

Examining the time dependence of DAMA's modulation amplitude

Chris Kelso^{1,a}, Christopher Savage^{2,3,b}, Pearl Sandick^{3,c}, Katherine Freese^{2,4,d}, Paolo Gondolo^{3,e}

¹ Department of Physics, University of North Florida, Jacksonville, FL 32224, USA

² Leinweber Center for Theoretical Physics, Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

³ Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA

⁴ Michigan Center for Theoretical Physics, Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

Received: 27 October 2017 / Accepted: 28 February 2018 / Published online: 16 March 2018

© The Author(s) 2018

Abstract If dark matter is composed of weakly interacting particles, Earth's orbital motion may induce a small annual variation in the rate at which these particles interact in a terrestrial detector. The DAMA collaboration has identified at a 9.3σ confidence level such an annual modulation in their event rate over two detector iterations, DAMA/NaI and DAMA/LIBRA, each with ~ 7 years of observations. This data is well fit by a constant modulation amplitude for the two iterations of the experiment. We statistically examine the time dependence of the modulation amplitudes, which “by eye” appear to be decreasing with time in certain energy ranges. We perform a chi-squared goodness of fit test of the average modulation amplitudes measured by the two detector iterations which rejects the hypothesis of a consistent modulation amplitude at greater than 80, 96, and 99.6% for the 2–4, 2–5 and 2–6 keVee energy ranges, respectively. We also find that among the 14 annual cycles there are three $\gtrsim 3\sigma$ departures from the average in our estimated data in the 5–6 keVee energy range. In addition, we examined several phenomenological models for the time dependence of the modulation amplitude. Using a maximum likelihood test, we find that descriptions of the modulation amplitude as decreasing with time are preferred over a constant modulation amplitude at anywhere between 1σ and 3σ , depending on the phenomenological model for the time dependence and the signal energy range considered. A time dependent modulation amplitude is not expected for a dark matter signal, at least for dark matter halo morphologies consistent with the DAMA signal. New data from DAMA/LIBRA–phase2 will certainly aid in determining whether any apparent time dependence is a real effect or a statistical fluctuation.

^a e-mail: ckelso@unf.edu

^b e-mail: chris@savage.name

^c e-mail: sandick@physics.utah.edu

^d e-mail: kfreese@nordita.org

^e e-mail: paolo.gondolo@utah.edu

1 Introduction

The nature of dark matter is one of the most compelling mysteries in both cosmology and particle physics. One of the foremost possibilities is that dark matter is a new fundamental particle that is yet to be discovered. There is currently an extensive experimental effort to try to detect the particle nature of dark matter through its interactions with Standard Model particles [1, 2]. Among the leading dark matter candidates are Weakly Interacting Massive Particles (WIMPs), a generic class of particles that includes, for example, the supersymmetric neutralino. These particles interact gravitationally and through the weak force, and their expected masses range from $\mathcal{O}(1)$ GeV to $\mathcal{O}(10)$ TeV. More than thirty years ago, Refs. [3, 4] proposed a mechanism for detecting weakly interacting particles, including WIMPs, via coherent elastic scattering with nuclei. Soon after, the dark matter detection rates in the context of a Galactic Halo of WIMPs were computed for the first time, and it was proposed that they could be differentiated from background by looking for an annual modulation of the signal [5].

The DAMA collaboration has constructed one of these “direct detection” experiments to search for the annual modulation of nuclear recoils due to dark matter scattering within NaI scintillation detectors. The original detectors were deployed in 1995 and consisted of 100 kg of NaI [6]. The experiment went through an upgrade in 2002 that included increasing the detector mass to 250 kg and improving the photo-multiplier tubes that detect the scintillation light from the nuclear recoils. Through the seven years that the original DAMA/NaI experiment ran, the total exposure was 1.08×10^5 kg-days [7, 8]. The upgraded DAMA/LIBRA experiment collected a larger exposure of 3.80×10^5 kg-days over the seven years it took data [9–11]. The DAMA collaboration has identified at a 9.3σ confidence level an annual modulation in their event rate that is consistent with

a dark matter signal [12–19]. Thus far, no plausible alternative explanation for the modulation signal observed in the DAMA detectors has been universally accepted [20]. Meanwhile, there are three upcoming NaI experiments that will test the DAMA modulation results: SABRE [21], which will consist of two identical experiments located in the northern and southern hemispheres; ANAIS-112 [22, 23], which, as of April 2017, is commissioning at the Canfranc Underground Laboratory; and COSINE-100 [24], a joint effort between the DM-Ice [25] and KIMS [26] collaborations, which has been running at Yangyang Underground Laboratory since September 2016. Furthermore, DAMA/LIBRA has been running in an upgraded phase 2 configuration since January 2011 [27], and may also shed light on past DAMA modulation results [19].

In this work we examine the amplitude of the modulation signal observed by the DAMA collaboration, focusing on the possibility of any time dependence of the amplitude over the 14 years of operation of the DAMA/NaI and DAMA/LIBRA experiments. We find that descriptions of the modulation amplitude as decreasing with time are preferred over a constant modulation amplitude at anywhere between 1σ and 3σ , depending on the phenomenological model for the time dependence and the signal energy range considered.

In Sect. 2 we remind the reader of the expected form for the modulation signal. In Sect. 3 we present our analysis of the modulation amplitudes and their potential time dependence. We discuss the results and draw conclusions in Sect. 4.

