Improving the signal estimation from On/Off observations in gamma-rays

Phys. Rev. D 103, 123001

Forskningsrådet

Giacomo D'Amico

University of Bergen

in collaboration with

Tomislav Terzić, Jelena Strišković, Michele Doro, Marcel Strzys and Juliane van Scherpenberg

2021 TEV Particle Astrophysics Conference

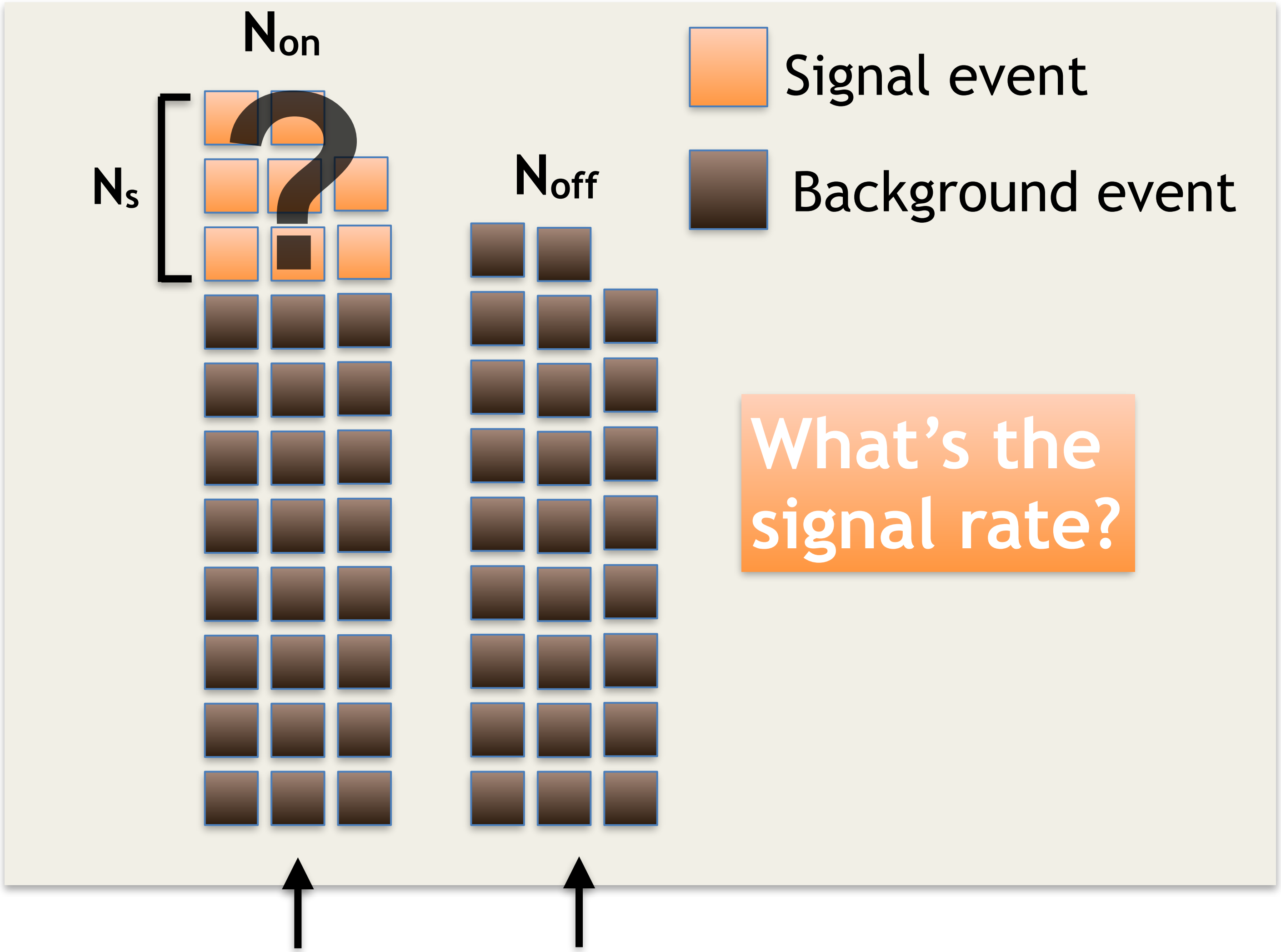
CHENGDU,CHINA Oct.25–Oct.29,2021

27 October 2021

OUTLINE

1. The On/Off measurement
2. The importance of discriminating variables
3. **The BASiL approach**
4. Performance of the BASiL approach compared to the standard way of estimating the signal rate and suppressing the background:
   - Using MC simulations of On/Off measurements from the MAGIC telescope
   - Using a real data sample
5. Conclusion

On/Off measurement



On measurement

Off measurement

In an **On/Off experiment**:

- a background-control (**Off**) region, which is supposedly void of any signal, is defined to estimate the **background rate** (b)
- the **On** source measurement instead provides an estimate of the **signal rate** (s) plus b, with the latter term supposed to be equal to that in the Off region.

The following **variables** are therefore introduced:

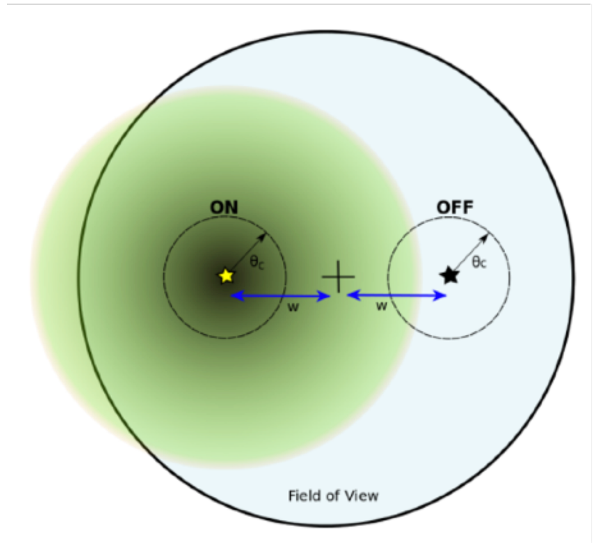
variable	description	property
N_{on}	number of events in the On region	measured
N_{off}	number of events in the Off region	measured
α	exposure in the On region over the one in the Off regions	measured
b	expected rate of occurrences of background events in the Off regions	unknown
s	expected rate of occurrences of signal events in the On region	unknown
N_s	number of signal events in the On region	unknown

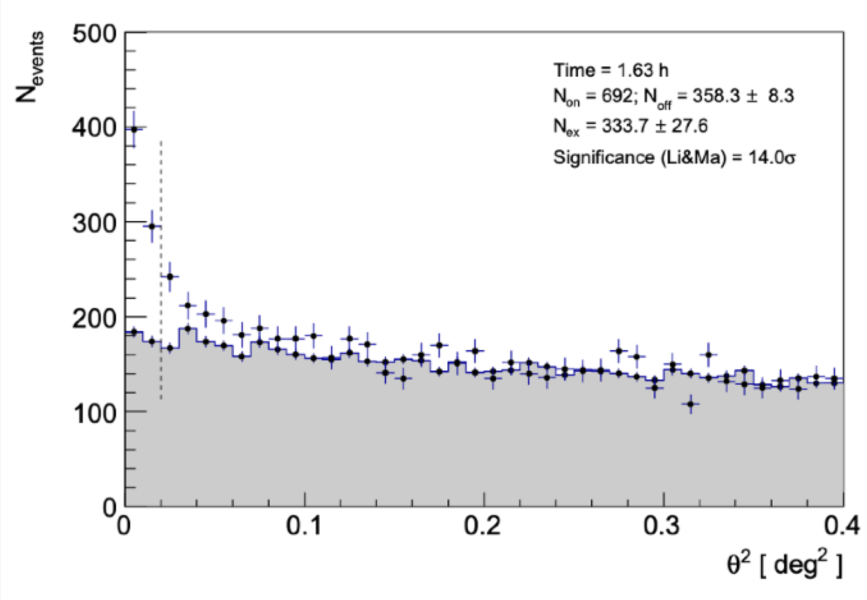
Probability mass function of observing N_{on} and N_{off} or
Likelihood function of the signal (s) and background (b) rate

