Signal estimation in on/off measurements including event-by-event variables

G. D'Amico,[1,*] T. Terzić,[2] J. Strišković,[3] M. Doro,[4] M. Strzys,[5] and J. van Scherpenberg[6]

[1]*Department for Physics and Technology, University of Bergen, Bergen NO-5020, Norway*
[2]*University of Rijeka, Department of Physics, 51000 Rijeka, Croatia*
[3]*Josip Juraj Strossmayer University of Osijek, Department of Physics, 31000 Osijek, Croatia*
[4]*Università di Padova and INFN, I-35131 Padova, Italy*
[5]*Institute for Cosmic Ray Research (ICRR), The University of Tokyo, Kashiwa, 277-8582 Chiba, Japan*
[6]*Max-Planck-Institut für Physik, D-80805 München, Germany*

Signal estimation in the presence of background noise is a common problem in several scientific disciplines. An "on/off" measurement is performed when the background itself is not known, being estimated from a background control sample. The "frequentist" and Bayesian approaches for signal estimation in on/off measurements are reviewed and compared, focusing on the weakness of the former and on the advantages of the latter in correctly addressing the Poissonian nature of the problem. In this work, we devise a novel reconstruction method, Bayesian analysis including single-event likelihoods (dubbed BASiL), for estimating the signal rate based on the Bayesian formalism. It uses information on event-by-event individual parameters and their distribution 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. Throughout the work, we maintain a general notation that allows us to apply the method generically and provides a performance test using real data and simulations of observations with the MAGIC telescopes, as a demonstration of the performance for Cherenkov telescopes. BASiL allows one to estimate the signal more precisely, avoiding loss of exposure due to signal extraction cuts. We expect its applicability to be straightforward in similar cases.

I. INTRODUCTION

In some experiments, where besides the signal the background is unknown, the signal itself can be obtained by a so-called on/off measurement: 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. A normalization factor α between the on and off exposure is normally introduced. If, for instance, the on and off regions have the same acceptance, then α is defined as the ratio of the effective observation time in the two regions: $\alpha = t_{\text{on}}/t_{\text{off}}$.[1] The measurement of the number of events in the on and off region results in independent positive count numbers N_{on} and N_{off}. If one exactly knew the signal flux s, the number of signal events N_s in the on

measurement would be a random variable following a Poisson distribution[2]

$$p(N_s|s) = \frac{(\xi \cdot s)^{N_s} e^{-\xi \cdot s}}{N_s!}, \quad s \geq 0, \qquad (1)$$

where

$$\xi = t_{\text{eff}} \cdot A_{\text{eff}}$$

is the exposure, with t_{eff} being the effective observation time and A_{eff} the effective area of the telescope. For simplicity of notation, throughout the paper we will assume $\xi = 1$ and we will refer to s and b as the signal and background rate, respectively. A summary of the variables used and their description can be found in Table I.

---

*giacomo.damico@uib.no
[1]A more detailed definition and discussion of α can be found in Ref. [1].

[2]The generic symbol $p()$ is used to indicate all probability density functions (PDFs) and probability mass functions (PMFs) (the former applies to continuous variables and the latter to discrete variables). Note that the order of arguments is irrelevant being $p(x, y|I)$ the "joint PDF (or PMF) of x and y under condition I."

TABLE I.   Summary of the variables with their description considered in an on/off measurement and used throughout the paper.

Variable	Description	Property	Probability distribution
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	Eq. (4) in which s is integrated out
s	Expected rate of occurrences of signal events in the on region	Unknown	Eq. (5)
N_s	Number of signal events in the on region	Unknown	Eq. (8)

The difficulty in estimating the signal rate s lies in the uncertainties connected to the determination of the number of signal events N_s from the measured counts N_{on} and N_{off}, especially in the case of small signal-to-noise ratio (SNR). It is also important to underline that, because of the Poissonian nature of the problem [see Eq. (1)], both s and N_s must be non-negative.

Assuming flat priors $p(s)$ and $p(b)$ (with $s > 0$ and $b > 0$) and by applying the Bayes theorem, we get that the PDF for the signal rate s is

$$p(s|N_{on}, N_{off}; \alpha) = \frac{\int db\, p(N_{on}, N_{off}|s, b; \alpha) p(b) p(s)}{\int ds\, db\, p(N_{on}, N_{off}, s, b; \alpha)}$$
$$\propto \int db\, p(N_{on}, N_{off}|s, b; \alpha). \quad (2)$$

Thus the PDF of the signal rate s is proportional to the likelihood function in which the background rate b is integrated out, leaving a marginal distribution of s.

The likelihood function can be expressed in the following way:

$$p(N_{on}, N_{off}|s, b; \alpha) = p(N_{on}|s, \alpha b) \cdot p(N_{off}|b)$$
$$= \frac{(s + \alpha b)^{N_{on}}}{N_{on}!} e^{-(s + \alpha b)} \cdot \frac{b^{N_{off}}}{N_{off}!} e^{-b}, \quad (3)$$

where we have made use of the independence of the measured values N_{on} and N_{off} and of the fact that both values come from a Poisson process with rate given respectively by $s + \alpha b$ and b.

Using the binomial identity,[3] we can factorize the likelihood in Eq. (3) in two Poisson distributions, one for N_s with expected value s and one for $N_{on} + N_{off} - N_s$ with expected value $b(1 + \alpha)$,

---

[3]For reasons that will be clear in a while, the bound variable in the binomial identity is called N_s, i.e.,

$$(s + \alpha b)^{N_{on}} = \sum_{N_s=0}^{N_{on}} \frac{N_{on}!}{(N_{on} - N_s)! N_s!} s^{N_s} (\alpha b)^{N_{on} - N_s}.$$

$$p(N_{on}, N_{off}|s, b; \alpha)$$
$$\propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(1 + 1/\alpha)^{-N_s}(N_{on} - N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}$$
$$\times \frac{(b(1 + \alpha))^{N_{on} + N_{off} - N_s}}{(N_{on} + N_{off} - N_s)!} e^{-b(1 + \alpha)}. \quad (4)$$

Here, factors that depend merely on N_{on} and N_{off} have been ignored.

The integral in Eq. (2) is now straightforward

$$p(s|N_{on}, N_{off}; \alpha) \propto \sum_{N_s=0}^{N_{on}} \frac{(N_{on} + N_{off} - N_s)!}{(1 + 1/\alpha)^{-N_s}(N_{on} - N_s)!} \cdot \frac{s^{N_s}}{N_s!} e^{-s}. \quad (5)$$

Note that now s and N_s are both random variables, which can have only non-negative values, in agreement with the Poissonian nature of the problem under study.

We take into account the following identities:

$$p(s|N_{on}, N_{off}; \alpha) = \sum_{N_s=0}^{N_{on}} p(N_s|N_{on}, N_{off}; \alpha) \cdot p(s|N_s) \quad (6)$$

and

$$p(s|N_s) = \xi \frac{(s \cdot \xi)^{N_s}}{N_s!} e^{-\xi \cdot s}. \quad (7)$$

The former results from marginalizing over the variable N_s. The latter is obtained from applying the Bayesian theorem with constant priors to the likelihood in Eq. (1). Recalling that $\xi = 1$, we can now compare Eqs. (5) and (6) to obtain the PMF of the variable N_s

$$p(N_s|N_{on}, N_{off}; \alpha) \propto \frac{(N_{on} + N_{off} - N_s)!}{(1 + 1/\alpha)^{-N_s}(N_{on} - N_s)!}. \quad (8)$$

Equation (8) allows one then to define the most probable value (the mode) as an estimation of the number of signal events. This procedure was in fact previously outlined by Loredo [see Eq. (5.13) of Ref. [2]] already in 1992.

The main goal of this work is to extend Eq. (8) by including the information of the individual events without limiting ourselves with a "global" method that makes use only of the number N_{on} and N_{off}. To do so, we will first use Monte Carlo (MC) simulations to compare the Bayesian approach [summarized in Eq. (5)] with the frequentist approach in Sec. II. We will also discuss why the former is preferable in this problem. Then in Sec. III, we will explain how to introduce single-event information in Eq. (8), and in Sec. IV we will investigate the effects on the precision in the estimation of the number of the signal rate, using as an example real data and simulations from the MAGIC imaging atmospheric Cherenkov telescopes (IACTs).

II. COMPARISON BETWEEN FREQUENTIST AND BAYESIAN APPROACH

In the previous section, we estimated the signal rate using a Bayesian approach. In literature, however, the on/off measurement problem is often solved in the frequentist approach.

In the frequentist approach [3–5], the background rate b, that is a nuisance parameter in the Bayesian approach, is not integrated out as done in Eq. (2). Instead from the likelihood in Eq. (3) one defines the following test statistic (usually referred to as the likelihood ratio):

$$\lambda(s) \equiv \frac{p(N_{on}, N_{off}|s, b = \hat{b}; \alpha)}{p(N_{on}, N_{off}|s = N_{on} - \alpha N_{off}, b = N_{off}; \alpha)}, \quad (9)$$

where[4]

$$\hat{b} = \frac{N + \sqrt{N^2 + 4(1 + \alpha)sN_{off}}}{2(1 + \alpha)}. \quad (10)$$

is the value of b that maximizes the likelihood in Eq. (3) for a given s, and $N \equiv N_{on} + N_{off} - (1 + 1/\alpha)s$.

The advantage of Eq. (9) is that according to Wilks's theorem [6], the function $-2\log \lambda(s)$ has an approximate χ^2 distribution with 1 degree of freedom, which can be used to extract confidence intervals. For example if we want the 68% confidence interval we impose $-2\log \lambda(s) = 1$ and it can be shown [3] that for a large number of counts N_{on} and N_{off} this condition is satisfied when[5]

$$s = (N_{on} - \alpha N_{off}) \pm \sqrt{N_{on} + \alpha^2 N_{off}}. \quad (11)$$

By imposing $-2\log \lambda(s) = 3.84$ one can get 95% upper limits, as it is done in Ref. [4], although with *ad hoc*

adjustments. These *ad hoc* adjustments are not surprising because the maximum likelihood approach described so far suffers from the following problems [2]:

(i) It only works well with counts number large enough. It is not suitable for low count numbers [3], while the Bayesian approach has no limitations on that.

