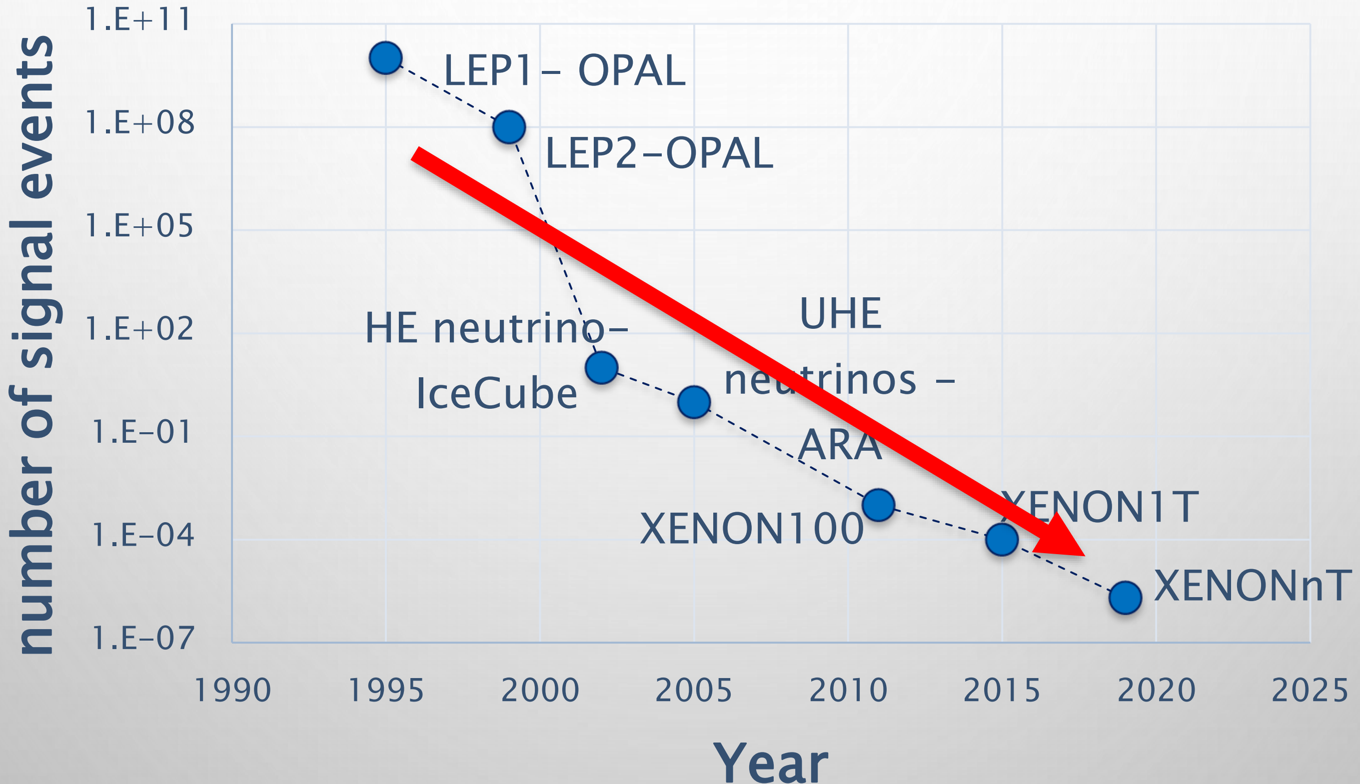
LIMIT SETTING: EVOLUTION VS. INTELLIGENT DESIGN

Hagar Landsman
Weizmann Institute of science

MY CAREER


number of signal events
1.E+11
1.E+08
1.E+05
1.E+02
1.E-01
1.E-04
1.E-07
1990
1995
2000
2005
2010
2015
2020
2025
Year
LEP1- OPAL
LEP2-OPAL
HE neutrino-
IceCube
UHE
neutrinos -
ARA
XENON100
XENON1T
XENONnT

