Excess estimation in On/Off measurements including single-event variables

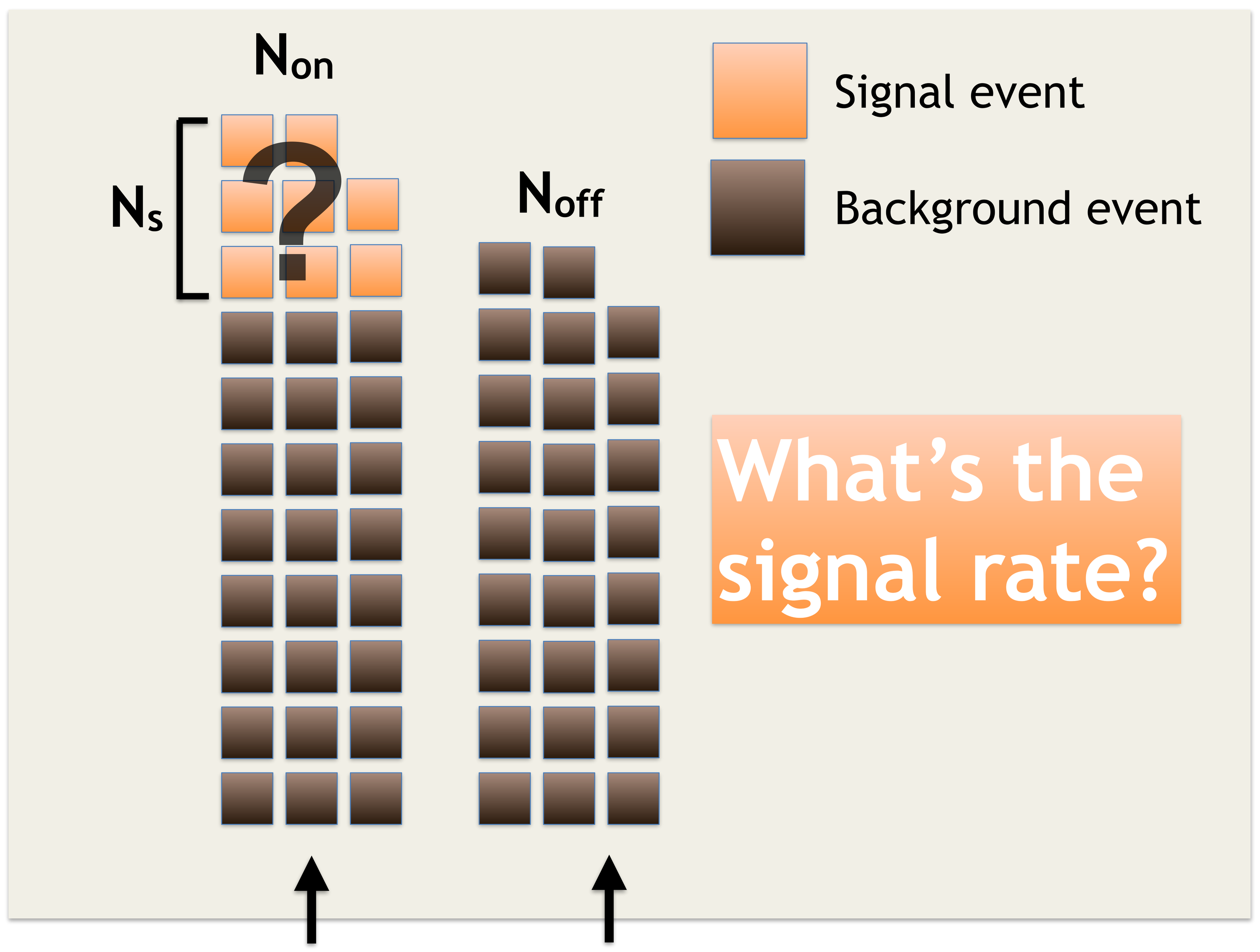
Giacomo D'Amico, University of Bergen, Norway

In collaboration with T. Terzić, J. Strišković, M. Doro, M. Strzys and J. van Scherpenberg

ONLINE ICRC 2021

UNIVERSITY OF BERGEN

On/Off measurement



$$\frac{(s+\alpha b)^{N_{on}}}{N_{on}!}e^{-(s+\alpha b)} \qquad \frac{b^{N_{off}}}{N_{off}!}e^{-b}$$