(ii) Only information about confidence intervals can be extracted. Legitimate questions such as "what is the probability of having N_s signal events in a sample of N_{on} events?" cannot be answered [this is indeed possible in the Bayesian approach as shown in Eq. (8)].

(iii) The frequentist result of Eq. (11) does not exclude negative rate, but a Poisson process conflicts with negative rate.[6]

To overcome the above issues *ad hoc* adjustments are required.[7] Another advantage of the Bayesian approach is that, once we have the PDF of the signal rate, all information is encoded in $p(s|N_{on}, N_{off}; \alpha)$ defined in Eq. (6). From this equation, we can obtain the mode that maximizes $p(s|N_{on}, N_{off}; \alpha)$, i.e., the most probable value, or the 68% credible interval[8] $[s_{left}, s_{right}]$ with s_{left} and s_{right} such that

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

$$p(s_{left}|N_{on}, N_{off}; \alpha) = p(s_{right}|N_{on}, N_{off}; \alpha). \quad (12)$$

If this last condition cannot be fulfilled[9] then $s_{left} = 0$ and upper limits (ULs) on the signal rate can be computed. The 95% UL s_{95} can be intuitively defined by

$$\int_0^{s_{95}} p(s|N_{on}, N_{off}; \alpha)ds = 0.95. \quad (13)$$

These definitions in the Bayesian formalism of credible interval and UL were already explored in the context of on/off measurements in γ-ray astronomy by the author of Ref. [7]. Although in the work in Ref. [7] Jeffreys (and not constant) priors[10] were assumed.

In Fig. 1 we show the comparison between the frequentist and Bayesian approach in estimating the signal rate. One can notice that $-2\log \lambda(s)$, defined in Eq. (9), always has the minimum value at $s = N_{on} - \alpha N_{off}$. Only for large

---

[4]Note that when the null hypothesis is assumed ($s = 0$), then $N = N_{on} + N_{off}$, $\hat{b} = \alpha N/(1 + \alpha)$ and Eq. (9) gives Eq. (17) of Ref. [3] for computing the detection significance.
[5]The factor after "±" in Eq. (11) is derived from the variance of the linear combination of independent random variables.

[6]One can argue that in the frequentist approach these negative rates are in the end put equal to zero and a negative flux will not be claimed. But while this comes naturally in the Bayesian approach, in the frequentist approach instead this needs to be done by "hand" with the introduction of *ad hoc* adjustments.
[7]The word "adjustments" is present in Ref. [4] 7 times.
[8]Not to be confused with the frequentist confidence interval.
[9]When dealing with low excess events the Bayesian credible intervals can be highly asymmetric around the estimated signal rate, as shown for instance in the right plots of Fig. 1 where $s_{left} = 0$.
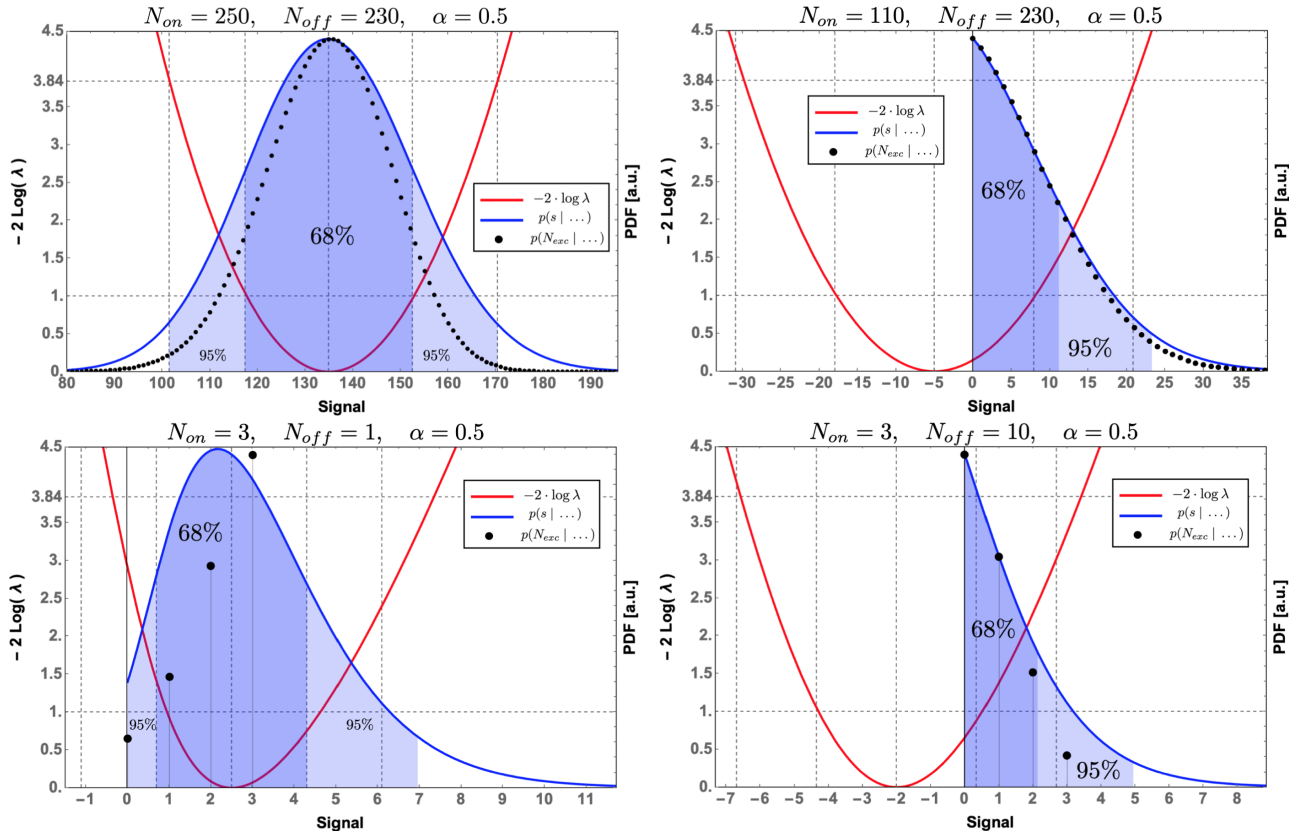[10]See for instance Ref. [8] for a review of the problem regarding the choice of the priors.

FIG. 1. Comparison for different values of $N_{\rm on}$ and $N_{\rm off}$ between $-2\log\lambda(s)$ (red line) defined in Eq. (9) and $p(s|N_{\rm on}, N_{\rm off}; \alpha)$ (blue line) and $p(N_s|N_{\rm on}, N_{\rm off}; \alpha)$ (black points), defined respectively in Eqs. (5) and (8). For the last distribution the x axis does not show the signal rate but the discrete variable N_s. The last two probability distributions have been rescaled for comparison with $-2\log\lambda(s)$. Vertical dashed lines are for $s = E(s) \pm k\sigma_s$ ($k = \pm\{0, 1, 2\}$), where $E(s) = N_{\rm on} - \alpha N_{\rm off}$ and $\sigma_s = \sqrt{N_{\rm on} + \alpha^2 N_{\rm off}}$.
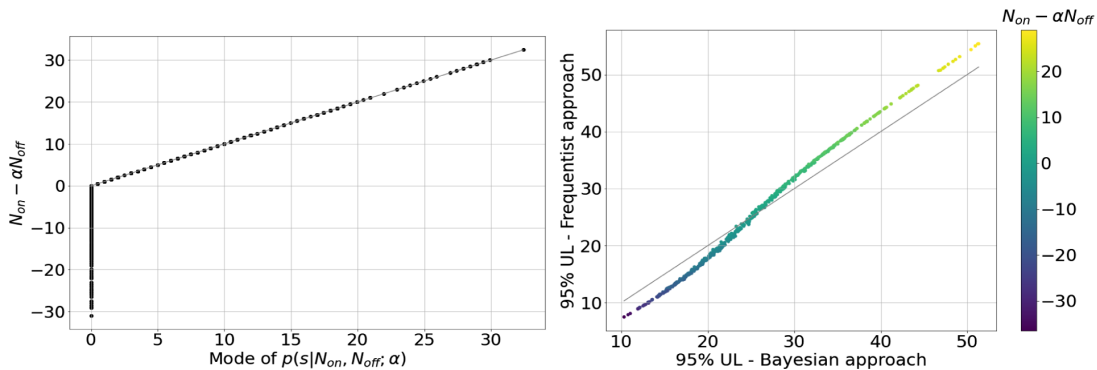

FIG. 2. Left: inferred signal from the Bayesian approach (x axis) and the frequentist approach (y axis) using MC simulations. Right: 95% UL obtained in the Bayesian approach (x axis) from Eq. (13) and in the frequentist approach (y axis) following the prescription in Ref. [4] using MC simulations; color coding indicates the value of $N_{\rm on} - \alpha N_{\rm off}$. MC simulations were produced assuming $s = 0$, $b = 200$, and $\alpha = 0.5$.

number of counts (see upper plots of Fig. 1) $-2\log\lambda(s) = 1$ when $s = (N_{\rm on} - \alpha N_{\rm off}) \pm \sqrt{N_{\rm on} + \alpha^2 N_{\rm off}}$. It is also interesting to notice that confidence and credible intervals agree for a large number of counts and when we are not close to the border $s = 0$. For low count numbers and when we are close to the border of the parameter space the frequentist

approach has the problems previously discussed for which one needs *ad hoc* adjustments.