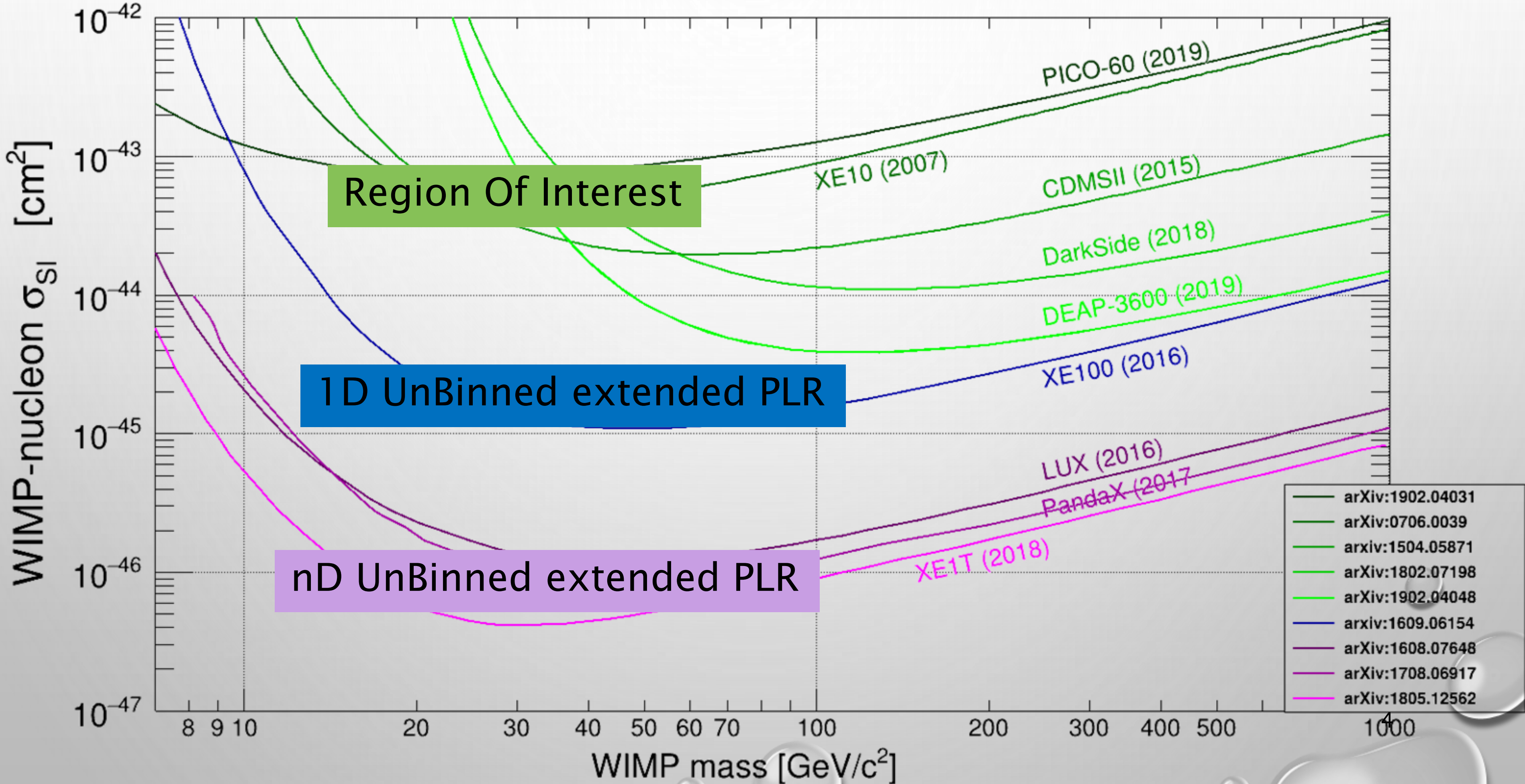
The Evolution of Limits Setting

Region Of Interest

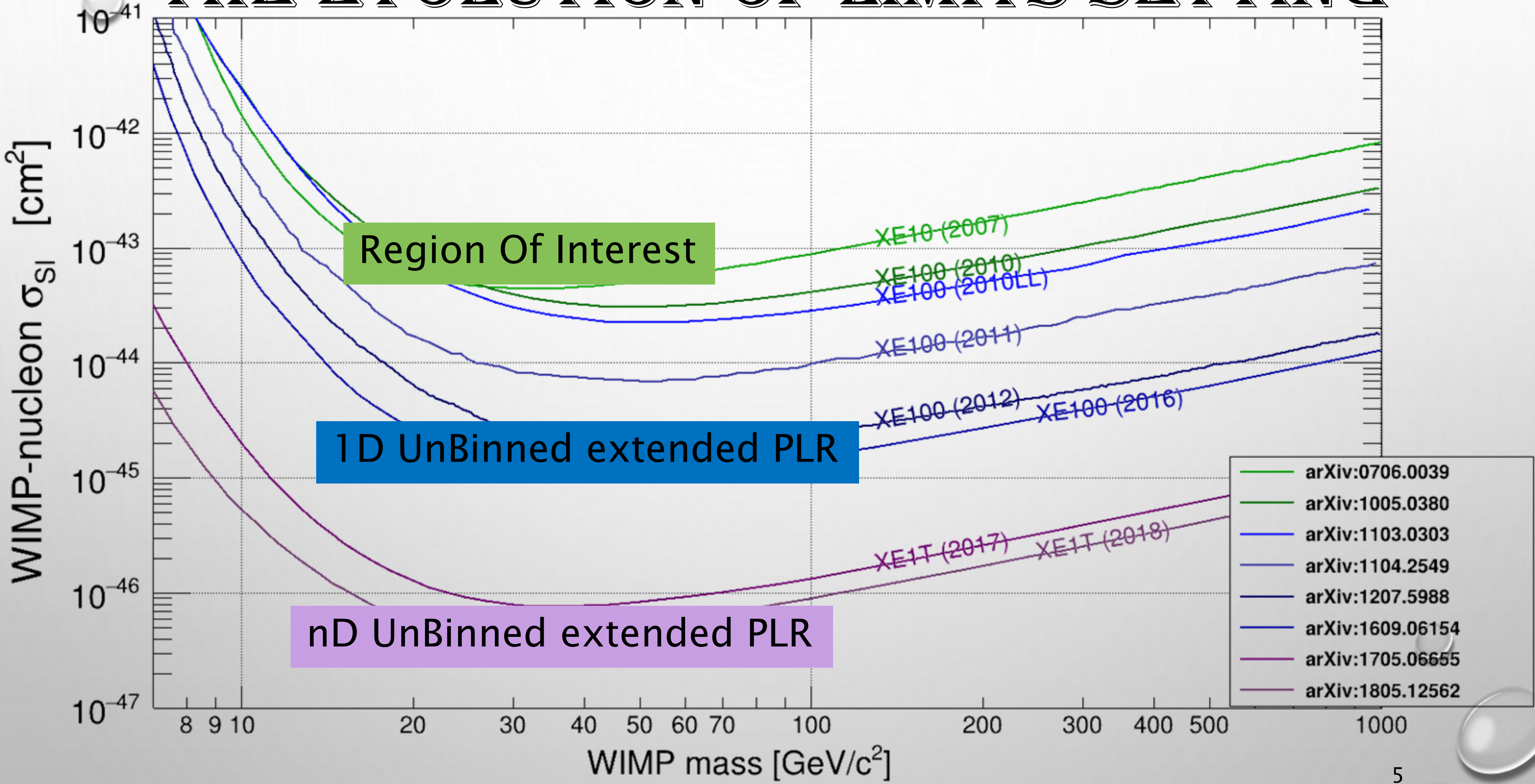
1D UnBinned extended PLR

nD UnBinned extended PLR

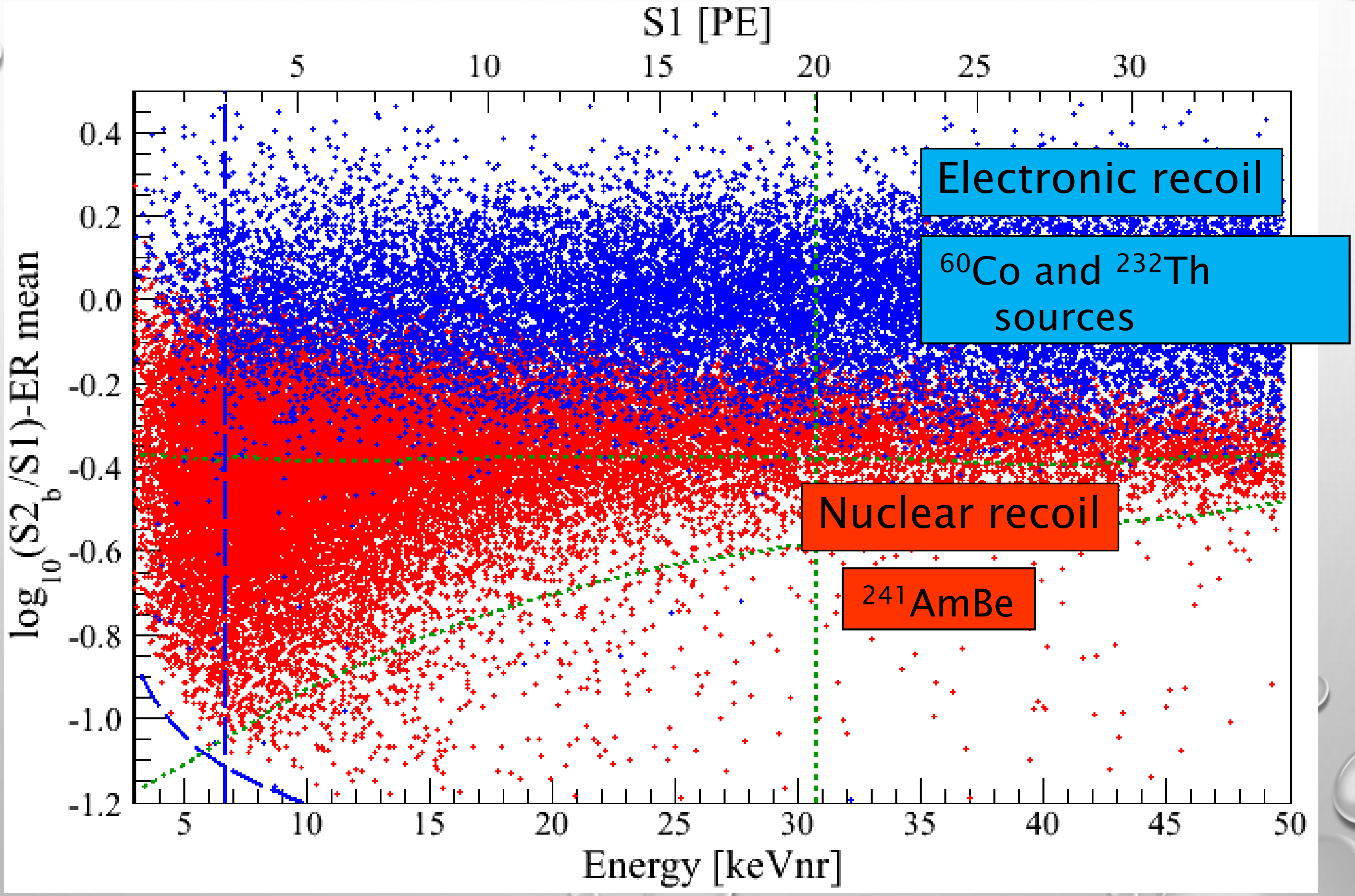
THE EVOLUTION OF LIMITS SETTING



XENON
The Evolution of Limits Setting



Detection principle
Discrimination variables



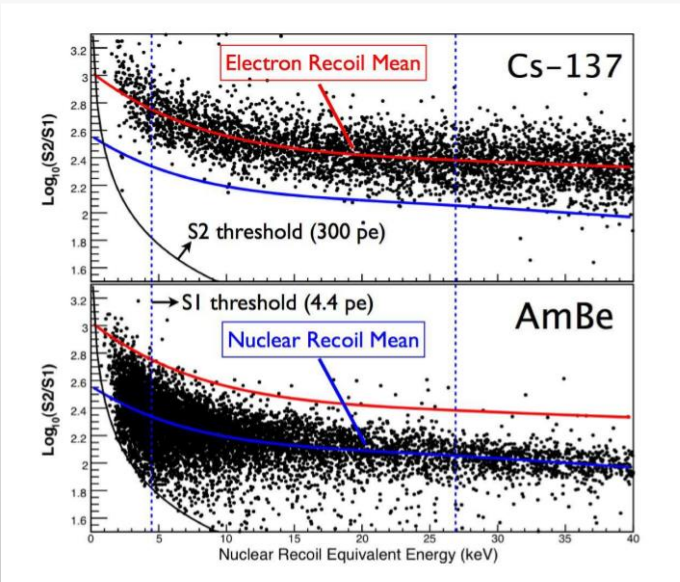
PRL **100**, 021303 (2008) PHYSICAL REVIEW LETTERS

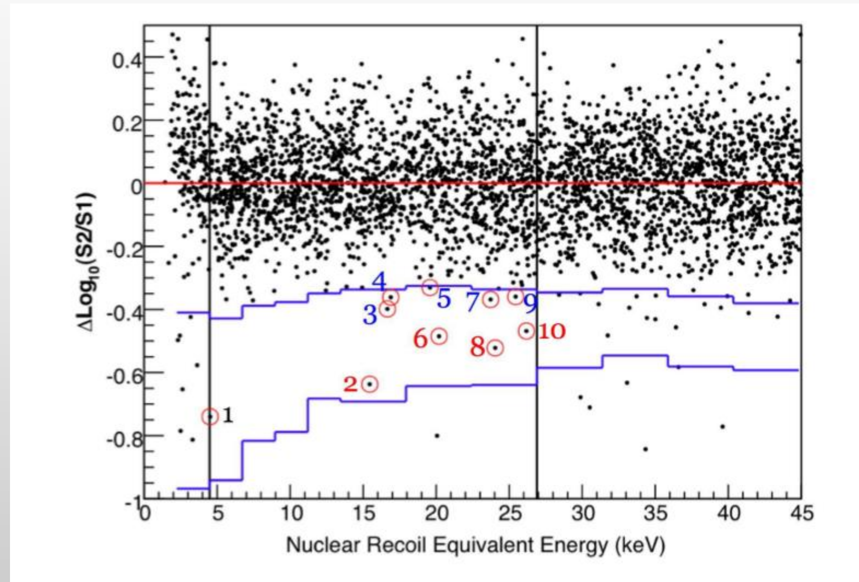
First Results from the XENON10 Dark Matter Experiment at the Gran Sasso National Laboratory

J. Angle,[1,2] E. Aprile,[3,*] F. Arneodo,[4] L. Baudis,[2] A. Bernstein,[5] A. Bolozdynya,[6] P. Brusov,[6] L. C.
C. E. Dahl,[6,8] L. DeViveiros,[9] A. D. Ferella,[2,4] L. M. P. Fernandes,[7] S. Fiorucci,[9] R. J. Gaitskell,[9] K.
R. Gomez,[10] R. Hasty,[11] L. Kastens,[11] J. Kwong,[6,8] J. A. M. Lopes,[7] N. Madden,[5] A. Manalaysay,[1,2]
D. N. McKinsey,[11] M. E. Monzani,[3] K. Ni,[11] U. Oberlack,[10] J. Orboeck,[2] G. Plante,[3] R. Santorelli,[3] J. M
P. Shagin,[10] T. Shutt,[6] P. Sorensen,[9] S. Schulte,[2] C. Winant,[5] and M. Yamashita[3]

