Data unfolding for model independent comparison of dark matter direct detection experiments

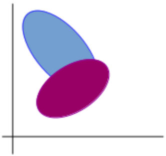
Matti Heikinheimo

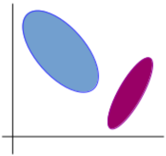
University of Helsinki & Helsinki Institute of Physics

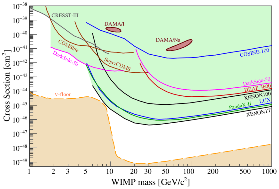
November 28, 2024

Model dependent vs model independent comparison of experiments

- To see if a given model, parametrized by some free parameters $\{\lambda_i\}$ is a plausible explanation for some data measured in two experiments A and B, one can fit the parameters to each data set and compare the confidence intervals.
- If the confidence intervals overlap, the model is compatible with both sets of data, otherwise not.
- An example of such analysis is the commonly used spin independent DM cross section exclusion.
- How to make the comparison, assuming as little as possible about the model?



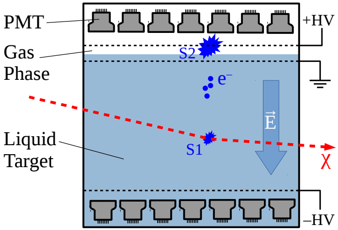




Direct detection

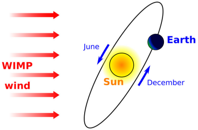
- Direct detection experiments look for DM scattering off the nucleus of the target material, by detecting the nuclear recoil (typically via scintillation light, electric signal or phonons).
- The event rate depends on the DM-nucleus scattering cross section, and the velocity distribution of DM:

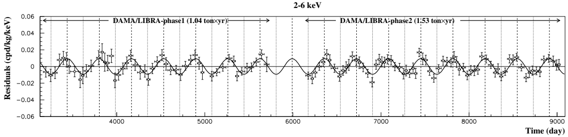
$$\frac{dR}{dE_r} = \frac{\rho_0 M}{m_N m_\chi} \int_{v_{\rm min}}^{v_{\rm esc}} \frac{f(v)}{v} \frac{d\sigma}{dE_r} d^3 v, \qquad v_{\rm min} = \sqrt{\frac{E_r m_N}{2\mu_{N\chi}^2}}.$$



Annual modulation and DAMA

- Most of the DM direct detection experiments count nuclear recoil -like events and attempt to minimize background.
- DAMA instead measures the modulation of the event rate in annual cycles.
- A modulation with maximum in June and minimum in December expected for the DM signal due to the motion of the earth with respect to the galactic rest frame.
- A constant background would cancel in the residual event rate $R(t) - \langle R \rangle$. Therefore the modulation amplitude would be associated with the DM event rate.
- However, to infer the expected DM signal in another experiment based on the DAMA observed modulation requires assumptions about the DM-nucleon interaction model and DM velocity distribution in the Milky Way.

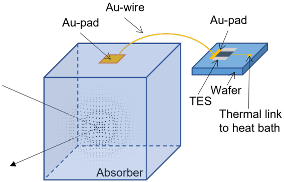


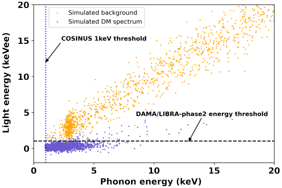


COSINUS and DAMA

- The first step in achieving as model independent as possible comparison with the DAMA data is to use the same target material NaI, to eliminate the dependence on the DM-nucleus interaction.
- The approach chosen in COSINUS is to measure two signals simultaneously, scintillation light and phonons (heat).
- DAMA only measures scintillation light. Therefore the DAMA implied nuclear recoil energy needs to be obtained by converting the (electron recoil calibrated) scintillation energy:

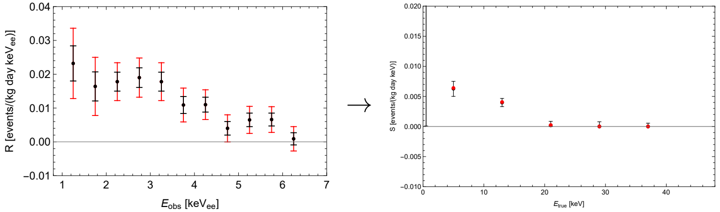
$$E_{\mathrm{nr}} = \frac{E_{\mathrm{ee}}}{Q^T}$$





