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

Sanat K. Sarkar
Temple University, Philadelphia, PA

This is a joint work with Nicolle Clements
Introduction and the Statistical Problem

Detecting and monitoring transient sources in the night sky is an area of astronomical research that receives much attention. Telescope scanners of the sky create gray scale images, which are converted pixel by pixel to numerical values. The number of pixels in each image can be several millions in size, which makes manual source detection impossible. The goal of researchers is to define thresholds above which pixels are believed to belong to real sources.

Source Pixel is a pixel in an image that is above some threshold and thus is part of a true source (transient object). Source is a collection of these source pixels that correspond to an astronomical object of interest.

Background Pixel is an image pixel that does not come from a source.
Typically, the data each night are assumed to come from a mixture Gaussian distribution. The observations from all the pixels are assumed to follow Gaussian distributions with a common variance $\sigma^2$. The mean is $\mu$ for a background pixel and $\mu + \theta_i$ for the $i$th source pixel.

To get around the nightly differences, astronomers typically subtract the background $\mu$ and divide by the noise level $\sigma$. 
Let $Z_i$ be the standardized value for the $i$th pixel. Then, given these $Z_i$, the main statistical problem is to

(i) consider testing the (null) hypothesis $H_i : \theta_i = 0$ against $\theta_i > 0$, simultaneously for all $i$,

(ii) determine a threshold $c$ (fixed or data-dependent) such that $Z_i > c$ will indicate the corresponding $H_i$ to be false (i.e., the pixel belongs to a source) and an appropriate measure of multiple testing error is controlled at a specified level $\alpha$. 
Different Measures of Multiple Testing Error

Table 1. The outcomes in testing $N$ null hypotheses (pixels)

<table>
<thead>
<tr>
<th></th>
<th>Rejected</th>
<th>Accepted</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Null</td>
<td>$V$</td>
<td>$U$</td>
<td>$N_0$</td>
</tr>
<tr>
<td>False Null</td>
<td>$S$</td>
<td>$T$</td>
<td>$N_1$</td>
</tr>
<tr>
<td>Total</td>
<td>$R$</td>
<td>$A$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

Commonly used measures involving type I errors (detecting background pixels as source pixels):

- $\Pr(V > 0) \rightarrow$ Familywise error rate (FWER)
- $E\left(\frac{V}{\max\{R,1\}}\right) \rightarrow$ False Discovery Rate (FDR)
General Multiple Testing Procedures

\( Z_i \rightarrow \) p-value \( P_i = 1 - \Phi(Z_i), i = 1, \ldots, N. \)

\( P_{(1)} \leq \cdots \leq P_{(N)} \rightarrow \) the ordered p-values.

Given a non-decreasing set of critical constants \( 0 < \alpha_1 \leq \cdots \leq \alpha_n < 1 \) (to be determined subject to a control of the FWER or FDR),

a **stepdown** procedure rejects \( H_i \) for all \( i \) such that \( P_i \leq P_{(\hat{K}_{SD})} \), where

\[
\hat{K}_{SD} = \max\{1 \leq i \leq N : P_{(j)} \leq \alpha_j \ \forall j \leq i\}
\]

if the maximum exists, otherwise accepts all the null hypotheses;

a **stepup** procedure rejects \( H_i \) for all \( i \) such that \( P_i \leq P_{(\hat{K}_{SU})} \), where

\[
\hat{K}_{SU} = \max\{1 \leq i \leq N : P_{(i)} \leq \alpha_i\},
\]

if the maximum exists, otherwise accepts all the null hypotheses;
Level $\alpha$ FWER Controlling Procedures

Bonferroni

A single-step method, rejects $H_i$ if $P_i \leq \alpha/N$.

Does not require any dependence assumption

Holm

A stepdown method with the critical constants $\alpha_i = \alpha/(N - i + 1)$.

A stepdown implementation of the Bonferroni, does not require any dependence assumption.

Hochberg

A stepup method with the critical constants $\alpha_i = \alpha/(N - i + 1)$.

More powerful than Holm, but requires some dependence assumption.
The notion of FWER is too stringent, particularly in large-scale multiple testing, not allowing some real source pixels to be correctly identified.

The False Discovery Rate, being a more powerful concept than the FWER, has now been the most commonly used notion of error rate in large-scale multiple testing.
Level $\alpha$ FDR Controlling Procedures

The Benjamini-Hochberg (BH, 1995) method is the most commonly used method of controlling the FDR. It is a stepup method with the critical constants $\alpha_i = i\alpha/N$, $i = 1, \ldots, N$. Thus, it operates as follows:

- 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 $P_{(K_{BH})}$.
The FDR of the BH method equals $\pi_0 \alpha$ when the $p$-values (the pixels) are independent, and is less than $\pi_0 \alpha$ when the $p$-values (the pixels) are positively dependent in a certain sense (Benjamini and Yekutieli, 2001; Sarkar 2002), where $\pi_0 = N_0/N$ 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 (pixels), the method due to Benjamini and Yekutieli (2001), the BY method, is often used. It operates as follows:

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

The BY method is basically the BH method with $\alpha$ replaced by $\alpha/C_N$, and so it is extremely conservative, particularly when $N$ is really large, and 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 $\pi_0$ 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 by relaxing control over at most a few false rejections (Sarkar, 2007; Sarkar and Guo, 2009, 2010).
In this talk, I’ll introduce two procedures improving the BY method under local dependency. This type of dependency occurs in astronomical applications. One of these procedures is based on results observed for adaptive BH methods.
Adaptive BH Methods

Use adaptive p-values $\hat{\pi}_0 P_i$, instead of $P_i$, in the BH method, for some suitable (conservative) estimate $\hat{\pi}_0$ of $\pi_0$.