Region Of Interest

XENON10 2005-2007



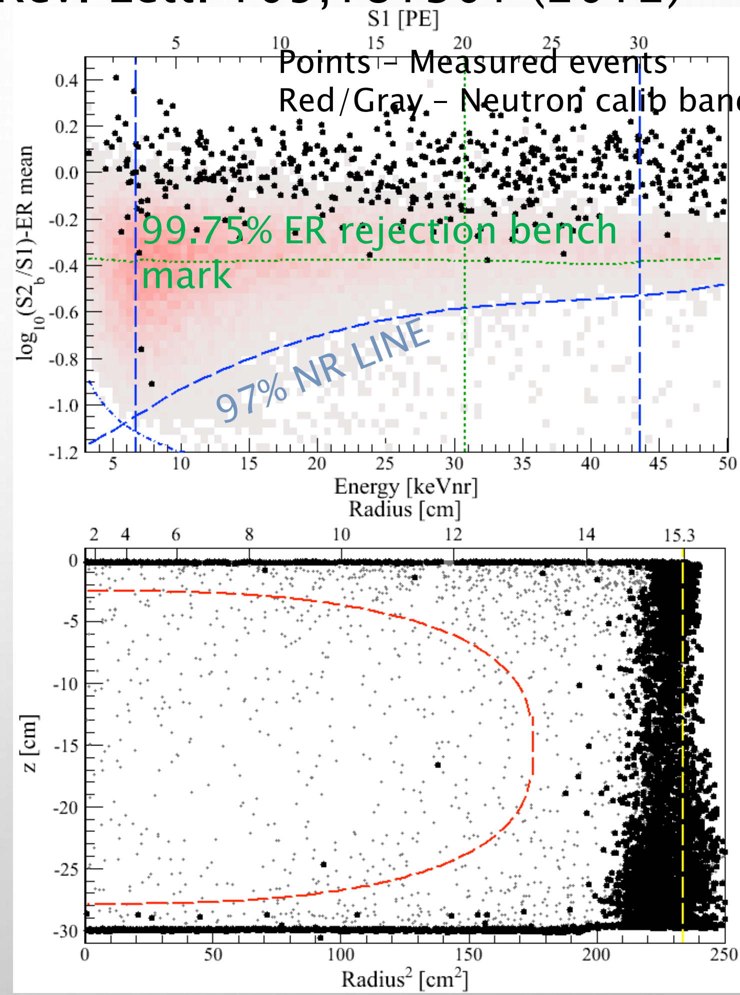


"However, the uncertainty of the estimated number of leakage events for each energy bin in the analysis of the WIMP search data is currently limited by available calibration statistics. Based on the analysis of multiple scatter events, no neutron induced recoil event is expected in the single scatter WIMP-search data set. To set conservative limits on WIMP-nucleon spin-independent cross section, we consider all ten observed events, with no background subtraction."

What do we learn from a TPC event?

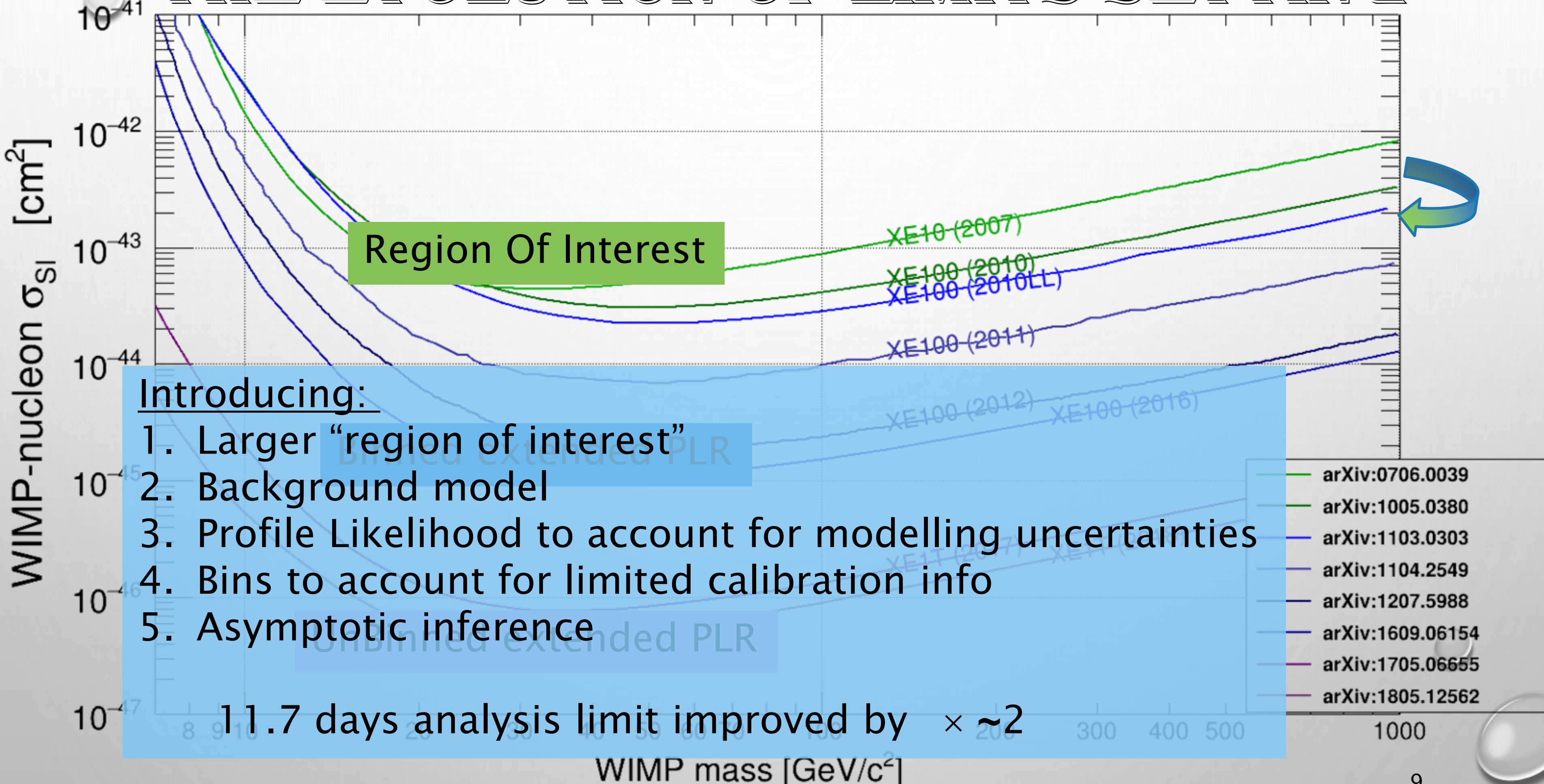
Illustrated by XENON100 2011/2012 data set
225 Live days   Phys. Rev. Lett. 109,181301 (2012)

- s1,s2:
  - Energy scale
  - Discrimination: ER vs. Nr (s1/s2)
- Vertex reconstruction
  - Fiducialization
  - Single vs. Multiple scatters
- Waveforms
- Event epoch time
- Slow control (detector stability)



Bold – below 99.75 ER line (likely to be NR)
dots – above 99.75 ER line (likely to be ER)

XENON
THE EVOLUTION OF LIMITS SETTING



Region Of Interest

Introducing:

1. Larger “region of interest”
2. Background model
3. Profile Likelihood to account for modelling uncertainties
4. Bins to account for limited calibration info
5. Asymptotic inference

11.7 days analysis limit improved by $\times$ ~2

THE LIKELIHOOD FUNCTION

Extended unbinned $\mathcal{L}$ — The guardians

[Poisson term] × [Model term] × [Ancillary term] × [Calib term]

$$\mathcal{L}_{ub} = \mathrm{Poiss}(N|N_s+N_b)\prod_{i=1}^{N}\frac{N_s f_s(x_i)+N_b f_b(x_i)}{N_s+N_b}.$$

$N_s\ (\sigma;\bar{\theta}_s, \bar{\theta}_g)$

$N_b^j\ (\bar{\theta}_b, \bar{\theta}_g)$

n_{data}

$f_s\ (\bar{x}|\bar{\theta}_s, \bar{\theta}_g)$

$f_b^j(\bar{x}|\bar{\theta}_b, \bar{\theta}_g)$

e. g.