2 Annual modulation

If dark matter is composed of a new, as-yet-unknown particle with weak interactions with ordinary matter, then dark matter particles in the galactic halo will pass through the Earth and occasionally scatter on nuclei, causing a nucleus to recoil. Dark matter direct detection experiments aim to observe these rare, low-energy nuclear recoils induced by collisions with a dark matter particle [4]. The diffuse dark matter halo has little bulk rotation so, from the perspective of the Earth as it orbits about the center of the galaxy (along with the rest of the galactic disk) and through the halo, there is a wind of dark matter approaching from the direction of the disk rotation. The form of the expected signal (measured in counts per mass per time per energy) in a direct detection experiment is often written as

$$\frac{dR}{dE}(E, t) = S_0(E) + S_m(E) \cos[\omega(t - t_0)] + \dots, \quad (1)$$

where $\omega \equiv 2\pi/\text{year}$, S_0 is the time-averaged differential recoil rate, S_m is the annual modulation amplitude, and higher-order terms in the expansion are suppressed. The

phase of the expansion t_0 is approximately the time of year at which the Earth is moving fastest into the dark matter wind, around the beginning of June for the Standard Halo Model (SHM), a common first-order approximation to the dark matter halo [5, 28]. For a recent review of status of the search for the annual modulation signal from dark matter including alternatives to the SHM, see Ref. [29].

3 Analysis

3.1 DAMA/NaI vs. DAMA/LIBRA

In this subsection, we gather the best fit modulation amplitudes for the DAMA/NaI and DAMA/LIBRA exposures, as well as the combined best fit amplitudes for the entire 14 years of operation for which the modulation amplitudes have been published. The DAMA collaboration provides the best fit modulation amplitudes for the 2–4, 2–5, and 2–6 keVee energy ranges. However, we would like to explore the signal in more detail, specifically the higher end of the energy range.

An estimation for the modulation amplitudes in the 4–6, 4–5, and 5–6 keVee energy ranges, which are not provided in the DAMA collaboration publications, can be obtained from the known modulation amplitudes available from the collaboration using the following procedure. For each energy interval, there is a modulation amplitude (S_m), as well as an average efficiency for the detector to detect a recoil (ε). If this interval is split into two ranges, one can write

$$\varepsilon S_m = f_1 \varepsilon_1 S_{m_1} + f_2 \varepsilon_2 S_{m_2}, \quad (2)$$

with the subscripts 1 and 2 referring to the two energy ranges, and $f_{1,2}$ representing the fraction of the energy interval that each of the two ranges covers. For example, S_m and S_{m_1} could be the published 2–5 and 2–4 keVee amplitudes, and S_{m_2} could be estimated using Eq. 2. This relationship comes from the fact that the DAMA experiment is fundamentally a counting experiment with the number of events in each range proportional to the total exposure. Similarly, the uncertainty (σ) can be expressed as

$$(\varepsilon\sigma)^2 = (f_1 \varepsilon_1 \sigma_1)^2 + (f_2 \varepsilon_2 \sigma_2)^2, \quad (3)$$

with $\sigma_{1,2}$ representing the uncertainty in each of the relevant energy ranges. This procedure assumes the data for the different energy ranges modulate with the same period and phase but statistically independent, i.e. the 2–5 keVee fit is the appropriately weighted average of fits to the 2–4 and 4–5 keVee ranges. In the case where the phase and period are fixed, this assumption is valid. When the phase and period are allowed to vary, however, the assumption is not strictly

Table 1 Best-fit modulation amplitudes for the DAMA/NaI, DAMA/LIBRA, and combined exposures. Energy-interval-averaged efficiencies are calculated from the overall efficiencies in Fig. 17 of Ref. [6] and Fig. 26 in Ref. [30]. Modulation amplitudes for the 2–4, 2–5, and 2–6 keVee energy ranges are fits from the DAMA collaboration, taken from Refs. [8, 11, 30]. (†) Modulation amplitudes for the 4–6, 4–5, and 5–6 keVee energy ranges are estimates derived from the existing analysis intervals, as described in the text (caveats apply)

	DAMA/NaI	DAMA/LIBRA	Combined
Exposure (kg-day)	1.08×10^5	3.80×10^5	4.87×10^5
Average efficiency			
2–4 keVee	0.54	0.61	0.59
2–5 keVee	0.63	0.65	0.65
2–6 keVee	0.70	0.69	0.69
4–6 keVee [†]	0.85	0.78	0.80
4–5 keVee [†]	0.80	0.74	0.75
5–6 keVee [†]	0.91	0.82	0.84
Modulation amplitude, fixed period and phase (/kg/day/keVee)			
2–4 keVee	0.0233 ± 0.0047	0.0167 ± 0.0022	0.0179 ± 0.0020
2–5 keVee	0.0210 ± 0.0038	0.0122 ± 0.0016	0.0135 ± 0.0015
2–6 keVee	0.0192 ± 0.0031	0.0096 ± 0.0013	0.0110 ± 0.0012
4–6 keVee [†]	0.0166 ± 0.0041	0.0041 ± 0.0016	0.0059 ± 0.0015
4–5 keVee [†]	0.0179 ± 0.0063	0.0048 ± 0.0022	0.0066 ± 0.0022
5–6 keVee [†]	0.0155 ± 0.0054	0.0034 ± 0.0022	0.0052 ± 0.0019
Modulation amplitude, free period and phase (/kg/day/keVee)			
2–4 keVee	0.0252 ± 0.0050	0.0178 ± 0.0022	0.0190 ± 0.0020
2–5 keVee	0.0213 ± 0.0039	0.0127 ± 0.0016	0.0140 ± 0.0015
2–6 keVee	0.0200 ± 0.0032	0.0097 ± 0.0013	0.0112 ± 0.0012
4–6 keVee [†]	0.0167 ± 0.0042	0.0034 ± 0.0016	0.0054 ± 0.0015
4–5 keVee [†]	0.0161 ± 0.0062	0.0043 ± 0.0022	0.0062 ± 0.0022
5–6 keVee [†]	0.0173 ± 0.0056	0.0026 ± 0.0022	0.0047 ± 0.0019

correct. In looking at the fits provided by the collaboration in Tables 3 and 4 of Ref. [11], we see that the largest difference in modulation amplitude between the fixed versus free fits is at the $\sim 0.5\sigma$ level. We thus conclude that the assumption that the energy bins modulate with the same period and phase is approximately correct to quite good accuracy. This caveat should be kept in mind, however, when examining our results for the derived data in the higher energy ranges.