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

- **Frequentist answer:** the 68% C.I. is given by

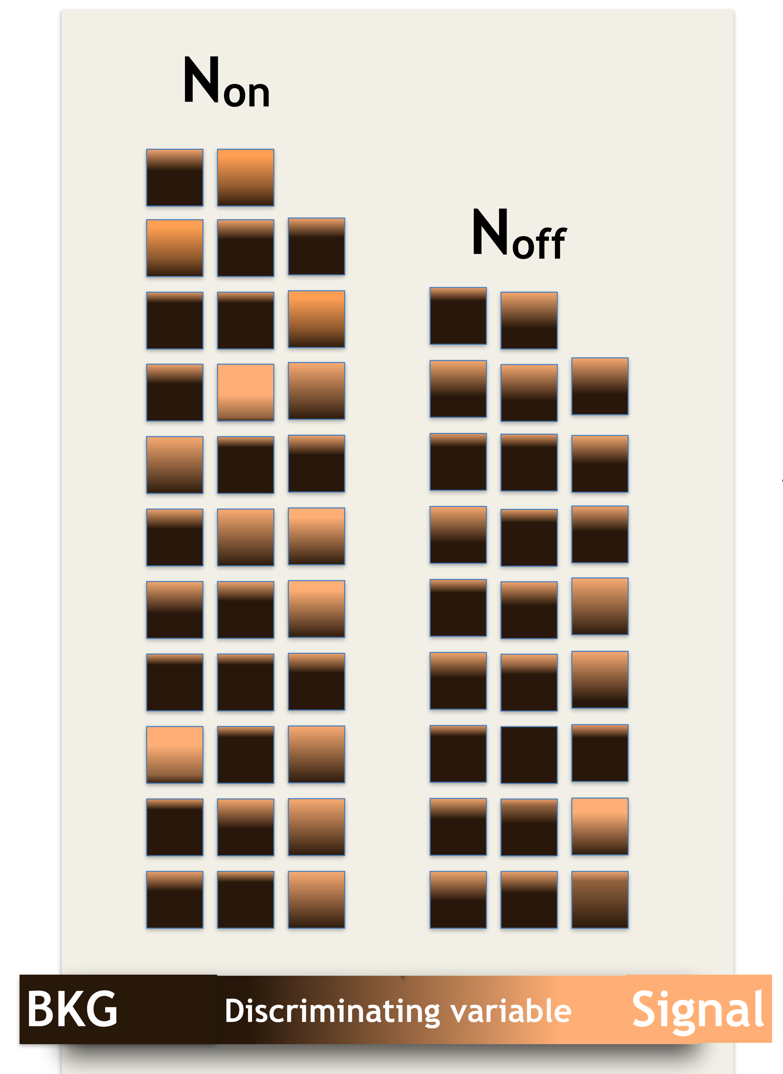
$$s = (N_{on} - \alpha N_{off}) \pm \sqrt{N_{on} + \alpha^2 N_{off}}$$

- **Bayesian answer:** the PDF of the signal rate is

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

(*)

The BASiL approach



Usually for each event a **discriminating variable** "x" is observed which can be used for suppressing the background. In this way only events surviving a fixed fiducial cuts (of the kind $x > x_{cut}$) will be considered. This approach has 2 disadvantages:

- also a fraction of the signal events will be excluded, which translates to a **reduced exposure** on the target
- after the selection, all events surviving a specific set of cuts are treated as **equally probable signal** (or background) events, regardless their "distance" from the cut

BASiL (Bayesian Analysis including Single-event Likelihoods) estimates the signal rate using information on **event-by-event individual parameters** and the way they distribute for the **signal and background population**.
Events are thereby weighted according to their **likelihood** of being a signal or a background event and background suppression can be achieved without performing fixed fiducial cuts.
We start by considering a "marked" Poisson process (with mark "x") whose likelihood is:

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

From the above likelihood, using the Bayesian approach, we got that the PDF of the signal rate is

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

(*)

combinatorial term

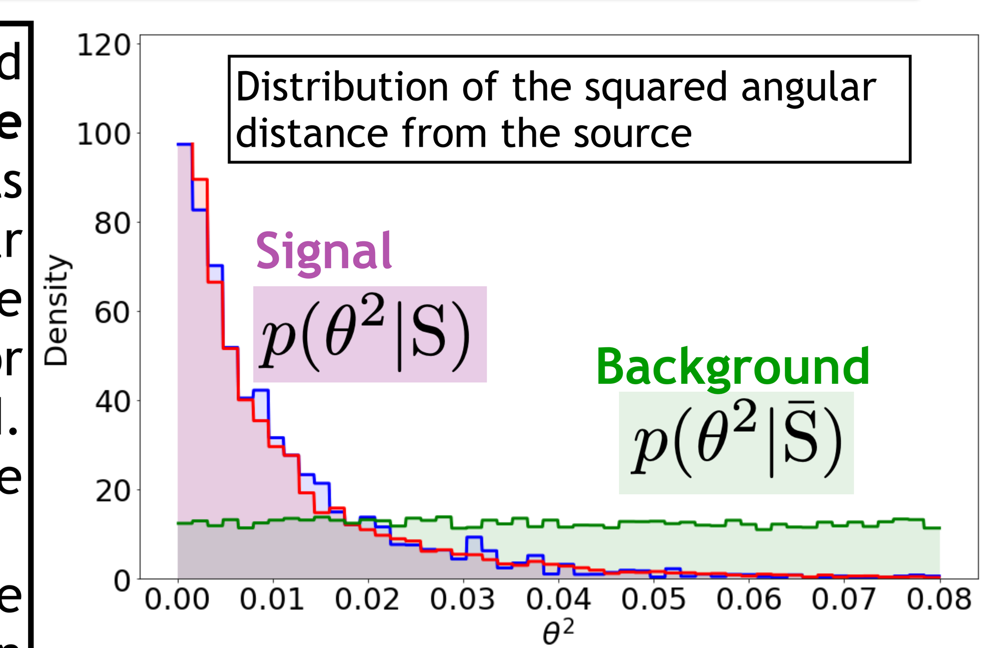
$$C(\vec{\mathbf{x}}, N_s) = \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\mathrm{S}) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\mathrm{S}})$$

