



Research



Cite this article: AbdusSalam S, Abel S, Bartlett D, Crispim Romao M. 2026 Symbolic regression and differentiable fits in beyond the standard model physics. *Phil. Trans. R. Soc. A* **384**: 20240593.

<https://doi.org/10.1098/rsta.2024.0593>

Received: 10 July 2025

Accepted: 20 October 2025

One contribution of 15 to a discussion meeting issue 'Symbolic regression in the physical sciences'.

Subject Areas:

high energy physics, particle physics, computational physics

Keywords:

supersymmetry, beyond standard model, symbolic regression

Author for correspondence:

Steven Abel

e-mail: s.a.abel@durham.ac.uk

Symbolic regression and differentiable fits in beyond the standard model physics

Shehu AbdusSalam¹, Steven Abel^{2,3},
Deaglan Bartlett⁴ and Miguel Crispim Romao³

¹Department of Physics, Shahid Beheshti University, Tehran, Iran

²Department of Mathematical Sciences, and ³Institute for Particle Physics Phenomenology, Durham University, Durham, UK

⁴Institut d'Astrophysique de Paris (IAP), CNRS, Paris, Île-de-France, France

SA, 0000-0001-8848-3462; SA, 0000-0003-1213-907X

We demonstrate the efficacy of symbolic regression (SR) to probe models of particle physics Beyond the Standard Model (BSM), by considering the so-called Constrained Minimal Supersymmetric Standard Model (CMSSM). Like many incarnations of BSM physics this model has a number (four) of arbitrary parameters, which determine the experimental signals, and cosmological observables such as the dark matter relic density. We show that analysis of the phenomenology can be greatly accelerated by using symbolic expressions derived for the observables in terms of the input parameters. Here we focus on the Higgs mass, the cold dark matter relic density and the contribution to the anomalous magnetic moment of the muon. We find that SR can produce remarkably accurate expressions. Using them we make global fits to derive the posterior probability densities of the CMSSM input parameters which are in good agreement with those performed using conventional methods. Moreover, we demonstrate a major advantage of SR, which is the ability to make fits using differentiable methods rather than sampling methods. We also compare the method with neural network (NN) regression. SR produces more globally robust results, while NNs require data that is focused on the promising regions in order to be equally performant.

This article is part of the discussion meeting issue 'Symbolic regression in the physical sciences'.

1. Introduction

A fundamental challenge in exploring theories Beyond the Standard Model (BSM) lies in confronting their predictions with experimental data. Traditional methods are well-established: we define a parameter space guided by physical reasoning and pragmatic constraints, then compute the low-energy spectra for each point. Initial filters ensure phenomenological viability, for example insisting on the stability of the lightest supersymmetric particle (LSP) in supersymmetric (SUSY) models, and then further calculation determines observables such as the dark matter relic density or the muon anomalous magnetic moment. To analyse the entire parameter space relatively efficiently (or indeed at all), sampling algorithms such as Markov Chain Monte Carlo (MCMC) or nested sampling methods [1–3] (e.g. MultiNest [4] and Dynesty [5]) can then be employed to derive posterior distributions for the model parameters.

While physically transparent, this approach faces a significant computational bottleneck: calculating each observable for a given set of input parameters involves complex, multi-step physical analyses, often encompassing subtle effects such as multi-loop corrections, co-annihilation and threshold effects. Let us illustrate by describing the chain of computation that yields the dark matter relic density in the Constrained Minimal Supersymmetric Standard Model (CMSSM). In any BSM model, we would like to determine how this observable depends on the fundamental parameters that are mostly set at a very high energy scale close to the scale of quantum gravity. In the case of the CMSSM there are actually three high scale parameters that we are free to choose, namely the gaugino masses, scalar masses and universal trilinear coupling. (These are typically set at the Grand Unified Theory scale of 2×10^{16} GeV.) The fourth free parameter is set at the electroweak scale, and is the ratio of Higgs Vacuum Expectation Values (VEVs): these four parameters are denoted respectively as $m_{1/2}$, m_0 , A_0 and $\tan\beta$. The relic density depends on the low scale (by which we mean electroweak scale) values of the couplings and mass spectrum, which in turn depend on the low scale values of all the parameters in the theory. The first step is therefore to find these low scale parameters by evolving the renormalization group (RG) equations, which are a coupled set of nonlinear differential equations that determine how all the parameters of the theory depend on the energy scale. Once these low scale values of the parameters have been obtained, one can then determine the mass eigenstates and eigenvalues, single out the lightest state (i.e. the LSP) which is our dark matter candidate, and then evaluate the scattering cross section as a function of temperature. Then this information gets fed into the assumed cosmological evolution in order to determine the density at which the LSP would have stopped annihilating in the early universe (the so-called freeze-out density). This involves solving yet another nonlinear differential equation, which finally yields the desired relic density. If one simply needs to know whether a particular choice of parameters gives a viable relic density or not, this is a great deal of effort!

Clearly closed form expressions for observables such as the relic density in terms of the model input parameters would be an invaluable short-cut to all of the labour outlined above. Although, in principle, closed form expressions linking the input parameters directly to the observables could be derived analytically, their complexity would typically render them impractical except in simplified or asymptotic regimes. Let us stay with the dark matter relic density for the moment, to appreciate just how difficult this would be in practice. What would it take for some omniscient being to derive symbolic expressions for the relic density using physical analysis? First they would have to determine closed form expressions for the low scale theory that can approximate all of the RG running. This is itself a very difficult task. At one-loop order, closed form expressions for the running parameters of the theory *can* be obtained when one of the CMSSM parameters, namely $\tan\beta$, is small because there is some quasi-fixed point behaviour in this region of parameter space. Alas, this possibility is mostly excluded experimentally, and moreover typically RG running must be performed to at least two-loop order to achieve sufficient accuracy. Nevertheless, let us suppose it were possible. Next one would have to derive analytic expressions for the mass spectrum from the low energy parameters. For say the scalar superpartners this would amount to being

able to find closed form expressions for the eigenvalues of 6×6 matrices. Next one would need to use these expressions along with the couplings to determine the LSP annihilation cross section. Typically, this *can* be expressed in a relatively short closed form to reasonable accuracy (albeit in terms of our masses and couplings which would at best be horrendous expressions involving not only the eigenvalues of the mass matrices but the diagonalization matrices as well), but again there are several regions of parameter space where subtle effects come into play such as so-called co-annihilation, where several states are competing to be the LSP, which enhances the annihilation rate. It would at this stage come as no surprise to our omniscient being to discover that these points in the parameter space of the CMSSM are the ones that give the most favourable relic densities. The most interesting points in parameter space are not typical of this particular BSM model in other words, so one requires not only the LSP cross sections, but also the cross sections involving other particles of similar mass (commonly referred to as Next-to LSPs or NLSPs) in the spectrum as well. Finally our omniscient being would need to analytically solve the nonlinear evolution of the densities of the LSP and NLSPs in the early universe, inserting these cross sections in order to find the final freeze-out density of the LSP. If it is not yet abundantly clear, finding analytic expressions for the relic density would amount to being able to solve a whole series of mathematical problems, a solution to any one of which is infeasible.

