PROCEEDINGS OF SCIENCE

Hands on XENON100 Dark Matter (DM) direct detection experiment: studying and modeling background and signal in a frequentist analysis framework

Bárbara-Rosario Montes Núñez
CIEMAT, Avda. Complutense, 22, E-28040, Madrid, Spain
E-mail: barbara.montes@ciemat.es

Brett Cornell
Caltech, 1200 E California Blvd, MC 367-17, Pasadena, CA 91125 USA
E-mail: cornell@caltech.edu

Alfredo Davide Ferella
INFN-Laboratori Nazionali del Gran Sasso, Assergi, 67010, Italy
E-mail: alfredo.ferella@lngs.infn.it

In this activity we try to reproduce the XENON100 direct dark matter search results from 224 live days of data taken between February 2011 and March 2012. The statistical approach to model signal and background in the profile likelihood analysis is presented and the procedure used by the XENON100 collaboration to interpret the results is partially implemented.

Gran Sasso Summer Institute 2014 Hands-On Experimental Underground Physics at LNGS - GSSI14, 22 September - 03 October 2014
INFN - Laboratori Nazionali del Gran Sasso, Assergi, Italy

PoS(GSSI14)017

In order to determine the nature of dark matter and discriminate among the large number of candidates and models, a great experimental effort in direct dark matter detection is currently underway. One such experiment which is operated in the Gran Sasso laboratory is XENON100, a double phase (liquid and gas) xenon time projection chamber [1]. The latest results published by XENON100, presented in figure 1 (left) [2], set the most sensitive limit for the spin independent cross section of WIMP-nucleon scattering at the time of publication. What follows is a brief study of the data handling, modeling and statistical methods used in that publication.

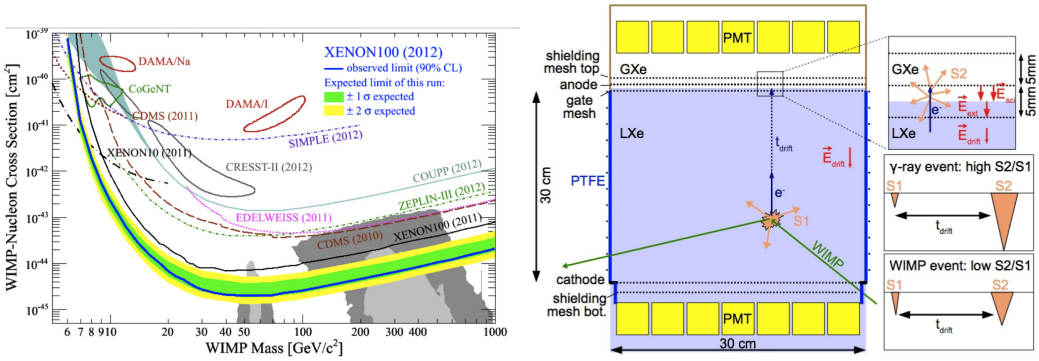

Figure 1: (Left) Latest XENON100 results from 2012 that set one of the best experimental limits on WIMP-nucleon cross section. (Right) Scheme of the WIMP detection in XENON100 experiment. Nuclear recoils are distinguished form electronic recoils by considering the ratio of secondary and primary scintillation light (S2/S1).

In XENON100, the detector operates by measuring the energy deposited when WIMPs elastically scatter from ordinary target xenon nuclei as it is shown in figure 1 (right). These recoils produce ionization and excitation of the other Xe atoms composing the sensitive volume. In the de-sexcitation process of these atoms, photons (primary scintillation light) and electrons are emitted. The primary scintillation photons (S1) propagate within the detector and are detected by two arrays of photomultiplier tubes (PMTs) placed on top and bottom of the detector, producing the S1 signal. Electrons are drifted upwards by an electric field and extracted from liquid to gas phase producing a secondary scintillation light (S2) proportional to the extracted ionization electrons. The ratio of the secondary to primary scintillation components depend on particle interaction, namely if the interaction occurred in the Xe nucleus or with one of the electrons. WIMP-electron interaction are kinematically very unfavorable. As a result, WIMP signals in XENON100 are expected to consist entirely of nuclear recoil interactions, while most backgrounds (such as γ's) will interact predominantly with the atomic electrons. This different detector response to nuclear and electronic recoils then essentially allows for event by event discrimination between signal and background.

This event by event discrimination is the bedrock of XENON100's analysis technique. We use calibration data (AmBe for nuclear recoils and Co60 gamma source for electronic recoils) in order to build a signal and background model, and science data taken from February 2011 and March 2012. The approach used here is described in more details in [3].

In figure 2 we show the nuclear recoil (NR), electronic recoil (ER) and dark matter (DM) distributions in the "linearized" log (S2/S1) (here called delta) vs S1 (in photoelectrons) parameter space.

