Astronomical Transient Detection using Grouped p-Values and Controlling the False Discovery Rate

Nicolle Clements, Sanat K. Sarkar and Wenge Guo

Identifying source objects in astronomical observations, in particular with reliable algorithms, is extremely important in large-area surveys. It is of great importance for any source detection algorithm to limit the number of false detections since follow up investigations are timely and costly. In this paper, we consider two new statistical procedures to control the false discovery rate (FDR) for group-dependent data - the two-stage BH method and adaptive two-stage BH method. Motivated by the belief that the spatial dependencies among the hypotheses occur more locally than globally, these procedures test hypotheses in groups that incorporate the local, unknown dependencies. If a group is found significant, further investigation is done to the individual hypotheses within that group. Importantly, these methodologies make no dependence assumption for hypotheses *within* each group. The properties of the two procedures are examined through simulation studies as well as astronomical source detection data.

Key words: Astronomy, false discovery rate, multiple testing, source detection, spatial autocorrelation.

1 Introduction

Detecting, classifying, and monitoring transient sources in the night sky, specifically Type Ia supernovae transients, is an area of astronomical research that receives much attention. Astronomical images represent the intensity of light, or roughly a count of the photons at every pixel. However, the number of pixels in each image can be several millions in size, which makes manual source detection impossible.

The term *source pixel* is commonly referred to as a pixel in an image that is above some threshold and thus is part of a true source (transient object). A *source* is a collection of these source pixels that correspond to an astronomical object of interest. The term *background pixel* is an image pixel that does not come from a source. A source, like a supernova transient, is a stellar explosion in the sky that can last for several weeks before fading away. If the host galaxy is reasonably close, then the supernova becomes quite bright. While there is no difficulty in detecting it at peak brightness, the scientific goal is to pick it up as it has just begun to rise and is still very faint. Also, there are many more distant galaxies than bright galaxies, so there are numerous supernovae that will just barely be seen even at peak brightness.

Typically, the data each night are assumed to come from a mixture Gaussian distribution, based on source and background pixels. One issue is that the mean and variance of this Gaussian distribution differs from night to night, due to varying observing conditions, such as cloud coverage and moonlight. The background pixels from the i^{th} night are assumed to be normally distributed with mean μ_i and variance σ_i^2. The source pixels from the i^{th} night and the j^{th} source are normally distributed with mean $\mu_i + \theta_j$, where θ_j can be very small. To detect these sources, we want to test the hypothesis $H_0 : \theta_j = 0$ vs. the alternative $H_1 : \theta_j > 0$. To get around the nightly differences, astronomers standardize the data, also known as computing the signal-to-noise ratio (SNR). One can search for transient sources that exceed some SNR threshold using the standardized data converted to p-values.

It is of great importance for any source detection algorithm to limit the number of false detections. This is because following up new detections is timely and costly. Astronomers want to spend as little of their time as possible viewing what turn out to be vacant regions of sky. Currently, there are several publicly available algorithms for source detection based on sliding cells, Voronoi tessellation, wavelets, and signal-to-noise filtering. Although these algorithms provide some limit to the number of false detections, they cannot provide proof or an upper bound to the number they falsely detect. To give astronomers a source detection procedure that controls a statistically meaningful measure incorporating Type I errors, i.e. false detections, would be a great asset.

2 Preliminaries and Background

The False Discovery Rate (FDR) proposed by Benjamini and Hochberg (1995), is the expected proportion of Type I errors among all the rejected null hypotheses. It is now a widely accepted notion of error rate to control in large-scale multiple testings arising in modern scientific investigations, including astronomical source detection. Suppose there are N pixels, with P_j,

$j = 1, \ldots, N$, being the p-values generated from the observations in those pixels. Then the Benjamini-Hochberg (BH) method controlling the FDR at a level α operates as follows:

The BH Method.

- Order the p-values from the smallest to the largest: $P_{(1)}, \ldots, P_{(N)}$.
- Find $k_{BH} = \max\{j : P_{(j)} \leq j\alpha/N\}$.
- Reject the null hypotheses whose p-values are less than or equal to $P_{(k_{\mathrm{BH}})}$.