Table 1 presents the results that have been released to date by the DAMA collaboration along with our estimated data for the higher energy ranges. The best-fit modulation amplitudes are listed under the assumption of a fixed period and phase, as well as allowing for a free period and phase. The modulation amplitudes displayed for the 2–4, 2–5, and 2–6 keVee energy ranges are the fits performed by the collaboration as found in Refs. [8, 11, 30], while the modulation amplitudes for the 4–6, 4–5, and 5–6 keVee energy ranges, as well as the efficiencies, are calculated as explained above, and we use the data in Refs. [6, 30] to calculate the average efficiencies.

In comparing the amplitudes for DAMA/NaI to DAMA/LIBRA as shown in Table 1, we find that the amplitudes decrease in every energy range. This points towards a possible inconsistency in the results between the two incarnations of the experiment. To test the hypothesis of a time-varying

modulation amplitude, we begin by examining the compatibility of the DAMA/NaI and DAMA/LIBRA fit results of Table 1, under the assumption of a common modulation amplitude. In Table 2, we present the minimum χ^2 and corresponding p -value for a χ^2 goodness-of-fit, as well as the best-fit common amplitude. We find the 2–6 keVee energy range is discrepant at the $\sim 3\sigma$ level for both the fixed and free period and phase. Our estimated data indicates that the majority of the cause of the poor fit is due to events with energies $\gtrsim 4$ keVee, as the 4–6 keVee interval is similarly discrepant. Note that the best fit modulation amplitudes in Table 2 match very well the values presented by the collaboration for the combined fits in the third column of Table 1, as expected.

We also collect the fits for the per-cycle modulation amplitudes as performed by the DAMA collaboration for the free period and phase in Table 3. Unfortunately the fits to the data under the assumption of a fixed period and phase have not been released to the public, and we are thus unable to analyze the data in that case. The mean cycle times and amplitudes are taken from Figure 3 of Ref. [11] and exposures are given in Refs. [8, 11]. The mean time for each cycle is relative to January 1, 1995, the first year of DAMA/NaI’s operation.

Table 2 Compatibility of the DAMA/NaI and DAMA/LIBRA fit results of Table 1, under the assumption of a common modulation amplitude. The minimum χ^2 and corresponding p -value for a χ^2 goodness-of-fit are shown, as well as the best-fit common amplitude. (†) Data for these energy intervals are derived as discussed in the text

	χ^2/dof	p	S_m
Fixed period and phase			
2–4 keVee	1.61/1	0.203 (1.3 σ)	0.0179 \pm 0.0020
2–5 keVee	4.55/1	0.033 (2.1 σ)	0.0135 \pm 0.0015
2–6 keVee	8.16/1	0.004 (2.9 σ)	0.0110 \pm 0.0012
4–6 keVee [†]	8.19/1	0.004 (2.9 σ)	0.0057 \pm 0.0015
4–5 keVee [†]	3.87/1	0.049 (1.9 σ)	0.0063 \pm 0.0021
5–6 keVee [†]	4.34/1	0.037 (2.1 σ)	0.0054 \pm 0.0020
Free period and phase			
2–4 keVee	1.84/1	0.176 (1.4 σ)	0.0190 \pm 0.0020
2–5 keVee	4.16/1	0.041 (2.0 σ)	0.0139 \pm 0.0015
2–6 keVee	8.89/1	0.003 (3.0 σ)	0.0112 \pm 0.0012
4–6 keVee [†]	9.01/1	0.003 (3.0 σ)	0.0050 \pm 0.0015
4–5 keVee [†]	3.18/1	0.074 (1.8 σ)	0.0057 \pm 0.0021
5–6 keVee [†]	6.01/1	0.014 (2.5 σ)	0.0045 \pm 0.0020

Table 3 The per-cycle modulation amplitudes as determined by the DAMA collaboration, allowing the period and phase to be freely fit in addition to the amplitude. Mean cycle times and amplitudes are taken

from Figure 3 of Ref. [11] and exposures are taken from Refs. [8,11]. The mean time for each cycle is given relative to January 1, 1995, the first year of DAMA/NaI's operation

	Exposure (kg-days)	$\langle t \rangle$ (year)	S_m (2–4 keVee) (/kg/day/keVee)	S_m (2–5 keVee) (/kg/day/keVee)	S_m (2–6 keVee) (/kg/day/keVee)
DAMA/NaI					
Cycle 1	4549	1.19	0.005 \pm 0.021	0.017 \pm 0.016	0.024 \pm 0.012
Cycle 2	14962	2.23	0.022 \pm 0.012	0.019 \pm 0.009	0.023 \pm 0.008
Cycle 3	22455	3.15	0.017 \pm 0.012	0.012 \pm 0.010	0.025 \pm 0.008
Cycle 4	16020	4.17	0.023 \pm 0.013	0.023 \pm 0.010	0.018 \pm 0.009
Cycle 5	15911	5.10	0.038 \pm 0.013	0.025 \pm 0.010	0.017 \pm 0.009
Cycle 6	16608	6.20	0.021 \pm 0.011	0.021 \pm 0.009	0.011 \pm 0.007
Cycle 7	17226	7.06	0.028 \pm 0.012	0.028 \pm 0.009	0.021 \pm 0.008
DAMA/LIBRA					
Cycle 1	51405	9.12	0.0294 \pm 0.0064	0.0181 \pm 0.0048	0.0098 \pm 0.0038
Cycle 2	52597	10.19	0.0194 \pm 0.0072	0.0134 \pm 0.0053	0.0089 \pm 0.0044
Cycle 3	39445	11.19	0.0172 \pm 0.0062	0.0173 \pm 0.0045	0.0122 \pm 0.0038
Cycle 4	49337	12.04	0.0192 \pm 0.0061	0.0181 \pm 0.0045	0.0125 \pm 0.0036
Cycle 5	66105	13.11	0.0105 \pm 0.0053	0.0091 \pm 0.0041	0.0086 \pm 0.0031
Cycle 6	58768	14.27	0.0150 \pm 0.0052	0.0070 \pm 0.0039	0.0102 \pm 0.0030
Cycle 7	62098	15.17	0.0166 \pm 0.0050	0.0097 \pm 0.0036	0.0092 \pm 0.0028