We ran MC simulations to compare the results obtained with the two approaches. In each MC simulation $N_{\rm on}$ is generated by the sum of two Poisson random numbers with expected count s and αb, respectively. $N_{\rm off}$ is instead
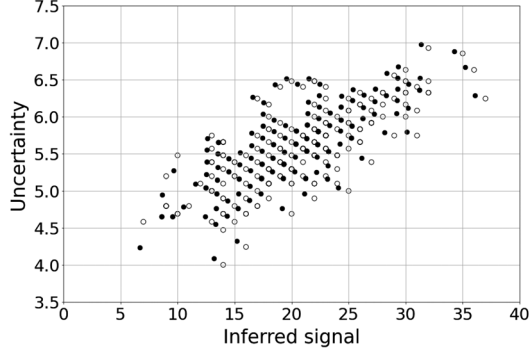
FIG. 3. Inferred signal (x axis) and its uncertainty (y axis) from MC simulations in which $s = 20$, $b = 5$, and $\alpha = 1$. In the Bayesian approach (filled circles) the inferred signal is the mode of the signal PDF [see Eq. (5)] with uncertainty given by $(s_{\text{right}} - s_{\text{left}})/2$ [see Eq. (12)]. In the frequentist approach (empty circles) the inferred signal is given by $N_{\text{on}} - \alpha N_{\text{off}}$ with uncertainty $\sqrt{N_{\text{on}} + \alpha^2 N_{\text{off}}}$.

generated from a Poisson distribution with expected count b. Once N_{on} and N_{off} are obtained the inferred signal and its uncertainty are computed according to Eq. (11) for the frequentist approach. In the Bayesian approach instead, the estimated signal is given by the most probable value, with uncertainty corresponding to the 68% credible interval defined in Eq. (12), i.e., $(s_{\text{right}} - s_{\text{left}})/2$.

Additionally, in the left plot of Fig. 2 one can see that, as long as $N_{\text{on}} - \alpha N_{\text{off}} > 0$, there is a perfect agreement between the two approaches in estimating the signal rate. However, this is no longer true for $N_{\text{on}} - \alpha N_{\text{off}} < 0$. In such case the Bayesian approach correctly (given the Poissonian nature of the problem) estimates a signal rate equal to zero. Similar considerations can be made for the 95% UL estimation shown in the right plot of Fig. 2, where both approaches are in good agreement: as long as $N_{\text{on}} - \alpha N_{\text{off}} > 0$ a slight overestimation of the UL value is observed for the frequentist approach [4] relative to the Bayesian one, while the opposite is true for $N_{\text{on}} - \alpha N_{\text{off}} < 0$.

In Fig. 3 we show the 68% confidence/credibility band (y axis of the plot) around the estimated signal rate (x axis of the plot) for both approaches, in which one can see that there is a good agreement between the results yielded by the frequentist and Bayesian approach.

III. PROBABILITY DENSITY FUNCTION OF THE SIGNAL RATE INCLUDING SINGLE-EVENT OBSERVABLES

In Eq. (8), we have defined the PMF of the number of signal events based on the number of events in the on and off regions. It is common to select these events based on signal extraction cuts on one or more event variables to increase the SNR. A very common example of this in

astronomy is a cut performed in a region around the source, so that all events outside such region are ignored. A more advanced example is the implementation of some classification algorithm[11] which yields for each event a discriminating variable that can be used for the background suppression.

A disadvantage of cutting data is that also a fraction of the signal events will be excluded, which translates to a reduced exposure on the target. Moreover, normally 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 cuts. We aim instead to fully exploit the information on how single-event variables distribute for a signal or a background population. Our goal is to show how by replacing a fixed signal extraction cut with a statistical weighting of the events according to specific information (that is, not excluding any event *a priori*), we obtain a more precise signal estimation. We call this novel method BASiL.

We start by including the information about the variables **x**, which we have observed for each event, in the inference of the signal rate s. The variable **x** might be a single observable (like a discriminating variable obtained by a classification algorithm) or a set of observables. Including $\vec{\mathbf{x}} = \{\mathbf{x}_1, …, \mathbf{x}_{N_{\text{on}}}\}$, Eq. (2) becomes

$$p(s|\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}; \alpha) \propto \int_0^\infty db\, p(\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}|s, b; \alpha). \quad (14)$$

Using now the chain rule of probabilities, we write the likelihood in the following way:

$$\begin{aligned} p(\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}|s, b; \alpha) \\ = p(\vec{\mathbf{x}}|N_{\text{on}}, s, \alpha b) \cdot p(N_{\text{on}}|s, \alpha b) \cdot p(N_{\text{off}}|b). \quad (15) \end{aligned}$$

The last two factors are Poisson distributions with expected counts $s + \alpha b$ and b, respectively. The first factor is the probability of observing the variables $\vec{\mathbf{x}}$ in a sample of N_{on} events with an assumed signal rate s and background rate αb. Given that all measured events are independent from each other, this probability is

$$\begin{aligned} p(\vec{\mathbf{x}}|N_{\text{on}}, s, \alpha b) \\ = \prod_{i=1}^{N_{\text{on}}} [p(\mathbf{x}_i|\gamma) \cdot p(\gamma|s, \alpha b) + p(\mathbf{x}_i|\bar{\gamma}) \cdot p(\bar{\gamma}|s, \alpha b)], \quad (16) \end{aligned}$$

where the term

---

[11]This is the case of the gamma-hadron separation for imaging Cherenkov telescopes, where each event is given a score called "hadronness" in MAGIC [9] or sometimes referred to as "gammaness" in the Cherenkov Telescope Array experiment.

$$p(\gamma|s, \alpha b) = 1 - p(\bar{\gamma}|s, \alpha b) = \frac{s}{s + \alpha b} \quad (17)$$

is the *prior* probability that one event is a signal event γ. We denote everything that is not signal as $\bar{\gamma}$.

The terms $p(\mathbf{x}|\gamma)$ and $p(\mathbf{x}|\bar{\gamma})$ are the PDFs of observing the variables $\mathbf{x}$ from a signal or background population, respectively. In the Bayesian formalism, they can also be referred to as the likelihood functions of being a gamma or background event respectively, having observed the variables $\mathbf{x}$ for that particular event. Depending on the kind of variable or problem under study, these likelihoods can be estimated from MC simulations, a different dataset, or be based on a theoretical model. One needs also to ensure that these likelihoods are normalized,

$$\int_{\mathcal{X}} p(\mathbf{x}|\gamma)d\mathbf{x} = \int_{\mathcal{X}} p(\mathbf{x}|\bar{\gamma})d\mathbf{x} = 1, \quad (18)$$

with $\mathcal{X}$ being the (multidimensional) parameter space in which the variables $\mathbf{x}$ are defined. The kind of likelihood in Eq. (15) is known in literature [10] as "marked" Poisson process, i.e., a Poisson process where each count is marked with a property (the variables $\mathbf{x}$ in our case) distributed according to a given PDF.

Equation (15) can now be rewritten as (for a full derivation see Appendix A)

$$p(\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}|s, b; \alpha)$$
$$\propto \sum_{N_s=0}^{N_{\text{on}}} \frac{(N_{\text{on}} + N_{\text{off}} - N_s)!}{(N_{\text{on}} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{x}, N_s)}{\binom{N_{\text{on}}}{N_s}}$$
$$\times \frac{s^{N_s}}{N_s!} e^{-s} \cdot \frac{(b(1+\alpha))^{N_{\text{on}}+N_{\text{off}}-N_s}}{(N_{\text{on}} + N_{\text{off}} - N_s)!} e^{-b(1+\alpha)}, \quad (19)$$

where the function C represents the combinatorial term

$$C(\vec{\mathbf{x}}, N_s) = \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i|\gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j|\bar{\gamma}) \quad (20)$$

with F_{N_s} being the set of all subsets of N_s integer numbers that can be selected from $\{1, \ldots, N_{\text{on}}\}$.

At this point, like done in Sec. I, we can easily marginalize the nuisance parameter b and obtain the final result for the PDF of the signal rate s,

$$p(s|\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}; \alpha) \propto \int db\, p(\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}|s, b; \alpha)$$
$$\propto \sum_{N_s=0}^{N_{\text{on}}} \frac{(N_{\text{on}} + N_{\text{off}} - N_s)!}{(N_{\text{on}} - N_s)!(1 + 1/\alpha)^{-N_s}}$$
$$\times \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{\text{on}}}{N_s}} \frac{s^{N_s}}{N_s!} e^{-s}. \quad (21)$$

Again one can recognize in this last expression the marginalization in Eq. (6), so that we can identify

$$p(N_s|\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}; \alpha)$$
$$\propto \frac{(N_{\text{on}} + N_{\text{off}} - N_s)!}{(N_{\text{on}} - N_s)!(1 + 1/\alpha)^{-N_s}} \frac{C(\vec{\mathbf{x}}, N_s)}{\binom{N_{\text{on}}}{N_s}}. \quad (22)$$

Given that the combinatorial term C, defined in Eq. (20), is the novelty of this method, it is worth elaborating its role by providing an example. Let us assume $N_{\text{on}} = 3$ 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. Thus, when $N_s = 0, 1, 2, 3$ the combinatorial term will be respectively[12]

$$C(\vec{x}, 0) = p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma}),$$
$$C(\vec{x}, 1) = p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\bar{\gamma})$$
$$+ p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma})$$
$$+ p(x_1|\bar{\gamma}) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma),$$
$$C(\vec{x}, 2) = p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\bar{\gamma})$$
$$+ p(x_1|\gamma) \cdot p(x_2|\bar{\gamma}) \cdot p(x_3|\gamma)$$
$$+ p(x_1|\bar{\gamma}) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma),$$
$$C(\vec{x}, 3) = p(x_1|\gamma) \cdot p(x_2|\gamma) \cdot p(x_3|\gamma).$$

From the above example it is clear how the combinatorial term is made up to account for all the possible combination of excess events among the total N_{on} events that can give the observed values $\vec{x}$. If, for instance,

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

i.e., all events are l time more likely to be a signal event, then

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

and

$$p(N_s|\vec{\mathbf{x}}, N_{\text{on}}, N_{\text{off}}; \alpha) \propto \frac{(N_{\text{on}} + N_{\text{off}} - N_s)!}{(N_{\text{on}} - N_s)!(1 + 1/\alpha)^{-N_s}} \cdot l^{N_s}. \quad (23)$$

By taking into account the information that all events are l times more likely to be a signal event, we have updated the PMF of the number of signal events introducing a factor l^{N_s}. Its maximum values are obtained for $N_s = N_{\text{on}}$ if $l > 1$, and for $N_s = 0$ if $l < 1$. For $l = 1$ we do not gain any

---

[12]In Appendix B, a general algorithm is shown for efficiently obtaining the term $C(\vec{\mathbf{x}}, N_s)$, given the list of likelihoods $p(\mathbf{x}_i|\gamma)$ and $p(\mathbf{x}_j|\bar{\gamma})$ as input.

information from the observed variable x, and we recover the result previously obtained in Sec. I.