The BH method controls the FDR at the desired level α, albeit conservatively, unless there is no real source pixel, only when the p-values are independent or positively dependent (in a certain sense). More specifically, the FDR of the BH method equals $\pi_0\alpha$ when the p-values are independent, and is less than $\pi_0\alpha$ when the p-values are positively dependent (Benjamini and Yekutieli, 2001; Sarkar 2002), where π_0 is the (true) proportion of background pixels. The difference between $\pi_0\alpha$ and the FDR gets larger and larger with increasing (positive) dependence among the p-values.

In absence of knowledge of any specific type of dependence structure among the p-values, the method due to Benjamini and Yekutieli (2001), the BY method, is often used. The BY method is an adjusted BH method with α replaced by α/C_N, where $C_N = \sum_{j=1}^{N} j^{-1}$. The BY method is extremely conservative, particularly when N is large, thus is not as powerful as one would hope in detecting true source pixels.

The idea of improving the BH method has been one of the main motivations behind much of the methodological developments taken place in modern multiple testing. This idea has flourished in a number of different directions; for instance, in (i) developing adaptive BH methods incorporating information about π_0 from the data into the BH method or taking an estimation based approach to controlling the FDR (Benjamini and Hochberg, 2000; Benjamini, Krieger and Yekutieli, 2006; Blanchard and Roquain, 2009; Gavrilov, Benjamini and Sarkar, 2009; Sarkar, 2008; Storey (2002); and Storey, Taylor and Siegmund, 2004); (ii) incorporating information about correlations or utilizing the dependence structure into the BH method (Efron, 2007; Romano, Shaikh and Wolf, 2008; Sun and Cai, 2009; and Yekutieli and Benjamini, 1999); and (iii) generalizing the notion of FDR to k-FDR by relaxing control over at most $k-1$ false rejections (Sarkar, 2007; Sarkar and Guo, 2009, 2010).

In the context of present astronomical applications, Hopkins et al. (2002) suggested a way of improving the BY method by incorporating local dependencies. They argue that astronomical images show some degree of correlation between pixels, but are not *fully* correlated. In other words, the brightness intensity of a given pixel is not influenced by all other $N-1$ pixels, rather it is only *partially* correlated with a smaller number (n) of pixels neighboring it. Any real transient signal should have the spatial shape of the stars covering some adjacent pixels, which is called the telescope 'point spread function' (PSF), and this n is related to the number of pixels representing the PSF. They propose to use the BY method with C_N replaced by $C_n = \sum_{i=1}^{n} i^{-1}$ to account for the local dependencies around the source pixels. This is clearly more powerful than the original BY method, but it can be shown that such adjustment to the BY method may fail to control the FDR when $\pi_0 \approx 1$.

Also in astronomical context, Friedenberg and Genovese (2009) considered detecting clusters of pixels, rather than individual pixels, and chose the probability of False Cluster Proportion (FCP) exceeding a certain value as the error rate to control. By relaxing the error rate control to clusters, rather than individuals, there is potential for more powerful procedures due to the reduction in data dimension. However, procedures with cluster-wise control may have some disadvantages compared to individual-wise control, as noted below.

Given the massive influx of data due to large-area surveys, it is crucial to be able to accurately *identify* and *classify* transient sources in real-time data collection. To do so, automated methods must strive to use all the data's available information to first identify and then classify objects (Savage, 2007). This means using not only clusters of outlying observations as the in the FCP, but also using individual pixels to systematically classify astronomical objects as either point-like (i.e. stars, quasars, supernova, etc.) or extended (i.e. galaxies, nebula, etc.). Currently, many classification methods generate a set of 'features' to determine the type of object discovered. Many of these features are estimated with pixel-wise information, such as source positions, fluxes in a range of apertures, and shapes using radial moments. Another nontrivial problem is deblending or splitting of adjacent sources, typically defined as a number of distinct, adjacent intensity peaks connected above the detection surface brightness threshold (Salzberg, 1995; Becker, 2006; Henrion, 2011). Deblending of nearby objects is nearly impossible with a cluster-wise approach. Because of these classification advantages after identifying new sources, we propose new methodology based on the idea of controlling the rate of false discoveries of individual pixels.