$\mathrm{Gaus}\left(\hat{N}_b^j \middle| N_b^j, \sigma^j\right)$

$\mathrm{Gaus}\left(\hat{\theta}\ \middle|\theta, \sigma_\theta\right)$

- Long list of observables x: S_1, S_2, (R,z,θ), t
- Long list of parameters: $\bar{\theta}_s, \bar{\theta}_g, \bar{\theta}_b$
  Some are correlated, some are not…

THE LIKELIHOOD FUNCTION

Extended unbinned $\mathcal{L}$ — The guardians

Poisson term × Model term × Ancillary term × Calib term

- Long list of observables x: S_1, S_2, (R,z,θ), t
- Long list of parameters: $\bar{\theta}_s, \bar{\theta}_g, \bar{\theta}_b$

Three choices:

1. Ignore –
   That's easy to implement
2. Binned model –
   Bins in discrimination space ("bands")
   Spatial bins (r,z,q)
   Temporal bins (E variations, background conditions, "runs")
3. Unbinned Model
   Higher dimensions for f_s, f_b
   Add nuisance parameters

$$\mathcal{L} = \mathcal{L}^{I} \times \mathcal{L}^{II} \times \mathcal{L}^{III}$$

THE LIKELIHOOD FUNCTION

Extended unbinned $\mathcal{L}$ — The guardians

[Poisson term × Model term] × [Ancillary term × Calib term]

Some (hopefully) good reasons to take it slowly:

➔Limited knowledge – risk of under/over coverage

- Limited calibration
- Lack of model
- Always risk of mis-modeling

➔Not needed

- The additional information / resolution is not needed

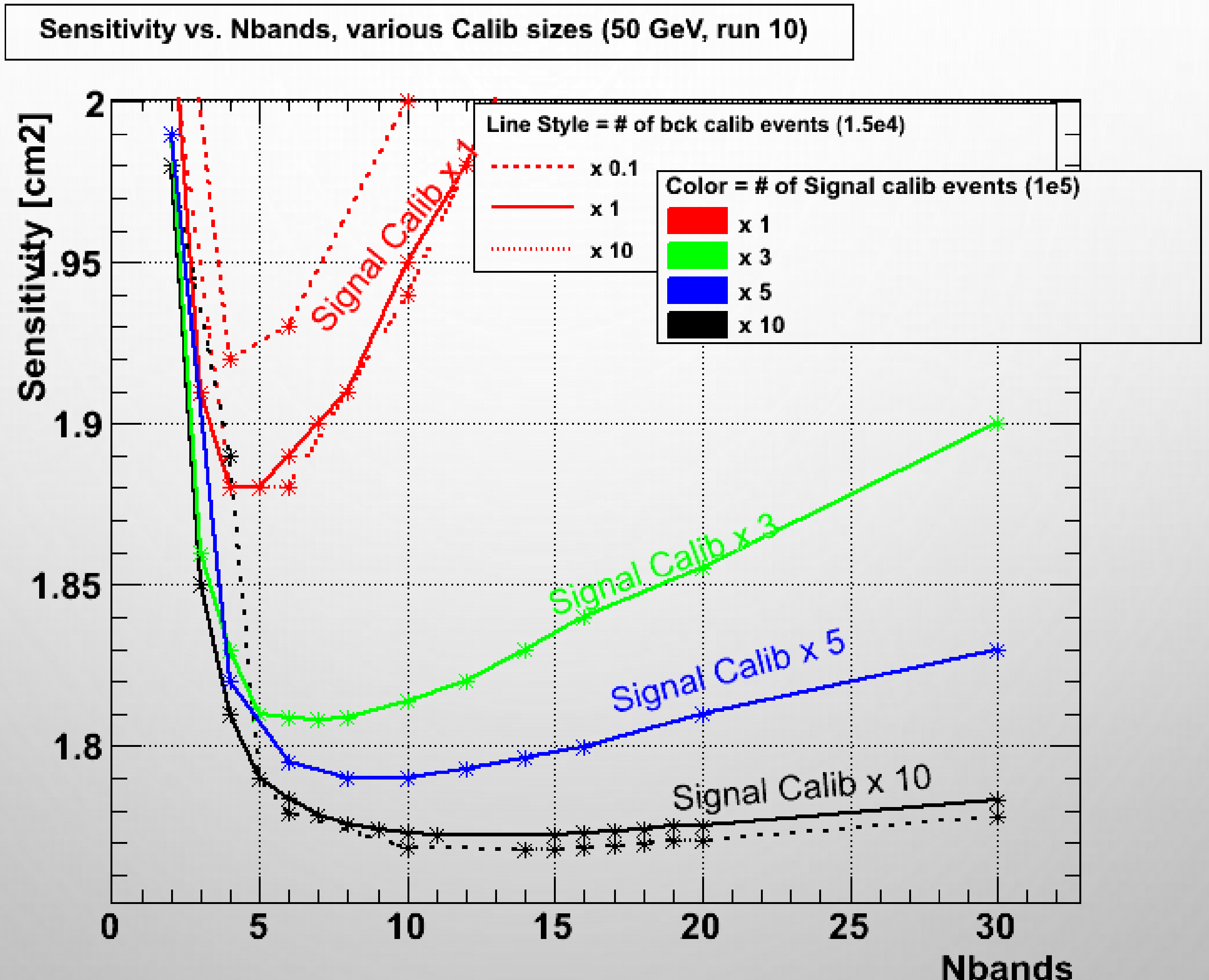
➔Save on resources

- Modeling and minimizing. Asymptoticness (checks or bypass)

  Required cpus, diskspace, people, nerves, sanity

Use the power of the guardians wisely!

How many bins to use?



XENON's 1st LIKELIHOOD FUNCTION

Parameter of interest:
- Ns – total number of signal events

Nuisance parameters:
- Nb – background events
- ε_s^i, ε_b^i – distribution along bands of sig/bck
- tLeff – deviation of Leff from median

$$\mathcal{L} = \prod_j^K \mathrm{Poiss}(n^j | \epsilon_s^j N_s + \epsilon_b^j N_b)$$

Poisson on data, per band.

$$\times \prod_{i=1}^{n^j} \frac{\epsilon_s^j N_s f_s(S1_i) + \epsilon_b^j N_b f_b(S1_i)}{\epsilon_s^j N_s + \epsilon_b^j N_b}$$

Distribution of events in each band

$$\times \prod_j^K \mathrm{Poiss}(m_s^j | \epsilon_s^j M_s) \times \prod_j^K \mathrm{Poiss}(m_b^j | \epsilon_b^j M_b)$$

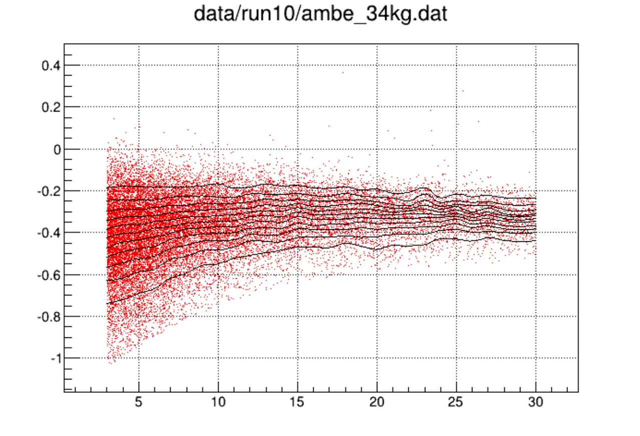
Poisson on calibration data

$$\times \exp(-(t - t_{\mathrm{obs}})^2/2)$$

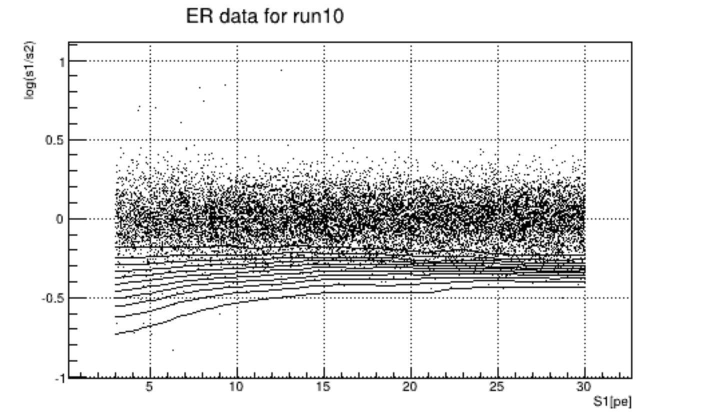
Leff penalty

XENON100 - PLR RESULTS

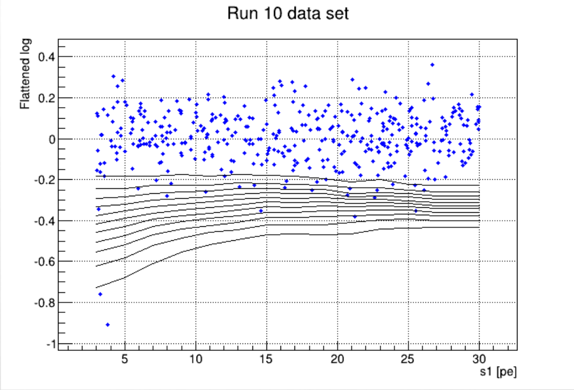
Using AmB data to construct bands.
Cross check using MC distribution



