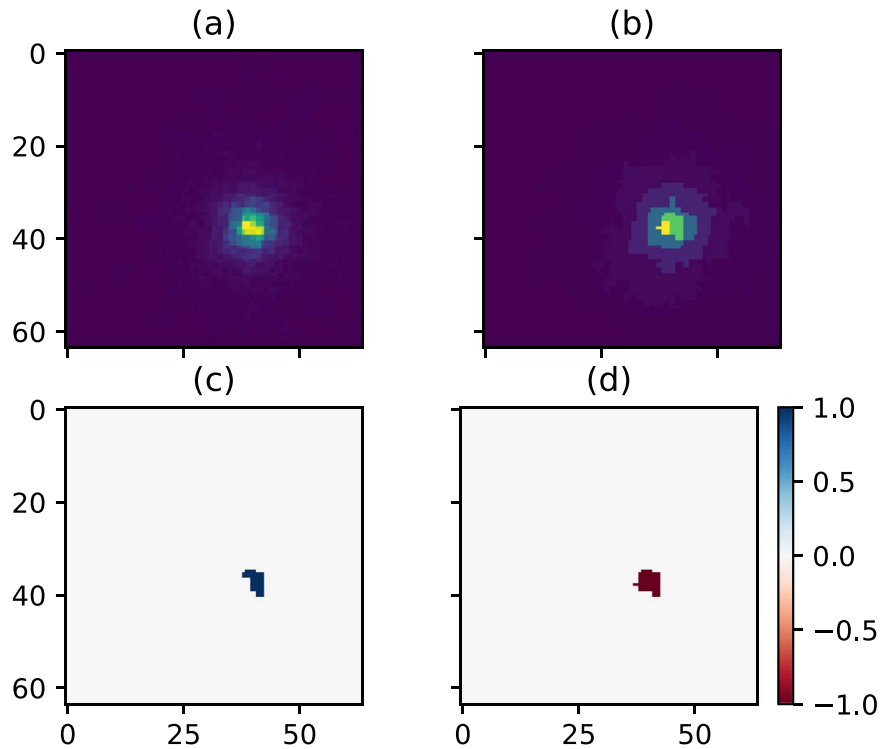
**Figure 6.** Results for Proxima Centauri. (a): the data image at time point 42 for the first band (200, 1000] in electronvolts. (b): the corresponding fitted value $\lambda_{i,t,w}$. (c): regions that show an increase (blue) and decrease (red) in intensity prior to this time point. Compared with the previous time interval, there was a significant increase in the source at this time point. (d): as in panel (c), but for the epoch after this time point. After this time point, the brightness in the source decreased. Notice that these two bottom plots share the color bar, where the value 1 denotes increasing and −1 denotes decreasing intensities.

### 3.2.1. No Change Point

There was no change point in this experiment, and the spatial variation of $\lambda_{i,t,w}$ is given in the bottom right plot of Figure 2. The results are reported as the blue curves in Figure 4. One can see that if the number of initial seeds is inadequate, the false-positive rate increases as the signal level increases above $\log_{10}$(average signal level) > 2.5. However, this does not happen when there are a large number of initial seeds; see the blue dotted curves in Figure 4. In fact, for this and the following two experiments, our method did not detect any false-positive change points. This suggests that when the images are under-segmented, the method tends to place more false change points to compensate for data variability not explainable by image segmentation.

### 3.2.2. Varying Intensity

In this experiment, $\lambda_{i,t,w}$ was obtained by multiplying (b) and (d) of Figure 2 together, so there are three change points over time. The results are similar to the no-change-point case, and shown as the green curves in Figure 4.

### 3.2.3. Varying Spectrum

In this last experiment, the energy bands were allowed to differ, and $\lambda_{i,t,w}$ was obtained by multiplying (c) and (d) of Figure 2 together. The results, reported as the red curves in Figure 4, are similar to the previous two experiments.

### 3.3. Empirical Conclusions

The following empirical conclusions can be drawn from the above experimental results.

1. This method works well in all cases when the signal level is sufficiently large. As a rule of thumb, for binning of the original data, it would be ideal to have 100 counts or more for each bin covering an astronomical source.
2. It is important to place enough initial seeds when applying SRG; otherwise the false-positive rate will increase with the signal level. See the second paragraph of Section 2.3.2 for some practical guidelines for initial seed selection.

## 4. Applications to Real Data

To illustrate its utility in astrophysics, we apply the proposed method to two real data sets, which are more complicated than those in the previous section. Specifically, we select these data sets with some obvious time-evolving variations to demonstrate the performance of our method.

### 4.1. XMM-Newton Observations of Proxima Centauri

Proxima Centauri is the nearest star to the Sun and as such is well suited for studies of coronal activity. Like our Sun, Proxima Centauri operates an internal dynamo, which generates a stellar magnetic field. In the standard model for stellar dynamos, the magnetic field lines wind up through differential rotation. When some of the magnetic field lines reconnect, the energy is released in a stellar flare. Such flares typically show a sudden rise in X-ray emission and a more gradual decay over several hours. In flares, flux and temperature are correlated such that a higher X-ray flux corresponds to a higher temperature and thus a higher energy of the average detected photons (see Güdel 2004, for a review of X-ray emission in stellar coronae and further references).
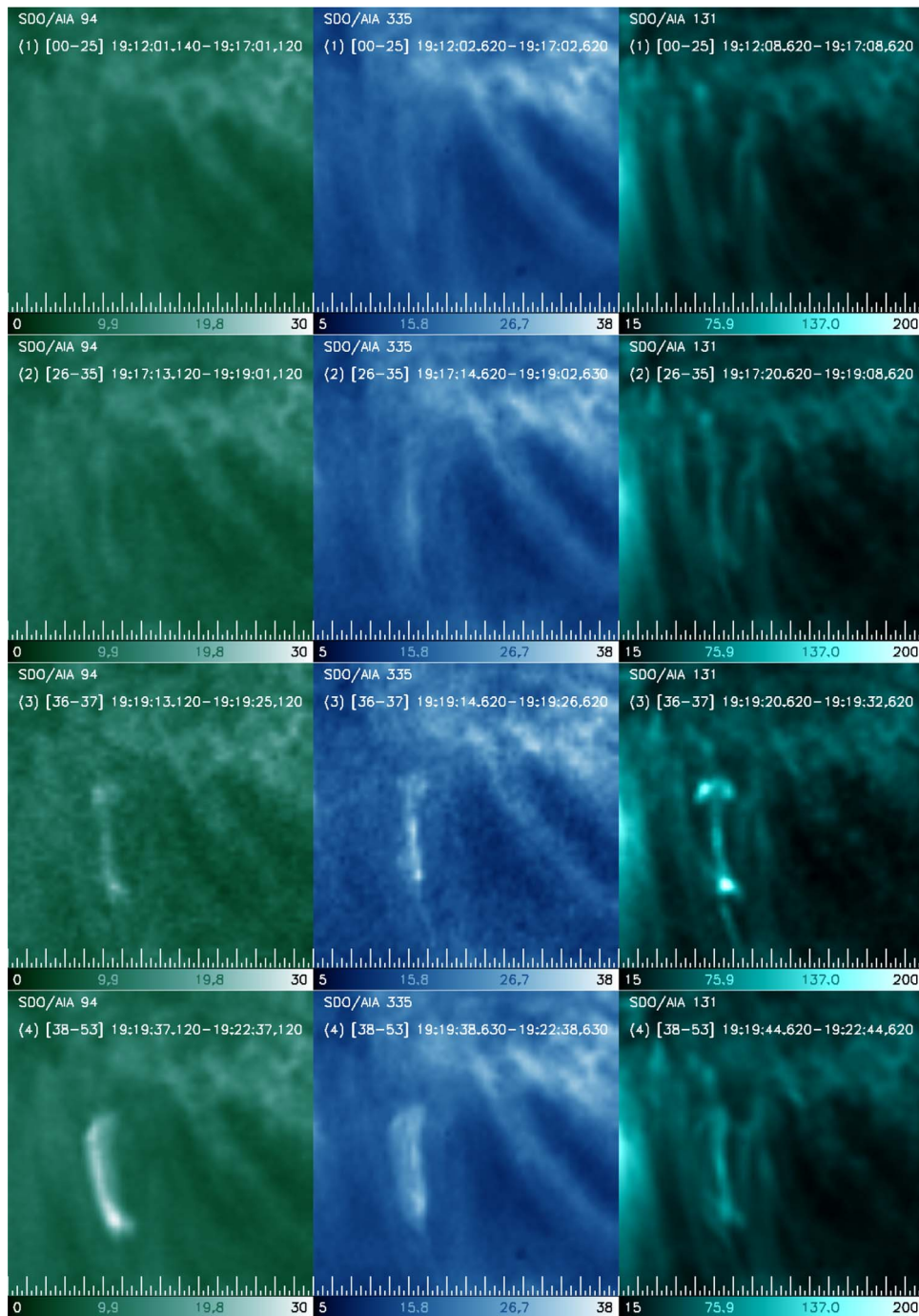
**Figure 7.** An isolated loop structure shown lighting up in three SDO/AIA passbands. Each row corresponds to the intensities in AIA filter images, averaged over the time duration found by our method, going from interval 1 (top) to interval 4 (bottom). The columns, going from left to right, show the 94, 335, and 131 Å filter band images. The filter name, time duration, and the image sequence indices are marked at the top of the image, and the intensity scale is marked at the bottom. The grid at the bottom of each image denotes the pixelation, with each image having a size of $64 \times 64$ pixels. Notice that the isolated loop becomes bright enough for detection in the third interval.

