Accelerating the BSM interpretation of LHC data with machine learning

Gianfranco Bertone,[1] Marc Peter Deisenroth,[2] Jong Soo Kim,[3]
Sebastian Liem,[1] Roberto Ruiz de Austri,[4] and Max Welling[5]

[1]*GRAPPA, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands*
[2]*Department of Computing, Imperial College London,*
180 Queen's Gate, SW7 2AZ London, United Kingdom
[3]*Center for Theoretical Physics of the Universe,*
Institute for Basic Science (IBS), Daejeon, 34051, Korea and
Instituto de Física Teórica UAM/CSIC, Madrid, Spain
[4]*Instituto de Física Corpuscular IFIC-UV/CSIC, Valencia, Spain*
[5]*Informatics Institute, University of Amsterdam,*
Science Park 904, 1098 XH Amsterdam, Netherlands
(Dated: November 10, 2016)

The interpretation of Large Hadron Collider (LHC) data in the framework of Beyond the Standard Model (BSM) theories is hampered by the need to run computationally expensive event generators and detector simulators. Performing statistically convergent scans of high-dimensional BSM theories is consequently challenging, and in practice unfeasible for very high-dimensional BSM theories. We present here a new machine learning method that accelerates the interpretation of LHC data, by learning the relationship between BSM theory parameters and data. As a proof-of-concept, we demonstrate that this technique accurately predicts natural SUSY signal events in two signal regions at the High Luminosity LHC, up to four orders of magnitude faster than standard techniques. The new approach makes it possible to rapidly and accurately reconstruct the theory parameters of complex BSM theories, should an excess in the data be discovered at the LHC.

Introduction: A vast effort is currently in progress to discover physics Beyond the Standard Model (BSM) at the Large Hadron Collider (LHC), motivated in part by the possible connection between new particles at the weak scale and the dark matter problem in astrophysics and cosmology [1–3]. The absence of clear evidence for BSM physics in current LHC data has been interpreted in the context of simplified models [4, 5] as well as of *full* models, such as various incarnations of the minimal Supersymmetric Standard Model (MSSM) [6–10].

Such studies, and even more the interpretation of a hypothetical excess in future data, are hampered by the computationally intensive task of sampling the high-dimensional parameter space of theoretical models, and comparing, for each sample, the predicted signal with actual data. For each set of input parameters one needs in fact to: (i) generate a Monte Carlo (MC) sample of collision events; (ii) run the sample through a detector simulation; and (iii) compare the predicted signal with data, often within signal regions (SRs) defined by experimental cuts on observable quantities, such as missing transverse energy, number of jets, momenta, and angles [7]. This procedure is computationally very expensive, and it constitutes the bottleneck for global analyses of BSM theories, especially for those with high-dimensional theory parameter spaces: in Ref. [6], for instance, it was estimated that ≈ 400 CPU-years would be needed to obtain a statistically convergent scan of a 15-dimensional supersymmetric model.

We demonstrate here that this bottleneck can be by-passed by introducing machine learning (ML) tools that can *learn* the mapping between theory and data, and then rapidly and accurately predict signal region efficiencies.

Gaussian processes: The number of events N_i in SR i can be written as $N_i = L\sigma\epsilon_i$, where L is the integrated luminosity, σ the production cross-section of the relevant process(es), and $\epsilon_i \in [0,1]$ is the SR efficiency (which is in turn the product of the detector efficiency times the acceptance, i.e. the fraction of events that passes analysis cuts). A *classification* ML method was introduced in Ref. [11] to predict whether or not a given point in the BSM theory parameter space is compatible with LHC data. Here, we are interested in the more general *regression* problem of estimating the continuous quantities ϵ_i given the input BSM parameters $\boldsymbol{\theta}$, i.e. in modeling the relationship $\epsilon_i = f_i(\boldsymbol{\theta})$.

We specifically implement here the Gaussian process (GP) regression model [12]. Instead of predicting a single value, a GP has the virtue of equipping predictions with consistent uncertainty estimates by means of a full posterior distribution. The crucial ingredient of GPs is the covariance function, which specifies the correlation structure between the function value at different points in the input parameter space. We use here for the covariance function an anisotropic squared exponential kernel [12]

$$k(\boldsymbol{\theta}, \boldsymbol{\theta}') = \sigma_f^2 \exp\left(\sum_j \frac{(\theta_j - \theta_j')^2}{2l_j^2}\right) \tag{1}$$

where the sum is over the BSM theory parameters. σ_f and l_j are hyperparameters: σ_f^2 encodes the intrinsic variance of the function we are modeling, and the l_j are characteristic length-scales which determine how quickly the function changes from point to point. Choosing the

optimal values of these hyperparameters to model our function is the *learning* task of GPs, and is done by the standard procedure of evidence maximization [12].