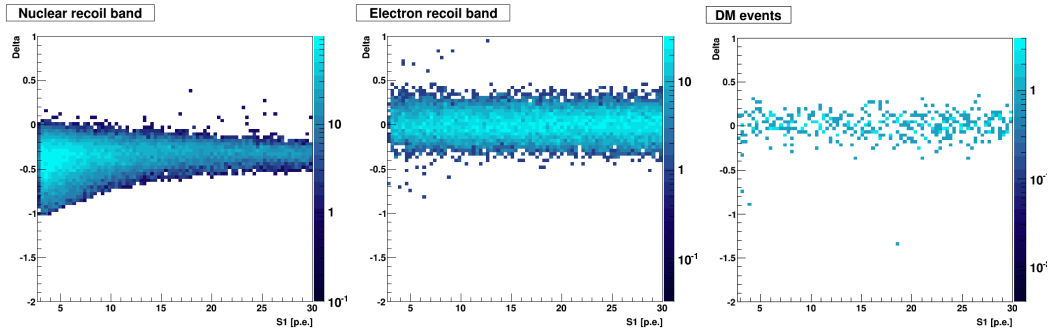

Figure 2: (Left) Nuclear recoil (NR), (center) electronic recoil (ER) and (right) dark matter (DM) distribution plots. In the plots we represent delta ("linearized" log of the ratio S2/S1) vs S1 signal in photoelectrons. Clear differences in shape and position are observed between NR and ER.

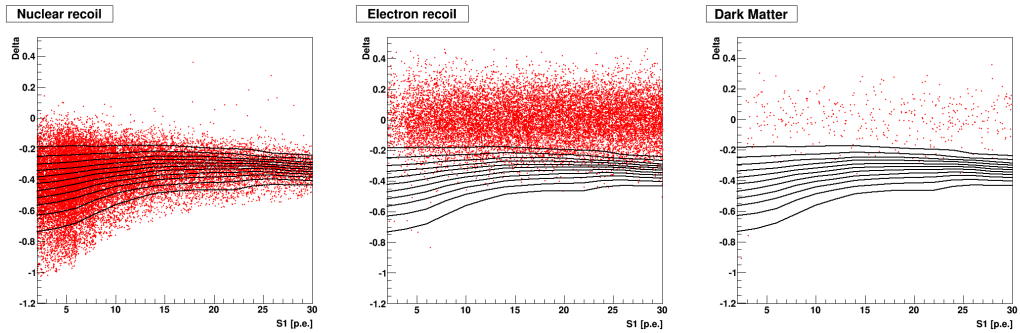

Figure 3: Modelling of the signal and background distributions by constructing delta bands (left). There are the same number of signal events in each band. The same bands are applied to the ER (center) and science data (right) distributions.

The first goal of this study is to calculate the number of signal events that maximizes the likelihood describing our science data. Therefore, we construct 12 bands in the delta vs S1 parameter space in order to have the same number of NR events in each band as it is shown in figure 3 (left). We then apply these bands to the ER, figure 3 (center), and science data, figure 3 (right). This adaptive binning model allows us to have consistent signal statistics per bin.

We can then calculate how many events we will have in each band for the different data sets (figure 4). The signal distribution (NR) is flat (by construction) compared to background (ER) and science data (DM).

Assuming gaussian errors, the number of events in each bin should follow a Poisson distribution. The mean of this distribution in principle varies bin by bin, since we are only interested in finding the maximum of the likelihood function and minimizers are more straightforward from an implementation standpoint than maximizers, we will focus on minimizing the negative log of this likelihood function (*nLL*). Given a set of n_i science data points in a certain bin i, shown in figure 4 (right), *nLL* is defined as follows:

$$nLL_i(N_s) = -n_i log\theta_i - \theta_i \tag{1}$$

The likelihood function depends on the expectation value of our signal and background in each

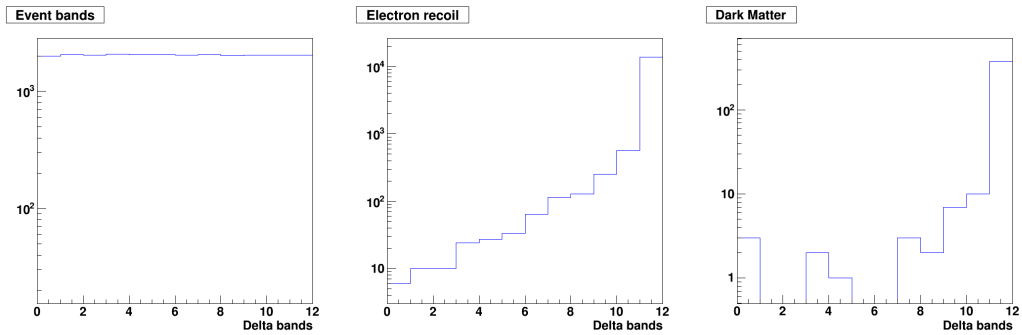

Figure 4: Number of events in each band. (Left) NR events show a flat distribution while for ER (center) and science data (right) the number of events in each bin follows a Poisson distribution.