Despite its proximity, Proxima Centauri and its corona are unresolved in X-ray observations; it is just the point-spread function (PSF) of the telescope that distributes the incoming flux over many pixels on the detector.

### 4.1.1. Data

We use a data set from XMM-Newton (Obs.ID 0049350101), where Proxima Centauri was observed for 67 ks on 2001-08-12.

Because of the high flux, the MOS cameras on XMM-Newton are highly piled-up, and we restrict our analysis to the data from the PN camera. We obtained the data from the XMM-Newton science archive hosted by the European Space Agency (ESA).[6] The data we received was processed by Observation Data Subsystem (ODS) version 12.0.0. Our analysis is based on the filtered PN event data from the automated reduction pipeline

---

[6] https://www.cosmos.esa.int/web/xmm-newton/xsa
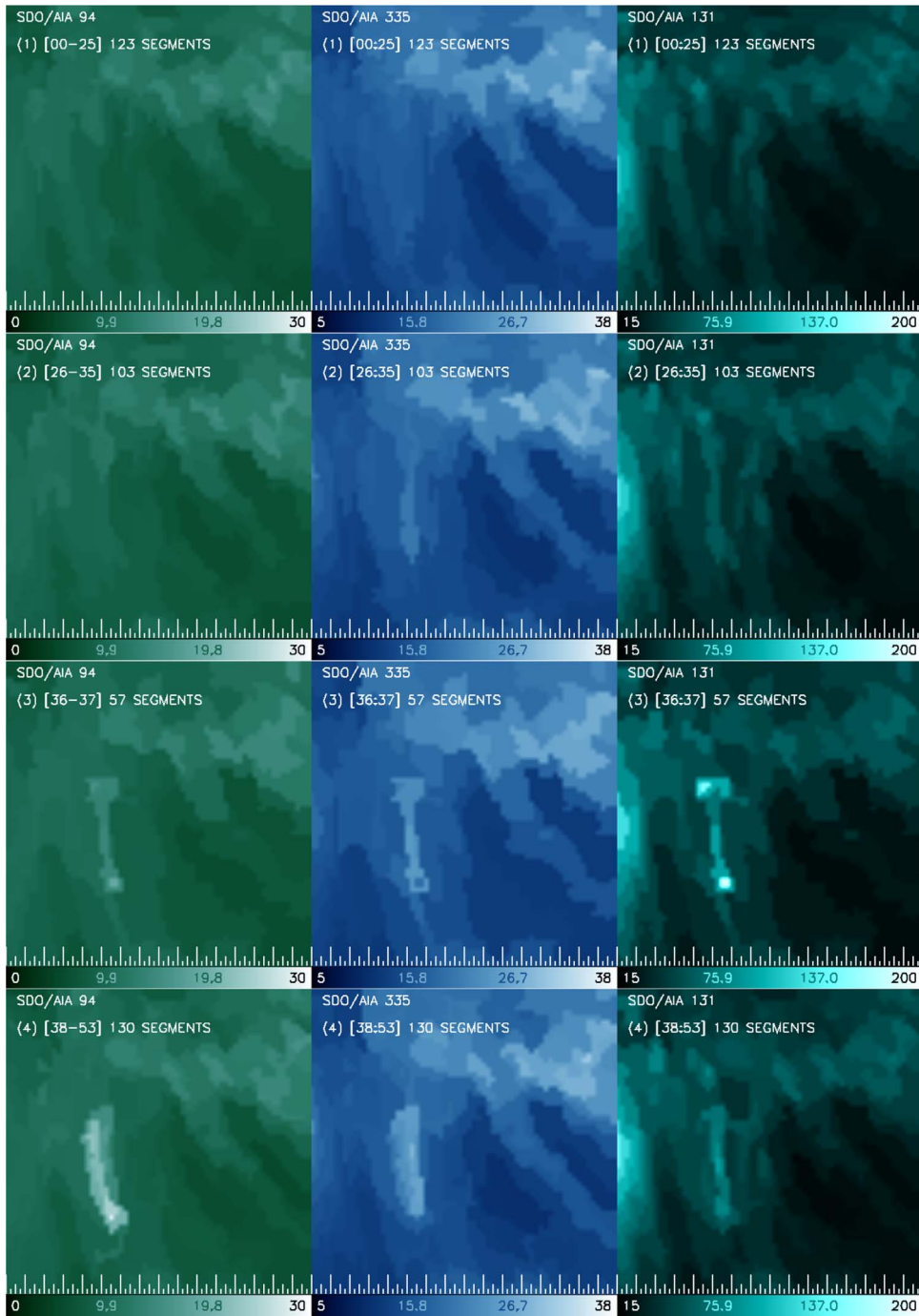
**Figure 8.** Intensities $\lambda_{i,t,w}$ as fit to the data from Figure 7 in spatial segments. The images are arranged in the same manner, and demonstrate that the loop structure is locatable and identifiable. The numbers of region segments found are also marked.

(PPS). Güdel et al. (2002) presented a detailed analysis and interpretation of this data set.

In our analysis, we only used a subset of photons with spatial coordinates within [25, 500, 27, 500] × [26, 500, 28, 500], and it was binned as images of size 64 × 64. We used the temporal bins of width 1100.4 s to generate 60 images. We binned the data into three energy bands, (200, 1000], (1000, 3000], and (3000, 10, 000] in electronvolts.

### 4.1.2. Results

Figure 5 presents the light curves for different bands as well as the locations of the detected change points. As there is only a single source of photons with a negligible background signal level, the detected change points coincide with the changes of the light curves. Many change points are detected for the abrupt increase and then decrease in brightness for all of the bands at the time points between 42 and 50. The time interval between 15 and 36 is detected as a homogeneous time interval, and the variation in light curves within this interval is viewed as common Poisson variations. A few change points are detected for the time interval before time point 15 and the interval after 50. A piecewise constant model is used to fit these gradual changes in intensities.