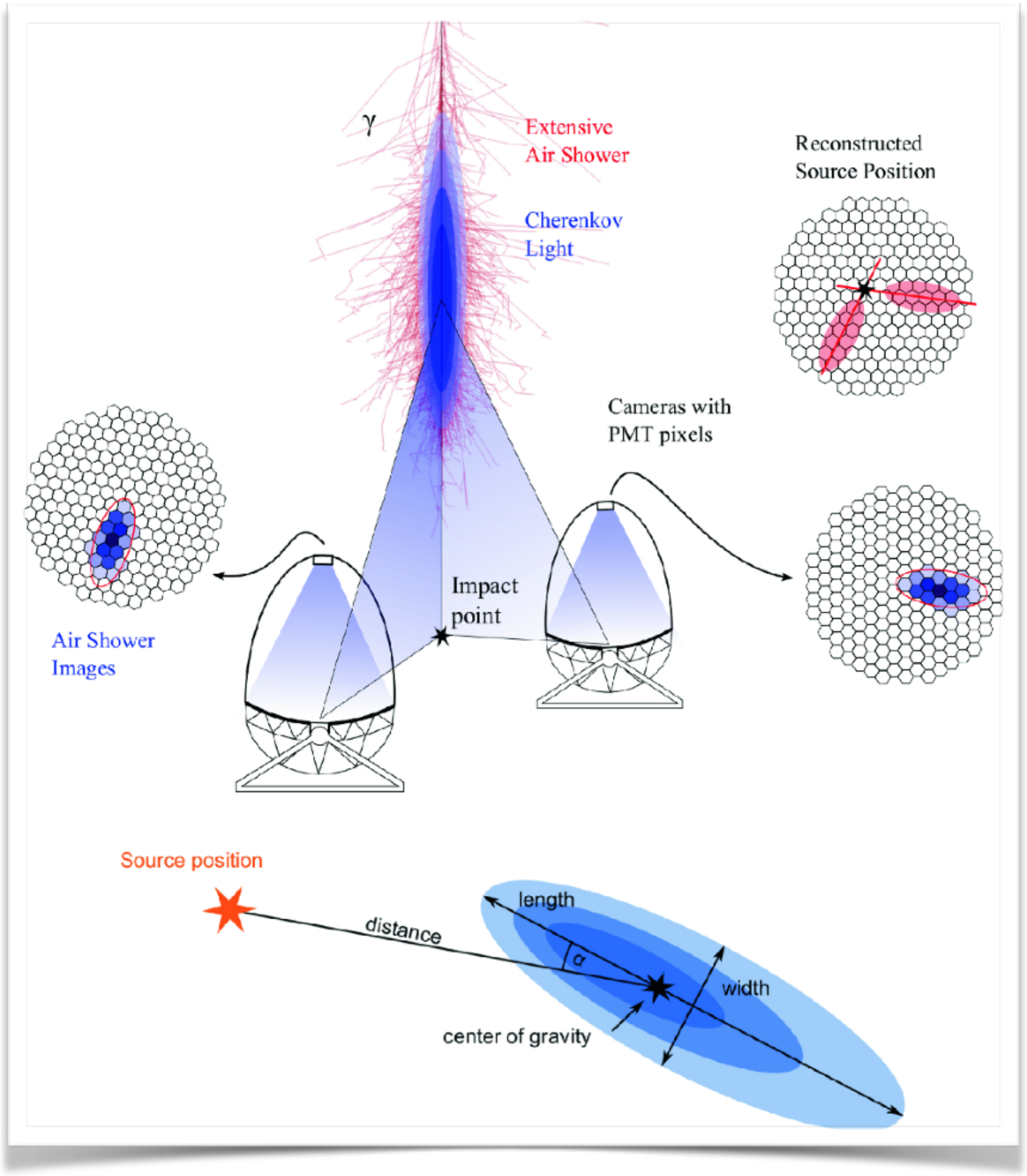
$$\frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \times \frac{b^{N_{off}}}{N_{off}!} e^{-b} = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = p(N_{on}, N_{off} \mid s, b; \alpha)$$

The case of Imaging Atmospheric Cherenkov Telescopes (IACTs)

- IACTs image the Cherenkov light emitted in the atmosphere by **extended atmospheric showers** generated by cosmic gamma rays (or cosmic rays) when entering the atmosphere.
- An **irreducible background** survives all possible image selection criteria and the signal estimation is performed through an **On/Off comparison** based, in which the Off sample is taken from a region in the sky where no signal is expected.
- Data are usually taken in the **wobble mode**: the source is placed with a certain offset with respect to the camera center during the observation, allowing a **simultaneous signal and background estimation**







SIGNAL ESTIMATION the frequentist approach

Signal estimation is usually done in the **frequentist approach:**

Likelihood function:

$$p(N_{on}, N_{off} \mid s, b; \alpha) = p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

↓

Likelihood ratio:

$$\lambda(s) \equiv \frac{p(N_{on}, N_{off} \mid s, b = \hat{b}\ ;\ \alpha)}{p(N_{on}, N_{off} \mid s = N_{on} - \alpha N_{off}\ ,\ b = N_{off}\ ;\ \alpha)}$$

value of b that **maximizes** the likelihood for a given s

$$\hat{b} = \frac{N^2 + \sqrt{N^2 + 4(1+1/\alpha)sN_{off}}}{2(1+\alpha)}$$

$$N = N_{on} + N_{off} - s(1+1/\alpha)$$

↓

$$-2 \log \lambda(s) \sim \chi^2$$

↓

Estimated signal rate:

$$s = (N_{on} - \alpha N_{off}) \pm \sqrt{N_{on} + \alpha^2 N_{off}}$$

SIGNAL ESTIMATION the Bayesian approach

An alternative way for estimating the signal rate is given in the **Bayesian approach:**

Bayes theorem:

$$p(s \mid N_{on}, N_{off}; \alpha) = \frac{\int db\, p(N_{on}, N_{off} \mid s, b; \alpha) p(b)\, p(s)}{\int ds\, db\, p(N_{on}, N_{off}, s, b; \alpha)} \propto \int db\, p(N_{on}, N_{off} \mid s, b; \alpha)$$

Introducing the binomial identity in the likelihood function

$$p(N_{on}, N_{off} \mid s, b; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1+\alpha))^{N_{on}+N_{off}-N_s}}{(N_{on}+N_{off}-N_s)!} e^{-b(1+\alpha)}$$

PDF of the signal rate

$$p(s \mid N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}$$

PMS of the number of signal event

$$p(N_s \mid N_{on}, N_{off}; \alpha) \propto \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}(N_{on}-N_s)!}$$

SIGNAL ESTIMATION comparison

In red -2 log λ(s)

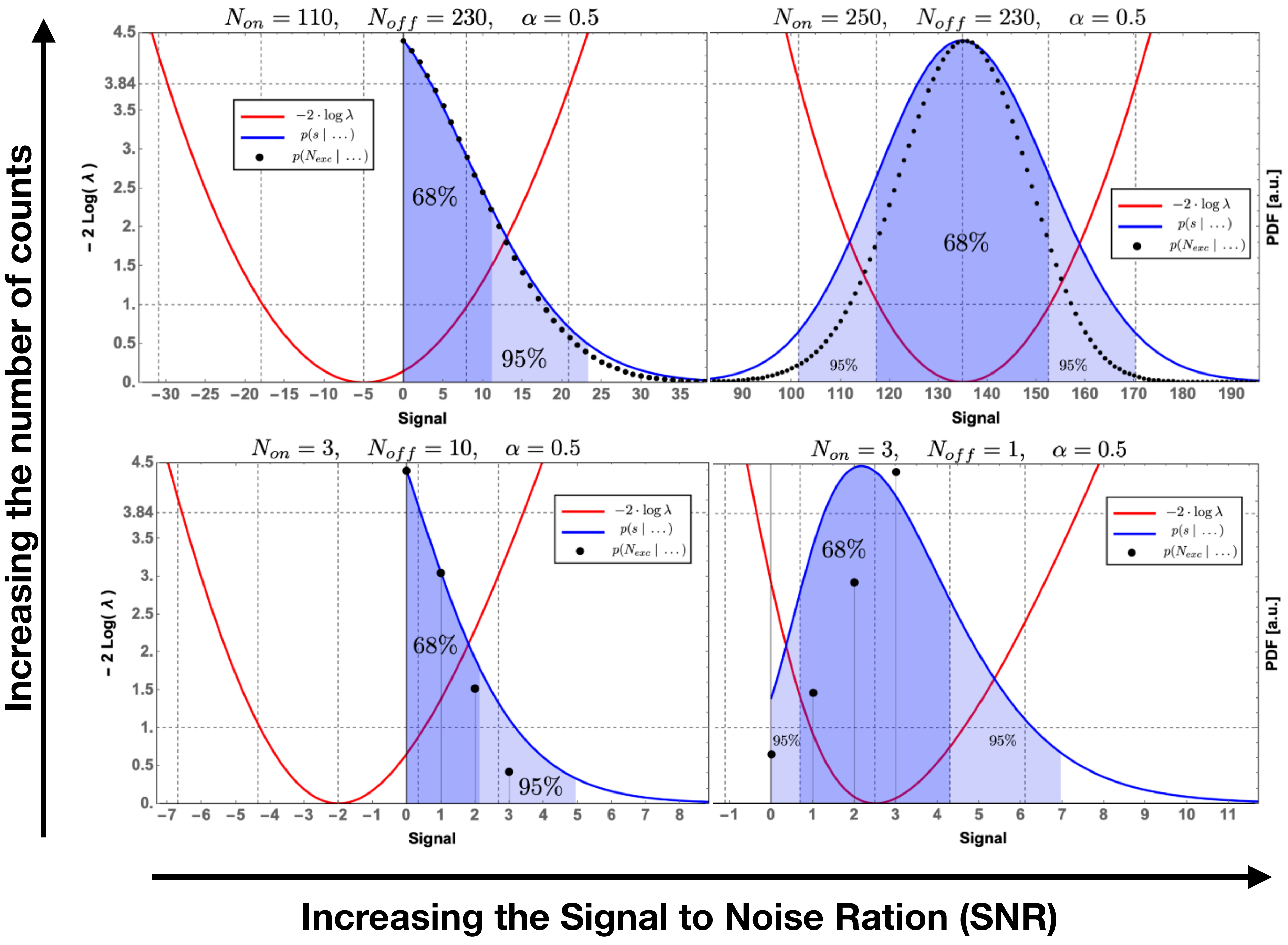
$$\lambda(s) \equiv \frac{p(N_{on}, N_{off} \mid s, b = \hat{\hat{b}}\ ;\ \alpha)}{p(N_{on}, N_{off} \mid s = N_{on} - \alpha N_{off}\ ,\ b = N_{off}\ ;\ \alpha)}$$

