Assessing Near-Future Direct Searches with Benchmark-free Forecasting: Supplementary Material

Thomas D. P. Edwards,* Bradley J. Kavanagh,† and Christoph Weniger‡

Gravitation Astroparticle Physics Amsterdam (GRAPPA),
Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics,
University of Amsterdam, Science Park 904, 1090 GL Amsterdam, The Netherlands
(Dated: October 15, 2018)

Appendix A: Technical details

First, we generate a large number of points in the parameter space of model $\mathcal{M}$, $\vec{\theta}_i \in \Omega_{\mathcal{M}}$ (typically of the order 10^5). The details of the distribution of the points do not matter in the limit of a large number of parameter points. Each point is projected onto the corresponding euclideanized signal, $\vec{x}_i = \vec{x}(\vec{\theta}_i)$. This projection depends on the details of the detector. If multiple experiments are used, the corresponding vectors are concatenated. Euclideanized signals are generated using `swordfish` [1] (see paper for technical details). This process essentially provides a sample of the model parameter space $\mathcal{M}$ embedded in the (usually higher dimensional) euclideanized signal space. In addition, the mapping between these spaces is known if $\mathcal{M}$ is sufficiently sampled. This sample of parameter points and corresponding euclideanized signals are the basis for the various estimation techniques used in this work.

Confidence contours. Given the sample of projected points, $\vec{x}_i$, it is now straightforward to generate expected contours around any parameter $\vec{\theta}_0 \in \Omega_{\mathcal{M}}$ in the model parameter space. Such regions are for instance shown in Fig. 1. To this end, we first calculate the projected signal $\vec{x}_0 = \vec{x}(\vec{\theta}_0)$. Then, using standard nearest neighbor finder algorithms [2], we identify the set of points $\vec{\theta}_i$ that are within a radius $r_\alpha(\mathcal{M})$ of point $\vec{\theta}_0$. Here, $r_\alpha(\mathcal{M})$ depends on the dimensionality of the parameter space $\Omega_{\mathcal{M}}$ as well as the significance level of interest, $r_\alpha(\mathcal{M}) = \sqrt{\chi^2_{k=d,\mathrm{ISF}}(1-\alpha)}$ where $\chi^2_{k=d,\mathrm{ISF}}$ is the inverse survival function of the Chi-squared distribution with $k = d$ degrees of freedom. For the 3-dim models that we consider in this paper, and a significance level of $\alpha = 0.045$, we have for instance $r_\alpha(\mathcal{M}) = 2.84$. Now, the points in the model parameter space $\Omega_{\mathcal{M}}$ that belong to the confidence region can be simply identified by back-projecting the nearest neighbors in euclideanized signal space (obviously this back-projecting just requires a lookup in the original list of model parameters). In this way, the generation of confidence regions around arbitrary signal points is efficiently achieved.

---

* t.d.p.edwards@uva.nl
† b.j.kavanagh@uva.nl
‡ c.weniger@uva.nl

Number of distinct signals. For Fig. 2 we are interested in the (maximum) number of points that can populate the model parameter space $\Omega_{\mathcal{M}}$ such that the model points can be discriminated in the sense of Eq. (1). This is equivalent to finding the maximum number of points in the euclideanized signal space that can populate the embedding of $\Omega_{\mathcal{M}}$ such that their mutual distance is at least $r_\alpha(\mathcal{M})$. We derive an estimate for this number with the following procedure: For each projected sample point $\vec{x}_i$, we calculate the number of nearest neighbors N_i within a distance $r_\alpha(\mathcal{M})/2$. This can be rather efficiently done using standard clustering algorithms. Now, we assign a weight to point $\vec{x}_i$, which is simply given by $w_i = 1/N_i$. It hence corresponds to the 'fractional contribution' of a single parameter point to a confidence region. The number of distinguishable signals is given by the sum

$$\nu^\alpha_{\mathcal{M},X}(\Omega_{\mathcal{M}}) = c_{ff} \sum_i w_i \,. \tag{A1}$$