The fitted images can be found in Figure 6. The source of most photons, a point source, is modeled by different piecewise
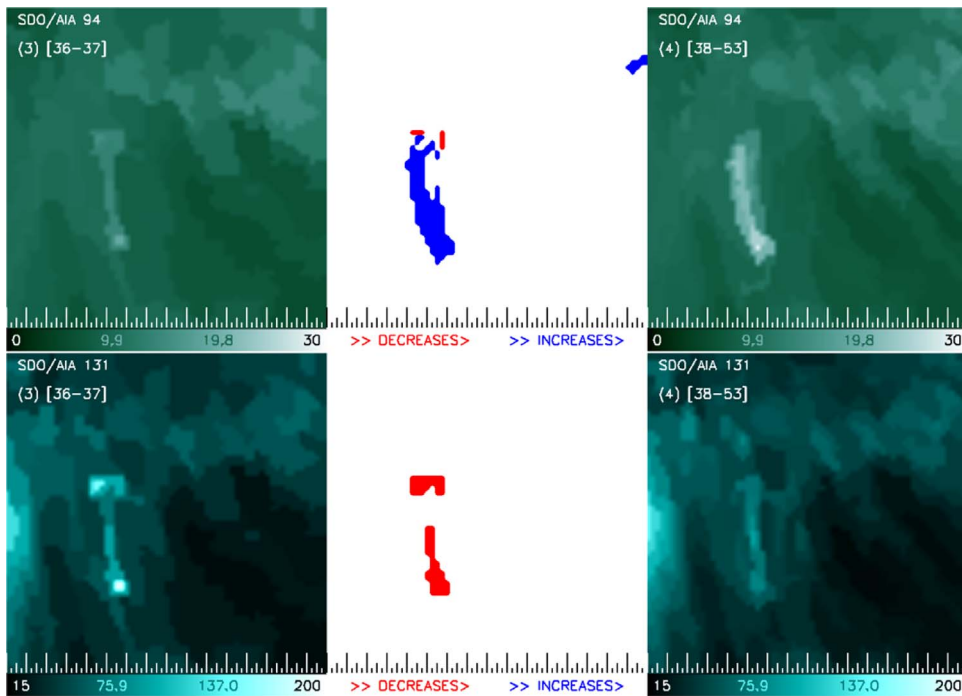
**Figure 9.** Demonstrating the isolation of key pixels of interest. Each set of three shows the fitted intensity in one passband in the third interval (*left*), followed by a bitmap of pixels (*middle*) showing where intensity increases (blue) and decreases (red), followed by the fitted intensity image in the same filter in the fourth time interval (*right*). The upper row shows the transition in the 94 Å filter, and the lower row shows the transition in the 131 Å filter. Notice that the loop continues to brighten at 94 Å, even as it starts to fade at 131 Å.

constant models at all of these time intervals. The center of the point source and the wings of the PSF region nearby were fitted by models with shapes like concentric circles.

To test our method for finding the regions of significant change, we also apply it here because we know what the answer should be. For this data set, as the observations for different bands change simultaneously, we combine all three bands to highlight the key pixels. That is, we highlight the regions for each band and take the intersection of these regions. Examples of the results based on the method in Section 2.4.2 can be found in Figure 6. The method indeed picks out the point source. With significance level $p = 10^{-2}$, the abrupt increase in brightness of the source at time points 41 and 42, as well as the sudden decrease at time point 43, can be detected successfully. By modifying the significance level, different sensitivities can be achieved.

## 4.2. Isolated Evolving Solar Coronal Loop

Images of the solar corona constitute a legitimate big data problem. Several observatories have been collecting images in extreme ultra-violet (EUV) filters and in X-ray passbands for several decades, and analyzing them to pick out interesting changes using automated routines; they have been largely unsuccessful. Catalogs like the Heliophysics Events Knowledgebase (Hurlburt et al. 2012; Martens et al. 2012) can detect and mark features of particular varieties, though these compilations remain beset by incompleteness (see, e.g., Barnes et al. 2017; Aggarwal et al. 2018; Hughes et al. 2019). In this context, our method provides a way to model solar features without being limited to a particular feature set to identify and locate regions in images where something interesting has transpired. As a proof of concept, we apply the method to a

simple case of an isolated coronal loop filling with plasma, as observed with the Solar Dynamics Observatory's Atmospheric Imaging Assembly (SDO/AIA) filters (Lemen et al. 2012; Pesnell et al. 2012). Considerable enhancements must still be made in order to lower the computational cost before the method can be applied to full-sized images at faster than observed cadence; however, we demonstrate here that a well-defined region of interest can be selected without manual intervention for a data set that consists of images in several filters.

### 4.2.1. Data

In particular, here we consider AIA observations carried out on 2014 December 11 between 19:12 UT and 19:23 UT, and focus on a $64 \times 64$ pixel region located $(+1'', -271'')$ from disk center, in which a small, isolated, well-defined loop appeared at approximately 19:19 UT. This region was selected solely as a test case to demonstrate our method; the appearance of the loop is clear and unambiguous, with no other event occurring nearby to confuse the issue; see Figure 7. We apply our method to these data, downloaded using the SDO AIA Cutout Service,[7] and demonstrate that the loop (and it alone) is detected and identified; see Figures 8, 9, and 10. AIA data are available in six filter bands, centered at 211, 94, 335, 193, 131, and 171 Å. Here, we have limited our analysis to three bands, 94, 335, and 131 Å, in which the isolated loop is easily discernible by eye (a full analysis including all of the filters does not change the results). Each filter consists of a sequence of 54 images, and while they are not obtained simultaneously,

---

[7] https://www.lmsal.com/get_aia_data/
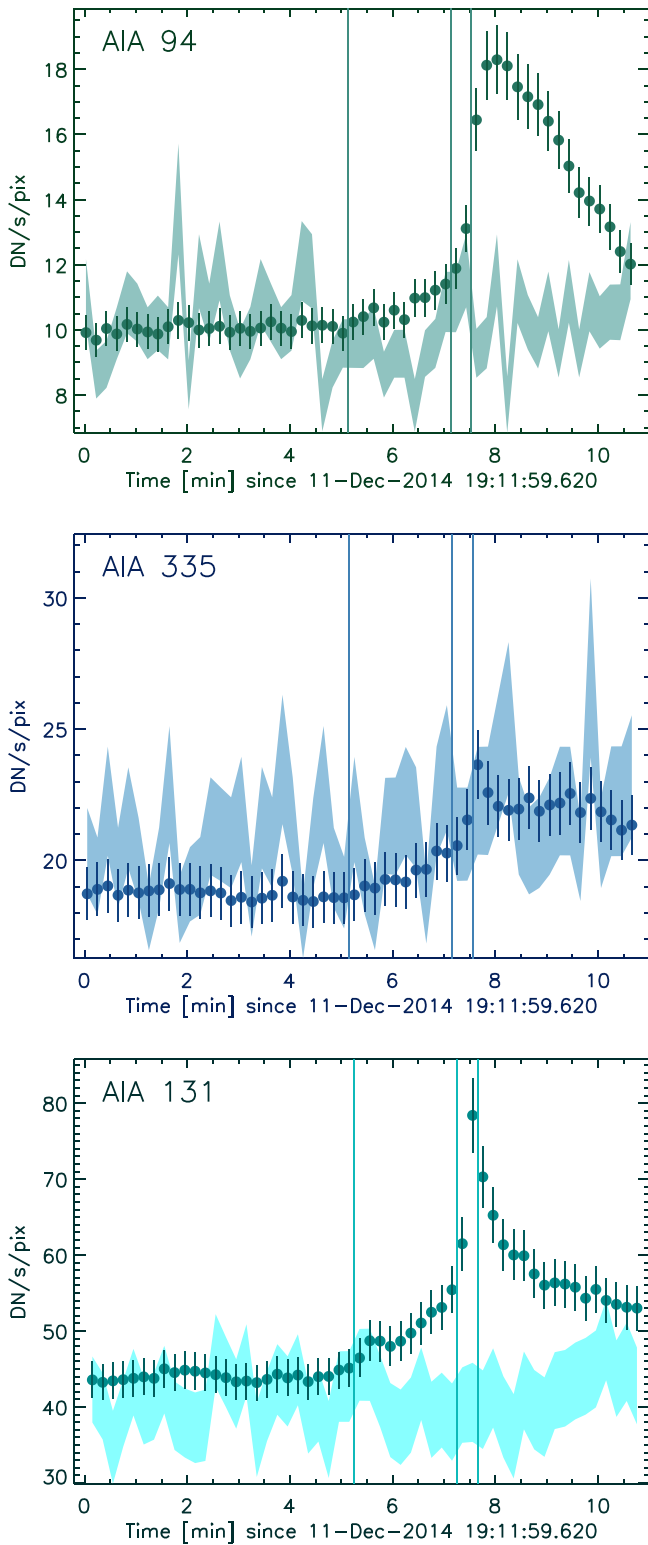
**Figure 10.** Light curves of the key pixels where changes are found, for the three filters used in the analysis: 94 Å (*top*), 335 Å (*middle*), and 131 Å (*bottom*). The averages of the observed intensities, weighted by the number of times each pixel is flagged as a key pixel, are shown as dots, along with the similarly weighted sample standard deviations, shown as vertical bars. The shaded regions represent the envelope of the sample standard deviation seen *outside* the flagged pixels. The vertical lines denote the change points found by our algorithm.