In blue the PDF of the signal rate

$$p(s \mid N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(1+1/\alpha)^{-N_s}(N_{on} - N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}$$

In black the PMF of the number of signal events N_s

$$p(N_s \mid N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(1+1/\alpha)^{-N_s}(N_{on} - N_s)!}$$



DISCRIMINATING VARIABLES

Discriminating variables "x" are observed which can be used for suppressing the background applying a **fixed fiducial cuts** (of the kind x > x_{cut}). This approach although has **2 disadvantages**:

- **reduced exposure** on the target
- all events surviving the cuts are treated as **equally probable signal**

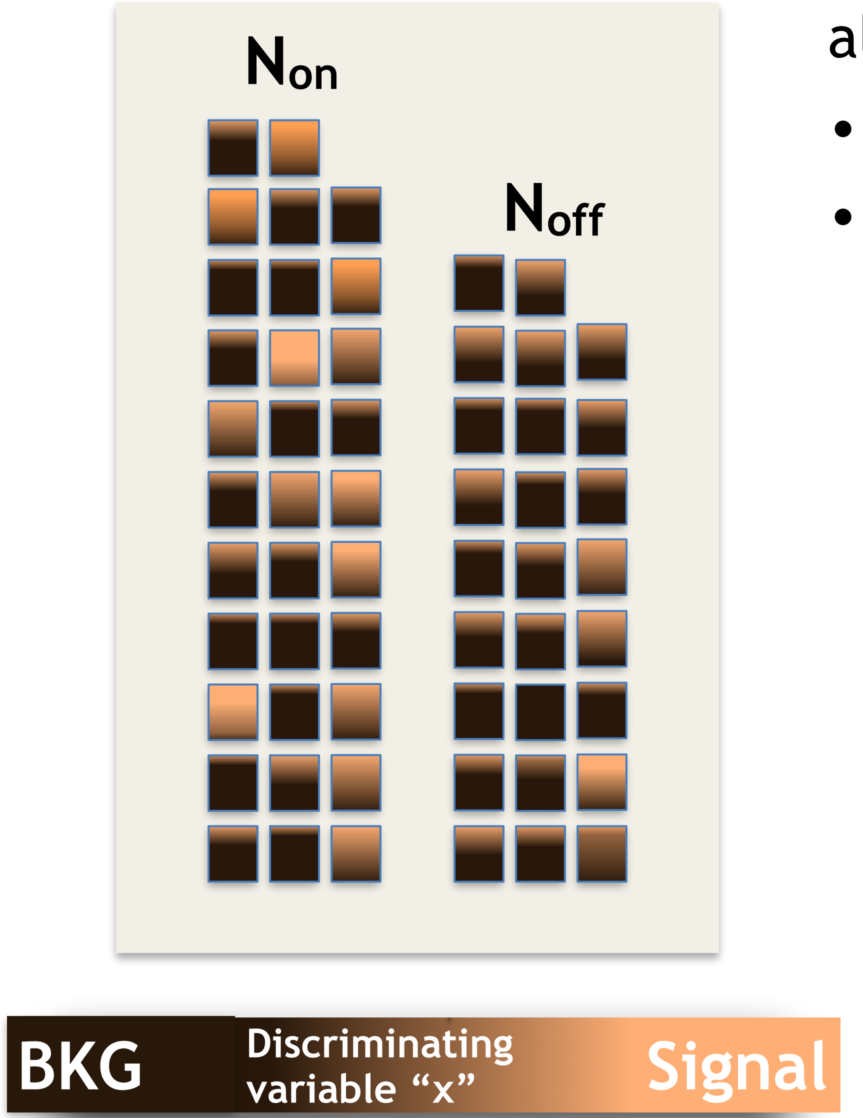


Figure from *Astroparticle Physics, 32(2), 73-88*

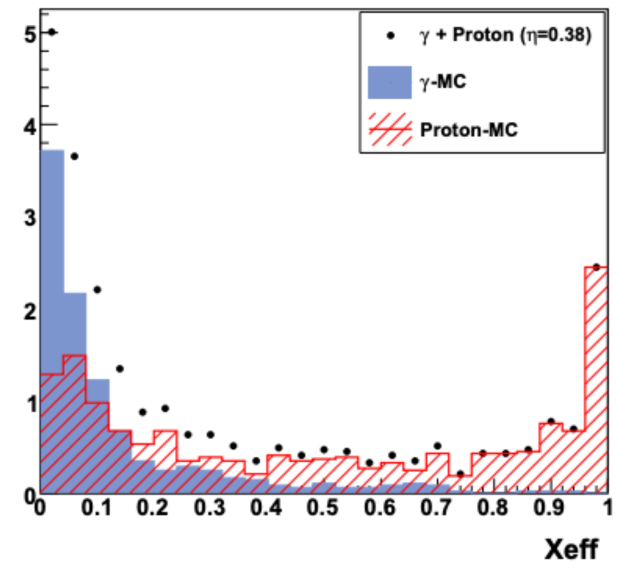


Figure 6: Example of the X_{eff}-*hadroness* (gamma-mistag) estimator distributions for two samples of simulated *Gamma* and *Hadron* events. A specific X_{eff} value is estimated for each event. The two samples are merged in a unique sample corresponding to $\eta = 0.38$ and analysed to validate the analysis method.