3 Proposed Methods

In this paper, we consider using a different idea of incorporating local dependencies and propose an alternative to Hopkins and the BY methods. Our idea is based on the arguments that if the dependencies among the pixels do occur more locally than globally, then by grouping the pixels using an appropriate group size we can make these groups independent of each other. This would be the best scenario where we can apply the BH (more powerful than the BY) method to detect the so called 'potential source groups', which we refer to as the groups containing at least one source pixel. Once a 'potential source group' is identified, we can go back to that group to detect which of the group's individual pixels belong to the source. Based on this general idea of pixel grouping, we propose the following two procedures, by choosing the group size, as in Hopkins et al. (2002), related to the PSF of the telescope. In particular, paralleling Hopkins et al.'s choice of n, the number of pixels representing the PSF, we chose our group size S to be this same quantity. Using this argument, the groups containing S 'partially correlated' pixels should behave independently.

PROCEDURE 1.

Step 1. Divide the data rectangle into D by D mutually exclusive groups. The group size is $S = D^2$ and the total number of groups is $N/S = G$ (say), with N being the total number of pixels (hypotheses).

Step 2. Find the minimum p-value in each of these G groups. Let $P_{\min}^{(g)}$ be that minimum for the gth group, $g = 1, \ldots, G$. Find $Q_g = SP_{\min}^{(g)}$, for $g = 1, \ldots, G$, which we will call the grouped p-values.

Step 3. Apply the BH method to these grouped p-values to detect the 'potential source groups'. That is, consider the (increasingly) ordered versions of the grouped p-values, $Q_{(1)}, \ldots, Q_{(G)}$, and identify those groups as being potential source groups for which the grouped p-values are less than or equal to $Q_{(k^*_{\mathrm{BH}})}$, where $k^*_{\mathrm{BH}} = \max\{g : Q_{(g)} \leq g\alpha/G\}$.

Step 4. Identify the jth individual pixel within the gth potential source group as being a source pixel if the corresponding p-value, say P_{gj}, is such that $SP_{gj} \leq k^*_{\mathrm{BH}}\alpha/G$.

Theorem 1. *Procedure 1 controls the FDR at α if the groups are independent or positively dependent in a certain sense.*

A proof of Theorem 1 is provided in Appendix. Our next procedure is based on the following idea, in addition to that of pixel grouping.

When adapting a multiple testing method to the number of true null hypotheses, say N_0, whether it is for controlling the FDR using the BH method or for controlling the familywise error rate (FWER) using the Bonferroni method (e.g., Finner and Gontscharuk, 2009; Guo, 2009; and Sarkar, Guo and Finner, 2010), the p-values are modified from P_j to $\tilde{P}_j = \hat{N}_0 P_j$, based on a suitable estimate $\hat{N}_0$ of N_0. One of these estimates is due to Storey, Taylor and Siegmund (2004):

$$\hat{N}_0 = \frac{W_N(\lambda) + 1}{1 - \lambda}, \tag{1}$$

where λ is a tuning parameter and $W_N = \sum_{j=1}^{N} I(P_j > \lambda)$ is the number of p-values exceeding λ and provides an information about the number of true null hypotheses in the data. For instance, in case of the Bonferroni method that rejects H_j if $NP_j \leq \alpha$, its adaptive version would reject the H_j if $\hat{N}_0 P_j \leq \alpha$. This would be potentially more powerful.

Notice that such an adaptive p-value is like a 'shrunken p-value', which gets shrunk towards a smaller value, and thus becomes more significant, if there is evidence of more signals in the data. So, when the p-values are locally dependent and tend to have similar local behaviors in terms of being either significant or non-significant, by doing similar adaptation separately within each group by estimating the number of true group specific signals, one could utilize the dependence within each group and potentially improve Procedure 1. With that in mind, we propose our second procedure as follows:

Procedure 2.

Step 1. Same as in Procedure 1.

Step 2. Find the minimum of the p-values in each of these G groups. Let P_{gj}, $j = 1, \ldots, S$, be the p-values in the gth group, and $P_{\min}^{(g)}$ be the minimum of these p-values, $g = 1, \ldots, G$. Find $\tilde{Q}_g = \hat{S}_g P_{\min}^{(g)}$, for $g = 1, \ldots, G$, where

$$\hat{S}_g = \min \left\{ \frac{\sum_{j=1}^{S} I(P_{gj} > \lambda) + 1}{1 - \lambda}, S \right\}, \tag{2}$$

which we will call the grouped adaptive p-values.