With the introduction of the combinatorial term in Eq. (22) we have devised the method to include event-by-event information for the computation of N_s. The power of this method clearly depends on the specifics of the datasets in which it is applied, and in turn, it depends on (i) the event parameters that are used, (ii) how they distribute for the signal and background population, and (iii) how performing is the signal extraction method that relies on a fixed fiducial cut. However, in order to be predictive and define a framework to assess the performance of the BASiL method, we apply it to a specific case, that of gamma-ray observation. For this purpose, we analyze real data from the MAGIC Collaboration.[13] Results reported in Ref. [12] will be used as a benchmark case.

IV. THE CASE OF IMAGING ATMOSPHERIC CHERENKOV TELESCOPES

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. For steady pointlike sources two variables are generally further used to suppress the background: the squared[14] angular distance from the source θ^2 and a particle identification variable, which in the case of MAGIC is computed by means of a random forest (RF) algorithm, and is dubbed hadronness (h) [9]. The RF event classifier takes the image parameters of the event as input and returns a value between 0 and 1. The smaller the value, the more the event looks like a gamma-ray event. The θ^2 parameter, related to the instrument point spread function depends on the telescope optics and mechanics and mostly on the shower physics and image reconstruction (see Ref. [13]). Therefore, the individual-event variables to consider are

$$\mathbf{x} = (\theta^2, h, E).$$

Because the distributions of θ^2 and h are energy dependent, we have also included the estimated energy E.[15]

---

[13]See Ref. [11].

[14]Signal events spread around the region of interest and for a pointlike source they distribute according to a two-dimensional Gaussian distribution. Such a two-dimensional Gaussian in the θ_x and θ_y space will correspond to an exponential function for the distribution of $\theta^2 = \theta_x^2 + \theta_y^2$.

[15]In principle one could also consider the time of arrival of individual events t as element of $\mathbf{x}$. While this may be useful in some scenarios, $\mathbf{t}$ can be neglected if we consider small enough time bins or similar conditions throughout the entire observation that allows us integrate out the time in our analysis.

The likelihoods of being a signal $p(\mathbf{x}|\gamma)$ or a background $p(\mathbf{x}|\bar{\gamma})$ event can be factorized into three terms,

$$p(\mathbf{x}|\gamma) = p(h|E, I, \gamma) \cdot p(\theta^2|E, I, \gamma) \cdot p(E|I, \gamma),$$
$$p(\mathbf{x}|\bar{\gamma}) = p(h|E, I, \bar{\gamma}) \cdot p(\theta^2|E, I, \bar{\gamma}) \cdot p(E|I, \bar{\gamma}), \quad (24)$$

where I stands for the conditions under which the observation has been performed (e.g., zenith angle, atmospheric opacity, etc.). As the correlation between θ^2 and h ranges between zero and 0.2, approximately, we can consider the two variables as independent. The same is not valid for the correlation between h, θ^2, and energy, which forces us to take into account the energy dependence of the distribution of θ^2 and h and apply the method in sufficiently small energy bins.[16]

We therefore focus on individual energy bins where the flux is assumed to be constant, so that $p(E|I, \gamma)$ and $p(E|I, \bar{\gamma})$ are uniform. This means that the factor $p(E|I, \gamma) = p(E|I, \bar{\gamma})$ is the same for all likelihoods in Eqs. (24) and can be therefore ignored. Thus, to get the likelihood for each event of being a signal or background event one needs to only compute the distribution in θ^2 and hadronness respectively from a signal and background population.

A sample of background events can be obtained by performing observations on regions of the sky (off regions) where no signal contamination is observed and with similar conditions I as the "on" sample, as explained in Sec. I. Obtaining a signal sample from the on region is less straightforward, because in the on region both a signal and an irreducible background contribution are present. For this reason in IACTs one has to rely on MC simulations of signal events to study the parameter distribution of a signal sample. Nonetheless, for a bright enough source like the Crab Nebula,[17] a very pure sample of γ-ray signal events can be extracted from the on measurement, which allows us to study its properties. Data are taken in the so-called wobble[18] mode [14], which yields N_{on} and N_{off} counts, then the excess is obtained by subtracting from the N_{on} counts the (relative small) background count αN_{off} and the procedure is repeated for different cuts in hadronness or θ^2.

Figure 4 shows the distribution in $\mathbf{x} = (h, \theta^2)$ of the signal excess from the Crab Nebula sample, MC-simulated signal, and background events. For brevity, we show only events with estimated energy between 189 and 300 GeV. Following the same prescription and using the

---

[16]If one wants to extrapolate the method to an unbinned analysis, a signal flux $p(E|I, \gamma)$ has to be assumed beforehand.

[17]The Crab Nebula is the brightest steady TeV gamma-ray emitter in the sky. It is a pulsar wind nebula which is used as a standard calibration for IACTs [12].

[18]In a wobble mode the source is placed with a certain offset with respect to the camera center during the observation. It allows simultaneous signal and background estimation.
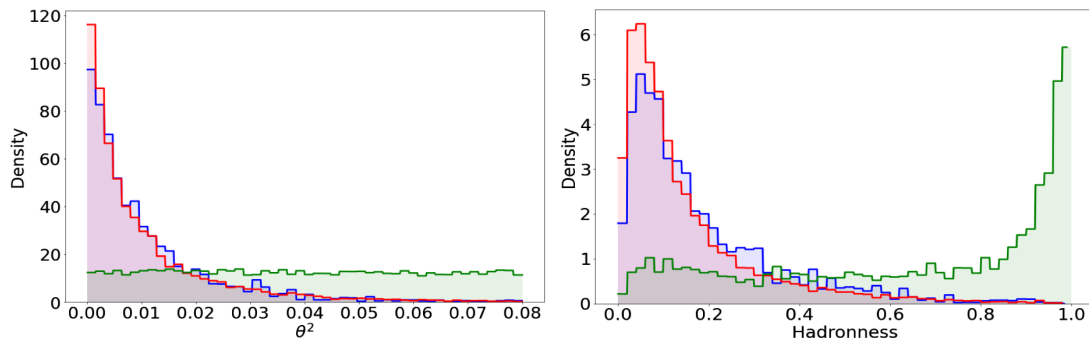
FIG. 4.   Distribution in θ^2 (left) and hadronness (right) for simulated (red) and observed (blue) signal excess and for background events (green) in the energy range 189–300 GeV. These distributions are obtained from Figs. 1 and 15 of Ref. [12]. For all histograms 10^5 events were generated and divided in 50 bins. All distributions (colored areas in the figure) are normalized to 1. For a discussion on the origin and effect of the MC/data discrepancies see Ref. [12].

same dataset, a similar analysis is performed for θ^2 which yields the distributions in the left plot of Fig. 4.

A few important facts appear from the plots in Fig. 4: (i) there is a mismatch between Monte Carlo data and real data, (ii) such difference is larger in hadronness than θ^2, especially at very low hadronness values, where several signal events are not classified as gammas with sufficient degree of confidence, and (iii) the signal selection based on θ^2 is efficient with a cut at about 0.02°, that entails 75% of the signal, while an optimal cut in hadronness is more complex to define, because it depends more strongly on the energy.

The optimization of the SNR can be done in several ways. The MAGIC collaboration elaborated a set of cuts specific for each energy bin, according to an "efficiency" parameter ϵ defined as the fraction of Monte Carlo signal events surviving a certain cut. In the following, we elaborate on this, and compare the outcome with the novel method which we propose.

Assuming a signal rate s and background rate b, MAGIC observations are simulated following these steps:

(i) Generate N_s and N_{bkg} from a Poisson distribution with expected value respectively s and $\alpha \cdot b$, and define the number of counts in the on region as

$$N_{on} = N_s + N_{bkg}.$$

(ii) Generate N_{off}, the total number of events in the off region, from a Poisson distribution with expected value b.

(iii) Generate θ^2 values for the events in the on region by randomly picking up N_s values from the signal distributions (blue histogram in the left plot of Fig. 4) and N_{bkg} values from the background distributions (green histogram in the left plot of Fig. 4).

(iv) Generate θ^2 values for the events in the off region by randomly picking up N_{off} values from the background distributions (green histogram in the left plot of Fig. 4).

(v) Finally, the same is done for generating hadronness values for the on and off measurements using this time the right plot of Fig. 4.

Having an on and off measurement, we get an estimation $\hat{s}$ of the signal rate using only the information about the total counts N_{on} and N_{off} and the single-event variables $\mathbf{x} = (h, \theta^2)$. This estimation is done using two different approaches, referred to as the "standard" and "BASiL" approach:

(1) The estimated signal rate is obtained from

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

where N_{on} and N_{off} are the numbers of events surviving the cut in θ^2 and/or hadronness for the on and off measurement, respectively. Cut values are obtained assuming a given γ-ray efficiency ϵ computed from the signal distributions (see blue histograms of Fig. 4). Being the most common way of suppressing the background and estimating s, we will refer to this approach as the standard one.