bin, θ_i, which is function of the number of signals, N_s. The signal and background distributions are also required and they will be calculated from our histograms (figure 4).

$$\theta_i(N_s) = N_s \cdot \varepsilon_s^i + N_b \cdot \varepsilon_b^i \tag{2}$$

The signal and background distributions are calculated as the number of events in the band normalized to total number of events:

$$\varepsilon_s^i = \frac{NR_i}{\sum_i NR_i} = \frac{NR_i}{NR} \tag{3}$$

$$\varepsilon_b^i = \frac{ER_i}{\sum_i ER_i} = \frac{ER_i}{ER} \tag{4}$$

The number of background events is obtained from data subtracting the number of signals:

$$N_b = Data - N_s = \sum_i Data_i - N_s \tag{5}$$

Taking all these aspects into account, we obtain the *nLL* as a function of N_s as shown in figure 5 (left) and calculate the minimum of the function, which corresponds to $N_s = 3.685$.

This, however, assumes that our histograms constructed from our calibration exactly follow our true background and signal distributions. This is not exactly true as poisson fluctuations will make the number events in each bin differ from the actual expectation value. In order to take this into account we introduce the nuisance parameters for both signal and background:

$$\varepsilon_s = [\varepsilon_s^1 \cdots \varepsilon_s^{12}] \tag{6}$$

$$\varepsilon_b = [\varepsilon_b^1 \cdots \varepsilon_b^{12}] \tag{7}$$

where ε is the occupancy probability for each bin. We are going to have one nuisance parameter for each bin, which means 12 parameters for signal and 12 for background, that account for the probability of having signal/background in that particular bin. Consequently, the expectation value is defined as:

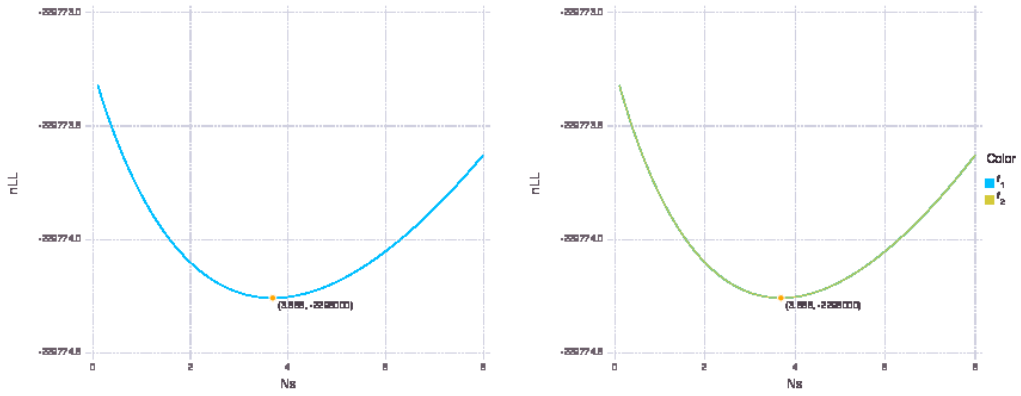

Figure 5: (Left) Likelihood plotted as function of signal events, N_s and calculated minimum. (Right) Comparison of both analysis before and after introducing nuisance parameters obtaining same results.