Thus to cope with the lengthy chain of analysis the scientific community has instead resorted to automated numerical computation, and developed computational ‘packages’ that will accept a given a set of input parameters, perform series of computations such as the one above automatically, and spit out the desired observables. Machine learning can be used to circumvent the computation chain entirely or to better focus on points of interest (for example in Refs [6–10]). However machine learning is generally non-interpretable and, for the problems we are interested in, usually not explainable either. In the current context, it is worth recalling that *in principle* a suitably trained neural-network that takes in the model parameters and spits out the relic density can be thought of as simply a gigantic nested function that has incorporated by trial-and-error the relations above that are impossible to derive through physical analysis. However the mathematical function underlying a neural-network has a huge degree of redundancy because the functional form is fixed by the network topology, and all that we can adjust are weights and biases, so not much can be learnt from it.

Non-interpretability in this context (that is the difficulty in understanding the dependence of the observables on the input parameters) is an irritation to physicists. This is because a precise chain of physical computation leads to the precise answer, and if we follow the physical logic we can anticipate *certain* rough correlations between input parameters and observables. For example, contributions from new (such as SUSY) particles generally die away as their masses grow due to the so-called decoupling theorem [11]. Therefore observables such as $(g - 2)_\mu$, the anomalous magnetic moment of the muon, generally decrease with increasing values of the three SUSY-breaking parameters $m_{1/2}$, m_0 , A_0 , and physicists have become rather good at locating regions of parameter space that will yield physically interesting results. This suggests that some degree of explainability should still be possible that incorporates this and similar physical dependencies.

Clearly, then, there is tension at play between the simplicity of analytic expressions and their effectiveness as regressors for physical observables. This motivates the use of *symbolic regression* (SR) [12], which in a sense provides an automated way to make this trade-off. SR was discussed in generality in Ref. [13], and for specific applications see Refs [14–19]. A comprehensive overview of SR methods is given in Refs [20–24].

What SR aims to provide is a set of analytic expressions for the observables, by learning the analytic formulae that can best reproduce their observed values and, like the neural-network, its rationale is merely to fit the data. In fact it is *morally* like a neural-network that is able to change its topology and activation functions in order to more efficiently fit the data. However this potentially gives it the power to yield explainable results in the realm of physics, because if simple correlations such as those alluded to above are important in some region of parameter space then they should manifest themselves in the symbolic expressions it produces. One could for example

imagine taking the symbolic expressions for $(g - 2)_\mu$ and expanding them around the decoupling limit to find analytic approximations that work in that region of parameter space and that have a physical interpretation. Of course this possibility is contingent on the asymptotic region carrying enough weight in the data provided for the symbolic regression.

The purpose of this paper is to explore the use of symbolic regression in the realm of BSM physics. First we will demonstrate that it is indeed a formidable tool for performing global fits on the CMSSM. To do this we present a set of analytical expressions for the Higgs mass, m_{H^0} , the supersymmetric contribution to the muon anomalous magnetic moment, $\delta(g - 2)_\mu$, and the dark matter relic density, $\Omega_{\text{DM}} h^2$. These expressions are functions of the four input parameters of the CMSSM. This part of the study is based on the work in Ref. [25]. The expressions have been made available at [26] alongside the code that produced them: the dataset used can also be found at [27]. In addition, we provide a ‘classifier’ function $C(m_{1/2}, m_0, A_0, \tan \beta)$ that can decide if a given point in parameter space is likely to be viable based on several physical criteria.

As an example of these, a particularly troublesome aspect of supersymmetric theories is that, because they introduce many new scalars, they are susceptible to developing instabilities. This happens along particular directions in field space that are relatively flat, along which the new scalar superpartners are able to get VEVs themselves, just like the Higgs (which would end in catastrophe). These directions, which are usually referred to as *charge and colour breaking (CCB) directions*, must be sufficiently lifted by the choice of parameters to prevent this occurring. Another constraint that can be included in the ‘classifier’ function is a fact that the LSP in the spectrum which will become our dark matter particle should of course be uncharged. This implies that it should be composed of uncharged gauginos and higgsinos, and not for example selectrons, and this again requires detailed knowledge of the RG running and subsequent low-energy spectrum. The classifier embodies all of these constraints in one function $C(m_{1/2}, m_0, A_0, \tan \beta)$, returning a value greater than 0.5 if a point is physically viable, and less than 0.5 when it is not. To list the criteria we use, they are absence of CCB minima, absence of charged LSP and absence of negative dark matter relic density. This information is provided by the packages for each data point, with the point being given a label value of 1 or 0 depending on its viability.

We will demonstrate that it is possible to use the expressions furnished by symbolic regression to perform global fits and determine the posterior probabilities of the CMSSM input parameters very rapidly. In comparison with the usual methods in which one has to complete the entire chain of computation for each ‘live point’ that one may be considering, we will see that fits carried out in this manner can be performed orders of magnitude more quickly.

We will also explore a second interesting potential benefit of symbolic regression which is that one may restrict the allowed symbolic expressions to be differentiable. One can do this either by including only differentiable operators in the pool of operators available to the symbolic regression, or by taking previously derived symbolic expressions and replacing non-differentiable operators with differentiable approximations. For example the binary operator $\max(a, b)$ (which appears to be quite important in our ultimate expressions at [26]), can be replaced everywhere by a ‘softmax’ function, for example

$$\max(a, b) \approx \text{softmax}(\kappa; a, b) = \frac{a + b}{2} + \frac{(a - b)}{2} \tanh(\kappa(a - b)), \quad (1.1)$$

which approaches the true max function exponentially quickly as $\kappa \rightarrow \infty$.

Having differentiable expressions is advantageous for several reasons. It allows for gradient-based optimization such as gradient-descent which can for example help in optimizing experimental set-ups. In addition having differentiable expressions from symbolic regression would allow the possibility of asymptotic analysis (as alluded to above), in which one analyses the behaviour of observables in limits of large/small parameters, possibly regaining explainability in some regions of parameter space. Error propagation calculations become more manageable, as derivatives can be used to estimate how measurement uncertainties in parameters feed through to the observable. Differentiable formulae for observables would also allow straightforward calculation of sensitivities (derivatives) with respect to parameters, and help identify which parameters

most influence the observable, guiding BSM model building. In the particular context of BSM one particularly important measure of the sensitivity is the so-called *fine-tuning*, which in some sense tells us how ‘likely’ the model is if we assume uniform prior probabilities of parameters. This has become an important (albeit somewhat controversial) indicator of the credibility of a BSM model. The most well-known of these is the Barbieri-Giudice measure of fine-tuning given in Ref. [28]. If we denote the parameters defining the model generically by p_i , then the Barbieri-Giudice sensitivity of the Higgs mass-squared $m_{H^0}^2$ to p_i is given by

$$\Delta_{p_i} = \frac{\partial \log m_{H^0}^2}{\partial \log p_i}. \quad (1.2)$$

The overall measure of fine-tuning (or overall measure of parameter sensitivity) is often taken as the maximum over all the parameters:

$$\Delta = \max_{p_i} \Delta_{p_i}. \quad (1.3)$$

Clearly having access to differentiable expressions for the Higgs mass $m_{H^0}^2(p_i)$ (and any other observables one may be interested in assessing the fine-tuning for) would be a crucial advantage for such questions.