Unfolding

- To facilitate the comparison we would like to make use of the DAMA observed modulation spectrum.
- This can be achieved by unfolding the DAMA spectrum, to obtain an estimate of the underlying true recoil spectrum.
- We can then directly compare this true spectrum to COSINUS observations, since COSINUS measures the true recoil energy via phonons (to a good approximation).



Unfolding

- Data unfolding refers to an approximate solution to the inverse problem: A set of observed events is produced via a response function from an underlying set of true events. The problem is to obtain an estimate for the true histogram given an observed histogram.
- We use iterative Richardson-Lucy unfolding[1] algorithm, formulated in terms of a response matrix A_{ij}, describing the probability for an event in true energy bin j to be observed in the DAMA energy bin i.
- Given the true histogram $\{x_j\}$, the expected observed histogram is given by

$$y_i = \sum_{j=1}^{N} A_{ij} x_j .$$

- The algorithm yields an estimate for x given y, A and a prior $x^{(0)}$:

$$x_j^{(k+1)} = \sum_{i=1}^{M} \frac{A_{ij} x_j^{(k)} y_i}{\sum_{l=1}^{N} A_{il} x_l^{(k)} \sum_{m=1}^{M} A_{mj}} .$$

[1] G. Zech,Nucl. Instrum. Meth. A **716** (2013), 1-9 doi:10.1016/j.nima.2013.03.026 [arXiv:1210.5177 [physics.data-an]].

Response matrix

- The response matrix for nuclear recoils off target nucleus T is given by

$$A_{ij}^{T} = \frac{1}{E_{\mathrm{nr}\,j}^{\mathrm{max}} - E_{\mathrm{nr}\,j}^{\mathrm{min}}} \int_{E_{\mathrm{nr}\,j}^{\mathrm{min}}}^{E_{\mathrm{nr}\,j}^{\mathrm{max}}} \epsilon_{\mathrm{DAMA}}^{T}(E_{\mathrm{nr}}; E_{\mathrm{ee}\,i}^{\mathrm{min}}, E_{\mathrm{ee}\,i}^{\mathrm{max}}) dE_{\mathrm{nr}}.$$

- Here the DAMA efficiency function is

$$\epsilon_{\mathrm{DAMA}}^{T}(E_{\mathrm{nr}}; E_{\mathrm{ee}}^{\mathrm{min}}, E_{\mathrm{ee}}^{\mathrm{max}}) = \frac{1}{2}\left(\mathrm{erf}\left(\frac{E_{\mathrm{ee}}^{\mathrm{max}} - Q^{T} E_{\mathrm{nr}}}{\sqrt{2}\sigma_{\mathrm{DAMA}}(Q^{T} E_{\mathrm{nr}})}\right) - \mathrm{erf}\left(\frac{E_{\mathrm{ee}}^{\mathrm{min}} - Q^{T} E_{\mathrm{nr}}}{\sqrt{2}\sigma_{\mathrm{DAMA}}(Q^{T} E_{\mathrm{nr}})}\right)\right)$$

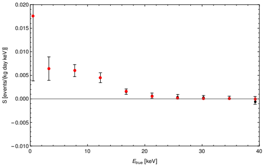
- where $Q^{\mathrm{Na}} = 0.3$ and $Q^{\mathrm{I}} = 0.09$ are the quenching factors and the resolution function is given by

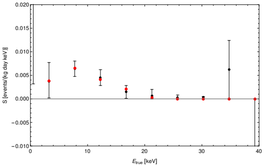
$$\sigma_{\mathrm{DAMA}}(QE) = (0.448\ \mathrm{keV_{ee}})\sqrt{QE/\mathrm{keV_{ee}}} + 0.0091 QE.$$

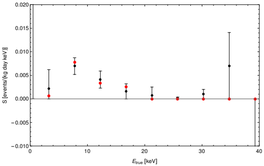
- The observed histogram is given by the data shown in the previous slide.
- For the prior we use a flat (constant) spectrum. The end result is not sensitive to this choice.