(2) In the BASiL approach $\hat{s}$ is instead defined from the mode of the PMF[19] defined in Eq. (22), where $\mathbf{x}$ can be either θ^2 and hadronness, or only one of them. The combinatorial term in Eq. (20) will be obtained using signal and background likelihood values from the signal distributions (blue histograms in Fig. 4) and background distribution (green histograms in Fig. 4).

It is important to stress that in both approaches the values of s, b, N_s, and N_{bkg} are not taken into account: only observed quantities (counts in the on and off regions, θ^2 and hadronness) are considered for signal-rate estimation.

---

[19]One could have also used the signal-rate PDF defined in Eq. (21), but such choice would not change the results since the distributions for s and N_s share the same mode (one is simply derived from the other by including Poisson statistics).
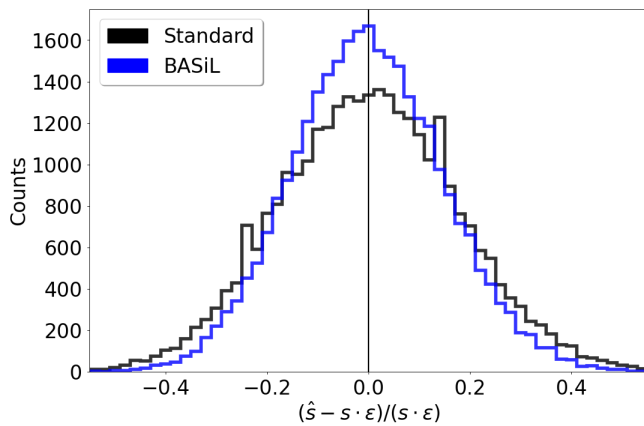
FIG. 5. Distribution of $(\hat{s} - s \cdot \epsilon)/(s \cdot \epsilon)$, with $\hat{s}$ the estimated signal rate obtained from $N_{\rm on} - \alpha N_{\rm off}$ in the standard approach (black) and from the mode of the distribution in Eq. (22) in the BASiL approach (blue). In the standard approach $N_{\rm on}$ and $N_{\rm off}$ are the number of events surviving the efficiency cut applied on the data. Efficiency is 95% and 75% respectively for hadronness and θ^2, which translates in a total efficiency $\epsilon = 67.6\%$ (see main text, Sec. IV for details). In the BASiL approach no cut is applied on the data, i.e., $\epsilon = 1$. Values of s and b used in the simulations are 10^2 and 3×10^3, respectively, with $\alpha = 1/3$.

More importantly, for this study we ignore the mismatch between the "real"-γ and the MC-γ distributions (respectively the blue and red histograms of Fig. 4). In Appendix D the study on how such mismatch between the MC and real data affects the estimation is shown. The conclusion is that this mismatch induces a bias in the estimation that leads one to underestimate the number of excess events.

At this point we defined the signal-estimation precision and bias respectively as the standard deviation and mean value of the rescaled distribution of $\hat{s}$, i.e.,

$$prec = \sigma\left(\frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon}\right), \qquad (25)$$

$$bias = \left\langle\frac{\hat{s} - s \cdot \epsilon}{s \cdot \epsilon}\right\rangle. \qquad (26)$$

Note that the efficiency cut is $\epsilon = 1$ in the BASiL approach, since no cut is applied on the data in this case. For a fair comparison in the standard approach $\hat{s}$ is put equal to zero whenever $N_{\rm on} - \alpha N_{\rm off} < 0$. In Fig. 5 an example of such distribution is shown, where the signal rate is estimated using the standard (black histogram) and BASiL approach (blue histogram).

A. Signal-estimation precision and bias
for different efficiency cut

We first study the evolution of the bias and precision defined in Eqs. (25) and (26) for different efficiency cuts considering only θ^2 or hadronness as single-event variable. For this study we assume a background intensity, in the on region $\alpha b = 1000$ and a SNR of 10%, i.e., $s = 100$. We then simulate observations following the steps previously described where in one case events have only θ^2 as an observed variable and only hadronness in the other case. Figure 6 reports the results.

It is worth noticing that in the hadronness case, starting from 100% efficiency, the precision in the standard approach improves immediately reaching its best value when the efficiency is about 80%–90%. The same improvement happens also in the θ^2 case, but it is smoother with a minimum around 70%–80%. An explanation of this effect can be easily found in Fig. 4, where it is clear that cutting in hadronness allows one to suppress more background than performing a similar cut (i.e., with the same efficiency) in θ^2. At low efficiency values the precision is dominated
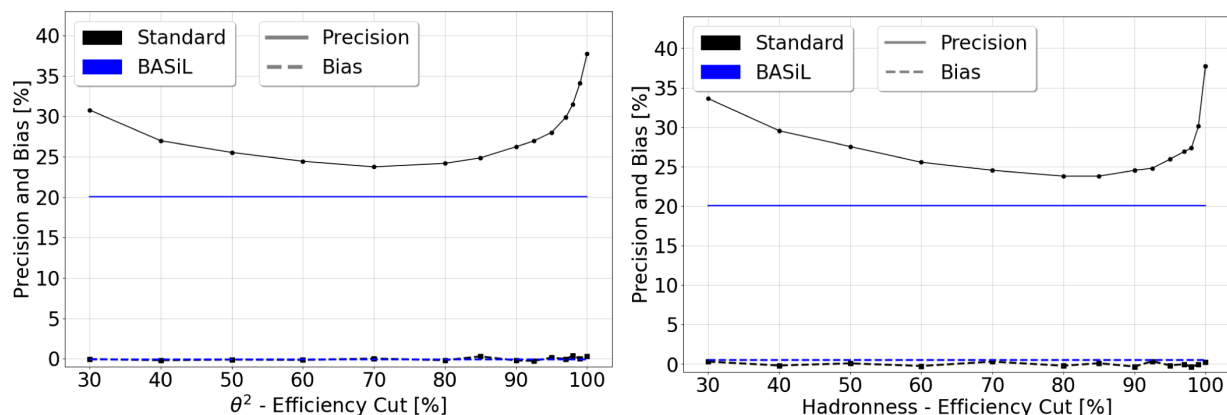

FIG. 6. Comparison between the standard (black) and BASiL (blue) approach for the evolution of the precision (full line) and bias (dashed line) in the signal-rate estimation assuming different gamma-ray efficiency cut in θ^2 (left) and hadronness (right). Note that for the BASiL approach the precision and bias do not depend on the efficiency, being $\epsilon = 1$ in such case. Nonetheless, for a visual comparison these values are shown as horizontal lines. The definition of precision and bias can be found in Eqs. (25) and (26), respectively. Observations are simulated assuming $s = 10^2$ and $\alpha b = 10^3$, with $\alpha = 1/3$.
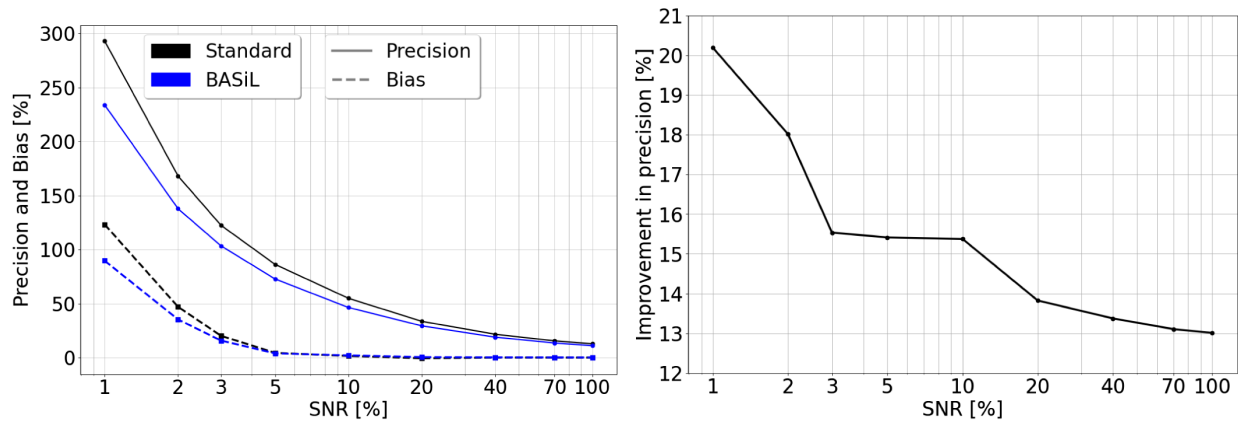
FIG. 7. Left: comparison between the standard (black) and BASiL (blue) approach for the evolution of the precision (full line) and bias (dashed line) assuming different SNR. In the standard approach the efficiency cut is fixed to 90% for hadronness and 75% for θ^2. Right: improvement for different SNR of the precision in the BASiL approach relative to the standard one. The definition of precision and bias can be found in Eqs. (25) and (26), respectively. Observations are simulated assuming $\alpha b = 100$, with $\alpha = 1/3$.

by the Poisson statistic of the excess number and therefore its evolution follows

$$prec \sim 1/\sqrt{N_s} \propto 1/\sqrt{\epsilon}.$$

In the BASiL approach instead 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 bias in the signal estimation is very close to zero, apart from small fluctuations, in both approaches. As discussed in Appendix D, the most important source of bias is due to the nonperfect agreement between the real and simulated signal distribution.

We therefore conclude that 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: for a SNR of 10% the improvement in precision is about ∼15% in both hadronness and θ^2.