Nuclear Recoils

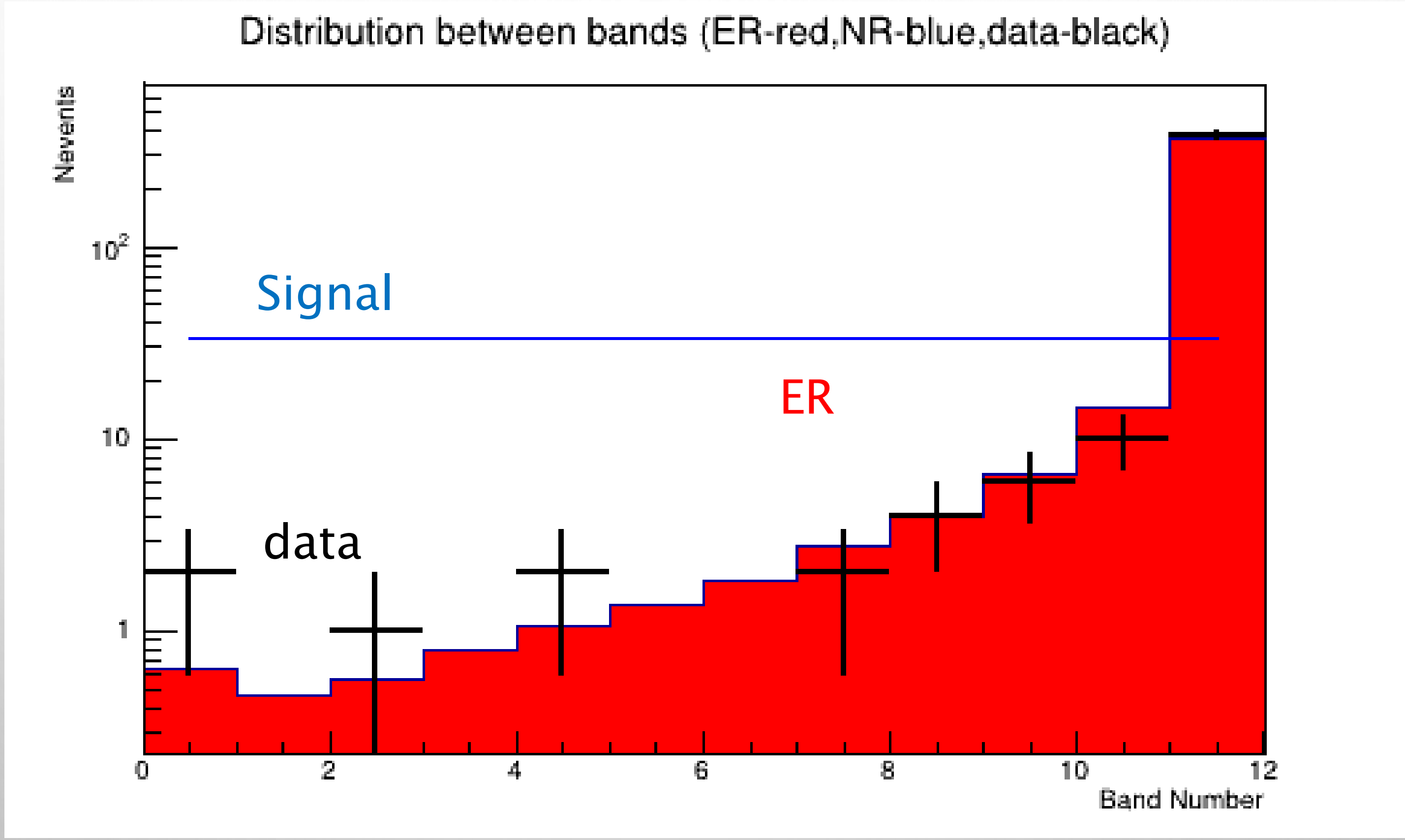


Electronic Recoils



Data

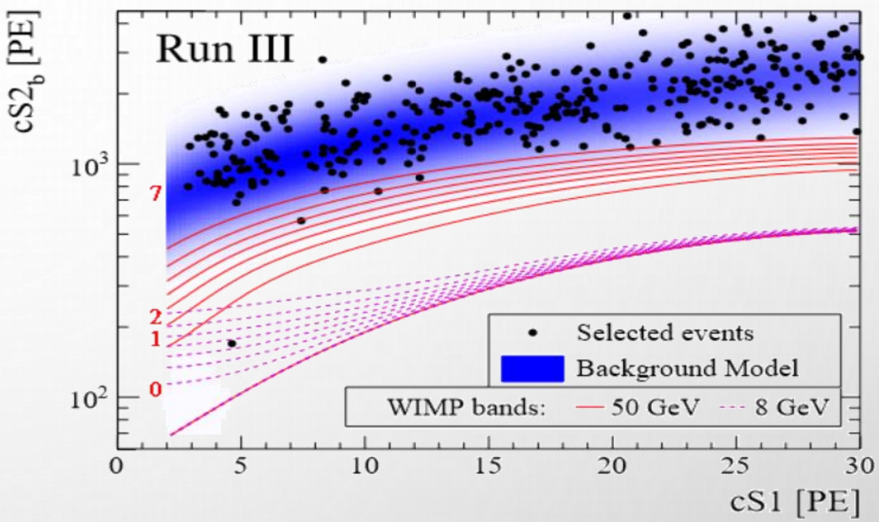
XENON100 - PLR RESULTS



XENON100 - PLR Results- Combination 2016 Analysis

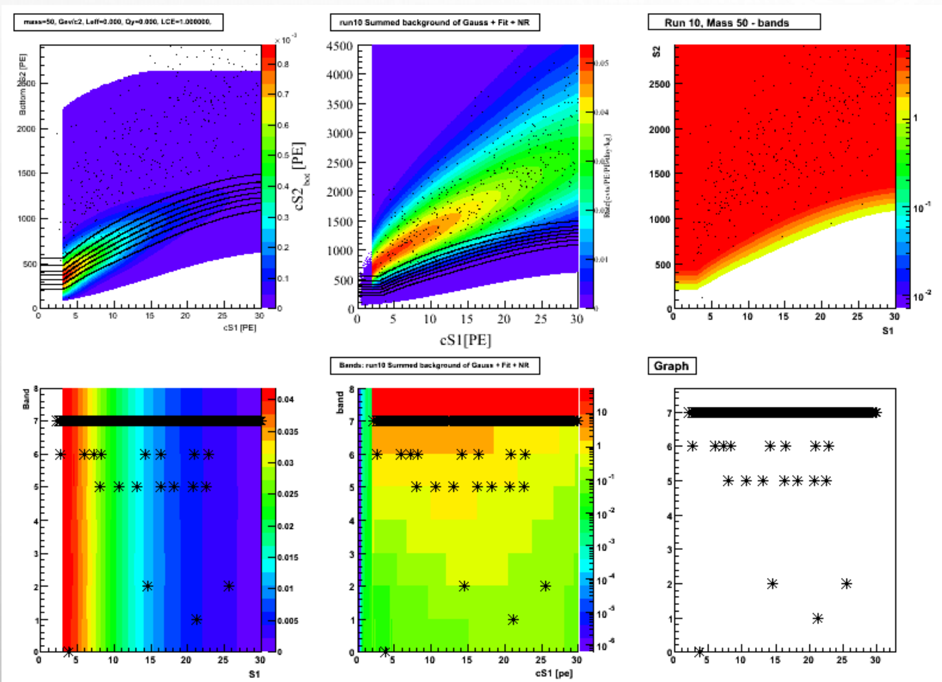
Introducing…

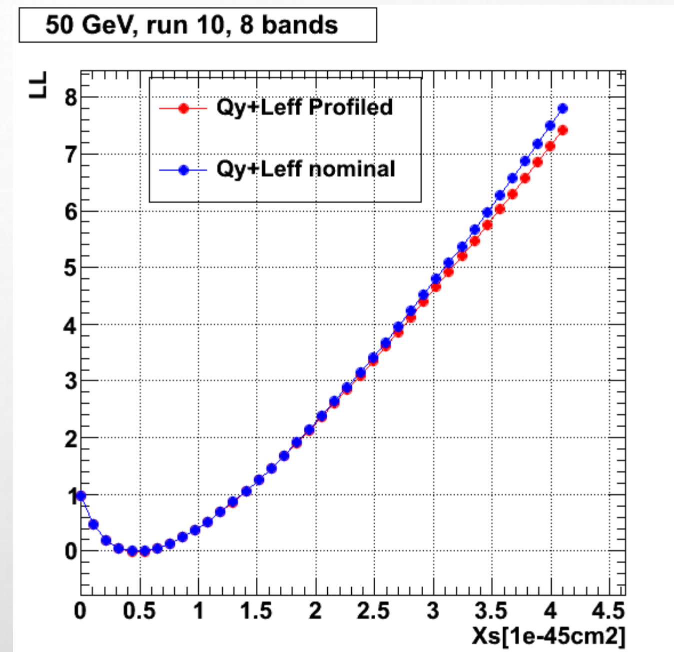
- Use MC for S2 modeling
  (instead of AmBe cal)
- Different bands for different masses
- Additional model nuisance parameter
  L_{eff} and Q_y



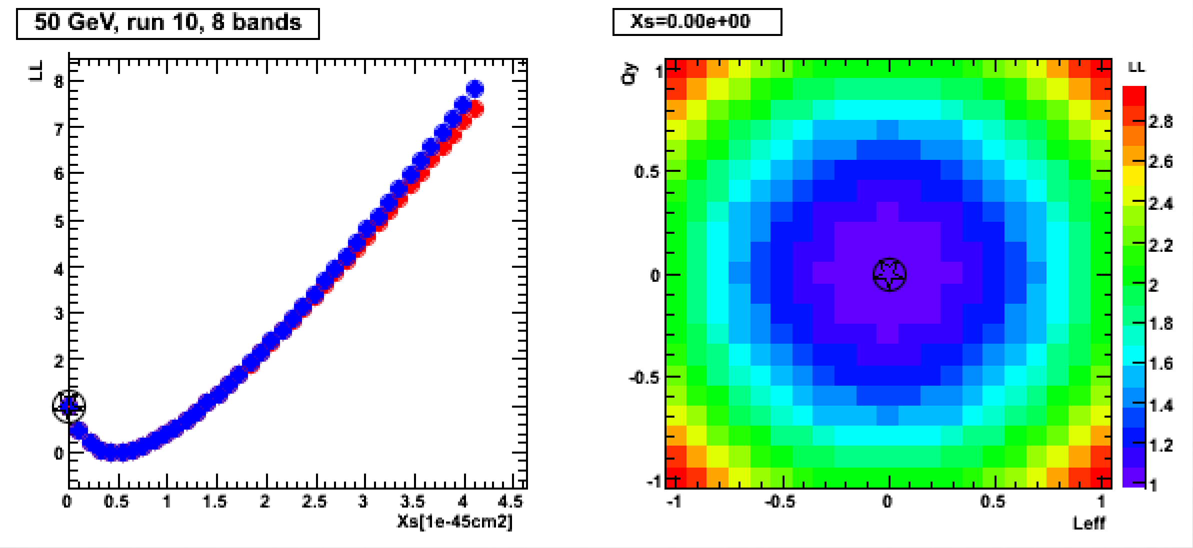
- Divide and conquer – PDF simulation production was done offline…and not during minimization
- Two variables: S1, and cS1 (and not just cS1)
  To better account for light collection efficiency variations