Here, c_{ff} is a filling factor correction related to the packing density of hyperspheres in a d-dim parameter space. For the 3-dim models in which we are often interested, this number is given by $c_{ff} = 0.74$ [3]. We find that this prescription provides an efficient and reliable way to estimate the number of distinct signals of a model. The main requirement in the calculation is that each of the potential confidence regions contains a sufficiently large (typically at least ten) number of samples. Adding more points to the original list would then not affect the result anymore. We tested the stability of our results by doubling the number of sampled points in various examples from the text. The results remained unchanged in the limit of ten points per distinct region.

Distinct signals compatible with $\mathcal{H}_0$. We are interested in the fraction of the distinct signal points that are consistent with a null hypothesis that is defined as a lower dimensional subspace of the full model parameter space, $\Omega_{\mathcal{S}} \in \Omega_{\mathcal{M}}$. Here, we call a point in $\Omega_{\mathcal{M}}$ 'consistent' with $\Omega_{\mathcal{S}}$ if the composite null hypothesis $\Omega_{\mathcal{S}}$ cannot be excluded against the alternative hypothesis $\Omega_{\mathcal{M}}$. To estimate this number, we first generate a large number of points in $\Omega_{\mathcal{S}}$. We then collect all points from the original sample of Ω_M whose minimum distance to any of the points from $\Omega_{\mathcal{S}}$ is smaller than the threshold values $r_\alpha(\mathcal{M},\mathcal{S})$. Here, the threshold is derived from a χ^2_k distribution with k degrees of freedom, where k is the difference in the dimensionality of $\Omega_{\mathcal{M}}$ and Ω_S (for the

examples in the paper, we usually have $k = 1$, and hence $r_\alpha(\mathcal{M}, \mathcal{S}) = 2$). The number of distinct signals that are compatible with the null hypothesis $\Omega_\mathcal{S}$ is then simply obtained by restricting the sum in Eq. (A1) to the points within the shell around $\Omega_\mathcal{S}$.

Parameter ranges and nuisance parameters. Finally, the contours in Fig. 3 and Fig. 4 are generated by identifying all points that are consistent with the indicated null hypotheses, as described in the previous paragraph. However, in these figures we also take into account the effects of DM halo uncertainties, as described in the main text (this is not easily possible when calculating the number of dimensions). To this end, we generate for each point $\vec{x}_i \in \Omega_\mathcal{M}$ several euclideanized signals corresponding to various randomly selected DM halo configurations. Again, the specific distribution of these points does not matter as long as the number is large enough to sufficiently cover the various halo configurations. In order to incorporate external constraints on the DM halo parameters, we add an additional contribution to the euclideanized distance calculation, which is just given by $(\eta_i - \bar{\eta})^2/\sigma_\eta^2$, where $\bar{\eta}$ and σ_η^2 are the mean and variance of the nuisance parameter, and η_i is the value of the nuisance parameter for a specific point i. The contribution to the $\vec{x}(\vec{\theta})$ is a simple concatenation of $(\eta_i - \bar{\eta})/\sigma_\eta$ with the Euclideanised signal.

For a large number of sampled points this approach exactly matches a profile log-likelihood analysis. We check this limit is saturated by increasing the number of sampled points until our results do not noticeably change.

Appendix B: Dark matter signal modeling

Halo Uncertainties. We incorporate halo uncertainties [4] by assuming Gaussian likelihood distributions for three parameters of the Maxwellian velocity distribution of DM: the Sun's speed $v_\odot = (242 \pm 10)$ km/s [5], the local circular speed $v_c = (220 \pm 18)$ km/s [6], and the Galactic escape speed $v_{esc} = (533 \pm 54)$ km/s [7]. We assume that these uncertainties are uncorrelated [8], though in general correlations coming from the modeling of the Milky Way halo can be included [9, 10]. We sample from these distributions as nuisance parameters in our signal calculation and include an additional penalization term to the euclideanized signal in Eq. (1).