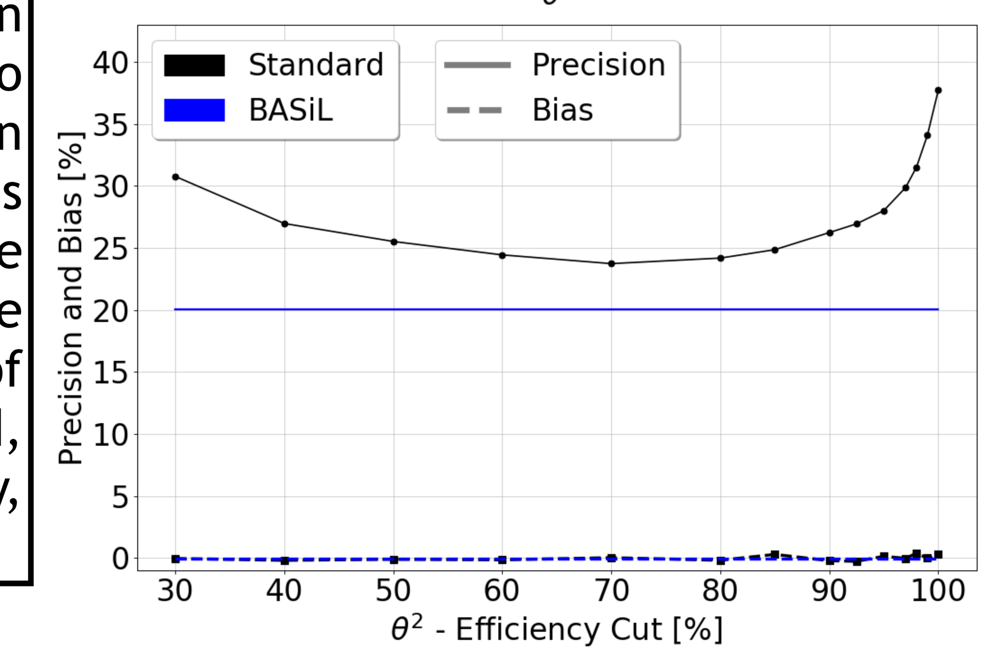
Once we have the PDF of the signal rate, its estimate is given by the mode of the PDF with uncertainty given by the 68% credible interval around the mode. In the BASiL approach therefore one can **avoid cutting** data by including in the PDF of the signal rate the **"combinatorial term"** that is made up to account for all the possible combination of signal events among the total N_{on} events that can give the observed set of discriminating variables "x". Such approach will be compared below with the **standard** one in which a cut is instead performed on the data and the signal estimate is given by $\mathbf{N_{on} - \alpha\, N_{off}}$.