2. Symbolically regressing the CMSSM

We now turn to the practicalities of performing SR in BSM studies. Let us first give a brief overview of the method we developed for our study which would we believe be applicable to most BSM models. We will then focus on the CMSSM to demonstrate its effectiveness.

(a) Generalities of BSM symbolic regression

For BSM studies we believe Operon (and its Python version PyOperon) [29], are optimal due to the efficient vectorized framework and low resource demands. This makes it particularly suitable for BSM studies because they have multi-dimensional parameter spaces, which in turn necessitates training the symbolic regressor on as much data as possible. In addition BSM models tend to contain important localized regions of interest containing physical poles and mass-degeneracies, and the SR expressions that are generated need to encompass this highly localized behaviour. Such regions incorporate those physical aspects such as the co-annihilation effect alluded to in the Introduction, that can be crucial in finding a good fit for an observable.

Consequently, successfully training a symbolic regressor on a BSM model will inevitably require fine multi-dimensional training data. The data points obviously have to be pre-run with available packages to generate values of observables which can then be incorporated into the loss-function of each individual in the population using the symbolic formulae generated by its expression trees for the observables in the study.

Operon is built upon Genetic Programming (GP), which assembles a population of symbolic expressions from a bank of pre-chosen functions, and then evolves it using repeated cycles of *selection*, *breeding* and *mutation*. Operon incorporates two objectives into the GP loop: one of these is the closeness of the formulae to the values generated by the packages, which can be expressed in terms of a number of metrics. The regression metric for this can be ‘mean square error’, ‘mean average error’, ‘ r^2 ’ or ‘normalized mean square error’. The other objective for the GP loop is the expression ‘length’, which is simply the string length of the mathematical expression. By evaluating the individuals in the population on these two metrics, Operon is able to rank them, and select them for breeding at the start of each cycle. It then performs breeding and mutation to generate a new population, and repeats the process to allow the population to evolve until it contains good symbolic expressions which lie along the so-called *Pareto front*. The final output of the Operon run is the population of individuals inhabiting the Pareto front.

Operon has a number of additional hyperparameters, for example the population size and the population initialization (see Ref. [29]). To optimize these hyperparameters as well as the regression metrics, we use Optuna loops [30]. As the ‘figure of merit’ for the choice of hyperparameters, one can use the relative error of the observables for the Pareto front individuals. This is preferable to other common regression metrics, such as r^2 , because the latter can be highly biased by large observable values, when the observable varies by orders of magnitude within the parameter space.

Let us briefly return to the question of those observables such as the dark matter relic density which may be hard to fit within a particular BSM model. As described above this can arise due to the viable values of the observable requiring some physical phenomenon to take place which turns out to occur only in certain restricted regions of parameter space. In such a situation one can find that the symbolic expressions are driven by the ‘wrong’ regions of parameter space in an attempt to fit the bulk of the data, but are not accurate in the regions of most interest. For an observable such as the dark matter relic density the problem is exacerbated by the fact that the generic value can be orders of magnitude larger than desired. In such cases, there are ever diminishing returns from larger budgets and maximal tree-sizes. This can be addressed by weighting the data so that it better favours the regions of interest for those observables. In particular, we would like the expressions produced to be able to capture the more subtle behaviour that leads to the desired cancellations. In order to do this one can operate by resampling the training data according to ‘flattening weights’. These level the distribution to ensure that the regressor is not overly dominated by fitting the observable around its most common values, as opposed to the values of interest. Indeed the weights can be adjusted to resample around the areas of most interest.

For difficult to fit observables like the $\Omega_{\text{DM}}h^2$, without resampling the training data in this manner, SR is unlikely to accurately map the physical region of interest (i.e. $\Omega_{\text{DM}}h^2 \sim \mathcal{O}(0.1)$ for the relic density). However, as we will see there can still be considerable contamination in the range of physical interest (i.e. points that the symbolic expressions indicate have reasonable dark matter relic density when actually they do not, and vice versa). This kind of problem is not present for observables that have acceptable *generic* behaviour within a particular BSM model, as for example the Higgs mass does within the CMSSM. In this sense, the latter can be considered to be an ‘SR-easy’ observable within the CMSSM while an observable like $\Omega_{\text{DM}}h^2$ that requires re-weighting is an ‘SR-hard’ one. The labels ‘easy’ and ‘hard’ are in this context obviously entangled with our own preference for what the observable should be, and crucially are BSM model dependent. Conversely, symbolic regression gives us a crude working criterion for determining how *natural* a particular BSM model is. Namely, the most natural BSM model with the least ‘fine-tuning’ in the sense of equation (1.2) is likely to be one in which all the observables are found to be SR-easy.

(b) Analytic expressions for the CMSSM

We now turn to the specific details of the CMSSM. As discussed the model is defined by four parameters $m_{1/2}$, m_0 , A_0 and $\tan\beta$ (in addition to the couplings and masses of the Standard Model upon which it is built). Our data was composed of 10^5 points sampled at random from the parameter space. We take uniform prior probabilities in the intervals $[0, 10]$ TeV for $m_{1/2}$ and m_0 , $[-6, 6]$ TeV for A_0 , and $[1.5, 50]$ for $\tan\beta$. The data was split into a training and a test set, and the observables m_{H^0} , $\Omega_{\text{DM}}h^2$ and $\delta(g-2)_\mu$ were determined for each data point using the particle spectrum generator SPheno [31] and the particle dark matter package MicrOMEGAs [32]. In addition the physical viability of each point was determined (and given a value 1 or 0) which feeds into the classifier regressor.