The major limitation of standard GPs is that training scales cubically with the size n of the training data set as it involves computing the inverse of $n \times n$ matrices. Therefore, in practice, there is a limitation on the amount of training data that can be used. To eliminate this limitation we make use of distributed GPs (DGPs), specifically the robust Bayesian Committee Machine [13] algorithm, which avoids large matrices by partitioning the training data into smaller data sets and distributing the computation across independent computing nodes.

Natural supersymmetry: As a proof-of-concept, we apply this new technique to the natural supersymmetry (SUSY) scenario, in which fine-tuning is low, and the electroweak scale is stabilized by a small subset of light SUSY states (e.g., [14–18]). We focus in particular on the minimal natural SUSY scenario of Refs. [19, 20], a realistic, yet low-dimensional theory, in which the gluinos, both stops, the left handed sbottom, and the higgsinos all have masses at TeV scale while the remaining states are decoupled. The six parameters of minimal natural SUSY are: the supersymmetric Higgs mixing parameter μ, the gluino mass parameter M_3, the ratio of the two Higgs vacuum expectation values $\tan\beta$, the third generation SU(2)-doublet squark soft-breaking parameter m_{Q_3}, the third generation SU(2)-singlet soft-breaking parameter m_{t_R}, and the top trilinear soft-breaking term A_t.

Data: The experimental scenario we consider is the planned high luminosity upgrade of the LHC (HL-LHC) [21] with 3000 fb^{-1} worth of data collected at 14 TeV center-of-mass energy. We focus on two mutually exclusive SRs defined in Ref. [22], for which the ATLAS collaboration provides background estimates.[1] These SRs are optimized for direct production of stops, the most relevant production channel for natural SUSY. The typical decay channels for the stop are: top or bottom quarks, W/Z/Higgs bosons, and the lightest neutralino. The detector signature is the presence of several jets (including b-jets), large missing transverse energy, and possibly leptons. We refer to the ATLAS note for the full definitions of the SRs, and we focus here on the 0-lepton and a 1-lepton SR.

Training and testing: For training and test data we analyzed 18 647 samples generated in Ref. [20], for which SR efficiencies were calculated using SPHENO 3.2.4 [23], PYTHIA 8.210 [24, 25] with default parton distribution function set [26], NLLFAST 3.1 [27–32], and CHECK-MATE 1.2.1 [33–36] with DELPHES 3.10 [37] and FAST-JET 3.0.6 [38–40].
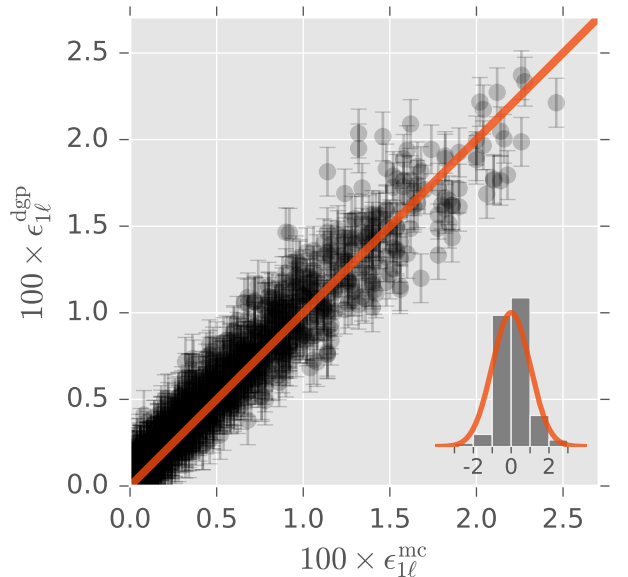
----

[1] In Ref. [22] they vary the cuts as a function of the stop mass. We use the cuts optimized for 1.1 TeV.