B. Signal-estimation precision and bias for different signal to background ratio

In the previous section we fixed the SNR to 10% and let the efficiency cut vary. We now want to do the opposite, i.e., study the precision and bias by varying the SNR. For this purpose we fixed the efficiency cut in hadronness and θ^2 to 90% and 75%, respectively, these values being the recommended ones [12] and the ones that, as one can see in Fig. 6, maximize the precision power of the standard approach. In the BASiL approach (in which $\epsilon = 1$) this time both hadronness and θ^2 will be considered when computing the single-event likelihoods of being a signal or a background event. Figure 7 displays the precision and bias for different values of SNR. As expected, in both approaches both values get worse as we decrease the signal (in the MC simulations the background is kept fixed to

$\alpha b = 100$). Such worsening is, however, less pronounced in the new approach, where the precision is about 20% better for a SNR of 1%. If the strength of the signal is instead equal to the background noise, i.e., SNR = 100%, then the improvement of the BASiL method relative to the standard one is ∼13% (see right plot of Fig. 7). One can also notice that at low values of SNR the bias increases: this is due to the fact that for weak signal rates estimates $\hat{s}$ that are close or equal to zero[20] become more frequent, and this inevitably shifts the mean value of the distribution of $(\hat{s} - s \cdot \epsilon)/(s \cdot \epsilon)$ through positive values.

We conclude that the BASiL method is capable of estimating the signal rate more precisely, without having to select data. This is due to the introduction of the combinatorial term defined in Eq. (20), which takes into account the likelihood of each event being a signal or background event.

C. A spectral energy distribution

After having evaluated the performance of the method by using MC simulations of events observed by the MAGIC telescopes, we now apply the method on a real dataset. For this purpose we used the data released by the MAGIC Collaboration [15], which includes 40 min of Crab Nebula observations chosen from the sample used for the performance evaluation in Ref. [12]. This dataset includes only events recorded at low zenith angles (<30°) and under good atmospheric conditions. All data were taken in the wobble mode with the standard offset of 0.4°. Off counts were obtained using three simultaneous off regions within the same field of view and with the same offset from

---

[20]Recall that for a fair comparison between the two approaches, in the standard approach $\hat{s}$ is put equal to zero whenever $N_{on} - \alpha N_{off} < 0$.
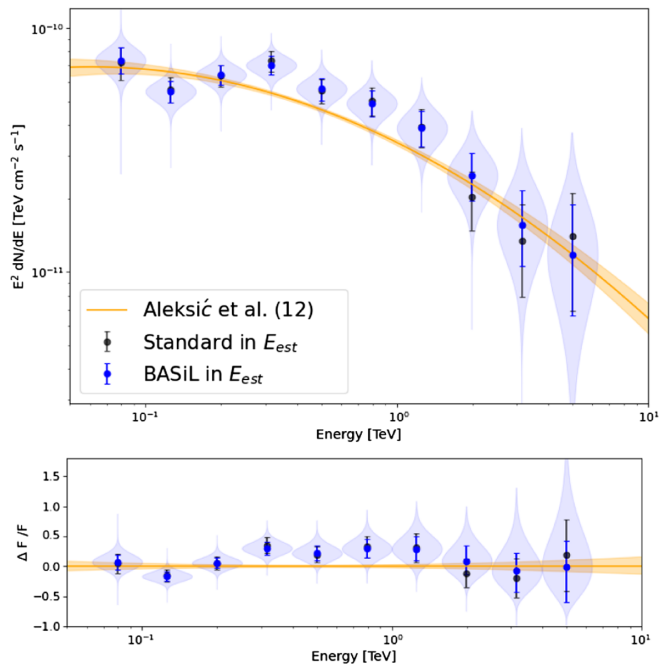
FIG. 8. SED in estimated energy of the Crab Nebula (in blue) obtained by processing 0.66 h of data released by the MAGIC Collaboration with the BASiL method. For comparison also the results (in black) obtained from the same data sample using the standard analysis procedure are reported in which efficiency cuts are applied. The blue bars of the data points are the credible interval obtained from Eq. (12), in which also uncertainties from the exposure are taken into account. "Violin" plots around each blue point represent the flux PDF. The obtained results are also compared to the Crab Nebula SED (in orange) from Ref. [12] in terms of the relative flux difference.

the camera center as the on region. Overall effective observation time is 39.2 min.

The standard data analysis (whose results are shown in black in Fig. 8) has been performed using the MAGIC Analysis and Reconstruction Software MARS [16] where a hadronness and θ^2 cut according to a high γ-ray efficiency (90% and 75% respectively) is applied. For the BASiL analysis instead no cut is applied on the dataset. Only a global $\theta^2 < 0.08$ deg^2 is considered to define four identical nonoverlapping regions from the center of the camera: one for the on region and three for the off regions. The resulting signal rate per energy bin is reported in Table II. It is worth comparing the signal estimation by using the BASiL approach (last column of Table II) with the one we would have obtained by simply performing the difference between the total counts in the on and off regions, i.e., $N_{on} - \alpha N_{off}$ (fourth column of Table II). One can notice that the BASiL approach manages to decrease by half the uncertainty in the signal estimation. This is in agreement with the result reported in Fig. 6, where the precision of the BASiL method (~20%) is about half the one obtained in the standard approach by not cutting data in hadronness or θ^2 (~40%).

TABLE II. Estimated signal rate in each energy bin used for computing the Crab Nebula SED reported in Fig. 8. The signal rate in the last column is estimated from the mode of the signal PDF defined in Eq. (21), while its uncertainties are computed using the credible interval defined in Eq. (12). From the second to the fourth column we report the counts N_{on} and N_{off} in the on and off regions, respectively, along with their difference $N_{on} - \alpha N_{off} \pm \sqrt{N_{on} + \alpha^2 N_{off}}$, i.e., the signal-rate estimation in the frequentist approach.

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	$3.9^{+2.3}_{-1.6}$

Combined with the exposure of the telescopes the values in the last column of Table II are then used to compute the spectral energy distribution (SED) points in Fig. 8. An advantage of the BASiL approach when estimating the source flux is its capability of providing a PDF contour plot associated with each energy bin (see Fig. 8). In this way not only error bars for each flux point are drawn, but a full PDF [corresponding to the PDF in Eq. (21)] is visualized, which encodes all information we have regarding the signal estimation for that energy bin.

In Fig. 9 we report the relative uncertainties in the flux estimation using the standard (in black) and BASiL (in blue) approach. The former is computed from
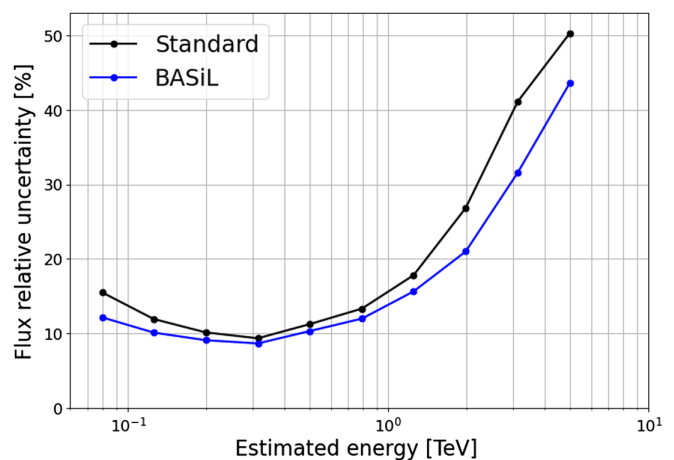

FIG. 9. Relative uncertainty in the flux estimation from Fig. 8 for the standard (black) and BASiL (blue) approach.

$$\frac{\sqrt{N_{\mathrm{on}} + \alpha^2 N_{\mathrm{off}}}}{N_{\mathrm{on}} - \alpha N_{\mathrm{off}}}, \tag{27}$$

with N_{on} and N_{off} the number of events surviving the θ^2 and hadronness cuts in the on and off regions, respectively.

The latter instead is obtained from

$$\frac{(s_{\mathrm{right}} - s_{\mathrm{left}})/2}{s^*}, \tag{28}$$

where s^* is the mode of the signal PDF and s_{right}, s_{left} are defined in Eq. (12). Uncertainties due to the exposure computation are added in quadrature, although they are negligible relative to the uncertainties in Eqs. (27) and (28). As one can see from Fig. 9, relative uncertainties in the flux estimation are smaller in the BASiL approach, especially at higher energies where the signal rate is weaker. In light of the analysis performed in the previous sections using MC simulations, this is totally expected and confirms our conclusion.

V. CONCLUSION AND OUTLOOK

In this paper we introduced a novel method for estimating the signal rate in experiments with imprecisely measured background. Common examples are astronomical measurements at high energies, in which messengers (e.g., γ rays, neutrinos) are detected on an event-by-event basis, but can equally successfully be applied in particle experiments. The BASiL method, as we dubbed it, relies on the Bayesian, rather than the more common frequentist approach. Its main feature is that it weights events according to their individual likelihood of being signal or background, considering all the information available. This weighting is best summarized by the PMF of the number of signal events in Eq. (22), in which the novelty of the method, i.e., the combinatorial term defined in Eq. (20), shows up. By doing so, BASiL avoids cutting data according to some (or a combination of) variable to suppress the background, which inevitably discards a part of the signal. Moreover, the new method, while yielding results consistent with the standard data analysis method (see Sec. IV C for a comparison on the example of Crab Nebula), improves the precision of the signal estimation, as demonstrated in Fig. 9. A convenient additional feature is a PDF contour associated

with each individual flux point. The improvement is particularly noteworthy in cases of small signal rates (see Sec. IV). Therefore, we expect BASiL to be especially useful for analysis of data from measurements of short transients or weak signals. Furthermore, certain investigations, such as searches for dark matter (see, e.g., [17–19]) or signatures of Lorentz invariance violation (LIV, see, e.g., [20–23]), base their analyses on characteristics of individual events (e.g., energy, detection time). Depending on the values of these individual characteristics, each event contributes differently to the sensitivity of the analysis. For example, in LIV searches, higher energy events contribute more to the sensitivity than events of lower energies. Standard data analysis methods, which rely on cuts to suppress the background, inevitably cut some signal events from the data sample, quite possibly the ones that would have contributed to the analysis sensitivity the most. Incorporating BASiL into analysis methods could be achieved by folding each event's contribution with its likelihood of being a signal or background event. In this way, every single event would contribute with a certain weight, increasing the analysis sensitivity. At the same time, the weights would ensure that the gain in sensitivity was not artificially created.