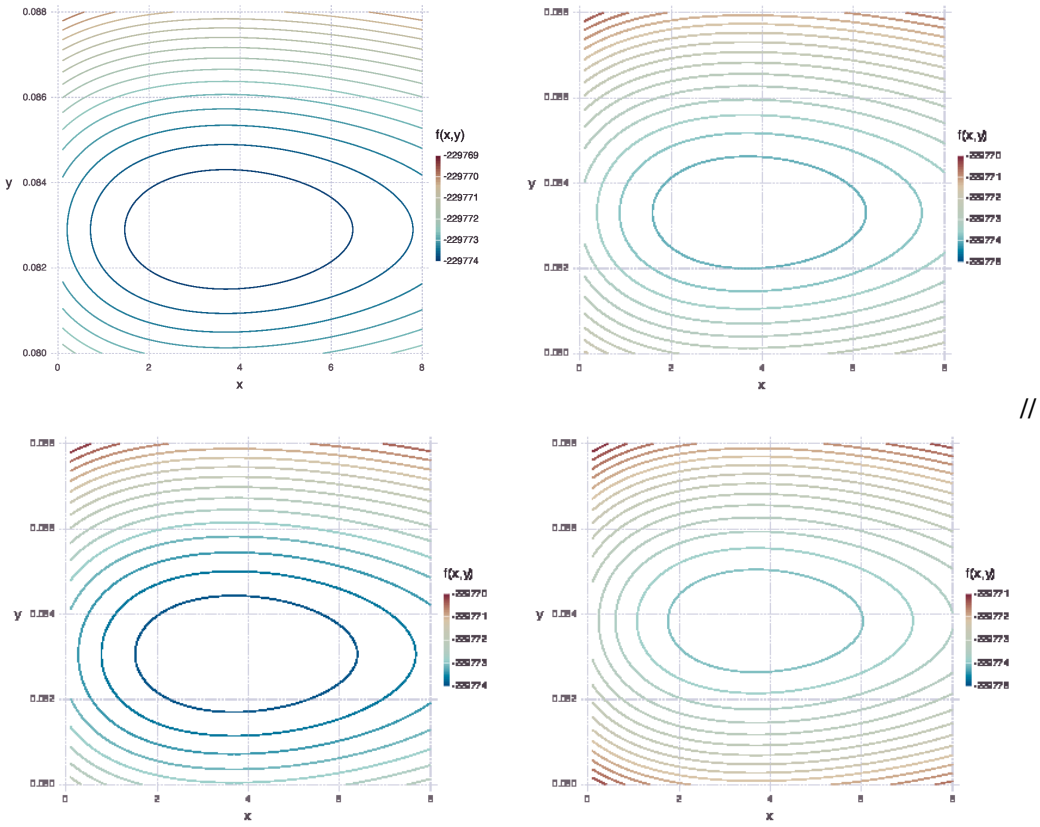

Figure 6: Contour plots showing probability of having signal in a bin versus number of signals for four different nuisance parameters.

$$\theta_s^i = \varepsilon_s^i \cdot NR \tag{8}$$

Considering these nuisance parameters we define the negative log likelihood:

$$nLL(N_s) \rightarrow nLL(N_s, \varepsilon_s, \varepsilon_b) = nLL(N_s) + nLL(\varepsilon_s) + nLL(\varepsilon_b) \tag{9}$$

$$nLL_i(\varepsilon_s^i) = -n_s^i log\theta_s^i - \theta_s^i \tag{10}$$

Our goal is therefore to calculate the expected number of signal events, profiling out our nuisance parameters. We calculate the minimum of the *nLL* for constant value of number of signals, N_s.

$$nLL(N_s, \varepsilon_s, \varepsilon_b) \rightarrow profLL(N_s) = nLL(N_s, \hat{\varepsilon}_s, \hat{\varepsilon}_b) \tag{11}$$

where $\hat{\varepsilon}_s, \hat{\varepsilon}_b$ =min(nLL) at constant N_s

$$nLL(N_s, \varepsilon_s = \varepsilon_{smin}, \varepsilon_b = \varepsilon_{bmin}) \tag{12}$$

The result is plotted on 5 (right), where we show the analysis before and after introducing the nuisance parameters. The results are very similar, which we explore in figure 6, where we have the number of signals plotted on the x axis and the probability of having a signal in a particular bin on the y axis for four nuisance parameters. There is no correlation between both number of signals and probability for any of them and, for this reason, we have obtained the same results from both analyses.

The total likelihood function considering all nuisance parameters and systematic errors from XENON100's profile likelihood analysis can be split into four different terms: poisson on data (per band), distributions of events in each band, poisson on calibration data and light scintillation efficiency (L_{eff}) penalty. This last term accounts for the deviation of L_{eff} from the median and it has not been included in the present study. In order to apply the described procedure to future work, the present analysis should be extended with the distribution of events in each band.

In summary, during our activity at the Gran Sasso Summer Institute, we put our hands on XENON100's profile likelihood procedure, and we started the corresponding calculation with actual XENON100 data. In spite of the fact that, due to our very limited time constraints, the calculation could not be completed, we are now very familiar with the actual implementation of an analysis of this type. This knowledge that we acquired in the Summer Institute will certainly be applied to future analysis in our own experimental framework.

References

[1] E. Aprile *et al.* [XENON100 Collaboration], Astropart. Phys. **35** (2012) 573 [arXiv:1107.2155 [astro-ph.IM]].

[2] E. Aprile *et al.* [XENON100 Collaboration], Phys. Rev. Lett. **109** (2012) 181301 [arXiv:1207.5988 [astro-ph.CO]].

[3] E. Aprile *et al.* [XENON100 Collaboration], Phys. Rev. D **84** (2011) 052003 [arXiv:1103.0303 [hep-ex]].

[4] G. Cowan, K. Cranmer, E. Gross and O. Vitells, Eur. Phys. J. C **71** (2011) 1554 [Erratum-ibid. C **73** (2013) 2501] [arXiv:1007.1727 [physics.data-an]].