FIG. 1. The DGP prediction, $\epsilon_{1\ell}^{\mathrm{dgp}}$, versus the MC prediction, $\epsilon_{1\ell}^{\mathrm{mc}}$, for the 1-lepton signal region. The black circles are 2000 test points in the parameter space of natural SUSY. The errors on $\epsilon_{1\ell}^{\mathrm{dgp}}$ are those predicted by the DGP itself. The orange line shows the desired behavior $\epsilon_{1\ell}^{\mathrm{dgp}} = \epsilon_{1\ell}^{\mathrm{mc}}$. The insert shows how the distribution of $(\epsilon_{1\ell}^{\mathrm{dgp}} - \epsilon_{1\ell}^{\mathrm{mc}})/\sigma_{1\ell}^{\mathrm{dgp}}$ (gray) compares with the standard normal distribution $N(0,1)$ (orange).

We used 16 647 of these samples to train DGPs for the two SRs, with one single level architecture with an ensemble of 256 GP. Training was fast due to the use of the DGP algorithm and took approximately 15 minutes on a desktop computer with a 4.0 GHz Intel 4790K processor. We then tested the predictions of the trained DGPs on the remaining 2000 points. In Fig. 1 we show the efficiency predicted by the DGP model in the 1-lepton SR, $\epsilon_{1\ell}^{\mathrm{dgp}}$, versus the values calculated with the full MC calculation, $\epsilon_{1\ell}^{\mathrm{mc}}$, for these 2000 test points. The DGP model accurately predicts the efficiencies, which cluster around the orange line defined by $\epsilon_{1\ell}^{\mathrm{dgp}} = \epsilon_{1\ell}^{\mathrm{mc}}$, with a spread consistent with the DGP error estimate, $\sigma_{1\ell}^{\mathrm{dgp}}$. We can quantify the agreement by calculating the χ^2; for both the 0-lepton and 1-lepton SRs we get $\chi^2 \approx 1300$, while naively expecting $\chi^2 = 2000 \pm 64$ given the 2000 degrees of freedom. The reason for these low values of χ^2 is that the DGP model slightly overestimates its error. We visualize this in the insert of Fig. 1 where we see that the distribution of $(\epsilon_{1\ell}^{\mathrm{dgp}} - \epsilon_{1\ell}^{\mathrm{mc}})/\sigma_{1\ell}^{\mathrm{dgp}}$ is more peaked than the standard normal distribution $N(0,1)$.

Reconstruction: The DGP model, thus, effectively acts as a surrogate model for the full simulation chain, opening up new opportunities in the interpretation of LHC data. For example, the DGP model can rapidly reconstruct the theory parameters of a BSM model in the case where an excess is observed on top of the Standard
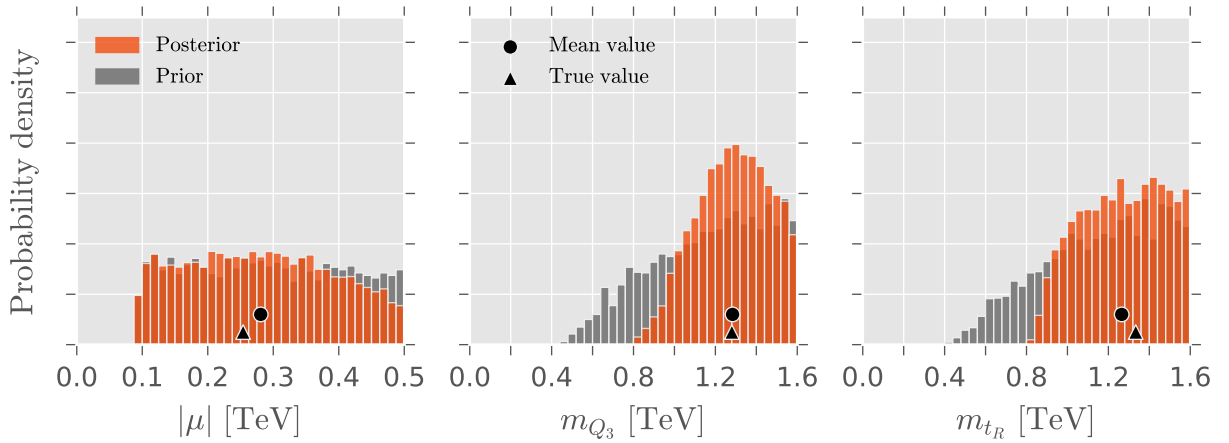
FIG. 2. Reconstruction of the natural SUSY parameters μ, m_{Q_3}, and m_{t_R} using a mock excess generated from the benchmark point. The DGPs was used to calculate the SR efficiencies; no event generation was performed. Gray histograms are the marginal priors where we only apply bounds on the Higgs boson mass $\in [121, 129]$ GeV and the chargino mass > 103.5 GeV. The orange histograms show the marginal posteriors when we also fit to the mock excess. We scan over both positive and negative μ but show only results for $|\mu|$ as they are symmetric.