the difference in time between the bands is ignorable on the timescale over which the loop evolves.

### 4.2.2. Results

The fitted images for the three-band case can be found in Figure 8. Notice that there is a loop-shaped object that is of interest. Based on the fitted result, this object starts to appear at time point c.36 and becomes brighter after that for the first band. In the second band, this object appears at time point c.26 and stays bright throughout the duration considered. However in the third band, the object becomes bright at time point c.36 and vanishes soon after time point c.38. The proposed method is able to catch these changes in different bands and to detect the corresponding change points.

After detecting these change points, we find the key pixels that contribute to the change points using the methods described in Section 2.4.1. This method is appropriate to highlight the regions that change rapidly after the change point because different bands may not change in the same direction for this data set. Here we apply this method on a single band, 94, as an example. We use $p = 10^{-15}$. See Figure 9 for an illustration. We find that the method could highlight the loop-shaped object, which starts to appear at time point c.36, and also detect the region that becomes much brighter after time point c.38. We also compute the light curves of the intensities in a region comprising the set of pixels formed from the union of all key pixels found at all change points in all of the filters; see Figure 10. Note that the event of interest is fully incorporated within the key pixels, with no spillover into the background, and the change points are retroactively found to be reasonably located from a temporal perspective in that they are located where a researcher seeking to manually place them would do so. The first segment is characterized by steady emission in all three bands, the second segment shows the isolated loop beginning to form, the third segment catches the time when it reaches a peak, and the last segment tracks the slow decline in intensity.

## 5. Summary

We have developed an approach to model photon emissions by astronomical sources. Also, we propose a practical algorithm to detect the change points as well as to segment the astronomical images, based on the MDL principle for model selection. We test this method on a series of simulation experiments and apply it to two real astrophysical data sets. We are able to recover the time-evolving variations.

Based on the results of simulation experiments, it is recommended that the average number of photon counts within each bin should be from 100 to 1000 for pixels belonging to an astrophysical object, so that the proposed method is able to find change points and limit false positives.

For future work, it will be helpful to quantify evidence of the existence of a change point by deriving a test statistic based on Monte Carlo simulations or other methods. Another possible extension is to relax the piecewise constant assumption and allow piecewise linear/quadratic modeling so that the method is able to capture more complicated and realistic patterns.

## Appendix
## Statistical Consistency

Here we prove that the MDL scheme is statistically consistent (see Section 2.2.5), thereby ensuring that the estimates of region segmentations and the Poisson intensities are reliable measures of the data. In the following, we assume that the size of the time bins $\Delta T_t = 1$, $\forall 1 \leqslant t \leqslant N_T$, as $N_T$ increases to infinity. That is to say, first, we study that as these underlying non-homogeneous Poisson processes are extending at the same rate, the size of the bins remains fixed, which leads to an increasing number of independent observations for any given part of this Poisson process. And by keeping the size of the bins fixed, we get rid of the case that the Poisson parameters keep varying as $N_T$ increases. Second, by setting $\Delta T_t = 1$, the Poisson parameter is numerically equal to the Poisson rate, which ease the arguments. The proof can be extended if we relax this assumption.

Given the above assumption, the photon counts have the following Poisson model,

$$y_{i,t,w} \overset{i.i.d.}{\sim} \text{Poisson}(\lambda_{i,t,w}). \qquad (A1)$$

Given change points $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_K)$, we set $\tau_0 = 0$ and $\tau_{K+1} = N_T$, and let $\nu_k = \tau_k / N_T$, $k = 0, 1, \ldots N_T$ to be the normalized change points. The consistency results are based on $\nu_k$'s being fixed as $N_T$ increases.

The observed images within the same interval follow the same corresponding piecewise constant model given change points $\boldsymbol{\tau}$. Let $x_{i,t-\tau_{k-1},w}^{(k)} = y_{i,t,w}$, $\tau_{k-1} + 1 \leqslant t \leqslant \tau_k$, $\boldsymbol{X}_{t-\tau_{k-1}}^{(k)} = \boldsymbol{Y}_t = \{y_{i,t,w} | i = 1, \ldots, N_I, w = 1, \ldots, N_W\}$, $\tau_{k-1} + 1 \leqslant t \leqslant \tau_k$, and $\boldsymbol{X}^{(k)} = \{\boldsymbol{X}_t^{(k)} | 1 \leqslant t \leqslant T_k\}$ where $T_k = \tau_k - \tau_{k-1}$.

Let $\lambda_{i,t,w}$'s within the $k$th interval follow the same corresponding 2D piecewise constant model with $m^{(k)}$ constant regions. The partition of the images within interval $k$ is specified by a region assignment function $R^{(k)}(.): \{1, \ldots, N_I\} \rightarrow \{1, \ldots, m^{(k)}\}$. The Poisson parameter $\mu_{h,w}^{(k)}$ is the value for the $w$th band in the $h$th

region of the $k$th interval. Let $\mu^{(k)} = \{\mu_{h,w}^{(k)} | h = 1, \ldots, m^{(k)}, w = 1, \ldots, N_W\}$. That is to say,

$$\lambda_{i,t,w} = \sum_{k=1}^{K+1} I_{\{t \in (\tau_{k-1}, \tau_k]\}} \mu_{R^{(k)}(i),w}^{(k)}. \qquad (A2)$$

In all, for pixel $i \in R_h^{(k)}$,

$$x_{i,t,w}^{(k)} \overset{i.i.d.}{\sim} \text{Poisson}(\mu_{h,w}^{(k)}). \qquad (A3)$$

For each $1 \leqslant t \leqslant T_k$, the log-likelihood function for region assignments $R^{(k)}$ and Poisson parameters $\mu^{(k)}$ is

$$\tilde{l}_k((R^{(k)}, \mu^{(k)}); \boldsymbol{X}_t^{(k)}) = \sum_{w=1}^{N_W} \sum_{h=1}^{m^{(k)}}$$
$$\times \sum_{i, s.t. R^{(k)}(i)=h} [x_{i,t,w}^{(k)} \log(\mu_{h,w}^{(k)}) - \mu_{h,w}^{(k)} - \log(x_{i,t,w}^{(k)}!)]. \qquad (A4)$$