Step 3. Apply the BH method to these grouped adaptive p-values to detect the 'potential source groups'. That is, consider the (increasingly) ordered versions of the grouped adaptive p-values, $\tilde{Q}_{(1)}, \ldots, \tilde{Q}_{(G)}$, and identify those groups as being potential source groups for which the grouped adaptive p-values are less than or equal to $\tilde{Q}_{(\tilde{k}^*_{\text{BH}})}$, where $\tilde{k}^*_{\text{BH}} = \max\{g : \tilde{Q}_{(g)} \leq g\alpha/G\}$.

Step 4. Identify the jth pixel within the gth potential source group as being a source pixel if the corresponding p-value P_{gj} is such that $\hat{S}_g P_{gj} \leq \tilde{k}^*_{\text{BH}} \alpha / G$.

Another adaptive method could also be considered by estimating the number of groups that do not contain any source signal, say G_0, and using the estimate $\hat{G}_0$ in place of G in Procedure 1, step 3 and 4. However, because of the sparse number of signals in astronomical data, the estimate $\hat{G}_0$ is often just as large or larger than G itself, providing no additional advantage over Procedure 1. This type of adaptive group estimation is better suited in data where π_0 is not so close to 1.

4 Simulation Study

We ran several simulation studies to examine the FDR control property and the power of our proposed procedures compared to existing methodology. One of the main advantages of the proposed procedures is that there is no dependence assumption of the p-values within each group. Thus, it is only fair to compare our procedures with existing methodology that has such relaxed assumptions (namely, BY and Hopkins).

Since the proposed procedures were developed to control the FDR under arbitrary dependence assumptions within each group, the simulation studies were done under two different dependent scenerios. In the first scenerio, each group's p-values are generated from a multivariate normal distribution with common correlation ($-\frac{1}{S-1} < \rho < 1$).

Second, the p-values were also generated from a multivariate normal distribution, but with an autoregressive type of correlation structure within each group, separately for each of the G groups. An autoregressive correlation structure indicates that data collected in a close spatial proximity tend to be more highly correlated than observations taken further apart. For example, let X_{ij} denote an observation in a particular group located in the i^{th} row and j^{th} column. Then, the correlation between two observations in that particular group can be written as $Corr(x_{ij}, x_{i'j'}) = \rho^{\max(|i-i'|,|j-j'|)}$, for any $0 \leq \rho \leq 1$. In other words, the correlation between two observations decreases in value as the absolute spatial distance between (i, i') or (j, j') increases.

Under these two correlation structures, we generated S dependent standard normal random variables independently for each of the G groups. Three of the G groups were chosen randomly for each simulation and one of the values $2, 3$ and 4 is added to the variables in each of these three groups. In other words, only three groups were assumed to contain all the signals. Simulation studies with varying number of signal groups (1 group to 10 groups, instead of 3 groups) were also computed, but since they yielded similar results, we have decided to restrict the discussion of our simulation studies to 3 signal groups. The group size S was chosen to be 25, using 5×5 groups ($D = 5$). The number of groups is $G = 900$, totaling $n = 22,500$ individual hypotheses per simulation. Since each simulation contained a fixed 3 groups of signal each of size 25, the proportion of true null hypotheses $\pi_0 = 1 - \frac{75}{22,500} = 0.996$. Using both correlation structures, we repeated this 1,000 times at each value of ρ.

Four methods were compared: Benjamini-Yekutieli, Hopkins', the proposed Two-Stage, and proposed Adaptive Two-Stage Procedure, using $\lambda = 0.5$. At each simulation, we estimate FDR by the proportion of falsely rejected hypotheses and the power is estimated by proportion of correctly rejected hypotheses. The average proportion of correctly and falsely rejected hypotheses over all repetitions is shown in Figure 1 for the fixed group correlation and in Figure 2 for the autoregressive case.