Figure from *Nucl. Instrum. Meth. A588:424-432,2008*

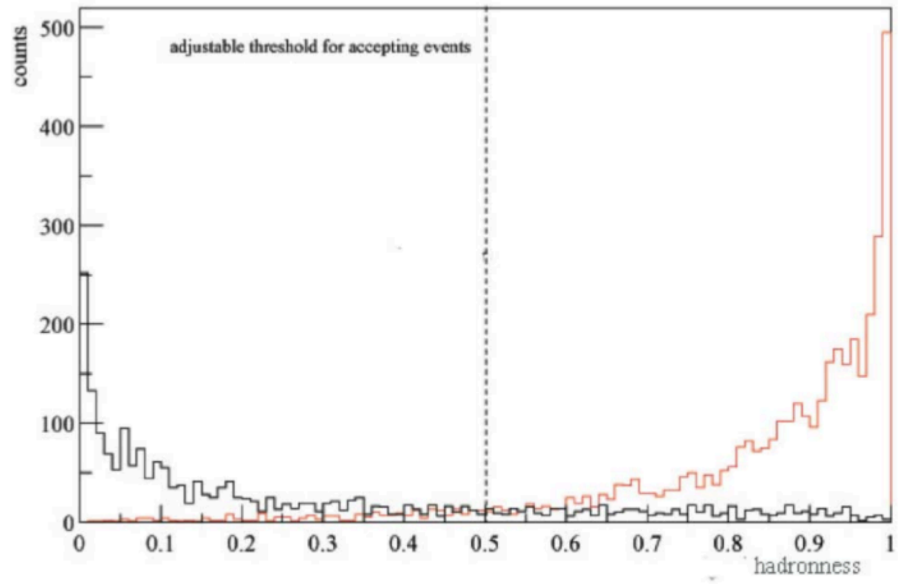
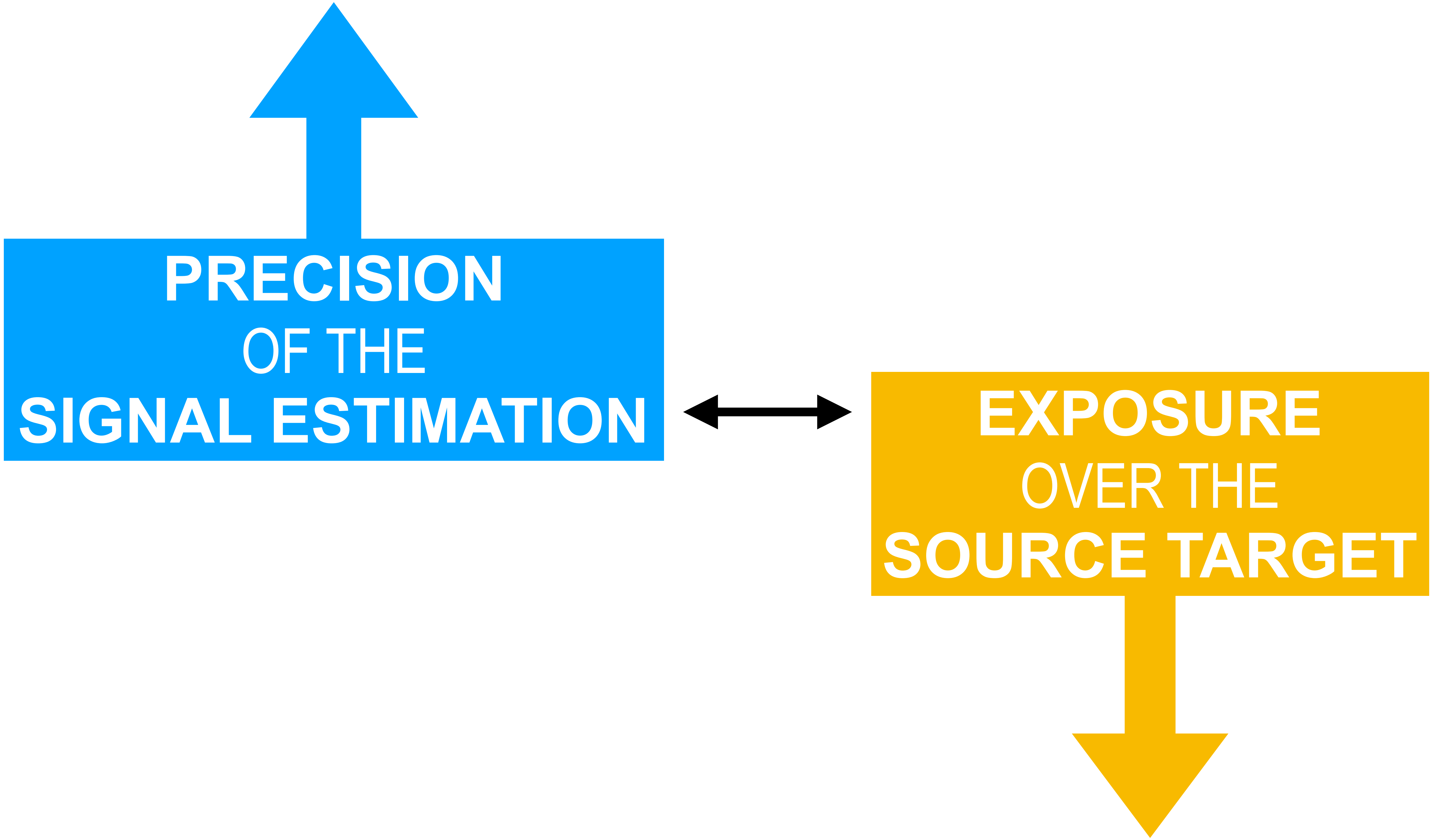


Fig. 3. Mean hadronness for two test samples of gammas (left peak, black) and hadrons (right peak, red). Hadronness is the final and only test statistic in γ/h separation.

DISCRIMINATING VARIABLES



DISCRIMINATING VARIABLES

PRECISION
OF THE
SIGNAL ESTIMATION

↔

EXPOSURE
OVER THE
SOURCE TARGET

?

The BASiL approach (Bayesian Analysis including Single-event Likelihood)

Our novelty method (BASiL) can do that!

Including the discriminating variables "x" in the likelihood and applying the Bayes theorem:

$$p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha) = \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-(s+\alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b} \cdot \prod_{i=1}^{N_{on}} \left[p(\mathbf{x}_i \mid \gamma) \cdot \frac{s}{s+\alpha b} + p(\mathbf{x}_i \mid \bar{\gamma}) \cdot \frac{\alpha b}{s+\alpha b} \right]$$

γ : Signal
$\bar{\gamma}$: BKG

PDF of the signal rate

$$p(s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}} \cdot \frac{s^{N_s}}{N_s!} e^{-s}.$$

PMF of the number of signal events

$$p(N_s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}}$$

Combinatorial term

$$C(\vec{\mathbf{x}}, N_s) = \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\gamma})$$

Combinatorial term

Given that the **combinatorial term** is the novelty of this method, it is worth to elaborate its role by providing an **example**

$$C(\vec{\mathbf{x}}, N_s) = \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\gamma})$$

Example 1:

Assume $N_{on} = 3$ events in our On region and that we have also measured x_1, x_2, x_3 respectively for each event, with x a variable whose distribution is $p(x|\gamma)$ for a signal population and $p(x|\bar{\gamma})$ for a background population

$$C(\vec{x}, 0) = p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma}),$$

$$\begin{aligned} C(\vec{x}, 1) = & p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma}) \\ & + p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma}) \\ & + p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma), \end{aligned}$$

$$\begin{aligned} C(\vec{x}, 2) = & p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma}) \\ & + p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma) \\ & + p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma), \end{aligned}$$

$$C(\vec{x}, 3) = p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma)$$

Example 2:

All events are l time more likely of being a signal event

$$p(\mathbf{x}_i|\gamma) = l \cdot p(\mathbf{x}_i|\bar{\gamma}) \quad \forall i \in \{1, \ldots, N_{on}\}$$

↓

$$C(\vec{\mathbf{x}}, N_s) \propto \binom{N_{on}}{N_s} l^{N_s}$$

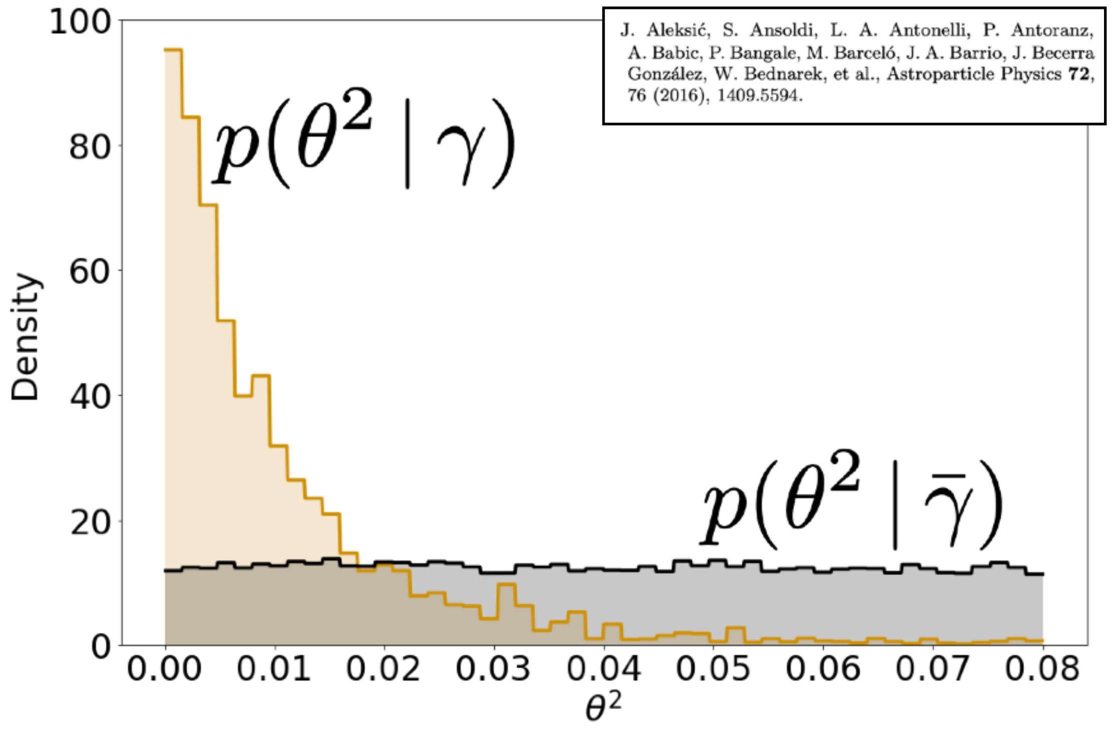
↓

$$p(N_s \mid \vec{\mathbf{x}}, N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(N_{on} - N_s)!(1 + 1/\alpha)^{-N_s}} \cdot l^{N_s}$$