As some of the terms in the log-likelihood function have nothing to do with the parameters estimated, we remove these terms and write down the log-likelihood to be

$$l_k((R^{(k)}, \mu^{(k)}); \boldsymbol{X}_t^{(k)}) = \sum_{w=1}^{N_W} \sum_{h=1}^{m^{(k)}}$$
$$\sum_{i, s.t. R^{(k)}(i)=h} [x_{i,t,w}^{(k)} \log(\mu_{h,w}^{(k)}) - \mu_{h,w}^{(k)}]. \qquad (A5)$$

Define $\psi_k = (R^{(k)}, \mu^{(k)})$ to be the parameter set for the $k$th interval, and $\mathcal{M}$ to be the class of models $\psi_k$ can take value from. Then the log-likelihood for the $k$th interval can be written as

$$L_T^{(k)}(\psi_k; \boldsymbol{X}^{(k)}) = \sum_{t=1}^{T_k} l_k((R^{(k)}, \mu^{(k)}); \boldsymbol{X}_t^{(k)}). \qquad (A6)$$

Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)$ be the normalized change-point location vector, and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_{K+1})$ be the parameter vector. Then vector $(K, \boldsymbol{\nu}, \boldsymbol{\psi})$ can specify a model for this sequence of images. The MDL is derived to be

$$\text{MDL}(K, \boldsymbol{\nu}, \boldsymbol{\psi}) = K \log(N_T) + \sum_{k=1}^{K+1} [m^{(k)} \log(N_I)$$
$$+ \frac{\log(3)}{2} \sum_{h=1}^{m^{(k)}} b_h^{(k)} + \frac{N_W}{2} \sum_{h=1}^{m^{(k)}}$$
$$\log(([T\nu_k] - [T\nu_{k-1}] + 1) a_h^{(k)})] - \sum_{k=1}^{K+1} L_T^{(k)}(\psi_k; \boldsymbol{X}^{(k)}). \qquad (A7)$$

Here the "area" (number of pixels) and "perimeter" (number of pixel edges) of region $R_h^{(k)}$ are denoted by $a_h^{(k)}$ and $b_h^{(k)}$.

In order to make sure that the change points are identifiable, we assume that there exists an $\epsilon_\nu > 0$ such that $\min_{1 \leqslant k \leqslant K+1} |\nu_k - \nu_{k-1}| > \epsilon_\nu$. Therefore, the number of change points is bounded by $K \leqslant [1/\epsilon_\nu] + 1$. And there exists a

constraint $A^K_{\epsilon_\nu}$ of $\boldsymbol{\nu}$ where

$$A^K_{\epsilon_\nu} = \{\boldsymbol{\nu} \in (0,1)^K | 0 < \nu_1 < ... < \nu_K < 1, \nu_k$$
$$- \nu_{k-1} > \epsilon_\nu, \forall\, 1 \leqslant k \leqslant K + 1\}. \tag{A8}$$

Then the estimation of the model based on MDL is given by

$$(\hat{K}_T, \hat{\boldsymbol{\nu}}_T, \hat{\boldsymbol{\psi}}_T) = \arg$$

$$\min_{K \leqslant [1/\epsilon_\nu]+1, \boldsymbol{\nu} \in A^K_{\epsilon_\nu}, \psi \in \mathcal{M}} \frac{1}{N_T} \mathrm{MDL}(K, \boldsymbol{\nu}, \boldsymbol{\psi}). \tag{A9}$$

Here MDL(K, $\boldsymbol{\nu}$, $\boldsymbol{\psi}$) is defined in Equation (A7), $\hat{\boldsymbol{\nu}}_T =$ $(\hat{\nu}_1,...,\hat{\nu}_{\hat{K}})$ and $\hat{\boldsymbol{\psi}}_T = (\hat{\psi}_1,...,\hat{\psi}_{\hat{K}+1})$, where $\hat{\psi}_k = (\hat{R}^{(k)}, \hat{\mu}^{(k)})$. And $\hat{\mu}^{(k)}$ is defined as

$$\hat{\mu}^{(k)} = \arg \max_{\mu^{(k)} \in \Theta_k(\hat{R}^{(k)})}$$

$$L_T^{(k)}((\hat{R}^{(k)}, \mu^{(k)}); \hat{X}^{(k)}) \tag{A10}$$

where $\hat{X}^{(k)} = \{Y_t | [T\hat{\nu}_{k-1}] < t \leqslant [T\hat{\nu}_k]\}$ denotes the estimated $k$th interval of the sequence of images.

We further define the log-likelihood formed by a portion of the observations in the $k$th interval by

$$L_T^{(k)}(\psi_k, \nu_d, \nu_u; X^{(k)})$$
$$= \sum_{t=[T_k\nu_d]+1}^{[T_k\nu_u]} l_k((R^{(k)}, \mu^{(k)}); X_t^{(k)}), \tag{A11}$$

where $0 \leqslant \nu_d < \nu_u \leqslant 1$ and $\nu_u - \nu_d > \epsilon_\nu$.

We denote

$$\sup := \sup_{0 \leqslant \nu_d < \nu_u \leqslant 1, \nu_u - \nu_d > \epsilon_\nu} \tag{A12}$$

to simplify the notation.

In this setting, an extension needs to be made such that $\nu_d$ and $\nu_u$ can be slightly outside [0, 1]. This means that the $k$th estimated interval could cover a part of the observations that belong to the $(k-1)$th and $(k+1)$th true intervals. Based on the formula (3.4) in Davis & Yau (2013), for a real-value function $f_T(\nu_d, \nu_u)$ on $\mathcal{R}^2$,

$$\sup_{\underline{\nu_d}, \overline{\nu_u}} f_T(\nu_d, \nu_u) \xrightarrow{a.s.} 0 \tag{A13}$$

is used to denote

$$\sup_{-h_T < \nu_d < \nu_u < 1 + r_T, \nu_u - \nu_d > \epsilon_\nu} f_T(\nu_d, \nu_u) \xrightarrow{a.s.} 0 \tag{A14}$$

for any pre-specified positive-valued sequences $h_T$ and $r_T$, which cover to 0 as $N_T \to \infty$.

The following assumptions on true Poisson parameters $\mu^{o(k)}_{h,w}$, $1 \leqslant h \leqslant m^{(k)}$, $1 \leqslant w \leqslant N_W$, $1 \leqslant k \leqslant (K^o + 1)$ are necessary.

**Assumption 1.**

$$0 < C_d := \min_{k,h,w} \mu^{o(k)}_{h,w} \leqslant C_u := \max_{k,h,w} \mu^{o(k)}_{h,w} < \infty. \tag{A15}$$

**Assumption 2.** For two true neighboring regions $R_p^{(k)}$ and $R_q^{(k)}$ at the $k$th interval,

$$\delta_1 := \min_{k,p,q,w} |\mu^{o(k)}_{p,w} - \mu^{o(k)}_{q,w}| > 0. \tag{A16}$$

**Assumption 3.** For any two neighboring intervals $(k - 1)$ and $k$

$$\delta_2 := \min_k \max_{i,w} |\mu^{o(k)}_{R^{(k)}(i),w} - \mu^{o(k-1)}_{R^{(k-1)}(i),w}| > 0. \tag{A17}$$

**Proposition Appendix A.1.** *For* $k = 1,...,K + 1$ *and any fixed* $R^{(k)}$, *there exists an* $\epsilon > 0$ *such that*

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} E|l_k((R^{(k)}, \mu^{(k)}); X_t^{(k)})|^{\nu+\epsilon} < \infty,$$

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} E|l_k'((R^{(k)}, \mu^{(k)}); X_t^{(k)})|^{\nu+\epsilon} < \infty,$$

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} E|l_k''((R^{(k)}, \mu^{(k)}); X_t^{(k)})| < \infty. \tag{A18}$$

This proposition holds for $\nu = 1, 2, 4$ due to the compactness of parameter space (Assumption 1) and bounded $E[(x^{(k)}_{i,t,w})^{\nu+\epsilon}]$.