The DAMA Collaboration has employed “run tests” as a check for consistency of the data between the different years of the experimental data. There are many different possible run tests that can be performed on a given set of data to check for consistency. In this context, a run is a group of consecutive identical elements in a two-valued random sequence constructed from the data. A run test checks for runs in the data, in terms of a two-valued characteristic one can assign to each data point. The collaboration has performed run tests to determine whether all data are consistent with the best fit

value for the amplitude, and found lower tail probabilities of 41, 29, and 23% for the 2–4, 2–5, and 2–6 keVee energy ranges, respectively. We have repeated the DAMA collaboration run tests and reproduced their results (see Table 4; Above and Below Best Fit). Here, we are interested in the consistency between the two iterations of the DAMA experiment, NaI and LIBRA, and we perform several other tests of interest as described below.

For the DAMA data, there are a total of 7 data points for each of the two iterations of the experiment (for each energy

bin). To check whether the NaI and LIBRA data are consistent with each other, we rank the full 14 measurements from low to high, keeping track of which experiment made the given measurement. The number of runs is then one plus the number of times in this list the subsequent element changes from one version of the experiment to the other. As an example, when the amplitude values for the 2–6 keVee energy range are sorted in this way, the list would be (LIBRA, LIBRA, LIBRA, LIBRA, LIBRA, NaI, LIBRA, LIBRA, NaI, NaI, NaI, NaI, NaI, NaI), giving a total of four runs in this list. In this version of the run test, the null hypothesis would be that the two distributions are equal, i.e. that the DAMA/NaI and DAMA/LIBRA experiments are measuring the same amplitude in a given energy bin. The alternative hypothesis is that the two distributions are not equal.

Under the null hypothesis, i.e. that the two populations of data are drawn from the same distribution, the probability of obtaining a certain number of runs can be calculated using

$$P(R = 2k) = 2 \frac{\binom{n_1 - 1}{k - 1} \binom{n_2 - 1}{k - 1}}{\binom{n_1 + n_2}{n_1}}, \tag{4}$$

where R is the number of observed runs, n_1 is the number of values from one population or experiment, and n_2 is the number of values from the other population or experiment, and k is an integer that gives the appropriate number of runs. Similarly, if the number of runs is odd, the probability of obtaining that number of runs can be written as

$$P(R = 2k + 1) = \frac{\binom{n_1 - 1}{k} \binom{n_2 - 1}{k - 1} + \binom{n_2 - 1}{k} \binom{n_1 - 1}{k - 1}}{\binom{n_1 + n_2}{n_1}}. \tag{5}$$

The lower tail probability can then be computed by summing the probability of two runs (there must be at least two runs) up to the actual number of runs observed. To be concrete, there are four runs in the case of the 2–6 keVee energy bin, so the probability would then be summed for obtaining, two, three, and four runs. The results for this run test applied to the data in Table 3 are collected in Table 4 under the heading of Ranked by Experiment. Of all of the possible run tests, this one is the most relevant to the question at hand. This test directly examines if the two iterations of the experiment are consistent with one another. We find that the null hypothesis for the 2–6 keVee energy range would be rejected at the 2σ level, while the 2–4 and 2–5 keVee energy ranges have p -values of 30 and 21%, respectively.

Other possible run tests focus on fluctuations above and below a reference point, i.e. the median, mean, best fit point,

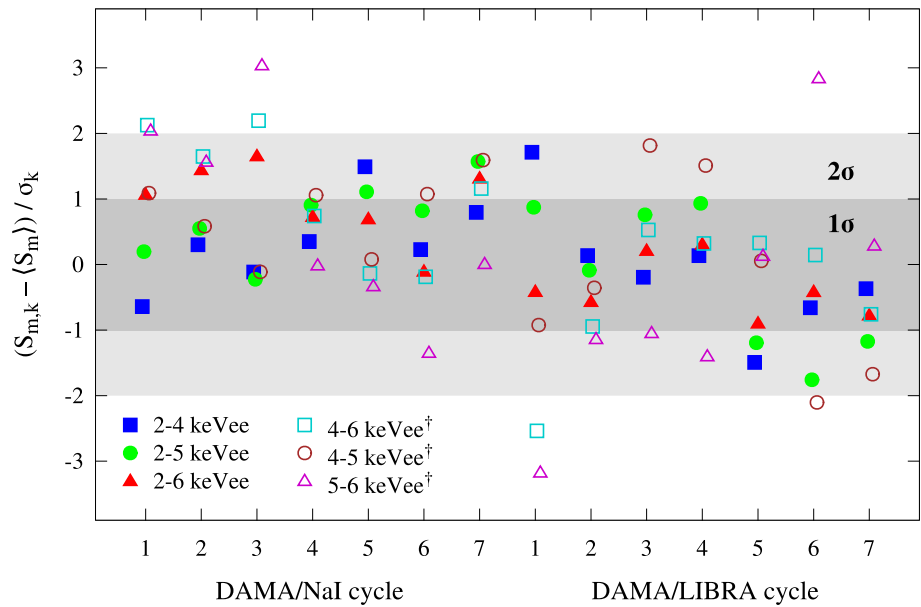
Table 4 Results for the run test as presented by the collaboration and for the tests we have performed as described in the text. The corresponding p -value is for a hypothesis test where the null hypothesis is taken to be the case where the two iterations of the experiment are drawing measurements from the same distribution