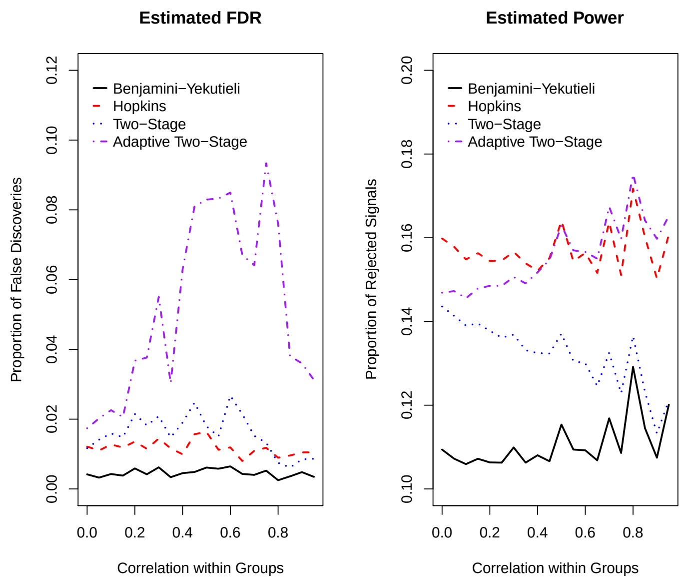


Fig. 1 Simulated FDR and Power for fixed group correlation structure

When examining the simulated power in the right side of Figure 1, both the Two-Stage and Adaptive Two-Stage Procedures outperform the BY method with the fixed group correlation structure. In other words, the Two-Stage Procedures correctly identify a higher proportion of hypotheses containing a signal. The Adaptive Two-Stage Procedure has competitive power with Hopkins' procedure and surpasses it when the within group fixed correlation becomes large ($\rho > 0.5$).

The simulated FDR in the left side of Figure 1, reveals a stable Two-Stage Procedure, with the estimated FDR < 0.05 across all fixed group correlations. However, the Adaptive Two-Stage Procedure seems to lose control of the FDR with moderately correlated data within groups ($0.5 < \rho < 0.8$). Although unfortunate, this result is not surprising. Other adaptive methodology also become unstable with large correlation among hypotheses.

Next, we look at the performance of the proposed procedures under the autoregressive within group correlation structure. When examining the simulated power in the right side of Figure 2, both the Two-Stage and Adaptive Two-Stage Procedures outperform the BY method under this group correlation structure. The simulated FDR in the left side of Figure 2, reveals a

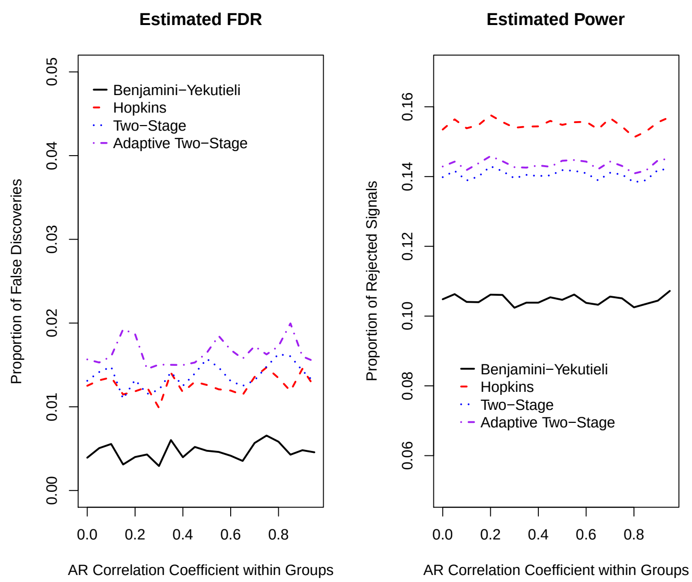


Fig. 2 Simulated FDR and Power for autoregressive correlation structure

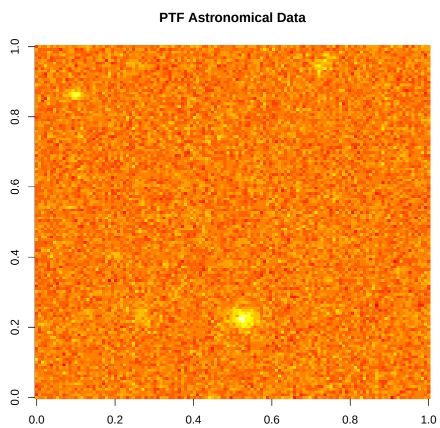
stable Two-Stage Procedure and Adaptive Procedure, with the estimated FDR < 0.05 across all autoregressive group correlations values of ρ.