**Proposition Appendix A.2.** *For* $k = 1,...,K + 1$ *and any fixed* $R^{(k)}$,

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} |\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \mu^{(k)}); X^{(k)})$$
$$- L_k((R^{(k)}, \mu^{(k)}))| \xrightarrow{a.s.} 0,$$

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} |\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T'^{(k)}((R^{(k)}, \mu^{(k)}); X^{(k)})$$
$$- L_k'((R^{(k)}, \mu^{(k)}))| \xrightarrow{a.s.} 0,$$

$$\sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} |\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T''^{(k)}((R^{(k)}, \mu^{(k)}); X^{(k)})$$
$$- L_k''((R^{(k)}, \mu^{(k)}))| \xrightarrow{a.s.} 0, \tag{A19}$$

where

$$L_k((R^{(k)}, \mu^{(k)})) := E(l_k((R^{(k)}, \mu^{(k)}); X_t^{(k)})),$$
$$L_k'((R^{(k)}, \mu^{(k)})) := E(l_k'((R^{(k)}, \mu^{(k)}); X_t^{(k)})),$$
$$L_k''((R^{(k)}, \mu^{(k)})) := E(l_k''((R^{(k)}, \mu^{(k)}); X_t^{(k)})). \tag{A20}$$

The estimated locations of change points are used to define the likelihood in practice. Therefore, the two ends of the $k$th interval might contain observations from the $(k - 1)$th and $(k + 1)$th true intervals, though the estimated change points are close to the true change points. It is necessary to control the effect at the two ends of the fitted interval.

**Proposition Appendix A.3.** *For* $k = 1,...,K + 1$ *and any fixed* $\psi$ *and any sequence of integers* $\{g(N_T)\}_{N_T \geqslant 1}$ *that satisfies* $g(N_T) > cN_T^w$ *for some* $c > 0$ *when* $N_T$ *is large enough, then*

$$\frac{1}{g(N_T)} \sum_{t=N_T-g(N_T)+1}^{N_T} l_k(\psi; X_t^{(k)}) \xrightarrow{a.s.} E(l_k(\psi; X_t^{(k)})),$$

$$\frac{1}{g(N_T)} \sum_{t=N_T-g(N_T)+1}^{N_T} l_k'(\psi; X_t^{(k)}) \xrightarrow{a.s.} E(l_k'(\psi; X_t^{(k)})). \tag{A21}$$

Based on Lemma 1 in Davis & Yau (2013), Proposition A.3 holds when Proposition A.1(2) holds and the Assumption $4^*$ in Davis & Yau (2013) is satisfied. And Assumption $4^*$ is satisfied because an independent process, like the current setting, must be mixing.

It is necessary to discuss the identifiability of models in $\mathcal{M}$. First we define $R^b$ as an over-segmentation compared with $R^s$ if $R^b(i) = R^b(j)$ leads to $R^s(i) = R^s(j)$.

**Proposition Appendix A.4.** *For the $k$th interval, the true model $\psi_k^o \in \mathcal{M}$ satisfies $\psi_k^o = \mathrm{argmax}_{\psi \in \mathcal{M}} E(l_k(\psi; X_t^{(k)}))$. Also, $\psi_k^o$ is uniquely identifiable, which means that if there exists a $\mu^*$ such that $l_k((R^o, \mu^o); X_t^{(k)}) = l_k((R^o, \mu^*); X_t^{(k)})$ almost everywhere for $X_t^{(k)}$, then $\mu^o = \mu^*$. And suppose that there exists another model $\psi_k^b = (R^b, \mu^b)$ such that $l_k(\psi_k^b; X_t^{(k)}) = l_k(\psi_k^o; X_t^{(k)})$ almost everywhere, then $R^b$ must be an over-segmentation compared with $R^o$. And $\mu^b$ satisfies $\mu_{R^b(i),w}^b = \mu_{R^o(i),w}^o, \forall i, w$.*

**Proof.** Suppose on the contrary that there exists a model $\psi^* = (R^*, \mu^*)$ that satisfies $\psi^* = \mathrm{argmax}_{\psi \in \mathcal{M}}(E(l_k(\psi; X_t^{(k)})))$, and $\psi^*$ is neither the true model nor an over-segmentation of the true model. Then there exist two pixels $i_0$ and $j_0$, such that they are neighboring pixels, $R^{o(k)}(i_0) \neq R^{o(k)}(j_0)$ and $R^*(i_0) = R^*(j_0)$. Therefore, by Assumption 2, we have

$$|\mu_{R^{o(k)}(i_0),w}^{o(k)} - \mu_{R^{o(k)}(j_0),w}^{o(k)}| \geqslant \delta_1 > 0. \tag{A22}$$

Define

$$\bar{\mu}_{h,w}(R) := \frac{1}{a_h(R)} \sum_{i, R(i)=h} \mu_{R^o(i),w}^{o(k)}, \tag{A23}$$

where $a_h(R)$ denotes the number of pixels in region $h$ given segmentation $R$. And in a special case,

$$\bar{\mu}_{h,w}(R^{o(k)}) = \mu_{h,w}^{o(k)}. \tag{A24}$$

Then for all possible $\mu^* \in \Theta_k(R^*)$, we have

$$
\begin{aligned}
&E(l_k((R^*, \mu^*); X_t^{(k)})) \\
&= E\left( \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} \sum_{i, s.t. R^*(i)=h} [x_{i,t,w}^{(k)} \log(\mu_{h,w}^*) - \mu_{h,w}^*] \right) \\
&= \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} \sum_{i, s.t. R^*(i)=h} [\mu_{R^{o(k)}(i),w}^{o(k)} \log(\mu_{h,w}^*) - \mu_{h,w}^*] \\
&= \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} a_h(R^*) [\bar{\mu}_{h,w}(R) \log(\mu_{h,w}^*) - \mu_{h,w}^*] \\
&\leqslant \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} a_h(R^*) \max_{\mu_{h,w}} [\bar{\mu}_{h,w}(R) \log(\mu_{h,w}) - \mu_{h,w}] \\
&= \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} a_h(R^*) [\bar{\mu}_{h,w}(R^*) \log(\bar{\mu}_{h,w}(R^*)) - \bar{\mu}_{h,w}(R^*)] \\
&= E(l_k((R^*, \bar{\mu}(R^*)); X_t^{(k)})).
\end{aligned}
\tag{A25}
$$