	Number of runs	p
Ranked by experiment		
2–4 keVee	7	0.383 (0.30 σ)
2–5 keVee	6	0.208 (0.81 σ)
2–6 keVee	4	0.0251 (2.0 σ)
Above and below best fit		
2–4 keVee	7	0.413 (0.22 σ)
2–5 keVee	6	0.287 (0.56 σ)
2–6 keVee	6	0.226 (0.75 σ)
Above and below median		
2–4 keVee	5	0.0775 (1.4 σ)
2–5 keVee	7	0.383 (0.30 σ)
2–6 keVee	6	0.209 (0.81 σ)
KS test		
2–4 keVee	–	0.21 (0.80 σ)
2–5 keVee	–	0.053 (1.6 σ)
2–6 keVee	–	0.0082 (2.4 σ)

etc. In this case, the measurements are put in an array in the order in which they were measured. Then a new list is formed where at each element, a ‘+’ sign or ‘–’ is placed depending on if the measurement is above or below the reference point. The number of runs are then counted in this list and the probabilities calculated exactly as in the previous case with n_1 and n_2 now representing the number of points above and below the reference value. The null hypothesis for this run test is that the data points are randomly fluctuating about the reference value. A run test of this type with reference to the best fit point is employed by the DAMA collaboration, and we reproduce their p -values. The results for these run tests are presented in Table 4. These tests all point towards consistency with the assumption of randomly fluctuating about the reference value. Although the Ranked By Experiment test is the most relevant to our discussion, the run tests in aggregate produce inconclusive results.

Due to the inconclusive results of the run tests, we also utilized the Kolmogorov–Smirnov (KS) test for the two data sets. The KS test is another non-parametric test in which the null hypothesis is that two data sets are drawn from the same distribution. As can be seen in Table 4, we find that the null hypothesis has p -values of 21 and 5% in the 2–4 and 2–5 keVee energy ranges, respectively, and that it is rejected at the 2.4σ level in the 2–6 keVee energy range. The KS test and the most relevant run test both point towards a possible inconsistency in the 2–6 keVee energy range. One possible explanation for these results, which will be explored in the

Fig. 1 The deviations of the modulation amplitude in each annual cycle for a given energy range compared to the average in that range. $S_{m,k}$ and σ_k are the amplitude and uncertainty in the k th annual cycle, respectively, and $\langle S_m \rangle$ is the average amplitude in the particular energy range for the entire 14 year period



remainder of this paper, is that the modulation amplitude is changing (seemingly decaying) with time.

3.2 Annual cycles

Given the apparent decrease in the modulation amplitude from DAMA/NaI to DAMA/LIBRA (the amplitude decreases in every energy range in Table 1), it is interesting to investigate in detail the time dependence of the modulation amplitude. Here, we postulate several phenomenological models for the time dependence of the modulation amplitude, and perform likelihood fits to investigate whether or not any model is favored by the data.

A visual representation of the deviations of the amplitude in each annual cycle from the average is shown in Fig. 1. For each energy interval considered, we plot the deviation in each annual cycle as $\frac{S_{m,k} - \langle S_m \rangle}{\sigma_k}$ where $S_{m,k}$ and σ_k are the amplitude and uncertainty in the k th annual cycle, respectively, and $\langle S_m \rangle$ is the uncertainty-weighted average amplitude in the particular energy range over the entire 14 year period. We observe that all the published data (in the 2–4, 2–5, and 2–6 keVee energy ranges) fall within 2σ of the average. A Shapiro–Wilk test of normality of the deviations gives p -values of 73, 10, and 58% for these published energy ranges, respectively. Low p -values point to a non-normal distribution of the published deviations. The most striking feature in the plot is the fact that among the 14 annual cycles, there are three $> 2.8\sigma$ departures from the average in the 5–6 keVee energy range. If the data in this range were gaussian distributed, then the probability of this occurring is 2×10^{-5} (4.2σ). This seemingly unlikely situation is an indication that there may be a problem with the data in this energy range.

To describe a potential time dependence of the modulation amplitude, we have performed likelihood tests for the following models: a constant modulation amplitude, separate constants for NaI and LIBRA, linear in time, exponential in time, and a broken exponential (a common exponential scale, but independent normalizations for NaI and LIBRA). We estimate the per-cycle modulation amplitudes for the 4–6, 4–5, and 5–6 keVee energy ranges using Eqs. (2) and (3) and these values are shown in Table 5. Fits for these models to the per-cycle DAMA modulation amplitude data from Tables 3 and 5 are presented in Table 6. The minimum χ^2 and corresponding p -value for a χ^2 goodness-of-fit are shown. Best-fit parameters are found in the final column and parameters not shown are profiled over. We have highlighted in bold any hypotheses that are discrepant at $\geq 2.5\sigma$.

Based on the goodness of fit test, we find that all the models we investigated (including the single constant amplitude) provide reasonable fits to each of the energy ranges, with the exception of the 5–6 keVee range. This range is excluded at nearly 4σ by the goodness of fit test for every model we studied. This is the quantitative conclusion regarding the odd behavior in the 5–6 keVee range observed in Fig. 1.