The corresponding data in Flexible Image Transport System format are publicly available [15].

APPENDIX A: DERIVATION OF EQ. (19)

Below one finds the derivation of Eq. (19),

$$
\begin{aligned}
p(\vec{\mathbf{x}}, N_{\mathrm{on}}, N_{\mathrm{off}} | s, b; \alpha) &= p(\vec{\mathbf{x}} | N_{\mathrm{on}}, s, \alpha b) \cdot p(N_{\mathrm{on}} | s, \alpha b) \cdot p(N_{\mathrm{off}} | b) \\
&= \prod_{i=1}^{N_{\mathrm{on}}} \left(p(\mathbf{x}_i | \gamma) \frac{s}{s + \alpha b} + p(\mathbf{x}_i | \bar{\gamma}) \frac{\alpha b}{s + \alpha b} \right) \cdot \frac{(s + \alpha b)^{N_{\mathrm{on}}}}{N_{\mathrm{on}}!} e^{-s - \alpha b} \cdot \frac{b^{N_{\mathrm{off}}}}{N_{\mathrm{off}}!} e^{-b} \\
&= \sum_{N_s=0}^{N_{\mathrm{on}}} \sum_{A \in F_{N_s}} \prod_{i \in A} p(\mathbf{x}_i | \gamma) \cdot \prod_{j \in A^c} p(\mathbf{x}_j | \bar{\gamma}) \cdot \frac{s^{N_s} (\alpha b)^{N_{\mathrm{on}} - N_s}}{(s + \alpha b)^{N_{\mathrm{on}}}} \cdot \frac{(s + \alpha b)^{N_{\mathrm{on}}}}{N_{\mathrm{on}}!} e^{-s - \alpha b} \cdot \frac{b^{N_{\mathrm{off}}}}{N_{\mathrm{off}}!} e^{-b}
\end{aligned}
$$

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

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

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

APPENDIX B: A GENERAL ALGORITHM FOR COMPUTING THE COMBINATORIAL TERM

In Sec. III we introduced the combinatorial term C defined in Eq. (20), providing an example on how to compute it when we have $N_{\mathrm{on}} = 3$ events in our on sample. We now want to show how the combinatorial term can be computed in a more general case without limiting ourselves to small count numbers. Let us assume we know the likelihoods of being a signal and a background event for each event i with $i = 1, 2, …, N_{\mathrm{on}}$. The list of likelihoods

$$
p(x_1|\bar{\gamma}), p(x_2|\bar{\gamma}), …, p(x_{N_{\mathrm{on}}}|\bar{\gamma})
$$

and

$$
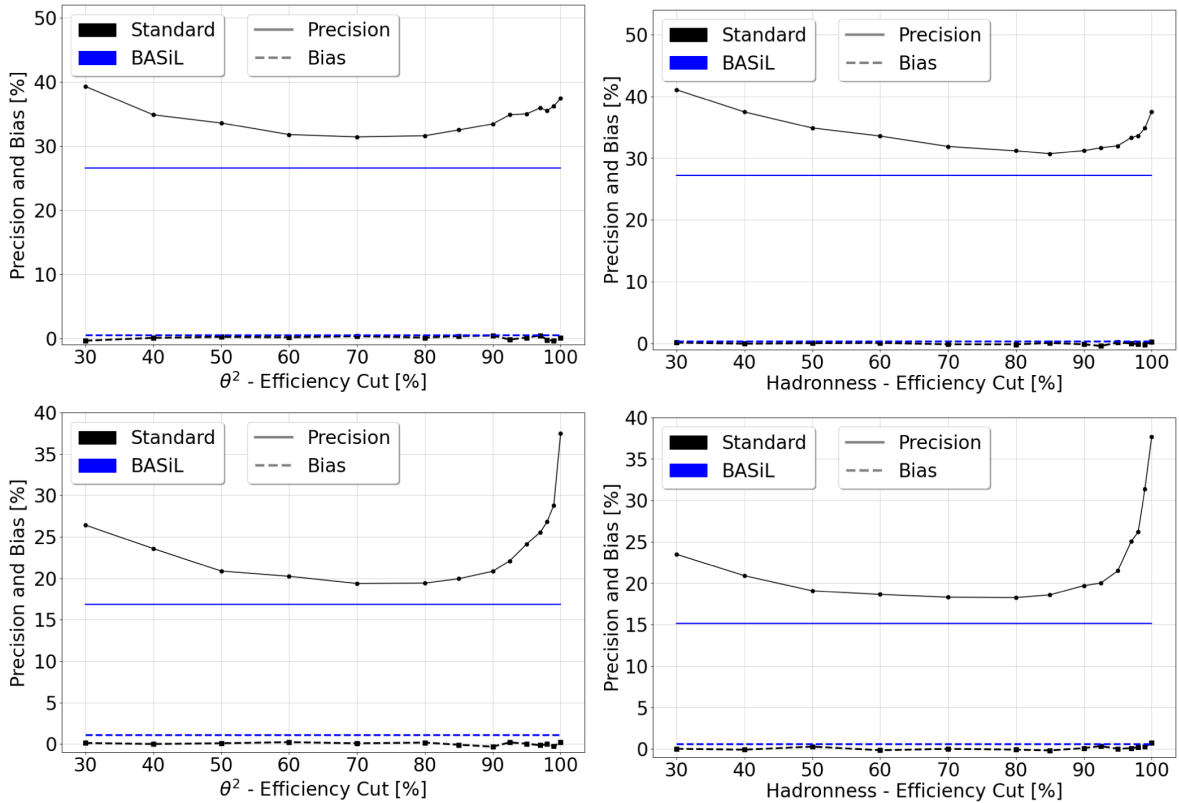p(x_1|\gamma), p(x_2|\gamma), …, p(x_{N_{\mathrm{on}}}|\gamma)
$$

FIG. 10.   Comparison between the standard (black) and BASiL (blue) approach for the evolution of the precision (full line) and bias (dashed line) in the signal-rate estimation assuming different gamma-ray efficiency cut in θ^2 (left) and hadronness (right). Note that for the BASiL approach the precision and bias do not depend on the efficiency, being $\epsilon = 1$ in such case. Nonetheless, for a visual comparison these values are shown as horizontal lines. Energy ranges considered are 75–119 GeV (top) and 754–1194 GeV (bottom). Observations are simulated assuming $s = 10^2$ and $\alpha b = 10^3$, with $\alpha = 1/3$.

can be respectively saved in two arrays. So that in one array we have all background likelihoods and in the other only signal likelihoods. It is important that the event order in both arrays must be the same. At this point an algorithm that takes as input these two arrays and provides on output the combinatorial term can be easily written as follows:

Algorithm 1. Combinatorial term.

---

input: array 1 of length n,
      array 2 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 < −array 1[i] ∗ C + array 2[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.

APPENDIX C: PERFORMANCE ON DIFFERENT REGIONS OF THE ENERGY SPECTRUM

We show in this Appendix the same analysis performed and described in Sec. IV, but considering events simulated at lower and higher energy ranges. We will focus in particular on the same energy bins used in Ref. [12], namely 75–119 and 754–1194 GeV. The most important feature that emerges by considering these two lower and higher energy bins is the fact that the signal/background separation better performs at higher energies. Such difference between low and high energy bins is caused by the fact that in IACTs the higher the energy of an event the larger and better its camera image will be.

In Fig. 10 we report the precision and bias for different efficiency cuts applied in θ^2 and hadronness. Similar conclusions made in Sec. IV for the medium energy bin also apply here: (i) the BASiL approach is capable of improving the signal-estimation precision by ∼15% in both energy bins, and (ii) the bias is always below the precision and close to zero. It is also worth noticing that, as expected, the precision increases as we go higher in energy.