Performance: BASiL better estimates the signal rate avoiding selection cut

From MC simulations of On/Off observations with the MAGIC telescope

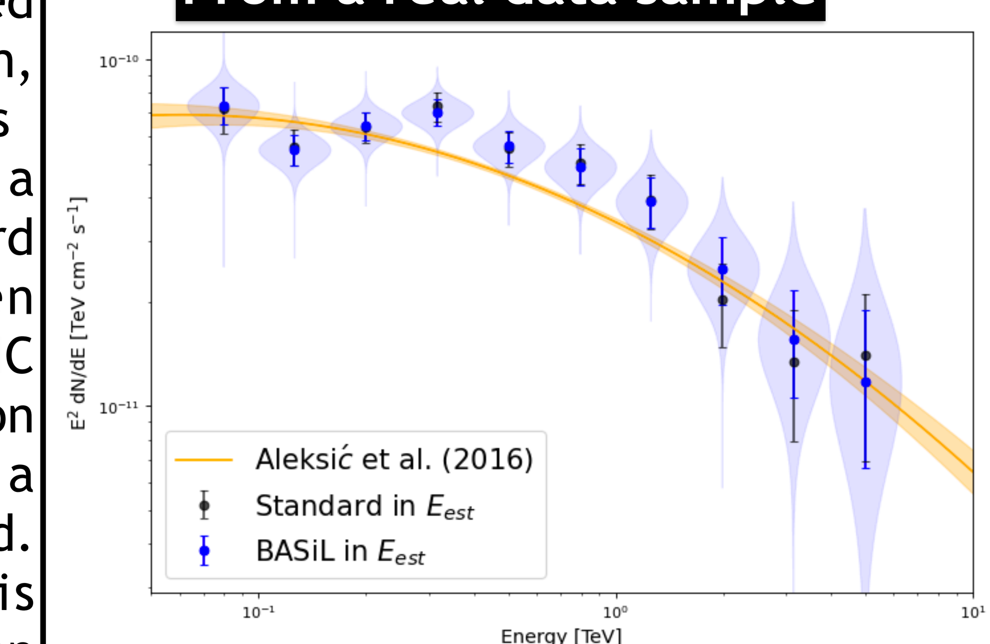
We study the evolution of the **bias** and **precision** in the estimation of the **signal rate** from MC simulations of On/Off measurements in which events are assigned an angular distance from the source sampled from the pink distribution if the event is a signal or from the green one if it is a background. Distributions are taken from the performance paper of the MAGIC telescope.
From 10^5 estimated signal rates using the **standard** and **BASiL** approach we then compute the precision and bias of the two methods. In the BASiL approach the precision does not depend on the efficiency and it is about 15 % better than the best precision we can achieve in the standard approach. The BASiL method, by including the likelihood of each event of being a signal or background, estimates the signal rate more precisely, while keeping the bias comparably low.

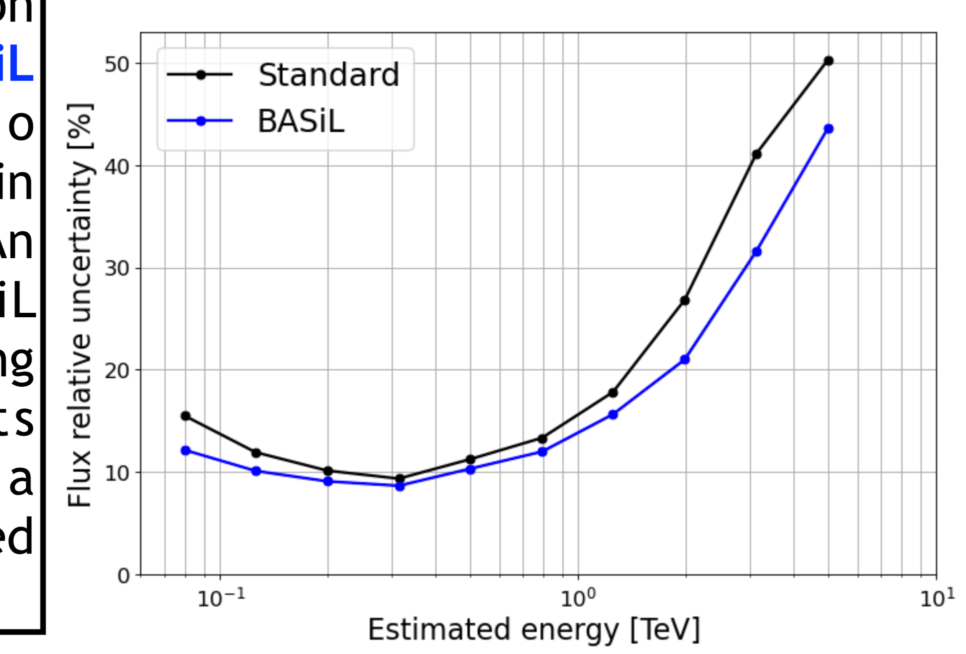




From a real data sample

We used the data released by the MAGIC collaboration, which includes 40 minutes of Crab nebula observations. The standard data analysis has been performed using the MAGIC Analysis and Reconstruction Software (MARS) where a cut on the data is applied. For the **BASiL** analysis instead no cut is applied on the data set. The **BASiL** approach manages to decrease the uncertainty in the signal estimation. An advantage of the BASiL approach when estimating the source flux is its capability of providing a PDF contour plot associated to each energy bin.





Selected references

This work: Phys.Rev.D 103 (2021) 12, 123001
T. J. Loredo, in Statistical Challenges in Modern Astronomy (1992) pp. 275-297
MAGIC coll. Astroparticle Physics 72, 76 (2016), 1409.5594.