6 Significance and Limit calculation

This section will describe the test statistic to evaluate if the measured alerts are background conform or can be attributed to a signal source. The test statistic is based on multiplying the p-values of different components. They can be separated into two categories, the pure number of alerts and the probability that doublets stem from background.

The Optical Follow-Up was designed with the expectation that most of the alerts will be triggered due to background events of atmospheric neutrinos. An excess of detected multiplets in comparison to the expected background average can indicate a possible signal contribution.

Given the average background expectation (Section ???) $\mu_{k,b}$ for multiplets with multiplicity k, the probability to see N_k or more alerts during one season is

$$P(N_k,\mu_k) = \sum P_{\text{Poisson}} = \sum_{i=N_k}^{\infty} \frac{\mu_k^i}{i!} \exp^{-\mu_k} . \tag{6.1}$$

The sum is aborted when a precision of 10^{-6} is reached. The probabilities for the different multiplicities k and seasons i are then multiplied to form the test statistic λ_{na} to evaluate the number of alerts

$$\lambda_{na}\left(N_k^i,\mu_{k,b}^i\right) = \prod_{k=2}^{\infty}\prod_{i} P(N_k^i,\mu_{k,b}^i). \tag{6.2}$$

The second component is only valid for doublets as higher multiplicities are automatically considered true alerts. The OFU system generates up to about 50 doublets per year of which only seven can be sent to Swift. The down selection is based on the OFU test statistic (Section ???) separating the more signal like and the more background like doublets. For each doublet an OFU test statistic value is drawn and compared to the background distribution to calculate the p-value $p_{j,i}^{ofu}$. The p-values of all doublets can not simply be multiplied as the number of doublets would influence the significance of the test statistic. E.g. three background doublets would form a smaller overall p-value $O(10^{-3})$ than one signal doublet $O(10^{-1})$. A combination of all $p_{j,i}^{ofu}$ was chosen to keep the total contribution in the same order of magnitude independent of the number of doublets. All $p_{j,i}^{ofu}$ are multiplied and the $N_2^{i\,th}$ square root is taken of the product with N_2^i being the number of considered doublets of season i

$$\lambda_{ofu} = \prod_{i} \sqrt[N_2^i]{\prod_{j=1}^{N_2^i} p_{j,i}^{ofu}} \tag{6.3}$$

Calculating $p_{j,i}$ is based on several steps. The OFU test statistic distribution differs for different zenith regions. To evaluate the correct test statistic a zenith value is drawn for each doublet according to the squared singlet rate (Figure 10, 11).

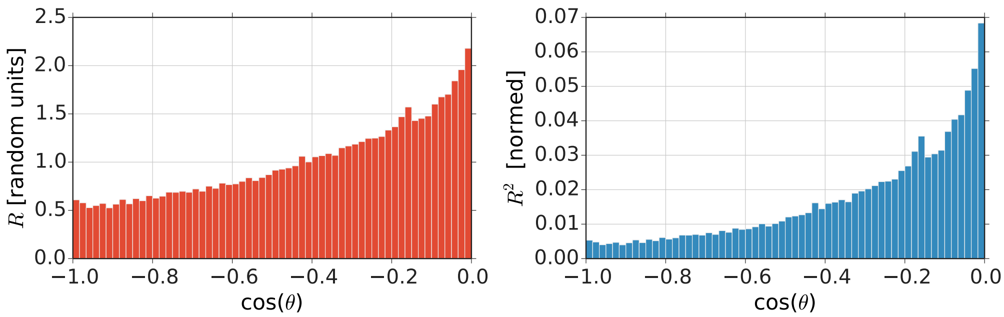


Figure 10: The plots show the cosinus zenith distribution for background of the IC86-BDT season. The left plot depicts the singlet rate in random units and the right plot is the probability density function of the doublet distribution (singletrate squared)

The northern sky was divided into twenty even regions in $\cos(\theta)$ and the probability density functions of the OFU test statistic were created for background and signal of all considered spectra for all seasons in each region. An

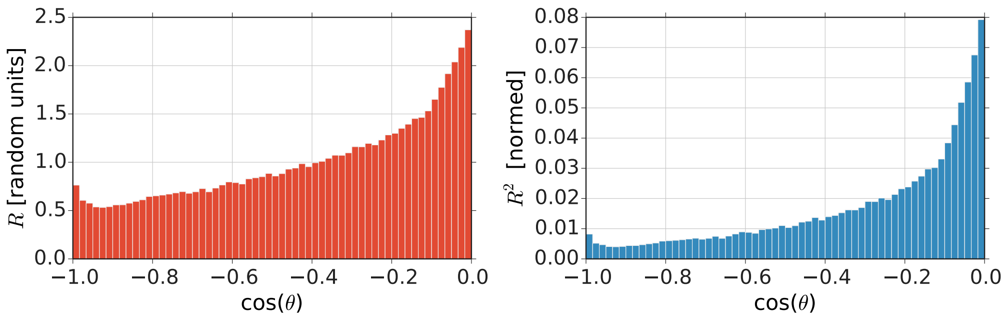
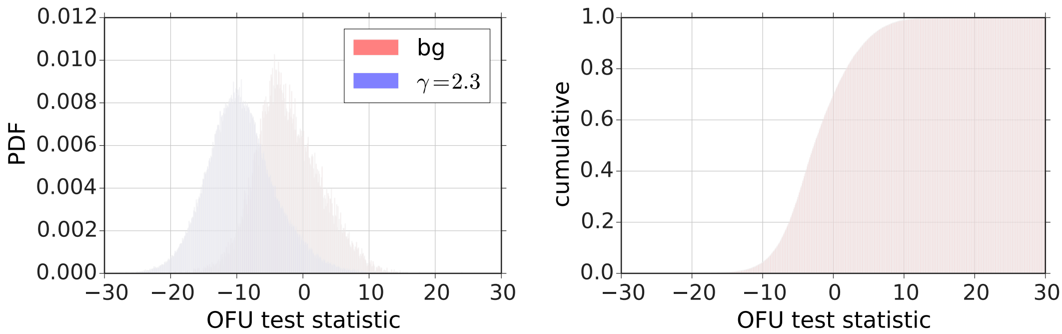