Model background. To demonstrate the feasibility of a full reconstruction procedure, we perform it on an mock data set generated assuming a future excess will be detected. Our benchmark is the following point in the natural SUSY parameter space: $\mu = 254.6$ GeV, $\tan\beta = 20$, $M_3 = 2000$ GeV, $m_{Q_3} = 1280.5$ GeV, $m_{t_R} = 1333.6$ GeV, $A_t = -2000$ GeV. The physical masses are 1145 and 1413 GeV for the stops, 1259 GeV for the sbottom, and 254 GeV for the neutralino mass. This benchmark point is not excluded by current searches [41–44], but should lead to a detectable signal in our signal regions: 33.2 ($\sim 3\sigma$ excess over the background) and 68.9 ($\sim 4\sigma$) events in the 0-lepton and 1-lepton regions respectively.

As a proof-of-concept, we scan over μ, m_{Q_3}, and m_{t_R} – which are the parameters that govern the masses of the squarks and neutralino involved (the gluino is heavy) – and fix the other parameters to their benchmark values. Our priors are uniform over $\mu \in [-0.5, 0.5]$ TeV, $m_{Q_3} \in [0.1, 1.6]$ TeV, and $m_{t_R} \in [0.1, 1.6]$ TeV. We further restrict the Higgs boson mass to the range $121 < m_h < 129$ GeV range[2], and the chargino mass to be above 103.5 GeV as per the LEP-2 limit [47]. Finally for the ATLAS mock data likelihood construction, we follow the prescription in Appendix A.2 of [6]. Notice that the 0-lepton and 1-lepton signal regions considered in this work are *exclusive*, thus the joint likelihood is the multiplication of the likelihoods for the two signal regions. The uncertainty on the mock signal includes a contribution arising from the uncertainty on the calculation of

the cross sections, as well as one arising from the uncertainty on the efficiencies calculated with the DGP model. We neglect correlated systematic uncertainties. We scan the theory parameter space with a modified version of SuperBayeS[3] [53–55].

We show in Fig. 2 the marginal prior (gray) and the marginal posteriors (orange), i.e. the probability distributions after taking the excess in the mock data into consideration. As one can see, the mock data in the two signal regions have a very limited impact on the determination of μ, but they lead to a measurement on m_{Q_3}, and to a more stringent lower limit on m_{t_R}. The posterior slightly disfavors larger values of $|\mu|$ because it determines the neutralino mass, and the missing energy cut requires that the mass difference between the neutralino and the produced squark is large. The parameters m_{Q_3} and m_{t_R}, together with the off-diagonal entries of the stop mass matrix, determine the mass of the three light squarks; $\tilde{b}_1$, $\tilde{t}_1$, and $\tilde{t}_2$. The analysis is sensitive to these masses through the production cross-sections, as the total cross-section must be large enough to produce the measured number of events. The total production cross-section is dominated by two contributions: one that depends on both m_{Q_3} and m_{t_R} (via the mass of $\tilde{t}_1$) and one that depends on m_{Q_3} only (via the mass of $\tilde{b}_1$). The constraint on the total production cross-section therefore translates more directly into a constraint on m_{Q_3}, explaining why we are able to reconstruct m_{Q_3} better than m_{t_R}.

---

[2] This range is the 2σ interval around the central value $m_h = 125.09 \pm 0.24$ GeV [45], with a theoretical error of 2 GeV added in quadrature[46].

[3] The core of SuperBayeS is the MULTINEST scanning algorithm [48, 49] which is interfaced with SPHENO 3.2.4 [23] and NLL-FAST 3.1 [27–32] to calculate observables. The likelihood is implemented in the ROOT framework [50] with the RooFit [51] and RooStats [52] packages.