Finally, we compared models (hypothesis tests) against a single, constant modulation amplitude taken as the null hypothesis. The results are displayed in Table 6 with the improvement in fit given by the $\Delta\chi^2$ along with the corresponding p -value. These tests quantify the level of improvement in the fits above the amount expected due to adding additional parameters, assuming the null hypothesis is correct. The improvements in fit for the 2–4 keVee energy range are marginal (1.0σ to 1.6σ), indicating consistency in this energy range. We determine, however, that all time-varying models we investigated are preferred over a constant ampli-

Table 5 The per-cycle modulation amplitudes for the 4–6, 4–5, and 5–6 keVee energy ranges, derived from the data in Table 3 as described in the text. Exposures and mean times for each cycle are the same as in Table 3

	S_m (4–6 keVee) (/kg/day/keVee)	S_m (4–5 keVee) (/kg/day/keVee)	S_m (5–6 keVee) (/kg/day/keVee)
DAMA/NaI			
Cycle 1	0.036 ± 0.014	0.033 ± 0.025	0.039 ± 0.016
Cycle 2	0.024 ± 0.011	0.015 ± 0.015	0.032 ± 0.016
Cycle 3	0.030 ± 0.011	0.005 ± 0.016	0.052 ± 0.015
Cycle 4	0.014 ± 0.011	0.023 ± 0.016	0.007 ± 0.016
Cycle 5	0.004 ± 0.011	0.008 ± 0.015	0.001 ± 0.016
Cycle 6	0.004 ± 0.009	0.021 ± 0.014	−0.011 ± 0.013
Cycle 7	0.017 ± 0.010	0.028 ± 0.014	0.007 ± 0.013
DAMA/LIBRA			
Cycle 1	−0.0053 ± 0.0044	−0.0003 ± 0.0073	−0.0099 ± 0.0053
Cycle 2	0.0008 ± 0.0054	0.0037 ± 0.0076	−0.0019 ± 0.0077
Cycle 3	0.0083 ± 0.0046	0.0176 ± 0.0062	−0.0001 ± 0.0067
Cycle 4	0.0073 ± 0.0043	0.0163 ± 0.0066	−0.0009 ± 0.0056
Cycle 5	0.0071 ± 0.0037	0.0068 ± 0.0062	0.0075 ± 0.0043
Cycle 6	0.0064 ± 0.0034	−0.0060 ± 0.0059	0.0176 ± 0.0038
Cycle 7	0.0035 ± 0.0031	−0.0016 ± 0.0048	0.0081 ± 0.0041

tude in the 2–6 keVee energy range at levels from 2.3σ to 2.7σ . As the 5–6 keVee data may be somewhat suspect, we find that the alternative models are still preferred over a constant amplitude in the 2–5 keVee range at the level of 2.0σ to 2.7σ . Although the data clearly indicate that some form of time dependence is preferred, none of the alternative models is clearly better than the others. Unfortunately, the data is not capable of distinguishing between the different functional forms proposed here to model the time dependence of the modulation amplitude.

In comparing the results from the constant amplitude from Tables 2 and 6, one might come to the conclusion that the results are contradictory. A constant amplitude is inconsistent at the $1-3\sigma$ level in Table 2, whereas the inconsistency is much less in 2–4, 2–5 and 2–6 keVee ranges in Table 6. One possible explanation for this result is that the inconsistencies are statistically diluted by the additional degrees of freedom that are included in Table 6. This explanation is supported by the fact that the values for the $\Delta\chi^2$ test for broken constant model shown in Table 6 are quite consistent with the values from Table 2. These two results are the most direct comparison between the two tables.

4 Discussion and conclusions

In this study, we have analyzed the annual modulation data released by the DAMA collaboration that they collected over 14 annual cycles and performed likelihood analyses for several phenomenological models to explore a possible time dependence in the modulation amplitude. Although a con-

stant modulation amplitude is already a good fit to the data, our results indicate that all of the models we examined are preferred over a constant amplitude at up to $\sim 3\sigma$. Although the data clearly prefer some form of time dependence, they are currently incapable of distinguishing between the different functional forms for the time dependence we investigated.

Though we have identified the fact that the modulation amplitude is discrepant at the $2-3\sigma$ level for each of the two versions of the experiment in all but the 2–4 keVee energy range, we do not propose a physical explanation for the phenomenon. More data will certainly aid in determining whether this is a real effect or a statistical fluctuation. The DAMA/LIBRA experiment has undergone a significant upgrade, and has been taking data in the DAMA/LIBRA-phase 2 configuration since January 2011 [27]. In November of 2010, all of the photo-multiplier tubes (PMT’s) in the DAMA/LIBRA experiment were replaced by high quantum efficiency PMT’s [31]. The anticipated lower thresholds (~ 1 keVee) will give access to a new signal range that should help to clarify the dark matter interpretation of the DAMA signal, as described in Ref [19].

It is also interesting to note that the half life for the exponential decay model in the 2–5 keVee energy range is about 11 years. If the collaboration releases data again in the near future, both the linear and exponential decay models would predict a noticeable decrease in the modulation amplitude, somewhere in the neighborhood of 50%. If this trend continues, this would have serious implications for the dark matter interpretation of the DAMA modulation.

In this study we have attempted to answer the question of whether the DAMA collaboration data demonstrates any

Table 6 Fits of the per-cycle DAMA modulation amplitude data of Table 3 and Table 5. The minimum χ^2 and corresponding p -value for a χ^2 goodness-of-fit are shown, followed by the $\Delta\chi^2$ and corresponding p -value for a hypothesis test where the conventional constant modulation amplitude is taken as the null hypothesis. Best-fit parameters are shown in the final column; parameters not shown are profiled over