225 DAYS, 50 GEV

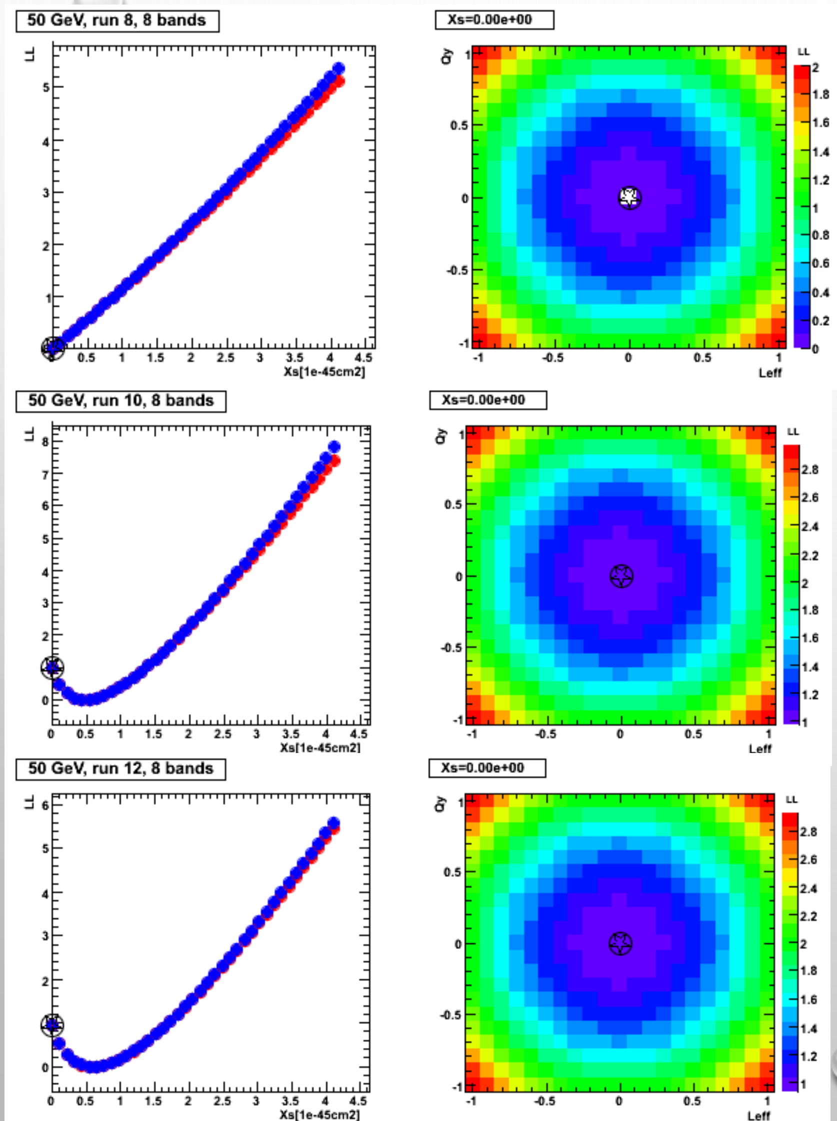


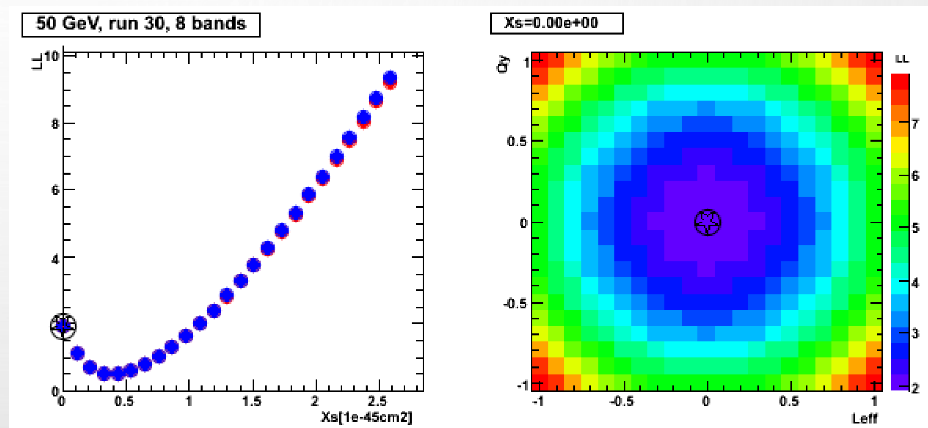


225 DAYS, 50 GEV



RUN COMBINATION

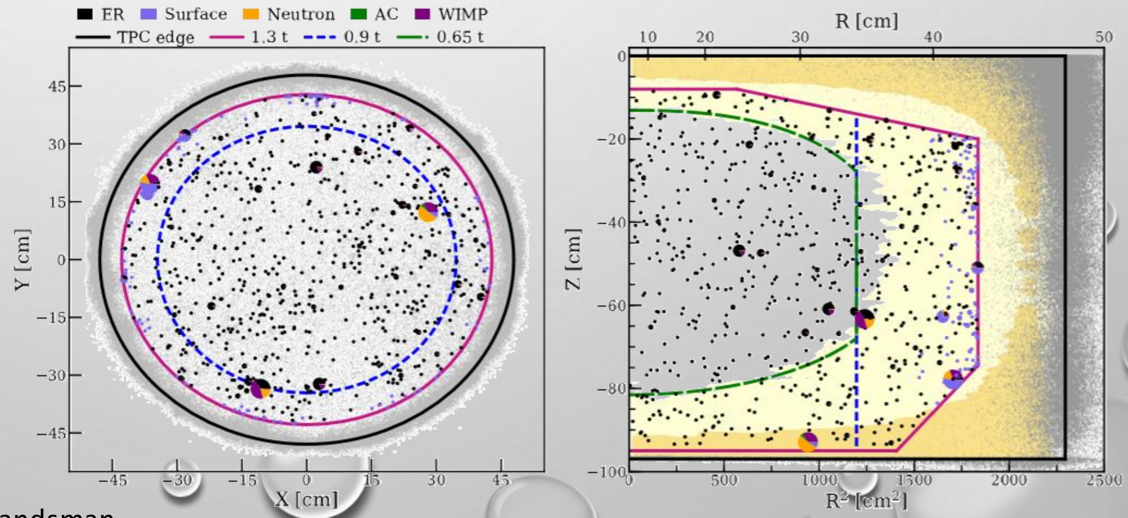




XENON1T - PLR results- 1ton-year 2018 analysis

Introducing…

- PDFs in higher dimensions (s1,s2,r). No bands in s2.
- Larger volume used
- 4 independent background models constraint by calibration and simulation
- More nuisance parameters
- More complete interaction model
- More sophisticated background model with some a-priori fits
- Safeguard to account for some mis-modeling



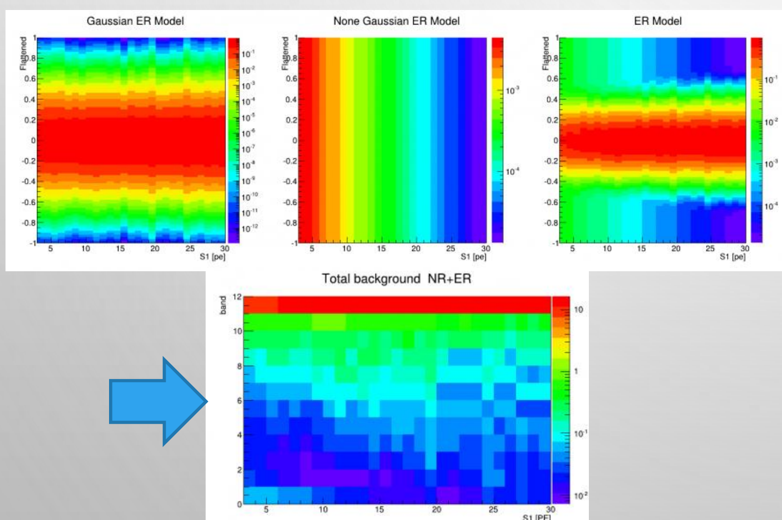
Some thought on signal model

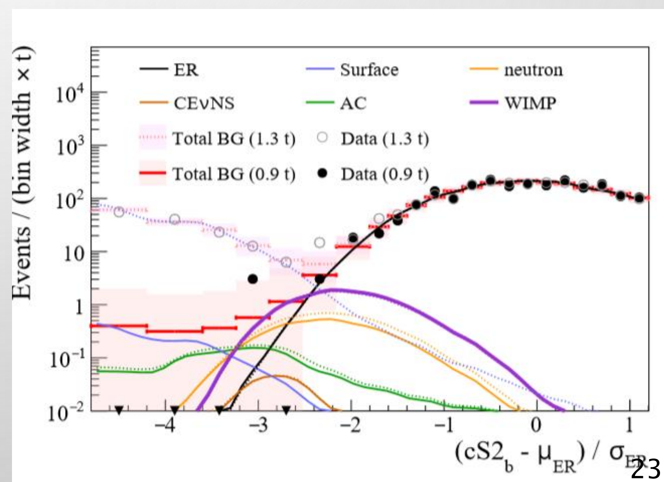
- Signal model sets: f_s and $n_s(\sigma)$ and (ε_s)
- Don't forget our parameter of interest is σ (not n_s)
- Energy scale: pe←→kev_{nr}
- Nuisance parameters in astrophysical model, interaction model, detector response
- No calibration sample available