Performance

Performance of the method from MC simulations

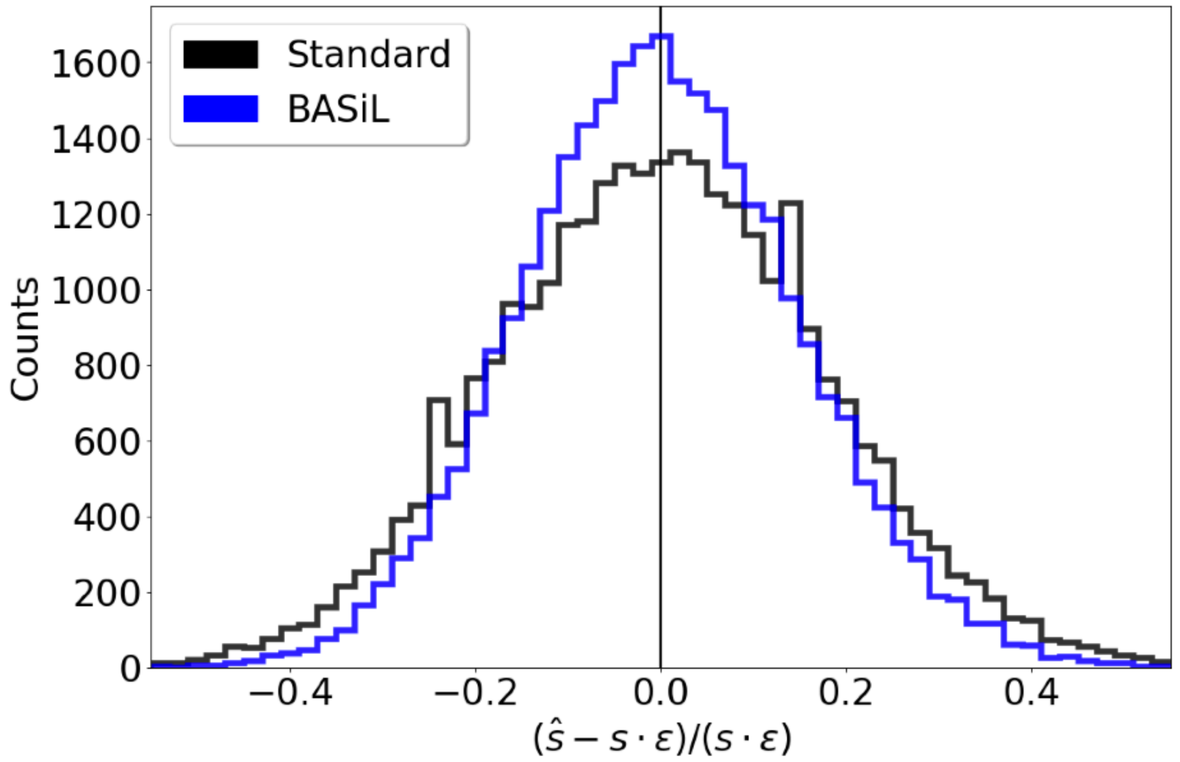
Events in an On/Off measurement are generated around the source center



FROM 10^5

SIMULATIONS

assuming b =3x10^3, α = 1/3 and s=10^2



the signal rate is then estimated using:

Standard

Selection cut applied $\epsilon \neq 1$

$$\hat{s} = N_{on} - \alpha N_{off}$$

BASiL

No selection cut applied $\epsilon = 1$

$$\hat{s} = \text{mode of } p(s|\dots)$$

EFFICIENCY CUT

ϵ = fraction of signal event surviving the cut

Precision and **bias** of the signal estimation

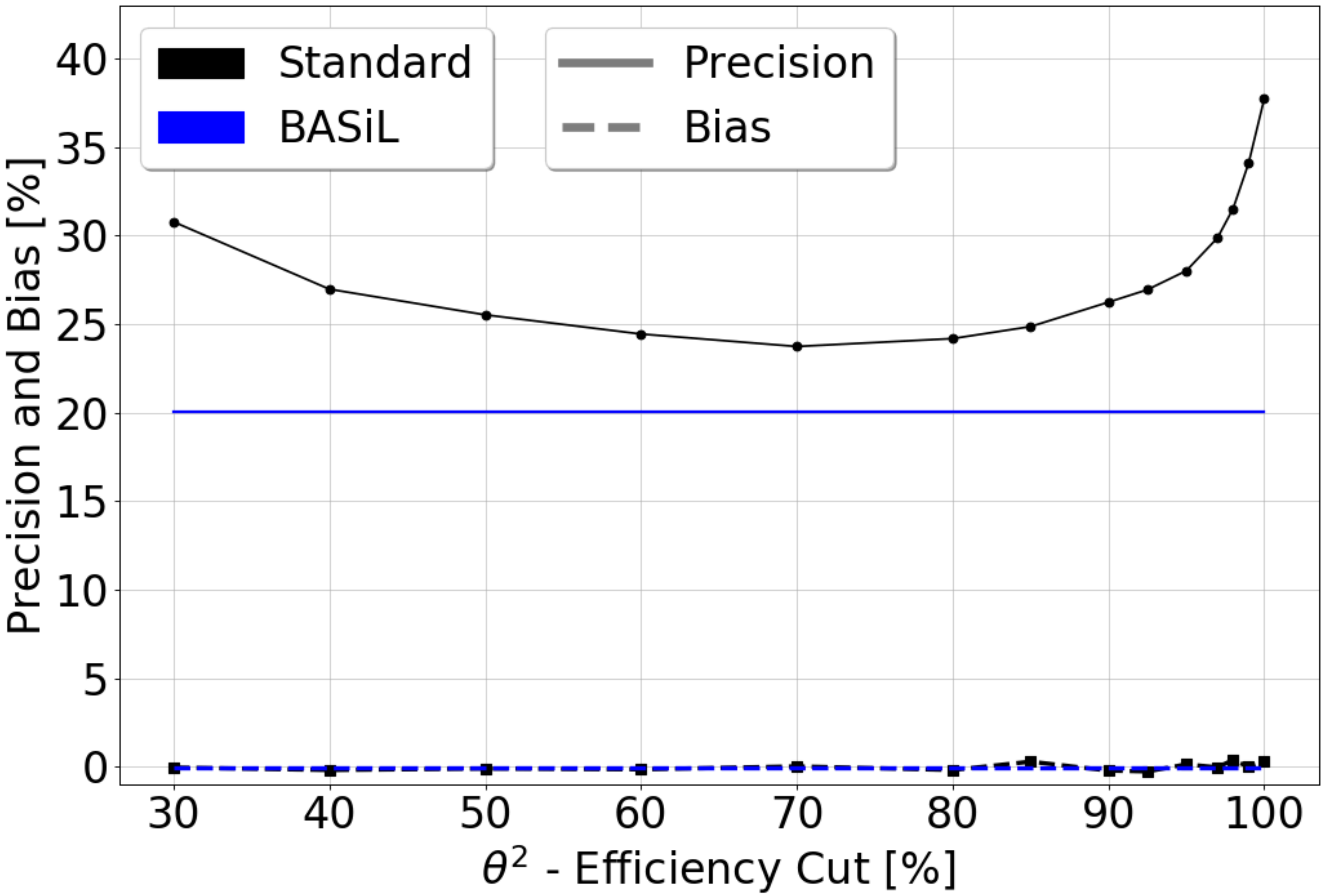
$$prec. = \sigma\left(\frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon}\right)$$

$$bias = \left< \frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon} \right>$$

Performance

Performance of the method from MC simulations

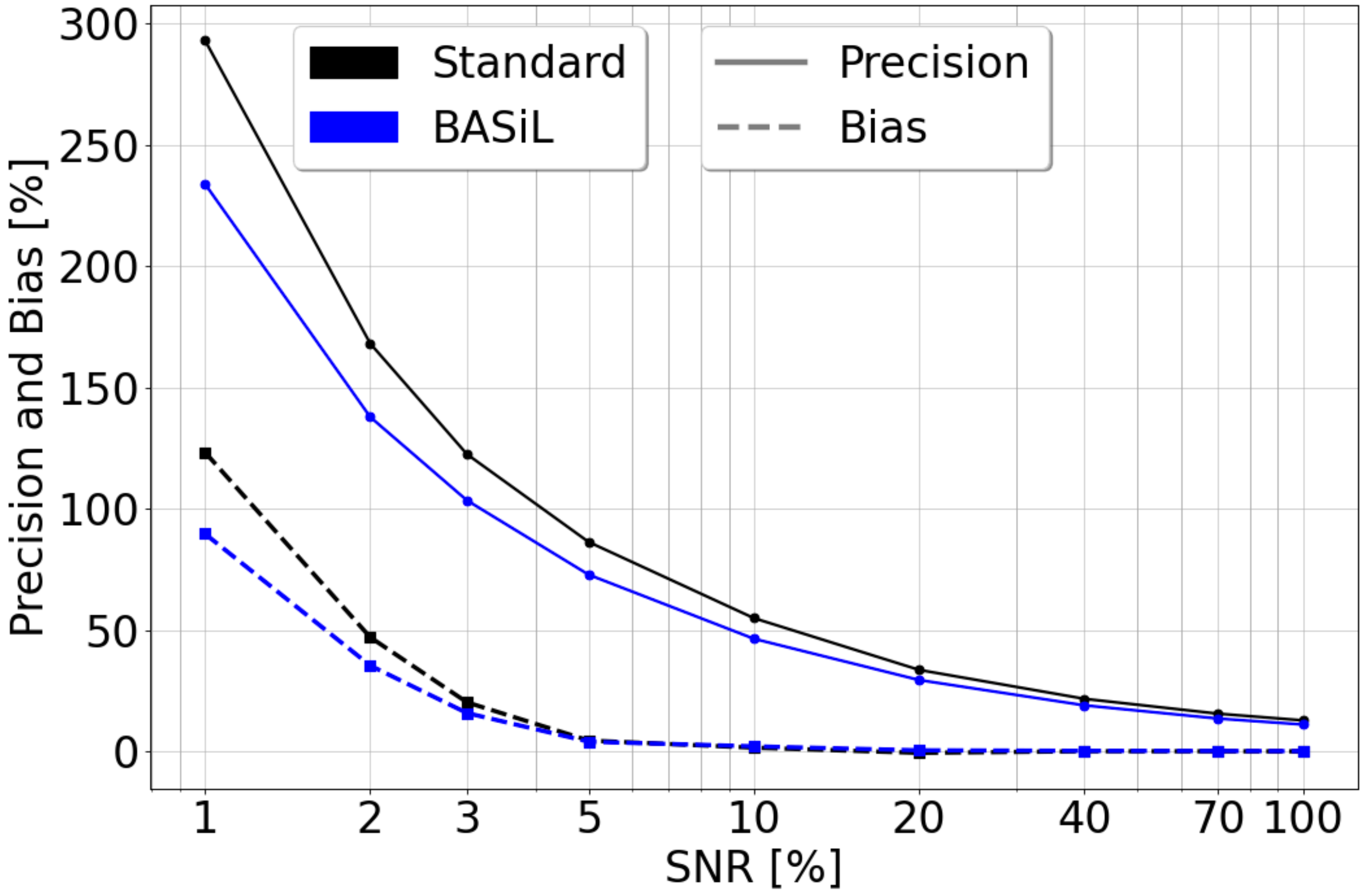
Evolution of the bias and precision for different efficiency cuts