In conclusion, the simulation study confirms that between the proposed Two-Stage Procedure and the BY method, both of which are theoretically known to control the FDR under arbitrary dependence within the groups, the former is clearly the better choice in terms of controlling the FDR under this dependence situation. Moreover, it is competitive with Hopkins', even though Hopkins' may not control the FDR. The simulation study also seems to indicate that the Adaptive Two-Stage Procedure controls the FDR when the correlation is fixed and small ($0 < \rho < 0.5$), but may become unstable as correlation gets more extreme. Impressively, the Adaptive Two-Stage Procedure under the autoregressive correlation scenario, seems to control the FDR over all positive values of ρ, which is yet to be proved theoretically.

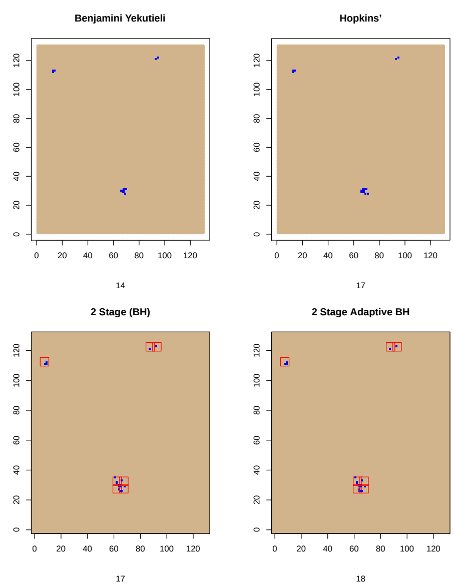
5 Application

The astronomical data used to illustrate our procedures comes from Palomar Transient Factory (PTF), one of the mid-size wide-field survey projects currently underway. Each image is 2048×4096 pixels, but a smaller sub-rectangle of noise (130×130) was chosen to apply the methods.

The data is approximately normally distributed with mean $\bar{x} = 721.7$ and variance $s^2 = 476.1$. A heat map of the image can be seen below in Figure 3a and the results in Figure 3b. The data were first standardized and converted to p-values. Results of four methods are presented: BY, Hopkins, Two-Stage BH (Procedure 1), and Adaptive Two-Stage BH (Procedure 2). Again, we have chosen $\lambda = 0.5$ in Procedure 2. Applying the BY procedure to the data rejects fourteen pixels and Hopkins rejects an additional three pixels. On the other hand, using our Two-Stage BH method, seven potential source groups are found to have seventeen source pixels and the Adaptive Two-Stage BH finds eighteen from those seven potential source groups.



(a) Image of Astronomical data from the Palomar Transient Factory



(b) The results of the four methods on the PTF Astronomical data. The blue points represent source pixels and the red boxes represent a potential source group. Below each plot is the total number of source pixels found using that method.

Fig. 3 Results from Palomar Transient Factory data

6 Concluding Remarks

We have proposed, in this research, two new FDR controlling methods to be used in group-dependent data - Two-Stage BH method and Adaptive Two-Stage BH method - and compared them with the existing methods of Benjamini-Yekutieli and Hopkins'. Both of our proposed methods compare favorably to the BY method in terms of the proportion of detected source pixels. When the group correlation is small ($\rho < 0.5$) or large ($\rho > 0.8$), both of these methods retain control of the FDR; however, when this correlation is moderate ($0.5 < \rho < 0.8$), the adaptive procedure seems to become unstable.

More investigation is needed to estimate the dependence structure of astronomical data to see if the local correlation is small enough to warrant use of adaptive methods. Further simulation studies should be done with larger repetitions, varying π_0, and incorporating other dependence structures.