One of these estimates of $\pi_0$, due to Storey, Taylor and Siegmund (2004) and often used, is the following:

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

where $\lambda$ is a tuning parameter and $W_N = \sum_{j=1}^{N} I(P_j > \lambda)$ is the number of $p$-values exceeding $\lambda$ and provides an information about the number of true null hypotheses (background pixels) in the data.
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 there are p-values that are dependent on each other and tend to have similar behaviors in terms of being either significant or non-significant, such adaptation could utilize the dependence within the p-values and potentially improve the BH method.

In fact, while the ultimate control of the FDR by an adaptive BH method has been theoretically proved only under independence of the p-values (pixels), there is numerical evidence that it works even under positive dependence as long as this dependence is not too high.

Similar results have been given recently for improving the FWER control of the Bonferroni method using adaptive p-values.
Two New Procedures with Applications to Transient Source Detection.

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 tens of adjacent pixels, which is called the telescope ‘point spread function’ (PSF), and this \( n \) is related to the number of pixels representing the PSF of the telescope. 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 \( \pi_0 \approx 1 \), which can certainly be the case when trying to detect sparse transient signals in astronomical data.
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 the so-called 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 $g$th group, $g = 1, \ldots, G$. Find $Q_g = S P_{\min}^{(g)}$, 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 $Q_{(K_{BH}^*)}$, where $K_{BH}^* = \max\{g : Q_{(g)} \leq g\alpha/G\}$.

Step 4. Identify the $j$th individual pixel within the $g$th potential source group as being a source pixel if the corresponding $p$-value, say $P_{gj}$, is such that $SP_{gj} \leq K_{BH}^*\alpha/G$. 
Theorem 1. Procedure 1 controls the FDR at $\alpha$ if the groups are independent or positively dependent in a certain sense.
When the $p$-values are believed to be locally dependent and tend to have similar local behaviors in terms of being either significant or non-significant, by considering adaptive $p$-values 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. (An Adaptive Version of Procedure 1)

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 $g$th group, and $P_{\text{min}}^{(g)}$ be the minimum of these $p$-values, $g = 1, \ldots, G$. Find $\tilde{Q}_g = \hat{S}_g P_{\text{min}}^{(g)}$, for $g = 1, \ldots, G$, where

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

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 $\tilde{Q}_{(K_{\text{BH}}^*)}$, where $K_{\text{BH}}^* = \max\{g : \tilde{Q}_{(g)} \leq g\alpha / G\}$.

Step 4. Identify the $j$th pixel within the $g$th potential source group as being a source pixel if the corresponding $p$-value $P_{gj}$ is such that $\hat{S}_g P_{gj} < K_{\text{BH}}^* \alpha / G$. 

22
Simulation Study

We ran a simulation study to examine the FDR control property and the power of our proposed procedures in comparison with Hopkins’ and the BY procedures.

We generated $S$ dependent random variables $X_i \sim N(\theta_i, 1)$, $i = 1, \ldots, S$, with a common non-negative correlation $\rho$, independently for each of the $G$ groups. Three of the $G$ groups were chosen randomly for each simulation to contain a signal $\theta_i = 2, 3,$ or 4 and all other $\theta_i = 0$. We repeated this 4,000 times by allowing $\rho$ to vary from 0 to 0.95 by steps of 0.05 using 200 repetitions each. The group size $S$ was chosen to be 25, using $5 \times 5$ groups. The number of groups was $G = 900$, totaling $N = 22500$ individual pixels. Since each simulation contained a fixed 3 groups of signal each of size 25, the proportion of true null hypotheses $\pi_0 = 0.996$. At each simulation, we applied the procedures at level $\alpha = 0.05$. 

Sarkar's research is supported by NSF Grant DMS-106344

23
Figure 1a is one of the simulated artificial images with $\rho = 0.3$. Visually, one should be able to see the 3 source groups being lighter in color than the background. Figure 1b presents the results of the four methods: BY, Hopkins’, Two-Stage BH (Procedure 1), and Adaptive Two-Stage BH (Procedure 2). We have chosen $\lambda = 0.5$ while applying Procedure 2.

The BY method finds three rejected pixels from two of the three simulated sources and Hopkins’ shows no improvement. Our proposed procedures, on the other hand, pick up significant pixels from all these three sources, Procedure 1 (Two-Stage BH) finds six and its adaptive version, Procedure 2, detects additional two pixels.
(a) Heat Map of Simulated Example.

(b) Results of four methods. Blue points indicate rejected pixels and red boxes indicate a significant group. The number of detected source pixels is below each plot.

Figure 1: Simulation Results.

Sarkar's research is supported by NSF Grant DMS-1006344
Proportion of False Rejections

Power

(a) Simulated FDR for each of the four methods.

(b) Simulated power for each of the four methods.

Figure 2: Simulated FDR and power.
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 $2048 \times 4096$ pixels, but a smaller sub-rectangle of noise ($130 \times 130$) 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 Procedure 1 (Two-Stage BH), seven potential source groups are found to have seventeen source pixels and its adaptive version, Procedure 2, 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.

Figure 3: Results from Palomar Transient Factory data.
Conclusion and Future Work

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 (< 0.5), both of these methods retain control of the FDR; however, when this correlation is large, 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 that are larger than 4,000 and varying $\pi_0$. 

Sarkar's research is supported by NSF Grant
DMS-1006344
References


Acknowledgements

- Eric Feigelson for acclimating us to transient detection methodology and the goals of astronomical research, and

- Peter Nugent for supplying the PTF data.