PyOperon was run on this sample of points using the re-weighting procedure for $\Omega_{\text{DM}}h^2$ described in §2a, which essentially flattens the weighting such that all values of $\Omega_{\text{DM}}h^2$ carry equal weight within the regression. We found that the re-weighting reduced the effective training dataset from 10^5 to 10^4 data points, and gave an approximately flat distribution of $\Omega_{\text{DM}}h^2$. (As we shall see this will improve the ability of the symbolic regressors to fit $\Omega_{\text{DM}}h^2$ to its inferred

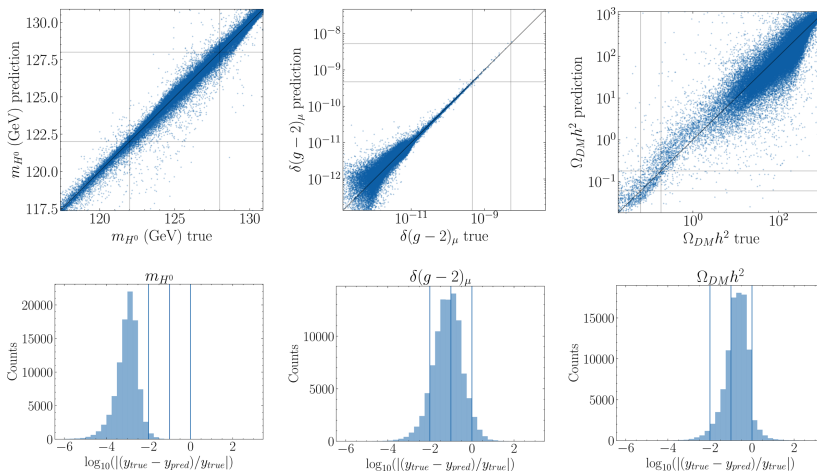


Figure 1. The performance of the SR expressions for the observables, on the test set. In the upper panels, we show the ‘true versus prediction’ scatter plots. The solid lines delineate the physically viable values. In the lower panels, we show the relative errors. Here the vertical lines delineate relative errors of 1%, 10% and 100%.

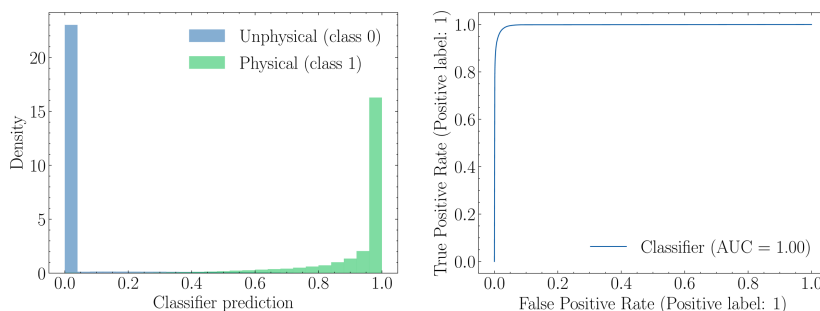


Figure 2. The classifier symbolic regressor performance, showing the output of the classifier when acting on the test data on the left and the ROC curve on the right.

value.) The result is the set of symbolic regressors for m_{H^0} , $\Omega_{\text{DM}}h^2$ and $\delta(g-2)_\mu$ and the classifier C which have been made available at [26].

To test the expressions, we present in the upper panels of figure 1 scatter plots (made using the test set) of the ‘true’ value of the observable (denoted generically y_{true}) versus the prediction (denoted y_{pred}) from the symbolic regressor. The relative errors of the predictions, defined as $|y_{\text{pred}} - y_{\text{true}}|/|y_{\text{true}}|$, are presented in the lower panels. The regression of the Higgs mass is evidently very accurate, with almost all points having relative error below 1%. For $\delta(g-2)_\mu$, the errors are also extremely small in the region of physical interest where $\delta(g-2)_\mu$ is enhanced, and actually smaller than the theoretical value of $(g-2)_\mu$ in the Standard Model itself. For $\Omega_{\text{DM}}h^2$ the *relative* errors are typically above 10%, but this is actually better than the 20% theoretical error that can be assigned to $\Omega_{\text{DM}}h^2$. Finally the performance of the classifier regressor is shown in the two panels of figure 2, as a distribution and as a ROC curve. From these plots, we see that (by accepting only points with say $C > 0.5$) we have an almost perfect classifier regressor. In other words, our function $C(m_{1/2}, m_0, A_0, \tan\beta)$ can be used to immediately reject unphysical points from the CMSSM.

It is worth commenting on the question of whether the symbolic regression has converged, or if better expressions could be found. To consider this question, we produced a 10 times larger dataset and generated a new set of symbolic expressions. It was found that these new expressions did not demonstrate any significant improvement in accuracy, while the method became prohibitively slow. It is worth noting that the scaling did not appear to be linear with dataset size.

3. Global fits

In order to put our symbolic expressions to the test let us now consider performing a global fit, that is a statistical analysis to determine the most probable values of the CMSSM model parameters. In this work, we shall use Dynesty [5]. This package performs nested sampling, a Bayesian technique used for parameter estimation and calculating the model evidence. It is a very efficient method of exploring complex, multi-dimensional parameter spaces, making it well-suited for BSM global fits. The process begins by defining a likelihood function that combines experimental and theoretical constraints into a single metric indicating how well a set of parameters fits the data. Using our previously specified flat priors for the parameters, Dynesty initializes a set of ‘live points’ sampled from these priors and iteratively replaces the lowest-likelihood point with new samples drawn from the prior constrained to higher likelihood regions. This iterative process systematically explores the parameter space, effectively mapping out the posterior distributions and estimating the Bayesian evidence.

This procedure can come with significant resource overheads due to the need in each iteration to find a new point with a likelihood greater than that of the lowest likelihood point in the set of live points. Hence, we will now show that replacing these costly full physics estimates of observables with symbolic expressions dramatically speeds up the process with only a modest loss of accuracy.

For these fits, we use $m_{H^0} = 125.04 \pm 0.12$ GeV [33] and $\Omega_{\text{DM}} h^2 = 0.12 \pm 0.02$ [34] which includes an aforementioned theoretical uncertainty of 20% in the relic density. The experimental value of the anomalous magnetic moment of the muon is $\delta(g-2)_\mu = (249 \pm 48) \times 10^{-11}$ [35]. This represents a discrepancy with its high precision Standard Model (SM) prediction which is of current interest [36]. Denoting the data for these three observables generically as $\underline{d} = \{\mu_i \pm \sigma_i\}$, where μ_i and σ_i represent the central values and uncertainties (with $i = 1, 2, 3$), then for each point $\underline{\theta} = \{m_{1/2}, m_0, A_0, \tan \beta\}$ we shall estimate the likelihood as

$$p(\underline{d}|\underline{\theta}) = \prod_{j=1}^3 \frac{1}{\sigma_j \sqrt{2\pi}} \exp\left\{-\frac{(y_{\text{pred}}^j - \mu_j)^2}{2\sigma_j^2}\right\}. \quad (3.1)$$

Here y_{pred}^j stands for the predictions given by the SR expressions for the observables. For comparison with the nested sampling results, v.2.1.4 of Dynesty was employed along with `DynamicNestedSampler` with 500 live points. (For the other parameters, we took no bootstrap, with `pfrac = 1`.)

In figure 3, we show the resulting posterior distributions of the CMSSM parameters, which have been fit to our three observables using Dynesty (with the plots produced with `GetDist` [37]). The 68% and 95% Bayesian probability contour lines (in red and labelled ‘Expressions’) are the posterior distributions from the fit which were determined using the SR expressions. Meanwhile the scatter plots (with colour denoting the Higgs mass, and labelled ‘Packages’) show the global fit made using conventional methods, in which our three observables were computed using the available packages.