It would also be interesting to study the astronomical source detection problem differently by adding a third dimension. Since astronomy data is often collected nightly, the assemblage can be thought of as a 'data cube' instead of a 'data matrix', where the first and second dimension are the spatial location and the third dimension is the date/time of observation. In other words, multiple testing procedures can be adapted to not only search for signals at every i^{th} row and j^{th} column location, but also at every time t. This set up could be explored in both a frequentist and Bayesian contexts.

The authors would like to thank Eric Feigelson for acclimating us to transient detection methodology and the goals of astronomical research, Peter Nugent for supplying the PTF data, and Peter Freeman for his commentary regarding the False Cluster Proportion methodology. The research of Sarkar and Guo were supported by NSF Grants DMS-1006344 and DMS-1006021 respectively.

7 Appendix

Proof of Theorem 1. We first prove the theorem assuming that the groups are independent. For that we need the following notations:

R: Number of source pixels detected,

V: Number of source pixels falsely detected,

RG: The index of the ordered (in terms of increasing values of grouped p-values) potential source group detected (which is also k^*_{BH}),

$RG^{(-k)}$: The index of the ordered potential source group detected based on the BH method applied to all the groups except the kth one and the critical values $g\alpha/G$, $g = 2, \ldots, G$, and

$J_0(g)$: The set of indices of the p-values in the gth group that correspond to background pixels.

Then,

$$\begin{aligned}
\text{FDR} &= E\left\{\frac{V}{\max\{R,1\}}\right\} = E\left[E\left\{\frac{V}{\max\{R,1\}}\Big| RG, R\right\}\right] \\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\sum_{r=1}^{N}\frac{1}{r}Pr\left\{SP_{kj} \le \frac{g}{G}\alpha, RG = g, R = r\right\} \\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\sum_{r=1}^{N}\frac{1}{r}Pr\left\{P_{kj} \le \frac{g}{N}\alpha, RG^{(-k)} = g-1, R = r\right\} \\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\sum_{r=1}^{N}\frac{g\alpha}{rN}Pr\left\{RG^{(-k)} = g-1, R = r\Big| P_{kj} \le \frac{g}{N}\alpha\right\} \\
&\le \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\sum_{r=1}^{N}\frac{\alpha}{N}Pr\left\{RG^{(-k)} = g-1, R = r\Big| P_{kj} \le \frac{g}{N}\alpha\right\} \\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\frac{\alpha}{N}Pr\left\{RG^{(-k)} = g-1\Big| P_{kj} \le \frac{g}{N}\alpha\right\} \qquad (3)\\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\sum_{g=1}^{G}\frac{\alpha}{N}Pr\left\{RG^{(-k)} = g-1\right\} \\
&= \sum_{k=1}^{G}\sum_{j\in J_0(k)}\frac{\alpha}{N} = \frac{N_0}{N}\alpha \le \alpha.
\end{aligned}$$

In (3), the fifth equality follows from the assumption that P_{kj} is distributed as $U(0,1)$ when it corresponds to a background pixel, the first inequality follows from the fact that $RG \le R$, and the seventh equality follows from the independence assumption of the groups. This proves the theorem under independence of the groups.

If the groups are not completely independent of each other, we will assume that they are positively dependent in the following sense:

The conditional expectation

$$E\left\{\phi(\mathbf{P}^{(-g)})\big|P_{gj} = u\right\}, \qquad (4)$$

where $\mathbf{P}^{(-g)}$ is the set of p-values corresponding to all pixels except those in the gth group. P_{gj} is the jth p-value corresponding to a background pixel in the gth group, and $\phi(\mathbf{P}^{(-g)})$ is an increasing (coordinatewise) function of all the p-values except those in the gth group, is non-decreasing in $u \in (0,1)$ for each g and j.