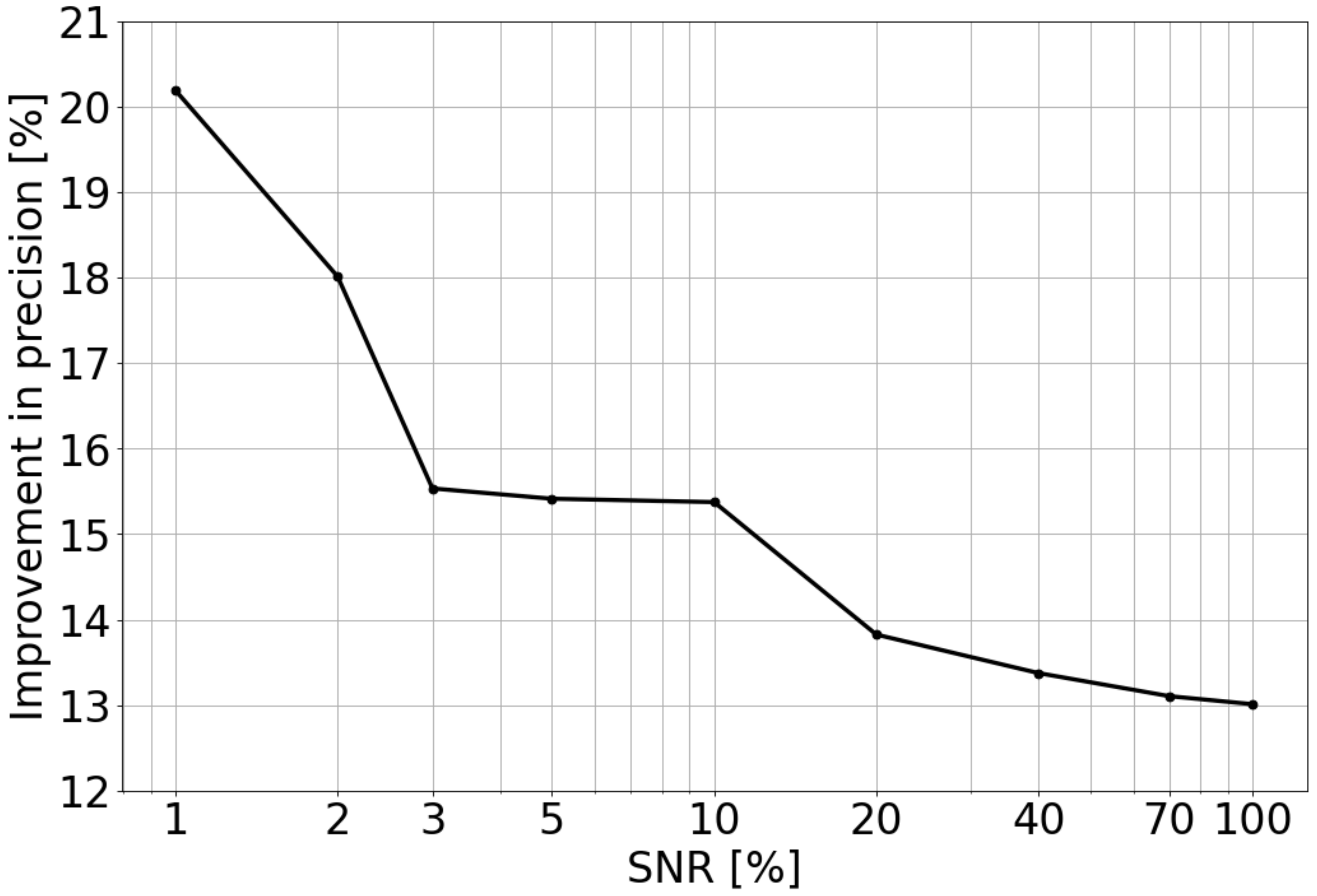


Performance

Performance of the method from MC simulations

Evolution of the bias and precision for different signal to noise ratios (SNRs)





Performance using data from the MAGIC telescope

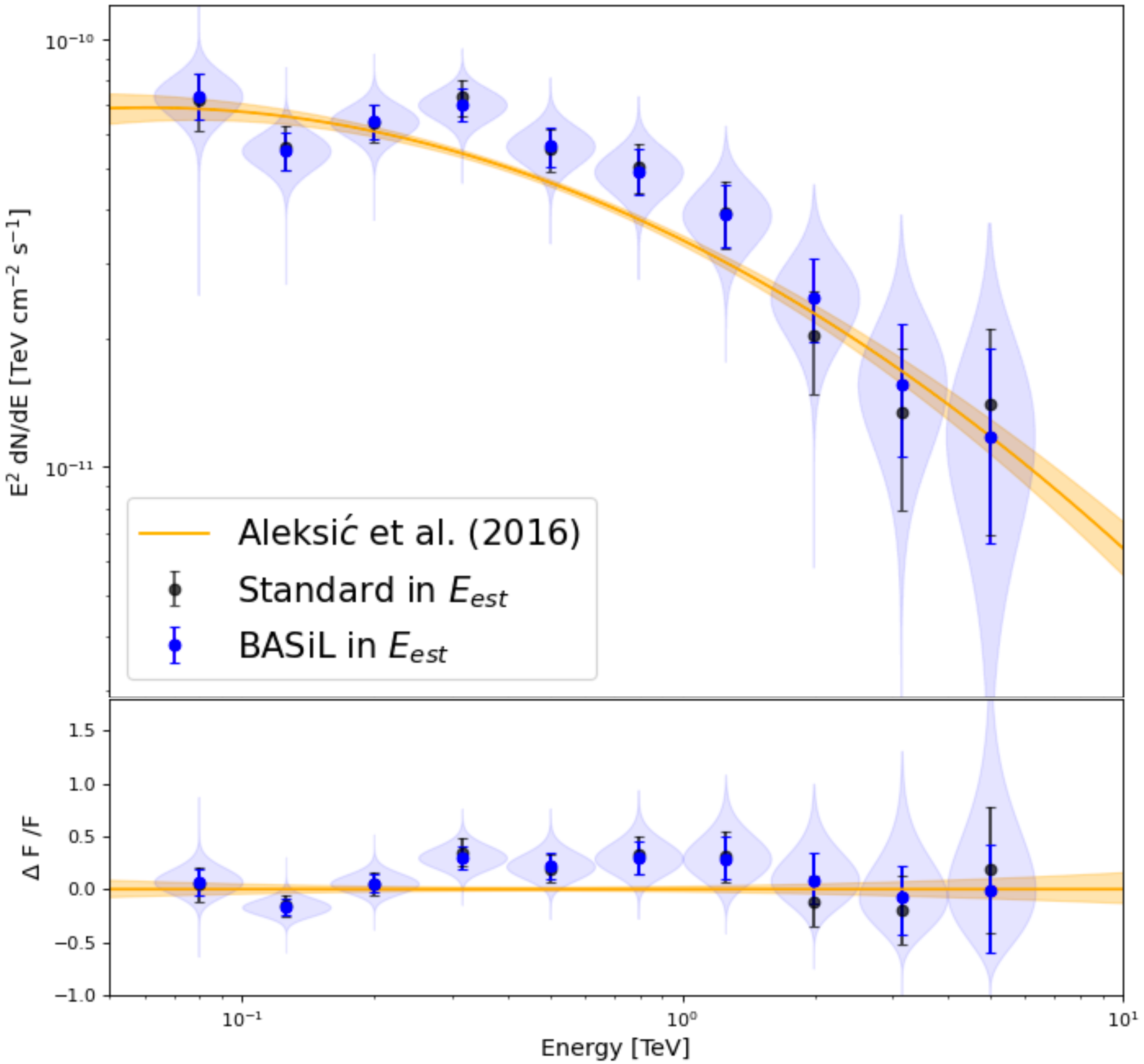
Performance of the method from a real data sample