We can conclude that the global fit analysis made using SR-derived expressions compares extremely well with the fit made using the conventional packages-based method. However we should comment on the apparent differences in the one-dimensional marginal plots. The most important point to make is that the plots are normalized by maximum height rather than area. This is the customary approach in order to compare the locations of maximum *a posteriori* probabilities. This makes the $\tan \beta$ one-dimensional marginal appear to be in worse agreement than it really is. (In area-normalized plots the two graphs would overlap over most $\tan \beta$ values with a somewhat larger spike appearing at low $\tan \beta$ in the plot made using expressions.) A second point is that in Ref. [25] a comparison was made between the fits made using packages with different sampling frameworks to each other. Namely we compared fits made using Dynesty and MultiNest with customary packages, and found that the fits made using these alternative frameworks

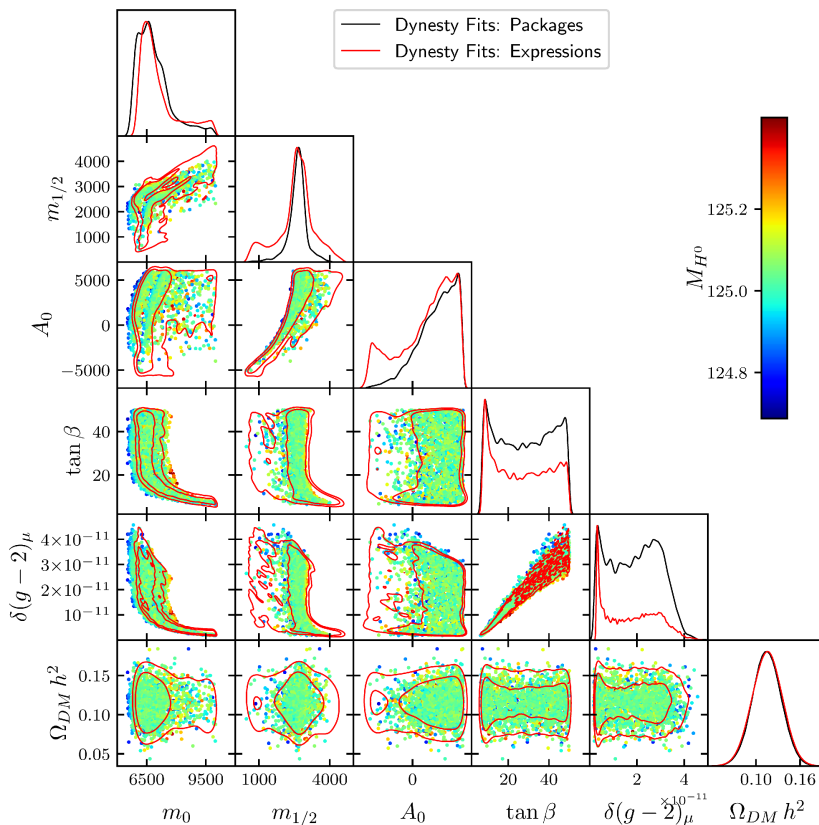


Figure 3. Posterior probability distributions for the CMSSM found using Dynesty, fitting the three observables m_{H^0} , $\delta(g-2)_\mu$ and $\Omega_{DM}h^2$. The package-based results are the scatter plots (and the black lines on the diagonal plots). The SR expression-based results are the red lines. Dimensionful parameters are in GeV. As is customary the one-dimensional marginals are normalized with respect to their maximum heights rather than areas in order to assess the agreement of maximum *a posteriori* probabilities.

showed similar minor discrepancies among themselves. Therefore we consider the results found using expressions to be *as reliable* in this sense as the results found using packages.

4. Differentiable symbolic regression

We now turn to the question of utilizing entirely differentiable expressions for symbolic regression. The most straightforward approach for this is to simply take the previously used approximations and replace their non-differentiable parts by differentiable approximations such as the softmax function in equation (1.1). Indeed for our application, the only non-differentiable operators of concern are min and max, which only appear in the classifier. Since we do not use this in the MCMC, we already have functions comprised of differentiable operators, and thus write these in JAX so that we can utilize its auto-differentiation. Of course, given that these are symbolic expressions, one could also analytically compute these derivatives.

Now that we have access to differentiable surrogate models for the ‘packages’, we can leverage differentiable optimization or sampling techniques to expedite our analysis. To demonstrate this, we replace the nested sampler Dynesty with a No-U-Turn Sampler (NUTS) [38], as implemented in NUMPYRO [39,40]. Using the same likelihood as before, we run four parallel chains with 5000 warm-up steps and 10^5 samples each. We verify that this results in Gelman–Rubin statistics [41], which are unity within 10^{-2} for each inferred parameter.

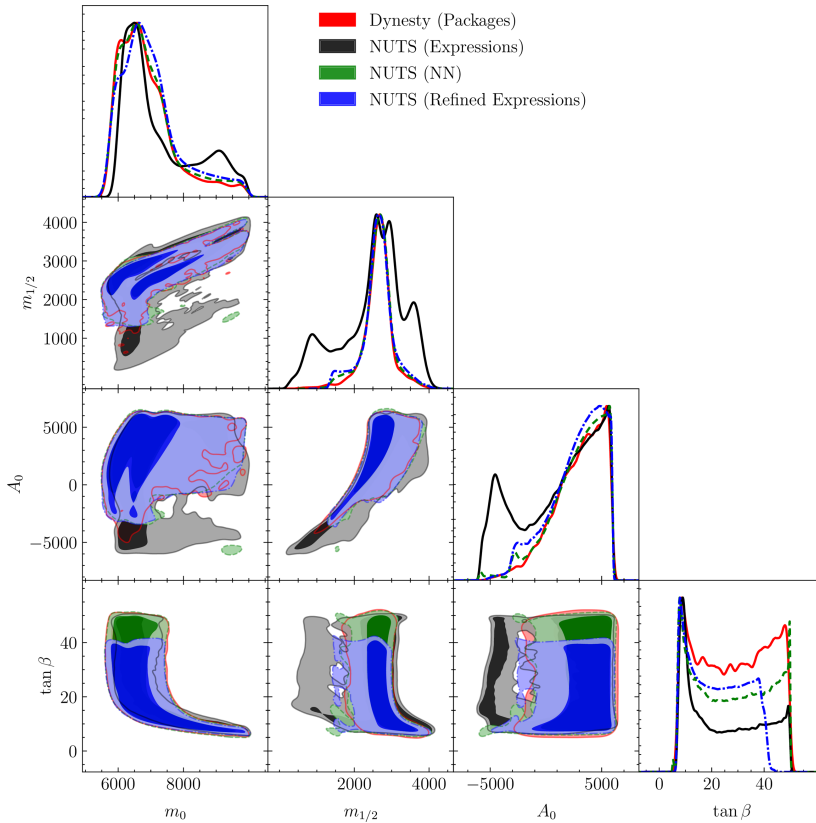


Figure 4. Posterior distribution of CMSSM parameters obtained using NUTS compared with those from Dynesty. The contours show the 68% and 95% confidence intervals. The red contours compute the outputs exactly, whereas the remaining ones use the expressions of Ref. [25] (black), a neural network (green) or refined symbolic expressions for a narrower prior range (blue) to bypass this calculation. The one-dimensional marginals are normalized with respect to their maximum heights rather than areas in order to assess the agreement of maximum *a posteriori* probabilities.