From (3), we note that

$$
\begin{aligned}
\mathrm{FDR} &\leq \sum_{k=1}^{G} \sum_{j \in J_0(k)} \sum_{g=1}^{G} \frac{\alpha}{N} Pr\left\{ RG^{(-k)} = g-1 \middle| P_{kj} \leq \frac{g}{N}\alpha \right\} \\
&= \sum_{k=1}^{G} \sum_{j \in J_0(k)} \sum_{g=1}^{G} \frac{\alpha}{N} \left[Pr\left\{ RG^{(-k)} \geq g-1 \middle| P_{kj} \leq \frac{g}{N}\alpha \right\} \right. \\
&\qquad \left. - Pr\left\{ RG^{(-k)} \geq g \middle| P_{kj} \leq \frac{g}{N}\alpha \right\} \right] \\
&\leq \sum_{k=1}^{G} \sum_{j \in J_0(k)} \sum_{g=1}^{G} \frac{\alpha}{N} \left[Pr\left\{ RG^{(-k)} \geq g-1 \middle| P_{kj} \leq \frac{g-1}{N}\alpha \right\} \right. \\
&\qquad \left. - Pr\left\{ RG^{(-k)} \geq g \middle| P_{kj} \leq \frac{g}{N}\alpha \right\} \right] \\
&= \sum_{k=1}^{G} \sum_{j \in J_0(k)} \frac{\alpha}{N} = \frac{N_0 \alpha}{N} \leq \alpha.
\end{aligned}
$$

The second inequality follows from the assumption (4) of positive dependence of groups. This completes our proof of Theorem 1.

References

1. Becker, A.C (2006). Transient Detection and Classification. *Astronomical Notes* **88**, 789-792.
2. Benjamini, Y & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, B* **57**, 289-300.
3. Benjamini, Y. & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics.* **25**, 60-83.
4. Benjamini, Y, Krieger, K. & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491-507.
5. Blanchard, G. & Roquain, E. (2009). Adaptive FDR control under independence and dependence. *Journal of Machine Learning* **10**, 2837–2871.
6. Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the Royal Statistical Society* **102**, 93-103.
7. Finner, H. & Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *Journal of the Royal Statistical Society, B* **71**, 1031–1048.
8. Friedenberg, D. & Genovese, C. (2009) Straight to the Source: Detecting Aggregate Objects in Astronomical Images with Proper Error Control. arXiv: 0910.5449.

9. Gavrilov, Y., Benjamini, Y. & Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control. *Annals of Statistics* **37**, 619–629.

10. Guo, W. (2009). A note on adaptive Bonferroni and Holm procedures under dependence. *Biometrika*, **96**, 1012-1018.

11. Henrion, M., Mortlock, D., Hand, D., Gandy, A. (2011). A Bayesian approach to star-galaxy classification. *Monthly Notices of the Royal Astronomical Society*, **412**, 2286-2302.

12. Hopkins, A. M., Miller, C. J., Connolly, A. J., Genovese, C., Nichol, R. C. & Wasserman, L. (2002). A new source detection algorithm using the false discovery rate. *The Astronomical Journal* **123**, 1086-1094.

13. Romano, J. P., Shaikh, A. M. & Wolf, M. (2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST* **17**, 417–442.

14. Salzberg, S., Chandler, R., Ford, H., Murthy, S., and White, R. (1995). Decision Trees for Automated Identification of Cosmic-Ray Hits in Hubble Space Telescope Images. *The Astronomical Society of the Pacific* **107**, 279-288.

15. Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Annals of Statistics* **30**, 239–257.

16. Sarkar, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Annals of Statistics* **35**, 2405-2420.

17. Sarkar, S. K. & Guo, W. (2009). On a generalized false discovery rate. *Annals of Statistics* **37**, 337-363.

18. Sarkar, S. & Guo, W. (2010). Procedures controlling generalized false discovery rate using bivariate distributions of the null p-values. *Statistica Sinica* **20**, 1227-1238.

19. Sarkar, S. K. (2008). On methods controlling the false discovery rate (with discussions). *Sankhya* **70**, 135–168.

20. Sarkar, S. K., Guo, W. & Finner, H. (2011). On adaptive procedures controlling the familywise error rate. To appear in the Journal of Statistical Planning and Inference.

21. Savage, R. & Oliver, S. (2007). Bayesian Methods of Astronomical Source Extraction. *The Astrophysical Journal* **661**, 1339-1346.

22. Storey, J.D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, B* **64**, 479-498.

23. Storey, J. D., Taylor, J. E. & Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society, B* **66**, 187-205.

24. Sun, W. & Cai, T. (2009). Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society, B* **71**, 393-424.

25. Yekutieli, D. & Benjamini, Y. (1999) Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference* **82**, 171-196.