In reconstructing the benchmark parameters from mock data we used 4 000 live points in MULTINEST, which required 10^5 likelihood evaluations before converging. The estimate of the two SR efficiencies with the DGP model took 0.06 seconds per evaluation on a single 4 GHz Intel 4790K core, *a factor $\sim 10^4$ faster than the $\mathcal{O}(10)$ minutes per evaluation required to generate the training set.* The whole scan took 66 hours on six CPUs.

Discussion and Conclusions: In this letter, we have introduced the use of Gaussian processes to accelerate the interpretation of LHC data in the framework of BSM theories. Their ability to estimate an error on their predictions makes them ideal for fast and robust approximate calculations. We have specifically demonstrated that the estimate of SR efficiencies can be accelerated by a factor 10^4, making it possible to rapidly and accurately reconstruct the natural SUSY theory parameters, should an excess in the data be discovered at the HL-LHC. The method can be generalized to any BSM theory, and it can be in principle extended to accelerate other time-consuming tasks, such as the calculation of the cross-sections, or of the likelihood itself.

Gaussian processes are currently a very active area of research and it is likely that the method will be further improved and refined. One particularly interesting extension to our implementation is multi-output prediction: instead of training one (D)GP for each SR separately, one can train a single (D)GP that predicts all the SR efficiencies simultaneously. The correlation between the SR efficiencies could then be used to make more precise predictions. Another direction is to move away from Gaussianity and use Student-t processes [56] which might model the underlying noise for the Monte Carlo generators better. Another intriguing development is Bayesian optimization [57], a form of active learning [58, 59], which aims to minimize the amount of training data needed, by letting the Gaussian process itself specify where to sample the theory parameter space next in an iterative fashion.

In the traditional approach to global analyses, new samples have to be generated in the parameter space of a BSM theory, and new simulations performed every time new data become available, as the sampling is driven by the likelihood. An important aspect of our new method is that detector simulations need to be performed *only once* for each BSM theory, to generate the training sample. Once this is done, the surrogate model can be reused by anyone, and applied to any data set.

---

[1] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rept. **267**, 195 (1996), arXiv:hep-ph/9506380 [hep-ph].

[2] G. Bertone, D. Hooper, and J. Silk, Phys. Rept. **405**, 279 (2005), arXiv:hep-ph/0404175 [hep-ph].

[3] G. Bertone, Nature **468**, 389 (2010), arXiv:1011.3532 [astro-ph.CO].

[4] D. Alves (LHC New Physics Working Group), J. Phys. **G39**, 105005 (2012), arXiv:1105.2838 [hep-ph].

[5] J. Abdallah *et al.*, Phys. Dark Univ. **9-10**, 8 (2015), arXiv:1506.03116 [hep-ph].

[6] C. Strege, G. Bertone, G. J. Besjes, S. Caron, R. Ruiz de Austri, A. Strubig, and R. Trotta, JHEP **09**, 081 (2014), arXiv:1405.0622 [hep-ph].

[7] G. Aad *et al.* (ATLAS), JHEP **10**, 134 (2015), arXiv:1508.06608 [hep-ex].

[8] K. J. de Vries *et al.*, Eur. Phys. J. **C75**, 422 (2015), arXiv:1504.03260 [hep-ph].

[9] M. Aaboud *et al.* (ATLAS), JHEP **09**, 175 (2016), arXiv:1608.00872 [hep-ex].

[10] V. Khachatryan *et al.* (CMS), Submitted to: JHEP (2016), arXiv:1606.03577 [hep-ex].

[11] S. Caron, J. S. Kim, K. Rolbiecki, R. Ruiz de Austri, and B. Stienen, (2016), arXiv:1605.02797 [hep-ph].

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (The MIT Press, 2005).

[13] M. P. Deisenroth and J. W. Ng, Proceedings of The 32nd International Conference on Machine Learning **37**, 1481 (2015), arXiv:1502.02843 [stat.ML].

[14] J. L. Feng, K. T. Matchev, and T. Moroi, Phys. Rev. Lett. **84**, 2322 (2000), arXiv:hep-ph/9908309 [hep-ph].

[15] R. Kitano and Y. Nomura, Phys. Rev. **D73**, 095004 (2006), arXiv:hep-ph/0602096 [hep-ph].

[16] H. Baer, V. Barger, and P. Huang, JHEP **11**, 031 (2011), arXiv:1107.5581 [hep-ph].

[17] M. Papucci, J. T. Ruderman, and A. Weiler, JHEP **09**, 035 (2012), arXiv:1110.6926 [hep-ph].