	Goodness-of-fit		vs. null hypothesis		Best-fit parameters
	χ^2/dof	p	$\Delta\chi^2/dof$	p	
Constant (null hypothesis): $S_m(t) = A$					
	A (dru)				
2–4 keVee	9.36/13	0.746 (0.3 σ)	–	–	0.0184 \pm 0.0020
2–5 keVee	13.7/13	0.396 (0.8 σ)	–	–	0.0139 \pm 0.0015
2–6 keVee	10.9/13	0.618 (0.5 σ)	–	–	0.0114 \pm 0.0012
4–6 keVee [†]	22.4/13	0.049 (2.0 σ)	–	–	0.0059 \pm 0.0014
4–5 keVee [†]	20.1/13	0.092 (1.7 σ)	–	–	0.0064 \pm 0.0022
5–6 keVee [†]	40.4/13	0.0001 (3.8σ)	–	–	0.0069 \pm 0.0018
Broken constant: $S_m(t) = A_N$ (NaI), A_L (LIBRA)					
	A_N, A_L (dru)				
2–4 keVee	8.10/12	0.777 (0.3 σ)	1.26/1	0.262 (1.1 σ)	0.0234, 0.0174
2–5 keVee	9.19/12	0.686 (0.4 σ)	4.49/1	0.034 (2.1 σ)	0.0211, 0.0125
2–6 keVee	3.62/12	0.989 (0.01 σ)	7.29/1	0.007 (2.7σ)	0.0192, 0.0101
4–6 keVee [†]	14.2/12	0.289 (1.1 σ)	8.25/1	0.004 (2.9σ)	0.0169, 0.0044
4–5 keVee [†]	15.4/12	0.220 (1.2 σ)	4.73/1	0.030 (2.2 σ)	0.0182, 0.0045
5–6 keVee [†]	37.0/12	0.0002 (3.7σ)	3.39/1	0.066 (1.8 σ)	0.0168, 0.0059
Linear: $S_m(t) = mt + b$					
	m (10 ^{–3} dru/year)				
2–4 keVee	6.85/12	0.867 (0.2 σ)	2.50/1	0.114 (1.6 σ)	– 0.87 \pm 0.55
2–5 keVee	6.54/12	0.886 (0.1 σ)	7.15/1	0.008 (2.7σ)	– 1.12 \pm 0.42
2–6 keVee	4.40/12	0.975 (0.03 σ)	6.51/1	0.011 (2.6σ)	– 0.87 \pm 0.34
4–6 keVee [†]	18.8/12	0.092 (1.7 σ)	3.58/1	0.058 (1.9 σ)	– 0.80 \pm 0.42
4–5 keVee [†]	13.5/12	0.335 (1.0 σ)	6.65/1	0.010 (2.6σ)	– 1.63 \pm 0.63
5–6 keVee [†]	40.3/12	0.00006 (4.0σ)	0.11/1	0.736 (0.3 σ)	0.18 \pm 0.55
Exponential: $S_m(t) = Ae^{-\beta t}$					
	β (/year)				
2–4 keVee	7.14/12	0.849 (0.2 σ)	2.22/1	0.136 (1.5 σ)	0.039 \pm 0.025
2–5 keVee	7.59/12	0.816 (0.2 σ)	6.10/1	0.014 (2.5σ)	0.061 \pm 0.023
2–6 keVee	3.64/12	0.989 (0.01 σ)	7.27/1	0.007 (2.7σ)	0.072 \pm 0.024
4–6 keVee [†]	14.9/12	0.245 (1.2 σ)	7.48/1	0.006 (2.7σ)	0.18 \pm 0.07
4–5 keVee [†]	15.3/12	0.223 (1.2 σ)	4.79/1	0.029 (2.2 σ)	0.14 \pm 0.06
5–6 keVee [†]	35.6/12	0.0004 (3.6σ)	4.79/1	0.029 (2.2 σ)	– 0.39 \pm 0.18
Broken exponential: $S_m(t) = A_N e^{-\beta t}$ (NaI), $A_L e^{-\beta t}$ (LIBRA)					
	β (/year)				
2–4 keVee	6.93/11	0.805 (0.2 σ)	2.43/2	0.297 (1.0 σ)	0.062 \pm 0.058
2–5 keVee	7.57/11	0.751 (0.3 σ)	6.12/2	0.047 (2.0 σ)	0.068 \pm 0.058
2–6 keVee	3.03/11	0.990 (0.01 σ)	7.88/2	0.019 (2.3 σ)	0.038 \pm 0.050
4–6 keVee [†]	13.6/11	0.253 (1.1 σ)	8.78/2	0.012 (2.5σ)	0.07 \pm 0.11
4–5 keVee [†]	15.1/11	0.178 (1.3 σ)	5.03/2	0.081 (1.7 σ)	0.08 \pm 0.14
5–6 keVee [†]	34.7/11	0.0003 (3.6σ)	5.73/2	0.057 (1.9 σ)	– 0.30 \pm 0.20

Bold values highlights any hypotheses that are discrepant at $\geq 2.5\sigma$
 A differential rate unit (dru) is equal to 1 count/kg/day/keVee
 (†) Data for these energy intervals are derived as discussed in the text

time dependence, and the answer to that question appears to be “yes,” at the 2 – 3 σ significance level in all but the 2–4 keVee energy range. A more conclusive answer will of course require additional data, which we look forward to in the near future, from the DAMA collaboration as well as from ANAIS-112, COSINE-100, and SABRE, each of

which will test the DAMA modulation results. Furthermore, given the questions raised here as well as the tension of the DAMA results with results from other dark matter direct detection experiments, we urge the DAMA collaboration and other experimental collaborations to make enough of their data public (i.e. providing time-stamped events) in order for

researchers to repeat their analyses to encourage scrutiny and, with any luck, a consistent explanation for the observed phenomena, dark matter or otherwise.

The Mathematica code that we have used to complete our analysis is freely available on the arXiv.