Data released by the **MAGIC** collaboration which includes **40 minutes** of **Crab nebula** observations

The corresponding data in FITS format are publicly available at https://github.com/open-gamma-ray-astro/joint-crab/tree/master/data/magic

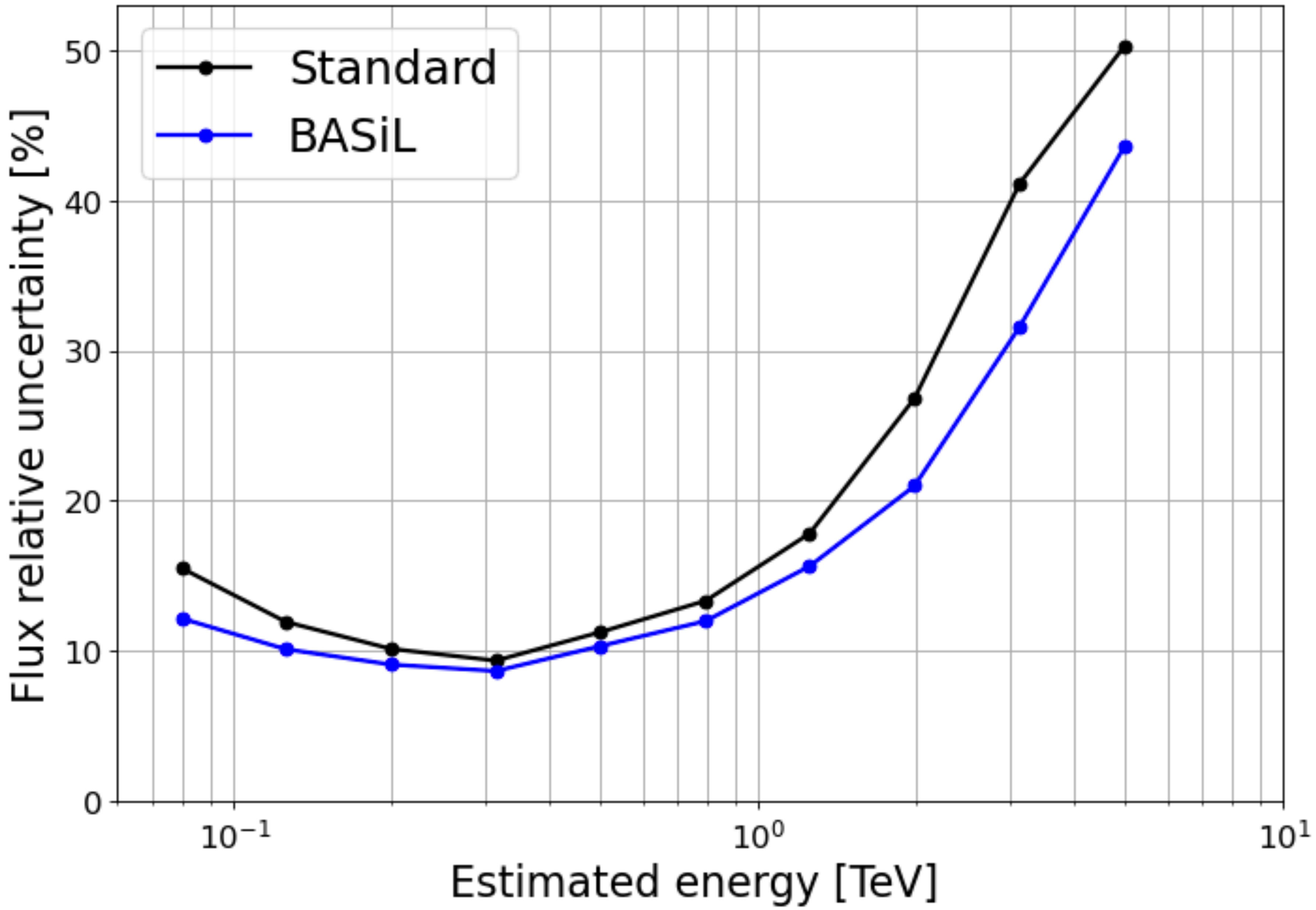
Counts in the On and Off regions and resulting signal rates per energy bin

E [GeV]	N_{on}	N_{off}	Signal (freq.)	Signal (BASiL)
63-100	1714	4494	216 ± 47	204 ± 23
100-158	933	2349	150 ± 35	187 ± 18
158-251	622	1327	180 ± 28	185 ± 16
251-398	439	846	157 ± 23	174 ± 14
398-631	335	593	137 ± 20	114 ± 11
631-1000	215	435	70 ± 16	$77.7^{+9.6}_{-8.9}$
1000-1585	132	256	47 ± 13	$41.8^{+7.0}_{-6.3}$
1585-2512	95	203	27 ± 11	$21.6^{+5.1}_{-4.5}$
2512-3981	56	140	9.3 ± 8.5	$8.6^{+3.4}_{-2.7}$
3981-6310	30	83	2.3 ± 6.3	$3.9^{+2.3}_{-1.6}$



Performance using data from the MAGIC telescope

Flux relative uncertainty (sensitivity plot)



Conclusions

- **Selection cuts** on discriminating variables inevitably discards also a fraction of the signal rate, **reducing the exposure** to the source target.
- Following the Bayesian approach, we showed a **new method** that not only **improves the precision** of the signal estimation, but it also **avoids the need of signal-extraction cuts**.
- We dubbed this method **BASiL** (Bayesian Analysis including Single-event Likelihoods): its main feature is that it **weights** events according to their individual **likelihood** of being **signal or background**, considering all the information available.
- The improvement is particularly noteworthy in cases of **small signal rates**.

THANK YOU FOR YOUR ATTENTION!

Backup slides

MORE ON THE COMPARISON BETWEEN FREQUENTIST AND BAYESIAN APPROACH

The two approaches generally **agree** (with ad-hoc adjustments for the frequentist approach)

Estimated signal rate

Frequentist

$$N_{on} - \alpha N_{off}$$

Bayesian

$$\text{mode of } p(s \mid N_{on}, N_{off}; \alpha)$$

Uncertainty

Frequentist

$$\sqrt{N_{on} + \alpha^2 N_{off}}$$

Bayesian

$$\int_{s_{left}}^{s_{right}} p(s \mid N_{on}, N_{off}; \alpha) ds = 0.68, \quad \text{with}$$
$$p(s_{left} \mid N_{on}, N_{off}; \alpha) = p(s_{right} \mid N_{on}, N_{off}; \alpha)$$