We plot the inferred posteriors for our expressions in figure 4, where we compare against the posterior obtained using Dynesty. One sees that we obtain similar agreement between the posteriors as in figure 3, with the same locations of the maxima, and reasonably similar marginalized one-dimensional posteriors. The match is not perfect, however, but it was also found by Ref. [25] (from where we take the Dynesty contours) that the posteriors obtained by Dynesty and Multinest (an alternative nested sampling algorithm) are also not identical.

Interestingly, we find that the expressions given in Ref. [26] appear to have some instabilities when we attempt to use their gradients in our MCMC. This manifests itself as a large number of divergences, which indicates a numerical instability during Hamiltonian trajectory integration. This is typically caused by regions of high curvature or poorly scaled parameters in the posterior distribution. We circumvented this problem by increasing the target acceptance probability of NUTS to be 0.95 instead of the default 0.8. This increases the probability of rejecting bad steps and thus fewer divergences are encountered. This essentially sacrifices efficiency for stability.

Upon more careful investigation, we identified the function approximating $\Omega_{\text{DM}} h^2$ as the cause of our divergences. We therefore generated a new symbolic regressor for this variable, which only contains differentiable operators. This was obtained by restricting PyOperon allowed operators to their differentiable subset, i.e. operators such as min and max were disallowed during the Optuna optimization loop.

Replacing the original expression with this expression enables us to keep the target acceptance probability at 0.8 without encountering the same number of divergences. We find sampling to be much more efficient; the run time on a MacBook Pro with an Apple M2 Pro chip (12-core CPU) and 16 GB of unified memory for the original expressions with NUTS is approximately 25 min, but for the same number of samples we only require 4.5 min for the new expression. (This could be accelerated further by utilizing the GPU, but we only used a CPU in these experiments.) This can be compared with the time required to sample this posterior with Dynesty, which took about 2 h on 72 cores using the packages, or 2 h on a single core using the expressions,¹ and thus we see a substantial improvement for either expression.

(a) Comparison with neural networks

For completeness, it is worth comparing at this point the performance of the SR-derived expressions to the performance of a neural network (NN), as these are also differentiable, and would also allow for a differentiable fit. To this end, we adopted a fully connected feed-forward neural network in JAX to produce differentiable likelihoods, so that posterior sampling could be performed in the same manner (NUTS) as for the differentiable symbolic expressions. The model consists of four hidden layers with 64 neurons each, mapping a four-dimensional input to a three-dimensional output. Parameters were initialized with a fixed random seed for reproducibility and trained using the Adam optimizer [42] with an initial learning rate of 10^{-3} . Given the different dynamic range and observational uncertainties of our outputs, instead of simply using a mean-squared-error (MSE) loss function, we weight each term in the MSE loss by the squared reciprocal of the observational error on that quantity. The learning rate was reduced by a factor of 0.5 if the validation loss did not improve for 200 epochs, with a minimum of 10^{-6} , and training was performed for up to 30 000 epochs.

We can make an interesting immediate observation about the NN performance. Namely, under the original priors and with the original training samples the NN failed to capture the same global picture that the SR expressions were able to provide out of the box. Therefore, in order to obtain meaningful posteriors for the NN regressors, substantially more data had to be generated in order to create a more tightly focused training domain than was required for the SR to perform. To take account of this, the NN was trained using a much larger dataset (10 million points), split 90% : 10% for training/validation, with a restricted prior window for the input parameters of

$$m_0 : [3.5, 10] \text{ TeV} \quad (4.1)$$

$$m_{1/2} : [0.02, 6] \text{ TeV} \quad (4.2)$$

$$A_0 : [6, 6] \text{ TeV} \quad (4.3)$$

$$\tan \beta : [5, 50] . \quad (4.4)$$

In addition, a ‘cut’ was enforced on the data to allow only those points that produced observables in promising ranges, namely

$$\Omega_{DM} h^2 \in [0, 0.3] \quad (4.5)$$

$$m_{H_0} \in [122, 128] \text{ GeV} \quad (4.6)$$

$$\delta(g-2)_\mu \leq 10^{-10} , \quad (4.7)$$

producing a dataset ‘zoomed in’ on the region of interest.

For a fair comparison, we then retrained the SR expressions on the same enlarged and ‘zoomed’ dataset that was used for the NN. One substantial difference in behaviour emerged, which was the greater sensitivity of the SR to those points which then lay outside the domain of the training data

¹For the expressions, the overhead of parallelizing the run outweighs the benefits as expressions are computationally cheap to evaluate.

which produces promising observables. To address this we trained a classifier to discard points lying outside the ‘promising observables’ window. This is operationally equivalent to enforcing a sharper prior and is consistent with the way that low-likelihood points would be treated during sampling. Interestingly, the NN showed much less sensitivity to these ‘undesirable’ regions, providing good fits regardless of the use of such a classifier.

Under these matched conditions the NN and SR posteriors became very similar and both agreed with the package-derived posterior (see figure 4, where the new SR result is shown in blue contours). For the three well-constrained parameters (m_0 , $m_{1/2}$ and A_0), we see that the refined SR contours match well those from either Dynesty or the NN, and the marginal posterior for $m_{1/2}$ is much more accurate than when obtained using the regressors of Ref. [25]. For both cases, the marginal distribution of $\tan\beta$ does not match perfectly, and for the refined SR expressions this becomes zero at slightly smaller values than for the truth. Hence, both the NN and refined SR expressions achieve good approximate posteriors, albeit with some small remaining artefacts.

The main differences in the performance of the NN versus the SR are therefore mainly practical: it seems that a NN approach requires a focused prior dataset in order to reach the same fidelity. By contrast, SR reaches comparable fidelity with substantially less focusing of the prior dataset and shows better global behaviour in this context.

5. Conclusion

Analytic expressions for observables are a powerful tool, however although some observables in a BSM model can be relatively easily treated with a full physical analysis, the vast majority of them cannot. In supersymmetric extensions of the Standard Model this is true even at the level of the mass spectrum (which in supersymmetry entails the diagonalization of 6×6 matrices), before it is even fed into observables such as the dark matter relic density. It is worth noting that the Higgs mass itself receives crucial one-loop and two-loop corrections.

In the absence of general analytical formulae, the typical approach for BSM physics is to perform a computation of the values of observables for any given value of the model’s input parameters, using available numerical packages. This results in a costly chain of computation that makes it difficult to effectively analyse the phenomenology of a BSM model such as the CMSSM over the whole parameter space, even using nested sampling techniques. There are indeed cases where it may take several minutes to compute the observable for a *single point* by conventional methods [10]. In this work, we have shown that symbolic regression can bridge this gap and provide analytical formulae where a fully analytical treatment of the physics would be too difficult.