  (calibration data can be used to constraint parameters)
- Need to artificially incorporate spatial and temporal detector instabilities

SOME THOUGHT ON BACKGROUND MODEL

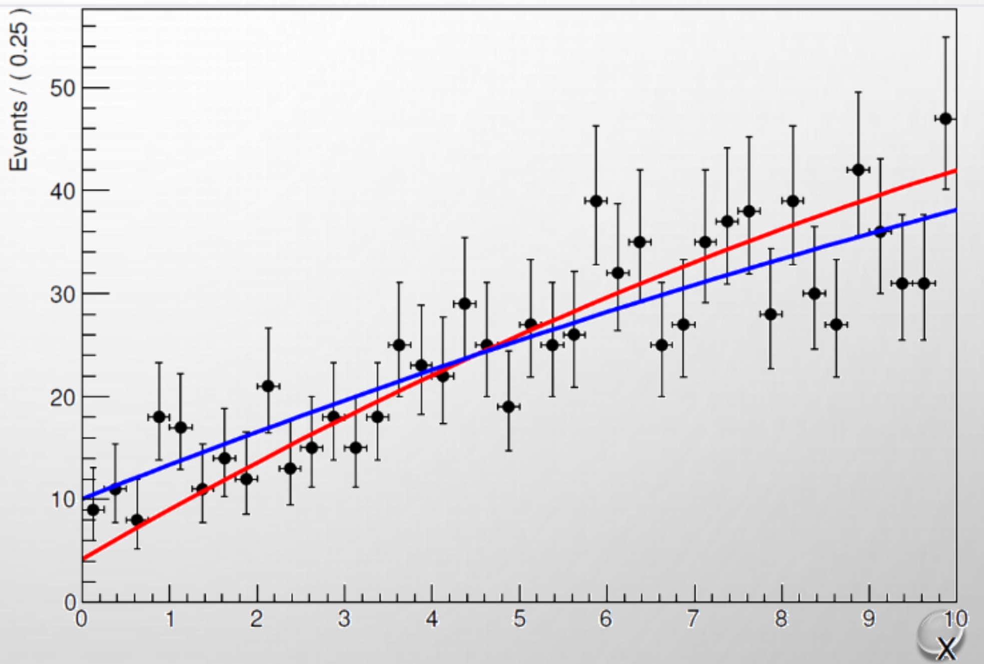
- Background model sets: f_b (ε_b) and sometimes N_b
- Several components of background: Fractions can be “frozen” or be nuisance
- Shape and magnitudes modeling
- Calibration samples may exist – statistic decreases with #variables
- Is our background model accurate “enough”?





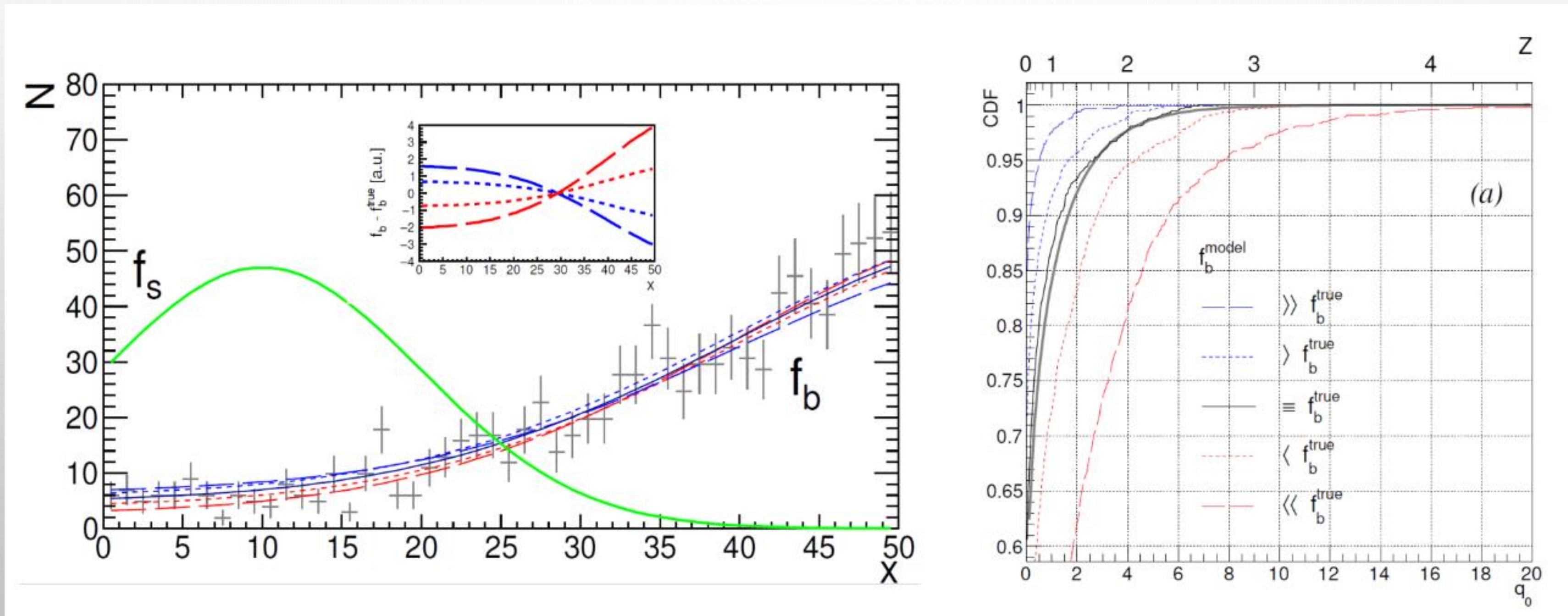
THE CURSE OF BACKGROUND MISMODELLING
THE PROBLEM

- Too many parameters
- Hidden parameters
- Partial underlying model
- ….Mistakes…



Might lead to enhanced false discovery rate or overly constrained limits

The curse of background mismodelling
The problem



Arxiv:1610.02643

The Curse of Background Mismodelling
The Problem

Arxiv:1610.02643

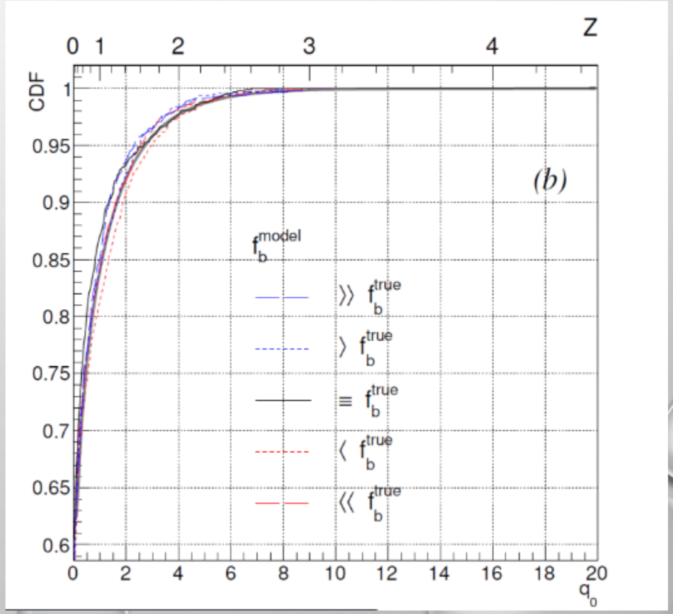
- Use the benchmark model
- Do not add extra nuisance parameters

$$f_b(x) \rightarrow (1-\varepsilon)f_b(x) + \varepsilon f_s(x)$$

$$L_{overall} = Poiss(N|N_s+N_b)\prod \frac{N_s f_s(x_i) + N_b(1-\varepsilon)f_b(x_i) + N_b \varepsilon f_s(x_i)}{N_s+N_b} \times L_{cal}(\varepsilon)$$

$$L_{cal}(\varepsilon) = \prod (1-\varepsilon)\, f_b(x_i) + \varepsilon f_s(x_i)$$

- Works for limits and discoveries
- Safeguards background components that are based on calibration
- We found out that a similar technique used for cross checks in the LHC, “spurious signal”



Where it hurts…

Many delicate points and challenges to address.

Here are just a few examples*:

Where it hurts…

EXAMPLE 1:"THE CURSE OF MISMODELLING"

The "safeguard" can provide some protection for models constructed based on calibration samples.