Finally the study on the signal-estimation precision and bias for different SNR is reported in Fig. 11 where again similar conclusions of Sec. IV apply. The improvement in
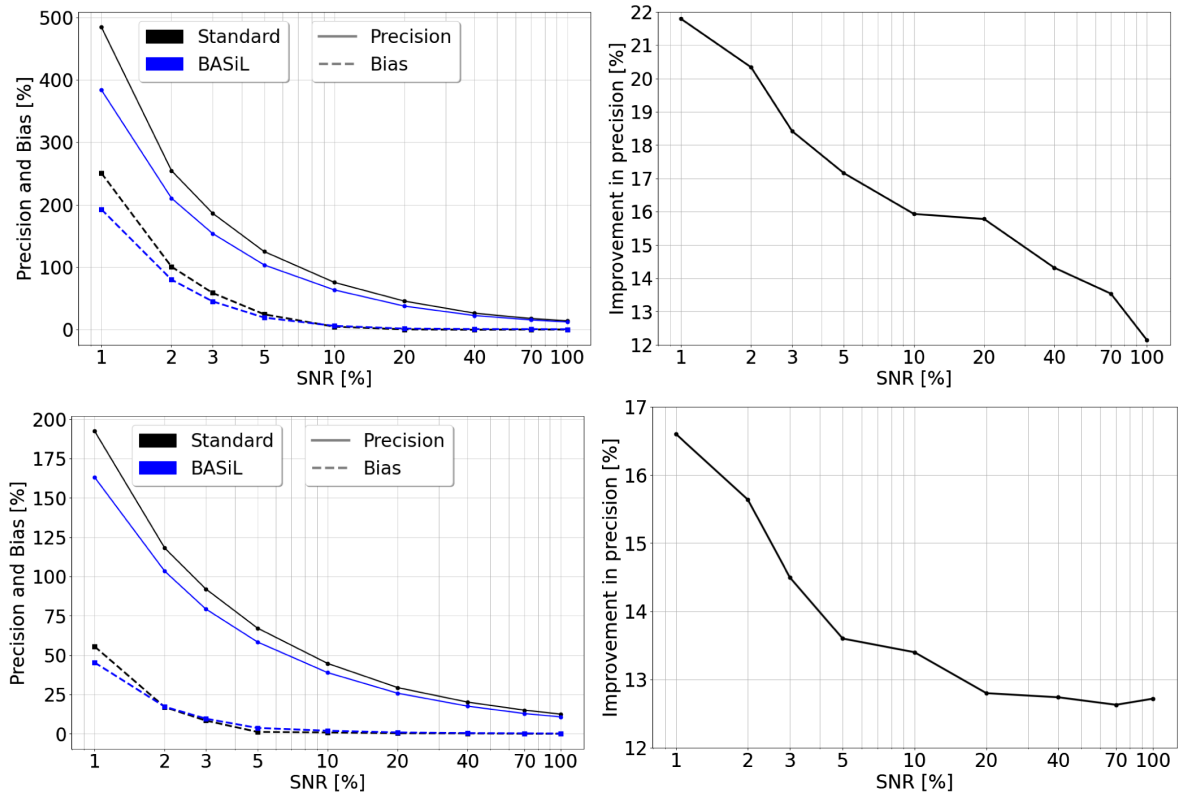

FIG. 11. Left: comparison between the standard (black) and BASiL (blue) approach for the evolution of the precision (full line) and bias (dashed line) assuming different SNR. In the standard approach the efficiency cut is fixed to 90% for hadronness and 75% for θ^2. Right: improvement for different SNR of the precision in the BASiL approach relative to the standard one. Energy ranges considered are 75–119 GeV (top) and 754–1194 GeV (bottom). Observations are simulated assuming $\alpha b = 100$, with $\alpha = 1/3$.
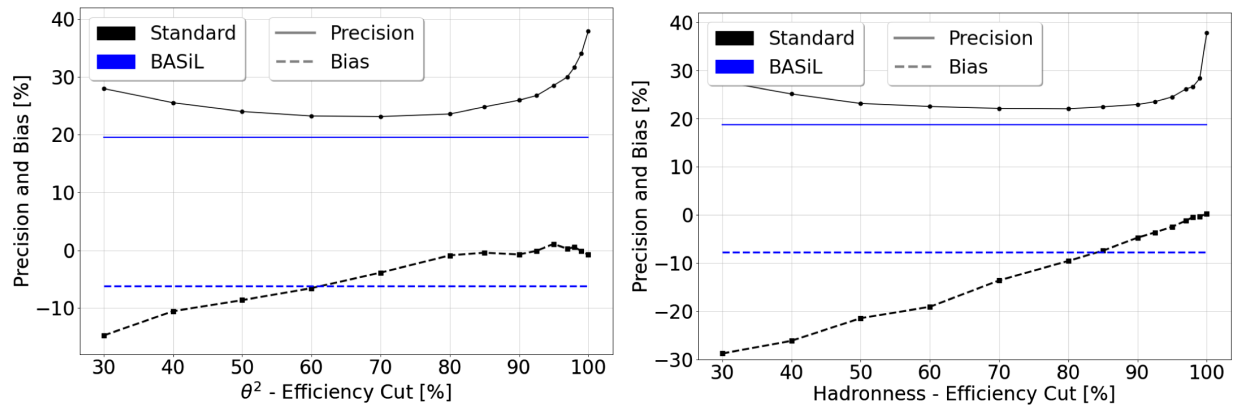
FIG. 12.   Same as Fig. 6 but with the inclusion of the mismatch between MC and real signal population, respectively blue and red histograms of Fig. 4.

the precision of the new approach increases as the SNR becomes smaller. Such improvement is more pronounced in the lower energy bin (compare the upper and bottom right plots of Fig. 11). This is due to the fact that in the higher energy bin the distinction between signal and background is more accurate and therefore signal-extraction cuts allow one to remove almost all of the background while losing very few signal events.

We conclude that the BASiL approach is capable of increasing the precision of the signal estimation in all energy ranges and that such improvement becomes more important when the observations are background dominated.

APPENDIX D: EFFECT OF THE MISMATCH BETWEEN MC AND REAL SIGNAL EVENTS

In Sec. IV we have studied the precision and bias in the estimation of the signal rate by simulating on/off measurements from the MAGIC telescopes. This estimation was performed using the standard and BASiL approach, in which the mismatch from the MC and the real-gamma population (respectively the blue and red histograms of Fig. 4) was ignored. We now want to study how such

mismatch can affect the signal estimation. In order to do so we are going to repeat the analysis reported in Sec. IV A, but this time when it comes to estimating N_s only the distributions from MC-γ events are considered: the real-γ distributions are only used in the simulation stage. The results of this analysis are reported in Fig. 12. By comparing this figure with Fig. 6, one can see that the precision is approximately the same. The main difference, as expected, comes from the bias which is no longer close to zero in both approaches. The bias resulting from the mismatch between MC and signal events increases as we cut more events. It is also more pronounced for the hadronness case: an explanation of this can be found in Fig. 4, where one can see that the mismatch between MC and signal events is more pronounced for the hadronness case. The BASiL approach produces a bias that is roughly equal to the one obtained in the standard approach by performing a cut in θ^2 and hadronness with an efficiency around 50%–60% and 70%–80%, respectively. It is also interesting to notice that such bias in both approaches is negative (the excess is underestimated), but anyway smaller in absolute value than the precision.

[1] D. Berge, S. Funk, and J. Hinton, Astron. Astrophys. **466,** 1219 (2007).

[2] T. J. Loredo, in *Statistical Challenges in Modern Astronomy* (Springer, New York, 1992), pp. 275–297.

[3] T. P. Li and Y. Q. Ma, Astrophys. J. **272,** 317 (1983).

[4] W. A. Rolke, A. M. López, and J. Conrad, Nucl. Instrum. Methods Phys. Res., Sect. A **551,** 493 (2005).

[5] J. R. Mattox, D. Bertsch, J. Chiang, B. Dingus, S. Digel, J. Esposito, J. Fierro, R. Hartman, S. Hunter, G. Kanbach *et al.*, Astrophys. J. **461,** 396 (1996).

[6] S. Wilks, Ann. Math. Stat. **9,** 60 (1938).

[7] M. L. Knoetig, Astrophys. J. **790,** 106 (2014).

[8] G. D'Agostini, arXiv:physics/9811045.

[9] J. Albert, E. Aliu, H. Anderhub, P. Antoranz, A. Armada, M. Asensio, C. Baixeras, J. A. Barrio, H. Bartko, D. Bastieri *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **588,** 424 (2008).

[10] J. Conrad, Astropart. Phys. **62,** 165 (2015).

[11] https://magic.mpp.mpg.de/.

[12] J. Aleksić, S. Ansoldi, L. A. Antonelli, P. Antoranz, A. Babic, P. Bangale, M. Barceló, J. A. Barrio, J. Becerra González, W. Bednarek *et al.*, Astropart. Phys. **72,** 76 (2016).

[13] P. Da Vela, A. Stamerra, A. Neronov, E. Prandini, Y. Konno, and J. Sitarek, Astropart. Phys. **98**, 1 (2018).

[14] V. P. Fomin, A. A. Stepanian, R. C. Lamb, D. A. Lewis, M. Punch, and T. C. Weekes, Astropart. Phys. **2**, 137 (1994).

[15] T. Hassan and C. Nigro, Github, https://github.com/open-gamma-ray-astro/joint-crab/tree/master/data/magic, https://github.com/open-gamma-ray-astro/joint-crab/tree/master/data/magic.

[16] R. Zanin, E. Carmona, J. Sitarek, P. Colin, K. Frantzen, M. Gaug, S. Lombardi, M. Lopez, A. Moralejo, K. Satalecka *et al.*, in *International Cosmic Ray Conference*, Vol. 33 of International Cosmic Ray Conference (Curran Associates, Inc., Red Hook, 2013), p. 2937.

[17] M. L. Ahnen, S. Ansoldi, L. A. Antonelli, C. Arcaro, D. Baack, A. Babić, B. Banerjee, P. Bangale, U. Barres de Almeida, J. A. Barrio *et al.*, J. Cosmol. Astropart. Phys. 03 (2018) 009.

[18] V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, C. Arcaro, D. Baack, A. Babić, B. Banerjee, P. Bangale,

U. Barres de Almeida *et al.*, Phys. Dark Universe **22**, 38 (2018).

[19] V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González *et al.*, Phys. Dark Universe **28**, 100529 (2020).

[20] J. Albert, E. Aliu, H. Anderhub, L. A. Antonelli, P. Antoranz, M. Backes, C. Baixeras, J. A. Barrio, H. Bartko *et al.* (MAGIC Collaboration), Phys. Lett. B **668**, 253 (2008).

[21] M. Martínez and M. Errando, Astropart. Phys. **31**, 226 (2009).

[22] M. L. Ahnen, S. Ansoldi, L. A. Antonelli, C. Arcaro, A. Babić, B. Banerjee, P. Bangale, U. Barres de Almeida, J. A. Barrio *et al.* (MAGIC Collaboration), Astrophys. J. Suppl. Ser. **232**, 9 (2017).

[23] V. A. Acciari, S. Ansoldi, L. A. Antonelli, A. Arbet Engels, D. Baack, A. Babić, B. Banerjee, U. Barres de Almeida, J. A. Barrio, J. Becerra González *et al.*, Phys. Rev. Lett. **125**, 021301 (2020).