We showed that the symbolic expressions that are produced can successfully short-circuit the chain of physical computation for a given BSM model (specifically the CMSSM for the current study) between the input parameters and the observables. This greatly reduces the computational resources that are needed for global fitting. Indeed our study found that two orders of magnitude less CPU time was needed to perform a global fit of the CMSSM when all three of the observables (namely the Higgs mass, dark matter relic density and anomalous magnetic moment of the muon) were simply evaluated using the SR-derived symbolic expressions. As we saw in figure 2, symbolic regression could also produce a remarkably effective symbolic classifier for implementing model constraints (such as absence of charge and colour breaking minima, and neutral dark matter in supersymmetric extensions of the Standard Model). This allows much more rapid rejection of unviable models. Moreover, in an era when there are still many competing BSM models, symbolic expressions provide an especially useful tool for model comparison, rather than having to correctly implement packages for each model independently.

It is worth commenting on the fact that although the inference with SR expressions is much faster, this does not take into account the fact that there is obviously computational cost in obtaining the training samples. To compare the computational cost of the global fits using expressions versus nested sampling, the number of times that, for example, the dark matter relic density has to be calculated is roughly the same, i.e. approximately a million times in both cases. An important

point, however, is that once the symbolic expressions have been generated they can be reused when for example there are new experimental results that warrant new posterior calculations. This therefore amortizes the costs of data generation and expression training.

A second great advantage of symbolic regression is the possibility to produce differentiable expressions for observables, allowing differentiable workflows for model analysis. Operationally there is a preference for the following procedure for differentiable symbolic regression: first perform unconstrained symbolic regression on the observables; then replace non-differentiable operators (like ‘max’) with differentiable approximations (like the ‘softmax’ function of [equation \(1.1\)](#)). The reason for preferring this procedure (as opposed to performing a symbolic regression with only differentiable operators) can be easily understood if one considers why non-differentiable operators are likely to turn up in the symbolic expressions of, for example, the classifier function. One objective of this function is to reject theories with charge- and colour-breaking minima, and this criterion is typically expressed as an inequality involving combinations of input parameters, which can be expressed as a max function in the classifier. Likewise satisfying experimental exclusion bounds would most likely result in max functions involving the masses, and so forth. In other words, during the analysis of a typical BSM model there are many occasions when the spectrum must be compared with physical values, and this can be implemented efficiently with non-differentiable operators. If one were to perform the symbolic regression excluding non-differentiable operators in the pool of available operators, then the regressor would be forced to try and rebuild operators such as ‘max’ with approximations such as the ‘softmax’ function of [equation \(1.1\)](#). But these approximations would greatly increase the length of the corresponding expressions, disfavouring such expressions in the symbolic regression. We believe this can lead to poor performance of an entirely differentiable symbolic regression procedure compared with a procedure in which non-differentiable operators are allowed, and simply replaced with approximations in the ultimate expressions. On the other hand, we also found instances, such as for $\Omega_{\text{DM}} h^2$, where it is advantageous to generate symbolic expressions using exclusively differentiable operators from the start. This appears to greatly reduce instability and increase efficiency during the differentiable optimization when producing global fits.

Finally, we found that the SR method compares favourably with a neural network approach. While a NN regressor is an appropriate and viable approach for this task, it produces compatible posteriors only when significant cuts are made on the prior window for the input parameters in order to ‘zoom in’ on regions that produce good observables. Symbolic regression, by contrast, proves to be more sample-efficient on this problem (in the sense that it does not require such prior cuts on the data), thus providing a robust global picture over the whole range of observables. The two approaches are therefore somewhat complementary: NNs are flexible and powerful when there is abundant training data that can be then focused, while SR offers analytic expressions that provide a better global picture, and which can be straightforwardly integrated into differentiable pipelines.

Data accessibility. All data has been made available at the repositories: [43] and <https://gitlab.com/miguel.romao/symbolic-regression-bsm>.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. S.Abd.: conceptualization, investigation, methodology, software, writing—original draft; S.Abe.: conceptualization, investigation, methodology, writing—original draft, writing—review and editing; D.B.: conceptualization, investigation, methodology, software, writing—original draft; M.C.R.: conceptualization, investigation, methodology, software, writing—original draft.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. S.Abe and M.C.R. are supported by the STFC under Grant No. ST/T001011/1. D.B. was supported by the Simons Collaboration on ‘Learning the Universe’ and is supported by Schmidt Sciences through The Eric and Wendy Schmidt AI in Science Fellowship. This work was performed using resources provided by the Cambridge CSD3, provided by Dell EMC and Intel using Tier-2 funding from EPSRC grant EP/T022159/1, and DiRAC funding from the STFC.

Acknowledgements. We thank Jonathan Patterson for smoothly running the Glamdring Cluster hosted by the University of Oxford, where some of the data processing was performed. We are extremely grateful to Bogdan Burlacu and Gabriel Kronberger for extensive guidance with Operon. We would like to thank Miles Cranmer and Pedro Ferreira for help and discussions. S.A.1 and S.A.2 thank CERN-TH and S.A.1 thanks the Institute for Theoretical Physics at Heidelberg University for hospitality extended during the initial stages of this work.

