The three vetoes have been tuned empirically for future use in analyzing $\chi^2$ statistics of $x$, which need to be investigated to either determine how better to $w = 1$, in consultation with its main contributors. To achieve a high detection rate of 99% at a set of high and quality planet candidates. The new set of vetoes, described here, will be the subject of future investigation and use in TPS. The new vetoes will improve the recovery rate of injected signals and the tuning of these vetoes will play a major role in increasing our search sensitivity to compensate, in some part, for Kepler’s inability to continue its nominal extended mission. Funding for the Kepler Mission has been provided by the NASA Science Mission Directorate.

### Detection Statistics

The Single Event Statistic (SES) is a simple matched filter in the whitened domain, which is the optimal detector for a deterministic signal in colored Gaussian noise.

$$\text{SES} = \frac{x_k - \hat{x}_k}{\sqrt{\sum_k \hat{x}_k^2}}$$

Here the $\hat{\cdot}$ denotes a whitened vector, $x$ is the data, and $\hat{x}_k$ is the trial transit pulse centered at time $n = j$. A grid is then constructed on the three dimensional parameter space of physically realistic pulsation duration ($\Delta t$), period ($T$), and epoch ($t_0$). The grid is constructed on the three dimensional parameter space in a way that strikes a balance between reducing signal loss due to signal-template mismatch and algorithm execution time. Each (0, T, j) triplet determines a set of P cadences, corresponding to the centers of the transits in the full pulse train that are combined to form a Multiple Event Statistic (MES).

The MES are then computed as:

$$\text{MES} = \sum_{k=0}^{N} \frac{x_k - \hat{x}_k}{\sqrt{\sum_k \hat{x}_k^2}}$$

The current threshold used for rejecting events based on the MES is 7.1. For events with a MES that exceeds this threshold, a Robust Detectors Statistic (RDS) is computed. To compute the RDS, a set of fit weights, $w$, are obtained by robustly fitting a whitened transit pulse train to the whitened data, after windowing to remove all but the in-transit cadences. The RDS are then computed as:

$$\text{RDS} = \frac{\chi^2_{w}}{\chi^2_{w}}$$

The current threshold used for rejecting events based on the RDS is 6.4.

### Pitfalls

In order for the three vetoes, discussed subsequently, and their associated component quantities to have the correct statistical properties, there are three subtleties that must be taken care of.

1. The whitening coefficients used when both the data and signal vectors are mathematically independent of the presence of any transits, depending on the data in the data.
2. The whitening coefficients are estimated in the wavelet domain through the use of a moving circular Modulus Absolute Deviation (MAD).
3. Although this method of estimating the whitening coefficients is robust against outliers, the effect of any signal in the data on the noise estimates, however small it may be, still perturbs the statistical properties of the vetoes.
4. Prior to doing the wavelet decomposition of the data, all in-transit cadences are gapped and filled using an adaptive Auto Regressive (AR) method to explicitly remove any signal that might be present in the data.

5. Due to the process of whitening the data in the wavelet domain, transits that might be present in the data get smeared out and overlap one another, thereby losing their independence.

- Consider that the data vector can in general be written:

$$\hat{x}(n) = \hat{w}(n) + \Delta h(n)$$

where $\hat{w}(n)$ is the noise signal, $\Delta h(n)$ is the signal amplitude, and $\hat{w}(n)$ is the signal vector. Instead of using the full data vector in the whitening, the contribution from each transit pulse should be separated out by gapping and filling in-transit cadences for all transits except the one whose center is at $n = m$, effectively breaking the data up into components:

$$\hat{x}(n) = \hat{w}(n) + \Delta h(n)$$

3. Due to the process of whitening the trial transit pulse train in the wavelet domain, the data, the pulse get smeared out and overlap one another, thereby losing their independence.

- Again, the full pulse train must be broken up so that each transit is treated individually.
- Each transit pulse is whitened separately and windowed to remove the effect of any out-of-transit cadences.

### After handling these pitfalls, a new version of the MES, which is not in general equal to the MES, can be computed as:

$$Z = \frac{\sum_{kk} x_k - \hat{x}_k}{\sqrt{\sum_k \hat{x}_k^2}}$$

### $\chi^2$ Vetoes

There are three $\chi^2$ vetoes used by TPS, two of which are described in detail in Seader et al.

$$\chi^2_{\text{vetoes}} = \sum_{kk} x_k - \hat{x}_k$$

- Measures the difference between the squared amplitude of the detector output and the squared SNR

$$x_k = \frac{\sum_{kk} x_k \hat{x}_k}{\sum_k \hat{x}_k^2}$$

- Expectation value is equal to one less than the number of in-transit cadences (ignoring signal/template mismatch)

$$x_k = \frac{1}{P - 1}$$

- This quantity is $\chi^2$ distributed with $P - 1$ degrees of freedom (ignoring signal/template mismatch)

### Detection Statistics

- Break $Z$ up into $P$ additive pieces, where $P$ is the number of transits

$$Z = \sum_{jj} z_j$$

- For each piece, compute the fractional expected value for comparison

$$q_j = \frac{x_j - \hat{x}_j}{\sqrt{\sum_k \hat{x}_k^2}}$$

- Form the differences between observed and expected values

$$\Delta q_j = z_j - q_j z$$

- Form a statistic from these differences

$$x^2_{\text{vetoes}} = \sum_{jj} (\Delta q_j)^2$$

- The expected value is less than the number of transits (ignoring signal/template mismatch)

$$x^2_{\text{vetoes}} = P - 1$$

- This quantity is $\chi^2$ distributed with $P - 1$ degrees of freedom (ignoring signal/template mismatch)

### Pitfalls

- Motivated by the argument of the joint probability of the set of SES from which the MES is formed from

$$x_k = \frac{\sum_{kk} x_k \hat{x}_k}{\sum_k \hat{x}_k^2}$$

- Formal a classical $\chi^2$ statistic by computing the observed set of SES to their corresponding expected values

$$x^2_{\text{vetoes}} = \sum_{kk} (x_k - \hat{x}_k)^2$$

- The number of degrees of freedom used for this version is $P - 1$

- This degrees of freedom overshoots the true expectation value, indicating that this version is likely not $\chi^2$ distributed.

- More work is needed to understand the statistical properties of this version and whether it should be modified or dropped from use.

- The threshold used for this veto was set too high in the past and led to an unacceptably high number of false dismissals – the threshold is much lower now.

### Detection Probabilities and False Alarm Probabilities

- So you can see that $\eta_{\text{veto}}$ is doing a majority of the work by vetoes 32,321 out of the total 35,616 false alarms. vetoes the three collectively.

- $\eta_{\text{veto}}$ uniquely vetoes an additional 1,867 transits $\eta_{\text{veto}}$ uniquely vetoes an additional 1,867 transits

### Results

- The current thresholds were chosen by analysis of results obtained by varying each of the three thresholds simultaneously.

- Using the current thresholds, the overall (MES, RDS, and $\chi^2$ vetoes) false alarm probability is just 4% while the detection probability is 96%.

- This Venn diagram shows the overlap among the false alarm targets that are vetoed by the various $\chi^2$ vetoes.

### References