Nuisance parameters can be added, but

- Require some model assumption
- Complicates analysis – heavier, slower

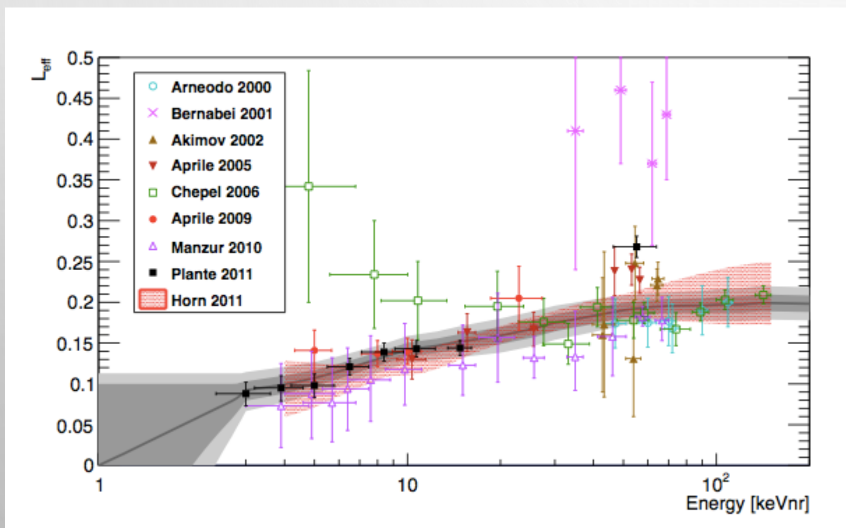
It is not enough

WHERE IT HURTS…

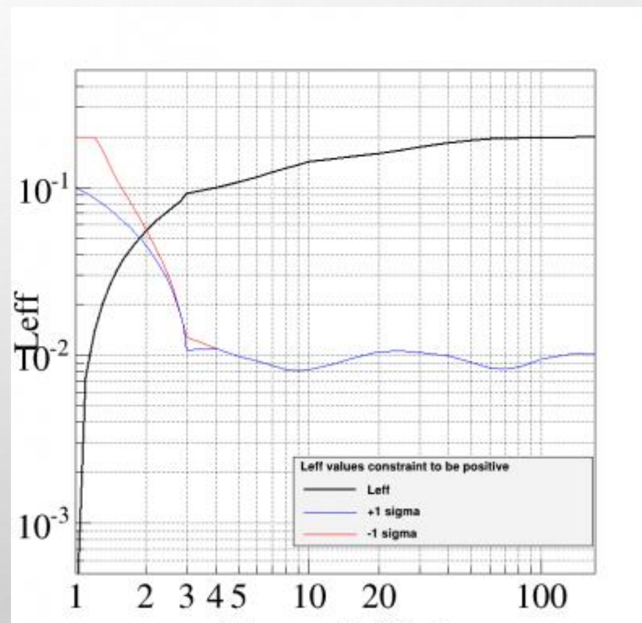
EXAMPLE 2: “THE CURSE OF THE UN-MODELLED”

- Include nuisance parameters without an underlying model
- Non physical regions
- Non symmetric nuisance uncertainties

E.G. L_{eff}



Phys. Rev. C 84, 045805 (2011)



WHERE IT HURTS…

EXAMPLE 3:“THE BLESSING OF ASYMPTOTICNESS”

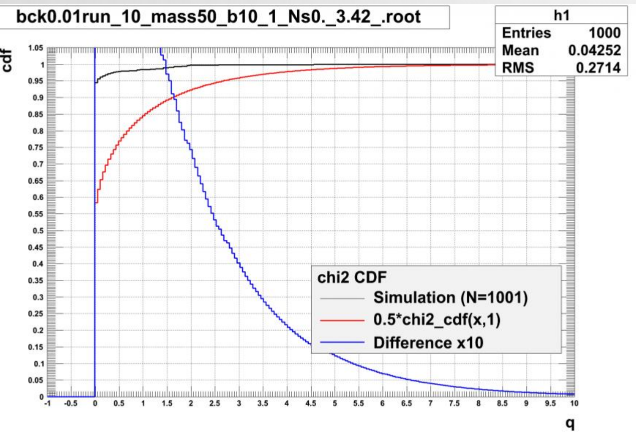
(Or “we ♥ wilks & arxiv1007.1727”)

Low bg

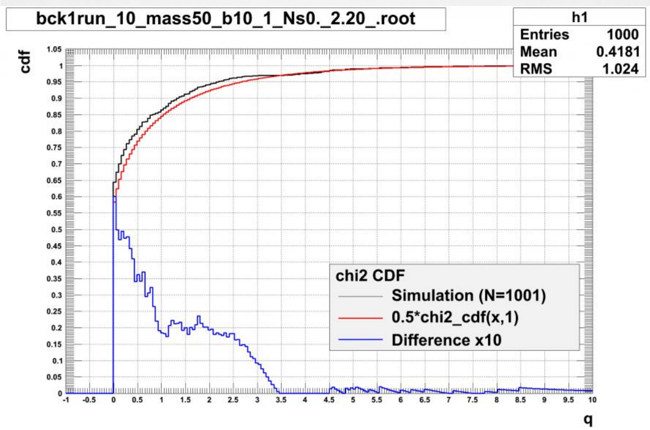
$$q_\sigma = \begin{cases} -2\ln\lambda(\sigma) & \hat{\sigma} < \sigma \\ 0 & \hat{\sigma} > \sigma \end{cases}$$

$$p_s = \int_{q_\sigma^{\mathrm{obs}}}^{\infty} f(q_\sigma|H_\sigma)\,\mathrm{d}q_\sigma .$$

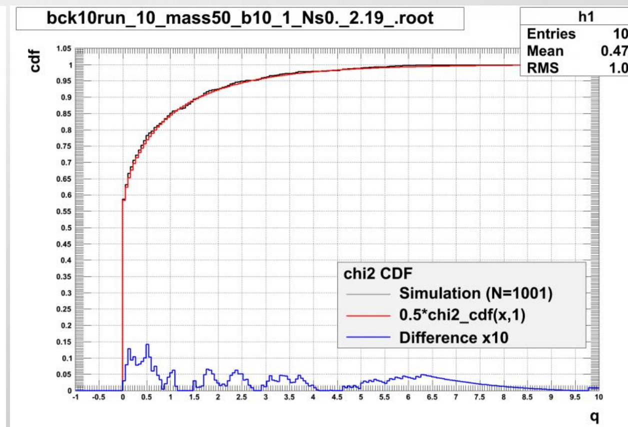
Ncalib=151



Ncalib=15128



Ncalib=151280



Need to verify asymptoticness and run MC if broken

Where it hurts…

<u>EXAMPLE 4:</u>"THE CURSE OF MULTIPLE DIMENSIONS"

Generating multidimensional (s1,s2,r,z…) pdf maps for "many" nuisance parameters variations

$$\lambda(\sigma) = \frac{\max\limits_{\sigma\ \text{fixed}} \mathscr{L}\left(\sigma; \mathcal{L}_{\text{eff}}, v_{\text{esc}}, N_b, \epsilon_s, \epsilon_b\right)}{\max \mathscr{L}\left(\sigma, \mathcal{L}_{\text{eff}}, v_{\text{esc}}, N_b, \epsilon_s, \epsilon_b\right)} \equiv \frac{\mathscr{L}\left(\sigma, \hat{\hat{\mathcal{L}_{\text{eff}}}}, \hat{\hat{v_{\text{esc}}}}, \hat{\hat{N}}_b, \hat{\hat{\epsilon}}_s, \hat{\hat{\epsilon}}_b\right)}{\mathscr{L}\left(\hat{\sigma}, \hat{\mathcal{L}_{\text{eff}}}, \hat{v_{\text{esc}}}, \hat{N}_b, \hat{\epsilon}_s, \hat{\epsilon}_b\right)}.$$

- Algorithm:
  - Prepare a model bank ahead of time
  - Or build the necessary model during minimization
  - (Possibly with smart book keeping and archiving)
- Nuisance parameter resolution
  - How large a step in modeling
  - Interpolate?
- Verifying asymptoticness or doing mc instead becomes painful
- Also: complicated codes

WHERE IT HURTS…

EXAMPLE 5:"THE CURSE OF HANDWAVING"

S1
[PE]

Correct S1 according to LCE(x,y,z): scale up or down the total PE measured to get number of PE we should have gotten with a uniform light collection efficiency

Data points in cS_1

WIMP model
[KeV]

for a given wimp mass, calculate differential rate in KeV, Translate to PE using average light yield. Poisson smear, Gauss smear, apply acceptances,

fs(cS1) pdf

Likelihood function

Loop on all events in each band. For each event, use its cS1 to check how likely it is to come from the signal pdf, or background pdf.

Problem: cS1 is not physical. Low PE cut, Poisson smearing should be done on s1!

Where it hurts…

EXAMPLE 6:“THE CURSE OF DIVERSITY”

$$p'_s = \frac{p_s}{1 - p_b}$$

where

$$1 - p_b = \int_{q_\sigma^{\rm obs}}^{\infty} f(q_\sigma | H_0)\, {\rm d}q_\sigma$$

e.g. Over coverage:

- Power constraint
- Cls (Roughly 90%CL→95%CL)
- Ce la vie

Where it hurts…

EXAMPLE 7:"THE CURSE OF PAGE LIMIT"

- Many details to the models, inference method…
- Information in papers is limited. Very often summarized to: "…as was done in [xx]."
- Would be nice to see more detailed likelihood functions…
- Would be nice to see more likelihood curves…
- Many consistency checks, verifications to be made . usually not even explicitly acknowledged.
- Follow up papers become more popular, but

…cannot make everyone happy….

PhysRevLett.120.132501

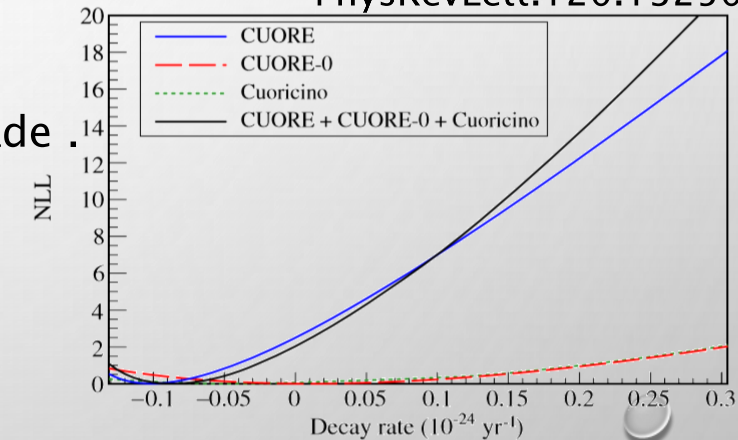


FIG. 4. Profile negative-log-likelihood curves for CUORE, CUORE-0, Cuoricino, and their combination.

GRAPHICAL SUMMARY

Back then....