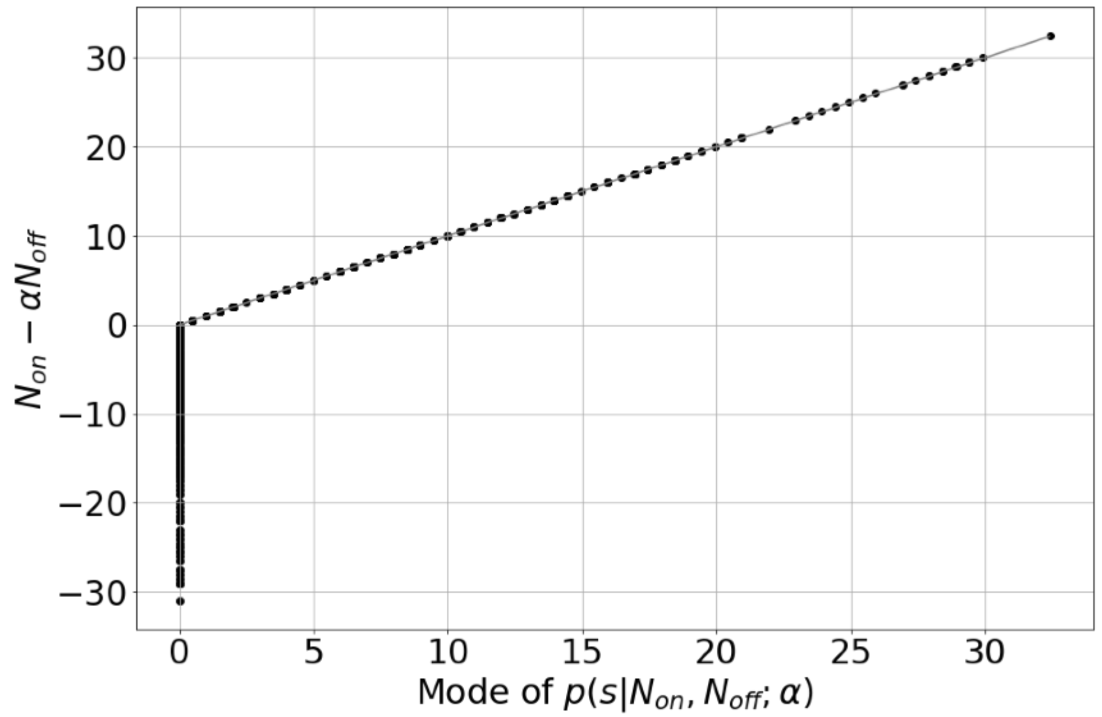
95% Upper Limit

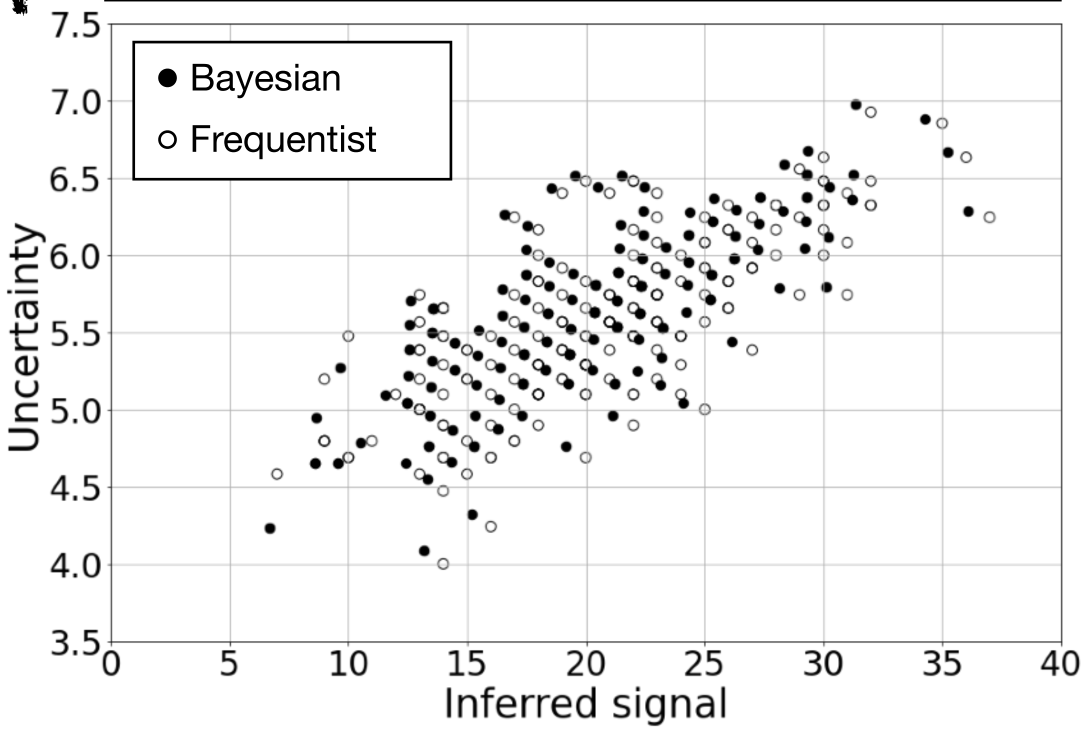
Frequentist

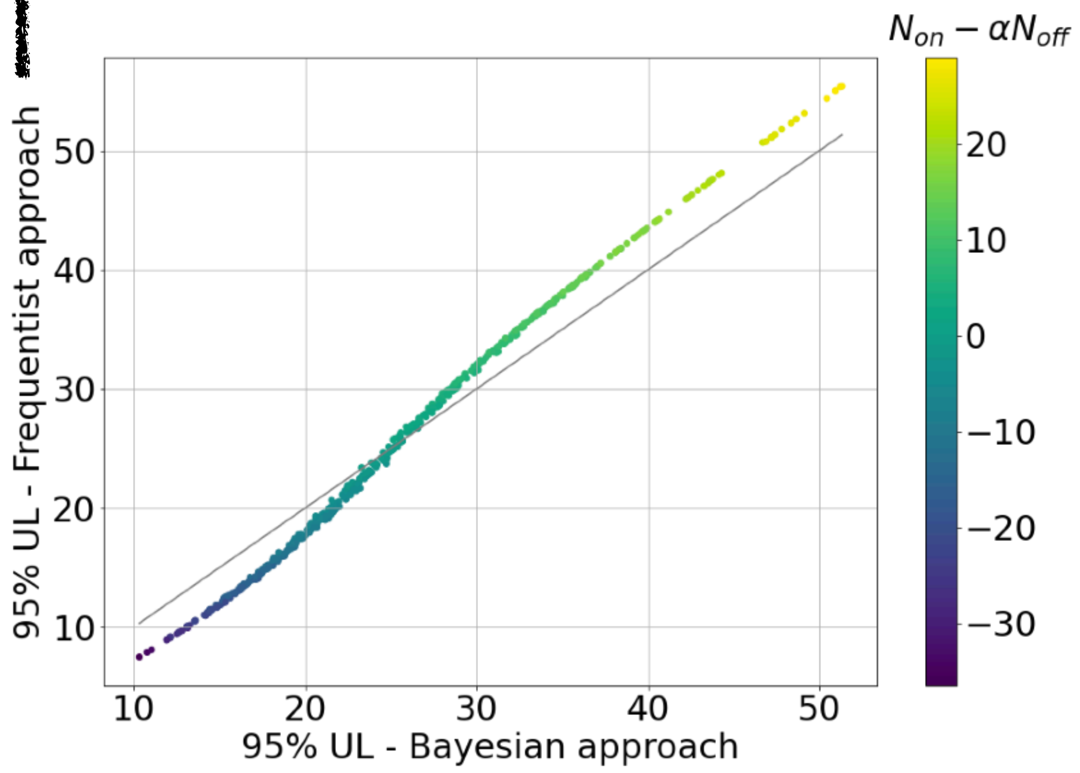
W. A. Rolke, et al. (2005)

Bayesian

$$\int_0^{s_{95}} p(s \mid N_{on}, N_{off}; \alpha) ds = 0.95.$$







Backup slides

Derivation of likelihood function including the discriminating variable “x"

$$p(\vec{\mathbf{x}}, N_{on}, N_{off} \mid s, b; \alpha) = p(\vec{\mathbf{x}} \mid N_{on}, s, \alpha b) \cdot p(N_{on} \mid s, \alpha b) \cdot p(N_{off} \mid b)$$

$$= \prod_{i=1}^{N_{on}} \left(p(\mathbf{x}_i \mid \gamma) \frac{s}{s+\alpha b} + p(\mathbf{x}_i \mid \bar{\gamma}) \frac{\alpha b}{s+\alpha b} \right) \cdot \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

$$= \sum_{N_s=0}^{N_{on}} \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\gamma}) \cdot \frac{s^{N_s} (\alpha b)^{N_{on}-N_s}}{(s+\alpha b)^{N_{on}}} \cdot \frac{(s+\alpha b)^{N_{on}}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

$$= \sum_{N_s=0}^{N_{on}} C(\vec{\mathbf{x}}, N_s) \cdot \frac{s^{N_s} (\alpha b)^{N_{on}-N_s}}{N_{on}!} e^{-s-\alpha b} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}$$

$$= \frac{\alpha^{N_{on}}/N_{off}!}{(1+\alpha)^{N_{on}+N_{off}}} \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(1+1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{N_{on}!/N_s!} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1+\alpha))^{N_{on}+N_{off}-N_s}}{(N_{on}+N_{off}-N_s)!} e^{-b(1+\alpha)}$$

$$\propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on}+N_{off}-N_s)!}{(N_{on}-N_s)!(1+1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{on}}{N_s}} \cdot \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1+\alpha))^{N_{on}+N_{off}-N_s}}{(N_{on}+N_{off}-N_s)!} e^{-b(1+\alpha)}$$

A general algorithm for computing the combinatorial term

```
algorithm Combinatorial_term
  input: array1 of length n,
         array2 of length n
  output: array C of length n+1

  n <- length of array
  C[n+1] <- [1, 0, ..., 0]
  FOR i=0 to n:
    D[n] <- [0, C[0], ..., C[n-1]]
    C <- array1[i] * C + array2[i] * D
  RETURN C
```

Here **array1,array2** have to be thought as the array containing the list of background and signal likelihoods, respectively. For instance, $C(\vec{\mathbf{x}}, 2)$ can be found in the third element of the array obtained in output from the algorithm above defined. Note that it may be useful when dealing with large count numbers to work with the logarithmic values of the likelihoods.

Backup slides

Including the effect of the mismatch between real and MC distributions for the discriminating variable