Also, we have

$$
\begin{aligned}
&E(l_k((R^*, \bar{\mu}(R^*)); X_t^{(k)})) \\
&= \sum_{w=1}^{N_W} \sum_{h=1}^{m^*} a_h(R^*) \max_{\mu_{h,w}} [\bar{\mu}_{h,w}(R^*) \log(\mu_{h,w}) - \mu_{h,w}] \\
&\leqslant \sum_{w=1}^{N_W} \sum_{i=1, i \notin \{i_0, j_0\}}^{N_I} \max_{\lambda_{i,w}} [\mu_{R^{o(k)}(i),w}^{o(k)} \log(\lambda_{i,w}) - \lambda_{i,w}] \\
&\quad + \sum_{w=1}^{N_W} [\mu_{R^{o(k)}(i_0),w}^{o(k)} \log(\mu_{R^*(i_0),w}^*) - \mu_{R^*(i_0),w}^* \\
&\quad + \mu_{R^{o(k)}(j_0),w}^{o(k)} \log(\mu_{R^*(j_0),w}^*) - \mu_{R^*(j_0),w}^*] \\
&< \sum_{w=1}^{N_W} \sum_{i=1, i \notin \{i_0, j_0\}}^{N_I} \max_{\lambda_{i,w}} [\mu_{R^{o(k)}(i),w}^{o(k)} \log(\lambda_{i,w}) - \lambda_{i,w}] \\
&\quad + \sum_{w=1}^{N_W} \max_{\lambda_{i_0,w}} [\mu_{R^{o(k)}(i_0),w}^{o(k)} \log(\lambda_{i_0,w}) - \lambda_{i_0,w}] \\
&\quad + \sum_{w=1}^{N_W} \max_{\lambda_{j_0,w}} [\mu_{R^{o(k)}(j_0),w}^{o(k)} \log(\lambda_{j_0,w}) - \lambda_{j_0,w}] \\
&= \sum_{w=1}^{N_W} \sum_{i=1}^{N_I} [\mu_{R^{o(k)}(i),w}^{o(k)} \log(\mu_{R^{o(k)}(i),w}^{o(k)}) - \mu_{R^{o(k)}(i),w}^{o(k)}] \\
&= \sum_{w=1}^{N_W} \sum_{h=1}^{m^{o(k)}} a_h^{o(k)} [\mu_{R^{o(k)}(i),w}^{o(k)} \log(\mu_{R^{o(k)}(i),w}^{o(k)}) - \mu_{R^{o(k)}(i),w}^{o(k)}] \\
&= E(l_k((R^{o(k)}, \mu^{o(k)}); X_t^{(k)})).
\end{aligned}
\tag{A26}
$$

Here the strict inequities must hold because of (A22).

Finally combining (A25) and (A26), we have

$$E(l_k((R^*, \mu^*); X_t^{(k)})) < E(l_k((R^{o(k)}, \mu^{o(k)}); X_t^{(k)})), \tag{A27}$$

which is a contradiction. This concludes the proof.

**Lemma 1.** *For any fixed $R^{(k)}$,*

$$
\begin{aligned}
&\sup_{\nu_d, \overline{\nu_u}} \sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} \left| \frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \mu^{(k)}), \nu_d, \nu_u; X^{(k)}) - (\nu_u \right. \\
&\left. - \nu_d) L_k((R^{(k)}, \mu^{(k)})) \right| \xrightarrow{a.s.} 0, \\
&\sup_{\nu_d, \overline{\nu_u}} \sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} \left| \frac{1}{N_T(\nu_k - \nu_{k-1})} L_T'^{(k)}((R^{(k)}, \mu^{(k)}), \nu_d, \nu_u; X^{(k)}) - (\nu_u \right. \\
&\left. - \nu_d) L_k'((R^{(k)}, \mu^{(k)})) \right| \xrightarrow{a.s.} 0, \\
&\sup_{\nu_d, \overline{\nu_u}} \sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} \left| \frac{1}{N_T(\nu_k - \nu_{k-1})} L_T''^{(k)}((R^{(k)}, \mu^{(k)}), \nu_d, \nu_u; X^{(k)}) - (\nu_u \right. \\
&\left. - \nu_d) L_k''((R^{(k)}, \mu^{(k)})) \right| \xrightarrow{a.s.} 0.
\end{aligned}
\tag{A28}
$$

See Proposition 1 and 2 in Davis & Yau (2013) for the proof.

**Lemma 2.** *Suppose that the true parameters for interval $k$ are $\psi^{o(k)} = (R^{o(k)}, \mu^{o(k)})$. And suppose that a region segmentation*

$R^{(k)}$ is specified for estimation. Let

$$\hat{\mu}_T = \hat{\mu}_T^{(k)}(\nu_d, \nu_u) := \arg \max_{\mu^{(k)} \in \Theta_k(R^{(k)})}$$

$$L_T^{(k)}((R^{(k)}, \mu^{(k)}), \nu_d, \nu_u; X_k),$$

$$\mu^{*(k)} := \arg \max_{\mu^{(k)} \in \Theta_k(R^{(k)})} L_k((R^{(k)}, \mu^{(k)})). \quad (A29)$$

*Then*

$$\sup_{\underline{\nu_d}, \overline{\nu_u}} |\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \hat{\mu}_T), \nu_d, \nu_u; X^{(k)})$$

$$-(\nu_u - \nu_d) L_k((R^{(k)}, \mu^{*(k)}))| \overset{a.s.}{\to} 0, \quad (A30)$$

where the supremum is defined in (A13). And if $R^{(k)} = R^{o(k)}$, we further have

$$\sup_{\underline{\nu_d}, \overline{\nu_u}} |\hat{\mu}_T^{(k)}(\nu_d, \nu_u) - \mu^{o(k)}| \overset{a.s.}{\to} 0. \quad (A31)$$

If $R^{(k)}$ is an over-segmentation of $R^{o(k)}$, then we have

$$\sup_{\underline{\nu_d}, \overline{\nu_u}} |\hat{\mu}_{T, R^{(k)}(i), w}(\nu_d, \nu_u)$$

$$-\mu^{o(k)}_{R^{o(k)}(i), w}| \overset{a.s.}{\to} 0 \ \forall i, w. \quad (A32)$$

**Proof.**

$$(\nu_u - \nu_d)(L_k((R^{(k)}, \mu^{*(k)})) - L_k((R^{(k)}, \hat{\mu}_T)))$$

$$\leqslant \sup_{\underline{\nu_d}, \overline{\nu_u}} |(\nu_u - \nu_d) L_k((R^{(k)}, \mu^{*(k)}))$$

$$-\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \mu^{*(k)}), \nu_d, \nu_u; X^{(k)})$$

$$+\frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \hat{\mu}_T), \nu_d, \nu_u; X^{(k)})$$

$$-(\nu_u - \nu_d) L_k((R^{(k)}, \hat{\mu}_T))|$$

$$\leqslant 2 \sup_{\underline{\nu_d}, \overline{\nu_u}} \sup_{\mu^{(k)} \in \Theta_k(R^{(k)})} \left| \frac{1}{N_T(\nu_k - \nu_{k-1})} L_T^{(k)}((R^{(k)}, \hat{\mu}_T), \right.$$

$$\times \nu_d, \nu_u; X^{(k)}) - (\nu_u - \nu_d) L_k((R^{(k)}, \mu^{(k)}))| \overset{a.s.}{\to} 0. \quad (A33)$$

The first inequity is obtained by the definition of the maximum likelihood estimator, and the last convergence comes from Lemma 1. As $\mu^{*(k)}$ maximizes $L_k((R^{(k)}, \mu^{(k)}))$ and $\nu_u - \nu_d > 0$, we have

$$|L_k((R^{(k)}, \mu^{*(k)})) - L_k((R^{(k)}, \hat{\mu}_T))| \overset{a.s.}{\to} 0. \quad (A34)$$

Combining (A33), (A34), and Proposition A.1(1), A30 holds. If $R^{(k)} = R^{o(k)}$, by Proposition A.4, $L_k((R^{(k)}, \mu^{(k)}))$ has a unique maximizer at $\mu^{o(k)}$, so (A31) holds. If $R^{(k)}$ is an over-segmentation compared with $R^{o(k)}$, by Proposition A.4, (A32) holds.

Now we give a preliminary result of the convergence when the number of change points is known.