Results

- Additionally, the algorithm depends on the specification for the number of true energy bins, which should be smaller than the number of bins in the observed histogram.
- And on the number of iteration steps, which should be large enough to obtain a good fit but not too large to avoid unwanted oscillatory behaviour in the estimated true spectrum.
- We have tested $N_{\mathrm{bins}} = 3, 5, 7, 9$, and 10, 100, 1000 iteration steps (from left to right below). Currently we are using 30 iteration steps.

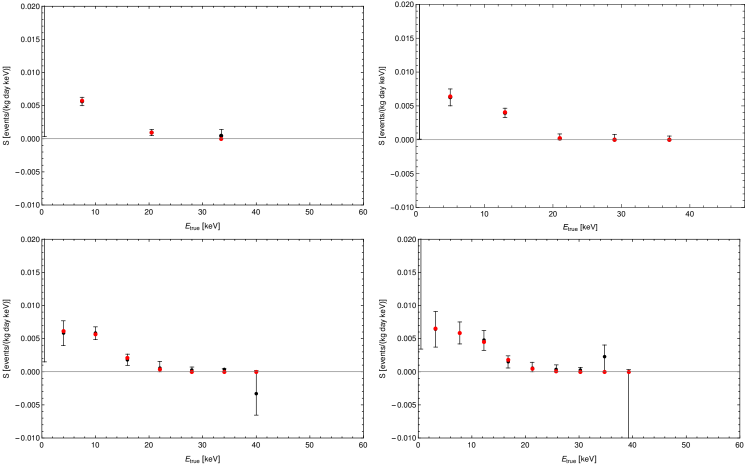






Results

Estimated true spectrum with $N_{\rm bins} = 3, 5, 7, 9$.

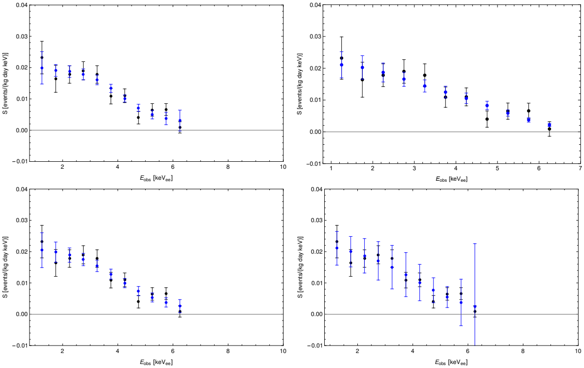


Uncertainty

- The uncertainty of the estimated true spectrum was obtained by generating variations of the observed DAMA data, by allowing the value in each bin to fluctuate with a Gaussian width given by the confidence limits reported by DAMA.
- 10k samples of such spectra where unfolded to obtain an ensemble of "true" spectra. The error bars shown in the previous slide correspond to 90% confidence limit of this ensemble.
- The red dots show the mean values of the sample, while the black dots are the results of unfolding the actual DAMA spectrum (without deviations).

Forward model test

- To check that the results make sense, the obtained true spectra were folded with the response matrix A, to obtain the corresponding expected histogram.
- The resulting spectra seem (by eye) compatible with the real DAMA spectrum.
- Statistical tests yield high compatibility: Kolmogorov-Smirnov p-value 0.997, χ^2 p-value 0.966.