Figure 11: The plots show the cosinus zenith distribution for signal with a spectral index of $\gamma = 2.3$ of the IC86-BDT season. The left plot depicts the singlet rate in random units and the right plot is the probability density function of the doublet distribution (singletrate squared)

example is shown in Figure 12a for the first zone from the pole for the IC86-BDT season. Using the zenith value the correct OFU test statistic probability density function is chosen to draw a value of the statistic. As the probability to have a Swift doublet is already included in calculating the expected number of doublets μ_2^i, the test statistic value must be smaller than the Swift cut of the season. The chosen value is then compared to the cumulative background distribution (Figure 12b) and the p-value is extracted.



(a) The probability density functions of the OFU test statistic (background and signal with a spectral index of $\gamma = 2.3$) for the first region starting from the pole for the IC86-BDT season.

(b) The cumulative distribution of the background OFU test statistic for the IC86-BDT season.

The complete test statistic is the combination of both components

$$\lambda = \lambda_{na} \cdot \lambda_{\text{ofu}} \tag{6.4}$$

The final significance and limit calculation is based on running pseudo experiments. To evaluate the agreement of the measured alerts with a pure background hypothesis, the average background expectation per season and multiplicity $\mu_{k,b}^i$ is used to draw a number of detected multiplets N_k^i according to a Poisson distribution per experiment. For all N_2^i doublets, $p_{j,i}^{ofu}$ is calculated. In total, 10^6 experiments are performed per season and multiplicity and the resulting test statistic values are compared to the measured result (λ_m). The p-value is the fraction of experiments with worse agreement to the background hypothesis ($\lambda < \lambda_m$).

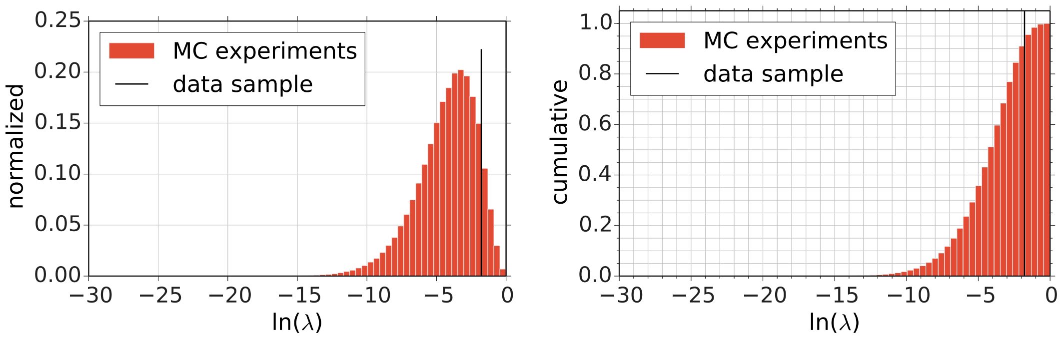
Similarly, a signal hypothesis can be tested. The GRB Toy Monte Carlo is used to estimate the average number of expected multiplets $\mu_{k,s}^i$ for a specific model and $N_{k,s}^i$ values are again generated according to the Poisson distribution. The number of alerts $N_{k,s+b}^i$ is then combination of background and signal alerts

$$N_{k,s+b}^i = N_{k,s}^i + N_{k,b}^i. \tag{6.5}$$

The influence of the OFU test statistic (Eq. 6.3) is then

$$\lambda_{\text{ofu}} = \prod_i \sqrt[N_{2,s+b}^i]{\prod_{j=1}^{N_{2,b}^i} p_{j,i,b}^{ofu} \cdot \prod_{j=1}^{N_{2,s}^i} p_{j,i,s}^{ofu}} \tag{6.6}$$

10^6 experiments are performed per season and multiplicities and GRB model and the test statistic (Eq. 6.4) evaluated. A model can be ruled out at 90% confidence level if 90% of the experiments show a worse agreement with the background only hypothesis than the actual experimental results. An example is shown in Figures 12a and 12b.



(a) 10^6 experiments were thrown and the test statistic λ was calculated according to equation 6.2. In read is the differential distribution of the Monte Carlo experiments while the black line represents the detected value.

(b) In read is the cumulative distribution of the Monte Carlo experiments while the black line represents the detected value. About 90% of the experiments have a worse agreement with the background only hypothesis.