**Theorem 1.** *(Theorem 1 in Davis & Yau (2013)). Let $\{Y_t | t = 1, \ldots, N_T\}$ be the observed images specified by $(K^o, \boldsymbol{\nu}^o, \boldsymbol{\psi}^o)$. And suppose that the number of change points $K^o$ is known. The*

change points and parameters are estimated by

$$(\hat{\boldsymbol{\nu}}_T, \hat{\boldsymbol{\psi}}_T) = \arg \min_{\boldsymbol{\lambda} \in A^m_{\epsilon_\lambda}, \psi \in \mathcal{M}} \frac{1}{N_T} \text{MDL}(K^o, \boldsymbol{\nu}, \boldsymbol{\psi}). \quad (A35)$$

*Then $\hat{\boldsymbol{\nu}}_T \overset{a.s.}{\to} \boldsymbol{\nu}^o$ and for each interval, the estimated $\hat{R}^{(k)}$ must be an over-segmentation comparing to the true region segmentation.*

We skip the proof of this theorem because it is quite similar to the proof of Theorem 1 in Davis & Yau (2013). Notice that we need to use Assumption 3 in the proof.

**Corollary 1.** *(Corollary 1 in Davis & Yau (2013)). Under the conditions of Theorem 1, if the number of change points is unknown and is estimated from the data, then*

1. *The number of change points cannot be underestimated. That is to say, $\hat{K} \geqslant K^o$ almost surely when $N_T$ is large enough.*
2. *When $\hat{K} > K^o$, $\boldsymbol{\nu}^o$ must be a subset of the limit of $\hat{\boldsymbol{\nu}}_T$ for large enough $N_T$.*
3. *In each fitted interval, the region segmentation must be equal to or be an over-segmentation comparing with the corresponding true region segmentation.*

See Corollary 1 in Davis & Yau (2013) for more details.

**Theorem 2.** *(Theorem 2 in Davis & Yau (2013)). Let $\boldsymbol{\nu}^o = (\nu_1^o, \nu_2^o, \ldots, \nu_{m^o}^o)$ be the true change points. And $(\hat{K}, \hat{\boldsymbol{\nu}}_T, \hat{\boldsymbol{\psi}}_T)$ is the MDL-based result. Then $\forall k = 1, 2, \ldots, K^o$, there exists a $\hat{\nu}_{t_k} \in \hat{\boldsymbol{\nu}}_T$ where $1 \leqslant t_k \leqslant \hat{K}$ such that*

$$|\hat{\nu}_{t_k} - \nu_k^o| = o(N_T^{-\frac{1}{2}}) \ a.s.. \quad (A36)$$

See the proof of Theorem 2 in Davis & Yau (2013).

**Lemma 3.** *Suppose the true region segmentation $R^{o(k)}$ is specified for the $k$th interval, then*

$$\hat{\mu}_T^{(k)}(\hat{\nu}_{k-1}, \hat{\nu}_k) - \mu^{o(k)} = O\left( \sqrt{\frac{\log\log(N_T)}{N_T}} \right) a.s.. \quad (A37)$$

*When the specific region segmentation $R^{(k)}$ is an over-segmentation compared with $R^{o(k)}$, then we have*

$$\hat{\mu}_{T, R^{(k)}(i), w}^{(k)}(\hat{\nu}_{k-1}, \hat{\nu}_k) - \mu^{o(k)}_{R^{o(k)}(i), w}$$

$$= O\left( \sqrt{\frac{\log\log(N_T)}{N_T}} \right) a.s. \ \forall i, w. \quad (A38)$$

See Lemma 2 in Davis & Yau (2013) for more details.
Then we come to the main result.

**Theorem 3.** *Let $\{Y_t | t = 1, \ldots, N_T\}$ be the observed images specified by $(K^o, \boldsymbol{\nu}^o, \boldsymbol{\psi}^o)$. The estimator $(\hat{K}_T, \hat{\boldsymbol{\nu}}_T, \hat{\boldsymbol{\psi}}_T)$ is defined by (A9). Then we have*

$$\hat{K}_T \overset{a.s.}{\to} K^o,$$
$$\hat{\boldsymbol{\nu}}_T \overset{a.s.}{\to} \boldsymbol{\nu}^o,$$
$$\hat{\boldsymbol{\psi}}_T \overset{a.s.}{\to} \boldsymbol{\psi}^o. \quad (A39)$$

See Theorem 3 in Davis & Yau (2013) for more details.

## ORCID iDs

Cong Xu ⓘ https://orcid.org/0000-0003-2762-6057
Hans Moritz Günther ⓘ https://orcid.org/0000-0003-4243-2840
Vinay L. Kashyap ⓘ https://orcid.org/0000-0002-3869-7996
Thomas C. M. Lee ⓘ https://orcid.org/0000-0001-7067-405X
Andreas Zezas ⓘ https://orcid.org/0000-0001-8952-676X

## References

Adams, R., & Bischof, L. 1994, ITPAM, 16, 641
Aggarwal, A., Schanche, N., Reeves, K. K., Kempton, D., & Angryk, R. 2018, ApJS, 236, 15
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Aue, A., & Lee, T. C. M. 2011, Ann. Statist., 39, 2912
Aue, A., & Horváth, L. 2013, Journal of Time Series Analysis, 34, 1
Barnes, G., Schanche, N., Leka, K. D., Aggarwal, A., & Reeves, K. 2017, in Proc. IAU Symp. 325, Astroinformatics, ed. M. Brescia et al. (Cambridge: Cambridge Univ. Press), 201
Carson, C., Belongie, S., Greenspan, H., & Malik, J. 1999, in Visual Information and Information Systems, 167, ed. D. P. Huijsmans & A. W. M. Smeulders (Berlin: Springer), 509
Davis, R. A., Lee, T. C. M., & Rodriguez-Yam, G. A. 2006, J. Am. Stat. Assoc., 101, 223
Davis, R. A., & Yau, C. Y. 2013, EJSta, 7, 381
Dey, V., Zhang, Y., & Zhong, M. 2010, in Int. Society for Photogrammetry and Remote Sensing (ISPRS) TC VII Symp., 100 Years ISPRS, vol. XXXVIII, ed. W. Wagner & B. Székely (Hannover: ISPRS), 31
Fan, M., & Lee, T. C. 2015, IET Image Processing, 9, 478
Felzenszwalb, P. F., & Huttenlocher, D. P. 2004, International Journal of Computer Vision, 59, 167
Freeland, S. L., & Handy, B. N. 2012, SolarSoft: Programming and Data Analysis Environment for Solar Physics, Astrophysics Source Code Library, ascl:1208.013
Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, Proc. SPIE, 6270, 62701V
Güdel, M. 2004, A&ARv, 12, 71
Güdel, M., Audard, M., Skinner, S. L., & Horvath, M. I. 2002, ApJL, 580, L73
Hughes, J. M., Hsu, V. W., Seaton, D. B., et al. 2019, JSWSC, 9, A38
Hurlburt, N., Cheung, M., Schrijver, C., et al. 2012, SoPh, 275, 67
Lee, T. C. M. 2000, J. Am. Stat. Assoc., 95, 259
Lee, T. C. M. 2001, International Statistical Review, 69, 169
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Martens, P. C. H., Attrill, G. D. R., Davey, A. R., et al. 2012, SoPh, 275, 79
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Rissanen, J. 1989, Stochastic Complexity in Statistical Inquiry Theory (River Edge, NJ: World Scientific Publishing Co., Inc.)
Rissanen, J. 2007, Information and Complexity in Statistical Modeling (New York: Springer)
Rousseeuw, P. J., & Croux, C. 1993, J. Am. Stat. Assoc., 88, 1273
Scargle, J. D., Norris, J. P., Jackson, B., & Chiang, J. 2013, ApJ, 764, 167
Wong, R. K. W., Kashyap, V. L., Lee, T. C. M., & van Dyk, D. A. 2016, AnApS, 10, 1107