Acknowledgements KF acknowledges support from the Swedish Research Council (Vetenskapsrådet) through the Oskar Klein Centre (Contract No. 638-2013-8993). KF acknowledges support from DoE Grant DE-SC007859 at the University of Michigan. PS is supported in part by NSF Grant No. PHY-1417367. PG was partially supported by NSF Award PHY-1415974. CK, PS, and CS thank the Department of Physics and Astronomy at the University of Utah for support. PG, CK, PS, and CS thank Nordita, the Nordic Institute for Theoretical Physics, for its hospitality. This research was supported in part by the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence “Origin and Structure of the Universe”. We would also like to thank Peter Smith for helpful comments on the draft of this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. Funded by SCOAP³.

References

- M. Drees, G. Gerbier (2012). [arXiv:1204.2373](https://arxiv.org/abs/1204.2373)
- J. Beringer et al., (Particle Data Group), Phys. Rev. D **86**, 010001 (2012)
- A. Drukier, L. Stodolsky, Phys. Rev. D **30**, 2295 (1984)
- M.W. Goodman, E. Witten, Phys. Rev. D **31**, 3059 (1985)
- A. Drukier, K. Freese, D. Spergel, Phys. Rev. D **33**, 3495 (1986)
- R. Bernabei et al., (DAMA), Il Nuovo Cimento A **112**, 545 (1999)
- R. Bernabei et al., (DAMA), Phys. Lett. B **480**, 23 (2000)
- R. Bernabei, P. Belli, F. Cappella, R. Cerulli, F. Montecchia, et al., Riv. Nuovo Cim. **26N1**, 1 (2003). [arXiv:astro-ph/0307403](https://arxiv.org/abs/astro-ph/0307403)
- R. Bernabei et al., (DAMA), Eur. Phys. J. C **56**, 333 (2008). [arXiv:0804.2741](https://arxiv.org/abs/0804.2741)
- R. Bernabei et al., (DAMA, LIBRA), Eur. Phys. J. C **67**, 39 (2010). [arXiv:1002.1028](https://arxiv.org/abs/1002.1028)
- R. Bernabei, P. Belli, F. Cappella, V. Caracciolo, S. Castellano et al., Eur. Phys. J. C **73**, 2648 (2013). [arXiv:1308.5109](https://arxiv.org/abs/1308.5109)
- A. Bottino, F. Donato, N. Fornengo, S. Scopel, Phys. Rev. D **68**, 043506 (2003). [arXiv:hep-ph/0304080](https://arxiv.org/abs/hep-ph/0304080)
- A. Bottino, F. Donato, N. Fornengo, S. Scopel, Phys. Rev. D **69**, 037302 (2004). [arXiv:hep-ph/0307303](https://arxiv.org/abs/hep-ph/0307303)
- P. Gondolo, G. Gelmini, Phys. Rev. D **71**, 123520 (2005). [arXiv:hep-ph/0504010](https://arxiv.org/abs/hep-ph/0504010)
- F. Petriello, K.M. Zurek, JHEP **0809**, 047 (2008). [arXiv:0806.3989](https://arxiv.org/abs/0806.3989)
- S. Chang, A. Pierce, N. Weiner, Phys. Rev. D **79**, 115011 (2009). [arXiv:0808.0196](https://arxiv.org/abs/0808.0196)
- M. Fairbairn, T. Schwetz, JCAP **0901**, 037 (2009). [arXiv:0808.0704](https://arxiv.org/abs/0808.0704)
- C. Savage, G. Gelmini, P. Gondolo, K. Freese, JCAP **0904**, 010 (2009). [arXiv:0808.3607](https://arxiv.org/abs/0808.3607)
- C. Kelso, P. Sandick, C. Savage, JCAP **1309**, 022 (2013). [arXiv:1306.1858](https://arxiv.org/abs/1306.1858)
- R. Bernabei et al., Eur. Phys. J. C **72**, 2064 (2012). [arXiv:1202.4179](https://arxiv.org/abs/1202.4179)
- E. Shields, J. Xu, F. Calaprice, Phys. Procedia **61**, 169 (2015)
- J. Amar et al., in *14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories (MG14) Rome, Italy, July 12–18, 2015* (2016). [arXiv:1601.01184](https://arxiv.org/abs/1601.01184), <https://inspirehep.net/record/1413721/files/arXiv:1601.01184.pdf>
- I. Coarasa et al. (2017). [arXiv:1704.06861](https://arxiv.org/abs/1704.06861)
- R. Maruyama et al. (COSINE-100), *COSINE-100 Dark Matter Experiment*. <http://cosine.yale.edu> (2017) (online; accessed 20 April 2017)
- E. Barbosa de Souza et al. (DM-Ice), Phys. Rev. D **95**, 032006 (2017). [arXiv:1602.05939](https://arxiv.org/abs/1602.05939)
- S.C. Kim et al., Phys. Rev. Lett. **108**, 181301 (2012). [arXiv:1204.2646](https://arxiv.org/abs/1204.2646)
- R. Bernabei et al., in *Proceedings of 15th Workshop on What Comes Beyond the Standard Models?* (2013), pp. 1–9 (2012). [arXiv:1301.6243](https://arxiv.org/abs/1301.6243), <http://inspirehep.net/record/1216452/files/arXiv:1301.6243.pdf>
- K. Freese, J.A. Frieman, A. Gould, Phys. Rev. D **37**, 3388 (1988)
- K. Freese, M. Lisanti, C. Savage, Rev. Mod. Phys. **85**, 1561 (2013). [arXiv:1209.3339](https://arxiv.org/abs/1209.3339)
- R. Bernabei et al., (DAMA), Nucl. Instrum. Methods A **592**, 297 (2008). [arXiv:0804.2738](https://arxiv.org/abs/0804.2738)
- R. Bernabei, P. Belli, A. Bussolotti, F. Cappella, V. Caracciolo, M. Casalboni, R. Cerulli, C.J. Dai, A. d’Angelo, A.D. Marco, et al., J. Instrum. **7**, P03009 (2012). <http://stacks.iop.org/1748-0221/7/i=03/a=P03009>