Required exposure

- For each true histogram (with $N_{\text{bins}} = 3, 5, 7, 9$) we find the optimal combined bin, by summing the 90%-errors in quadrature, to find the optimal number of true energy bins to combine so that the implied lower limit R_{bound} on the total event rate is largest.
- Essentially this means that we combine all the bins that show non-zero event rate, and ignore those that are compatible with zero.
- The required exposure for a 90% exclusion of the DAMA DM signal is then given by

$$\mathcal{E} = \frac{2.3}{\epsilon_{\text{COSINUS}} R_{\text{bound}}},$$

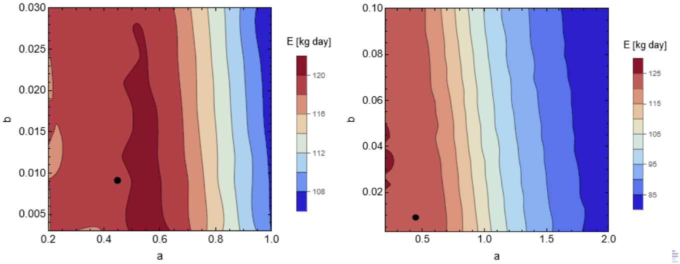
- where we have used a constant $\epsilon_{\text{COSINUS}} = 0.25$ efficiency above the threshold energy, and the factor 2.3 is the 90%-limit for the expected number of events assuming zero observed events (zero background).
- We have tested the effect of varying the threshold energy between 0.5 keV to 3 keV.

DAMA energy resolution

- We wish to investigate how our results depend on the assumed DAMA enegy resolution, which enters in the response matrix for the unfolding procedure.
- We parametrize the DAMA resolution function as

$$\sigma_{\mathrm{DAMA}}(Q^T E_{\mathrm{nr}}) = (a\ \mathrm{keV_{ee}})\sqrt{Q^T E_{\mathrm{nr}}/\mathrm{keV_{ee}}} + bQ^T E_{\mathrm{nr}},$$

  where the nominal values for the parameters are $a = 0.448$, $b = 0.0091$.
- The figures show the required exposure as a function of these parameters, for 3 (left) and 5 (right) bins, zero background and 1 keV threshold.

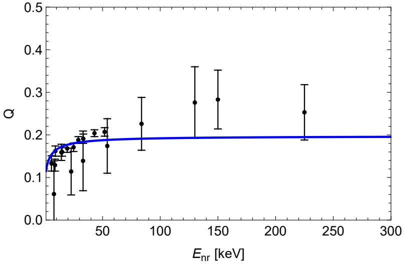


DAMA quenching factors

- To account for a possibly energy dependent quenching factor for Na in DAMA we use a parametrisation of the Lindhard model:

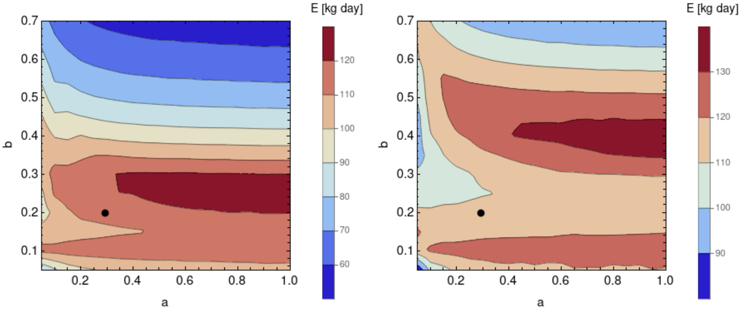
$$Q^{\mathrm{Na}}(E_{\mathrm{nr}}) = b\frac{ag}{1+ag}, \quad g = 3E_{\mathrm{nr}}^{0.15} + 0.7E_{\mathrm{nr}}^{0.6} + E_{\mathrm{nr}}$$

- A fit of this form to a sample of measured values for the quenching factor is shown below, returning best fit values of $a = 0.294$, $b = 0.197$.
- The nominal DAMA values for Na would imply $b = 0.3$, $a \to \infty$



DAMA quenching factors

- The figures show the required exposure as a function of these parameters, for 3 (left) and 5 (right) bins, zero background and 1 keV threshold.



Conclusions

- We have performed an unfolding analysis to facilitate comparison between the DAMA annual modulation signal and the expected COSINUS event rate exclusion.
- The required exposure for 90% (95%) exclusion of the DAMA dark matter signal is about 120 (170) kg days, assuming COSINUS energy threshold 1 keV and zero background.
- The analysis is quite robust to changes in model parameters, including DAMA energy resolution and quenching factors.
- A background model for COSINUS in preparation, our exclusion projection will be straightforward to update to include a non-zero background.