Detector specifications. We implement a simplified XENON1T for which we adopt an S1-only analysis, full details of which are given in Sec. IIIB of [1]. For the recoil spectrum dR/dS_1, we use 19 bins linearly spaced between 3 and 70 PE (corresponding to nuclear recoil energies $E_R \in [5, 40]$ keV). The number of bins was chosen for computational efficiency with no noticeable loss in accuracy. We have checked that including a 20% energy resolution [11] and increasing the number of bins should have no substantial effect on our results. Background distributions as a function of S_1 are described in Fig. 3 of [12] for which we assume 1% uncertainty on all components separately. We also sum over different Xenon isotopes, weighting by their naturally-occurring mass fractions [13].

For our future Xenon detector we scale up the observation time of XENON1T-2017 by a factor of 100, assuming that background rates stay constant. This exposure roughly corresponds to that expected for the full run of the XENONnT experiment [14], so we will refer to this future detector as *XENONnT*.

DarkSide20k: We directly use the recoil energy spectrum dR/dE_R as our signal, assuming that the only relevant isotope is Argon-40. We follow the specifications of the Darkside50 detector [15], with the nuclear recoil efficiency taken from Fig. 6 in [15]. The background is assumed to be flat across the entire energy range with an expected 0.1 events (with 10% uncertainty) over 1422 kg days of observation. We assume 19 linearly spaced bins between 32 and 200 keV.

For our future detector, we assume an exposure of 7.3×10^6 kg days, corresponding to a 1-year exposure with a 20-tonne detector. We assume that the background will remain at 0.1 events in the total exposure. This detector configuration roughly resembles *DarkSide20k* [16].

[1] T. D. P. Edwards and C. Weniger, (2017), arXiv:1712.05401 [hep-ph].

[2] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Journal of Machine Learning Research **12**, 2825 (2011).

[3] E. W. Weisstein, "Hypersphere packing," .

[4] A. M. Green, J. Phys. **G44**, 084001 (2017), arXiv:1703.10102 [astro-ph.CO].

[5] J. Bovy *et al.*, Astrophys. J. **759**, 131 (2012), arXiv:1209.0759 [astro-ph.GA].

[6] S. E. Koposov, H.-W. Rix, and D. W. Hogg, Astrophys. J. **712**, 260 (2010), arXiv:0907.1085 [astro-ph.GA].

[7] T. Piffl *et al.*, Astron. Astrophys. **562**, A91 (2014), arXiv:1309.4293 [astro-ph.GA].

[8] A. H. G. Peter, Phys. Rev. **D81**, 087301 (2010), arXiv:0910.4765 [astro-ph.CO].

[9] L. E. Strigari and R. Trotta, JCAP **0911**, 019 (2009), arXiv:0906.5361 [astro-ph.HE].

[10] M. Pato, L. E. Strigari, R. Trotta, and G. Bertone, JCAP **1302**, 041 (2013), arXiv:1211.7063 [astro-ph.CO].

[11] M. Schumann, L. Baudis, L. Btikofer, A. Kish, and M. Selvi, JCAP **1510**, 016 (2015), arXiv:1506.08309 [physics.ins-det].

[12] E. Aprile *et al.* (XENON), Phys. Rev. Lett. **119**, 181301 (2017), arXiv:1705.06655 [astro-ph.CO].

[13] J. L. Feng, J. Kumar, D. Marfatia, and D. Sanford, Phys.

Lett. **B703**, 124 (2011), arXiv:1102.4331 [hep-ph].

[14] G. Plante (XENON), `https://conferences.pa.ucla.edu/dm16/talks/plante.pdf` (2016).

[15] P. Agnes *et al.* (DarkSide), Phys. Lett. **B743**, 456 (2015),

arXiv:1410.0653 [astro-ph.CO].

[16] C. E. Aalseth *et al.*, Eur. Phys. J. Plus **133**, 131 (2018), arXiv:1707.08145 [physics.ins-det].