References

1. Skilling J. 2004 Nested sampling. *AIP Conf. Proc.* **735**, 395. (doi:10.1063/1.1835238)
2. Skilling J. 2006 Nested sampling for general Bayesian computation. *Bayesian Anal.* **1**, 833–859. (doi:10.1214/06-BA127)
3. Ashton G *et al.* 2022 Nested sampling for physical scientists. *Nat. Rev. Methods Primers* **2**, 1–37. (doi:10.1038/s43586-022-00121-x)
4. Feroz F, Hobson MP. 2008 Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *Mon. Not. R. Astron. Soc.* **384**, 449–463. (doi:10.1111/j.1365-2966.2007.12353.x)
5. Speagle JS. 2020 DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.* **493**, 3132–3158. (doi:10.1093/mnras/staa278)
6. Caron S, Kim JS, Rolbiecki K, de Austri RR, Stienen B. 2017 The BSM-AI project: SUSY-AI-generalizing LHC limits on supersymmetry with machine learning. *Eur. Phys. J. C.* **77**, 257. (doi:10.1140/epjc/s10052-017-4814-9)
7. Hammad A, Park M, Ramos R, Saha P. 2023 Exploration of parameter spaces assisted by machine learning. *Comput. Phys. Commun.* **293**, 108902. (doi:10.1016/j.cpc.2023.108902)
8. Abreu de Souza F, Crispim Romão M, Castro NF, Nikjoo M, Porod W. 2023 Exploring parameter spaces with artificial intelligence and machine learning black-box optimization algorithms. *Phys. Rev. D* **107**, 035004. (doi:10.1103/PhysRevD.107.035004)
9. Romão JC, Crispim Romão M. 2024 Combining evolutionary strategies and novelty detection to go beyond the alignment limit of the Z_3 3HDM. *Phys. Rev. D* **109**, 095040. (doi:10.1103/PhysRevD.109.095040)
10. Diaz MA, Cerro G, Dasmahapatra S, Moretti S. 2025 Bayesian active search on parameter space: a 95 GeV Spin-0 Resonance in the B-LSSM. *SciPost. Phys. Core* **8**, 068. (doi:10.21468/SciPostPhysCore.8.4.068)
11. Appelquist T, Carazzone J. 1975 Infrared singularities and massive fields. *Phys. Rev. D* **11**, 2856–2861. (doi:10.1103/PhysRevD.11.2856)
12. Koza JR. 1992 *Genetic programming: on the programming of computers by means of natural selection*. Cambridge, MA: MIT Press.
13. Udrescu SM, Tegmark M. 2020 AI Feynman: a physics-inspired method for symbolic regression. *Sci. Adv.* **6**, eaay2631. (doi:10.1126/sciadv.aay2631)
14. Butter A, Plehn T, Soybelman N, Brehmer J. 2024 Back to the Formula – LHC Edition. *SciPost. Phys* **16**, 037. (doi:10.21468/SciPostPhys.16.1.037)
15. Abel SA, Constantin A, Harvey TR, Lukas A. 2023 Cosmic inflation and genetic algorithms. *Fortsch. Phys.* **71**, 2200161. (doi:10.1002/prop.202200161)
16. Bartlett DJ, Desmond H, Ferreira PG. 2022 Exhaustive symbolic regression. *IEEE Trans. Evol. Computat.* **28**, 950–964. (doi:10.1109/TEVC.2023.3280250)
17. Koksang SM. 2023 Cosmological parameter constraints using phenomenological symbolic expressions: on the significance of symbolic expression complexity and accuracy. *Phys. Rev. D* **108**, 043539. (doi:10.1103/PhysRevD.108.043539)
18. Sousa T, Bartlett DJ, Desmond H, Ferreira PG. 2024 Optimal inflationary potentials. *Phys. Rev. D* **109**, 083524. (doi:10.1103/PhysRevD.109.083524)
19. Tsoi HF *et al.* 2024 Symbolic regression on FPGAs for fast machine learning inference. *EPJ Web of Conf.* **295**, 09036. (doi:10.1051/epjconf/202429509036)
20. deFranca FO *et al.* 2023 Interpretable symbolic regression for data science: analysis of the 2022 competition. (doi:10.48550/arXiv.2304.01117)
21. Cranmer MD, Xu R, Battaglia PW, Ho S. 2019 Learning symbolic physics with graph networks. *CoRR* **abs/1909.05862**. (doi:https://doi.org/10.48550/arXiv.1909.05862)

22. Cranmer MD, Sanchez-Gonzalez A, Battaglia PW, Xu R, Cranmer K, Spergel DN, Ho S. 2020 Discovering symbolic models from deep learning with inductive biases. *CoRR* **abs/2006.11287**. (doi:10.48550/arXiv.2006.11287)
23. Cava WGL, Orzechowski P, Burlacu B, França FO, Virgolin M, Jin Y, Kommenda M, Moore JH. 2021 Contemporary symbolic regression methods and their relative performance. *CoRR* **abs/2107.14351**. (doi:10.48550/arXiv.2107.14351)
24. Cranmer M. 2023 Interpretable machine learning for science with PySR and SymbolicRegression.jl (doi:10.48550/arXiv.2305.01582)
25. AbdusSalam S, Abel S, Romão MC. 2025 Symbolic regression for beyond the standard model physics. *Phys. Rev. D* **111**, 015022. (doi:10.1103/PhysRevD.111.015022)
26. AbdusSalam S, Abel S, Crispim Romão M. 2024 Symbolically regressing beyond the standard model physics. <https://gitlab.com/miguel.romao/symbolic-regression-bsm>
27. AbdusSalam S, Abel S, Crispim Romão M. 2024 1 Million cMSSM parameter space points with low- energy predictions from SPheno and microOMEGAS. (doi:10.5281/zenodo.11366471)
28. Barbieri R, Giudice GF. 1988 Upper bounds on supersymmetric particle masses. *Nucl. Phys. B* **306**, 63–76. (doi:10.1016/0550-3213(88)90171-X)
29. Burlacu B, Kronberger G, Kommenda M. Operon C++: an efficient genetic programming framework for symbolic regression (ed. R Allmendinger). In *Proc. of the 2020 genetic and evolutionary computation conference companion*, pp. 1562–1570. Association for Computing Machinery. (doi:10.1145/3377929.3398099)
30. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. 2019 Optuna: a next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, Anchorage, Alaska*. (doi:10.1145/3292500.3330701). <https://arxiv.org/abs/1907.10902>.
31. Porod W. 2003 SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e+e- colliders. *Comput. Phys. Commun.* **153**, 275–315. (doi:10.1016/S0010-4655(03)00222-4)
32. Bélanger G, Boudjema F, Pukhov A, Semenov A. 2002 MicrOMEGAS: a program for calculating the relic density in the MSSM. *Comput. Phys. Commun.* **149**, 103–120. (doi:10.1016/S0010-4655(02)00596-9)
33. CMS. 2025 Measurement of the Higgs boson mass and width using the four leptons final state. *Phys. Rev. D* **111**, 092014. (doi:10.1103/PhysRevD.111.092014)
34. Aghanim N. 2020 Planck 2018 results. VI. Cosmological parameters. *Astron. Astrophys.* **641**, A6. (doi:10.1051/0004-6361/201833910)
35. Aguillard DP *et al.* 2023 Measurement of the positive muon anomalous magnetic moment to 0.20 ppm. *Phys. Rev. Lett.* **131**, 161802. (doi:10.1103/PhysRevLett.131.161802)
36. Aoyama T *et al.* 2020 The anomalous magnetic moment of the muon in the Standard Model. *Phys. Rep.* **887**, 1–166. (doi:10.1016/j.physrep.2020.07.006)
37. Lewis A. 2019 GetDist: a python package for analysing Monte Carlo samples. <https://getdist.readthedocs.io/en/latest/>
38. Hoffman MD, Gelman A. 2011 The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *arXiv E-Prints* arXiv. (doi:10.48550/arXiv.1111.4246)
39. Bingham E. 2018 Pyro: deep universal probabilistic programming. *arXiv E-Prints* arXiv. (doi:10.48550/arXiv.1810.09538)
40. Phan D, Pradhan N, Jankowiak M. 2019 Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv E-Prints* arXiv. (doi:10.48550/arXiv.1912.11554)
41. Gelman A, Rubin DB. 1992 Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472. (doi:10.1214/ss/1177011136)
42. Kingma DP, Ba J. 2017 Adam: a method for stochastic optimization. *arXiv*. <https://arxiv.org/abs/1412.6980>
43. AbdusSalam S, Abel S, Crispim Romão M. 2024 1 Million cMSSM parameter space points with low- energy predictions from SPheno and microOMEGAS. Zenodo. (doi:10.5281/zenodo.11366471)