[18] H. Baer, V. Barger, P. Huang, and X. Tata, JHEP **05**, 109 (2012), arXiv:1203.5539 [hep-ph].

[19] M. Drees and J. S. Kim, Phys. Rev. **D93**, 095005 (2016), arXiv:1511.04461 [hep-ph].

[20] J. S. Kim, K. Rolbiecki, R. Ruiz, J. Tattersall, and T. Weber, (2016), arXiv:1606.06738 [hep-ph].

[21] "LHC commissioning schedule," http://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-schedule-update.pdf, online; accessed 25 October 2016.

[22] M. Aaboud *et al.* (ATLAS), *Prospects for benchmark*

Supersymmetry searches at the high luminosity LHC with the ATLAS Detector, Tech. Rep. ATL-PHYS-PUB-2013-011 (CERN, Geneva, 2013).

[23] W. Porod and F. Staub, Comput. Phys. Commun. **183**, 2458 (2012), arXiv:1104.1573 [hep-ph].

[24] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].

[25] N. Desai and P. Z. Skands, Eur. Phys. J. **C72**, 2238 (2012), arXiv:1109.5852 [hep-ph].

[26] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, D. Stump, W.-K. Tung, and C. P. Yuan, Phys. Rev. **D78**, 013004 (2008), arXiv:0802.0007 [hep-ph].

[27] W. Beenakker, R. Hopker, M. Spira, and P. M. Zerwas, Nucl. Phys. **B492**, 51 (1997), arXiv:hep-ph/9610490 [hep-ph].

[28] W. Beenakker, M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas, Nucl. Phys. **B515**, 3 (1998), arXiv:hep-ph/9710451 [hep-ph].

[29] A. Kulesza and L. Motyka, Phys. Rev. Lett. **102**, 111802 (2009), arXiv:0807.2405 [hep-ph].

[30] A. Kulesza and L. Motyka, Phys. Rev. **D80**, 095004 (2009), arXiv:0905.4749 [hep-ph].

[31] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, E. Laenen, and I. Niessen, JHEP **08**, 098 (2010), arXiv:1006.4771 [hep-ph].

[32] W. Beenakker, S. Brensing, M. n. Kramer, A. Kulesza, E. Laenen, L. Motyka, and I. Niessen, Int. J. Mod. Phys. **A26**, 2637 (2011), arXiv:1105.1110 [hep-ph].

[33] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall, and J. S. Kim, Comput. Phys. Commun. **187**, 227 (2015), arXiv:1312.2591 [hep-ph].

[34] J. S. Kim, D. Schmeier, J. Tattersall, and K. Rolbiecki, Comput. Phys. Commun. **196**, 535 (2015), arXiv:1503.01123 [hep-ph].

[35] D. Dercks, N. Desai, J. S. Kin, K. Rolbiecki, J. Tattersall, and T. Weber, (2016), CTPU-16-36,CSIC-16-117 [hep-ph].

[36] http://checkmate.hepforge.org/.

[37] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi (DELPHES 3), JHEP **02**, 057 (2014), arXiv:1307.6346 [hep-ex].

[38] M. Cacciari and G. P. Salam, Phys. Lett. **B641**, 57 (2006), arXiv:hep-ph/0512210 [hep-ph].

[39] M. Cacciari, G. P. Salam, and G. Soyez, JHEP **04**, 063 (2008), arXiv:0802.1189 [hep-ph].

[40] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. **C72**, 1896 (2012), arXiv:1111.6097 [hep-ph].

[41] G. Aad *et al.* (ATLAS), Eur. Phys. J. **C75**, 510 (2015), [Erratum: Eur. Phys. J.C76,no.3,153(2016)], arXiv:1506.08616 [hep-ex].

[42] M. Aaboud *et al.* (ATLAS), Phys. Rev. **D94**, 052009 (2016), arXiv:1606.03903 [hep-ex].

[43] M. Aaboud *et al.* (ATLAS), *Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2016-050 (CERN, Geneva, 2016).

[44] M. Aaboud *et al.* (ATLAS), *Search for the Supersymmetric Partner of the Top Quark in the Jets+Emiss Final State at sqrt(s) = 13 TeV*, Tech. Rep. ATLAS-CONF-2016-077 (CERN, Geneva, 2016).

[45] C. Patrignani, Chin. Phys. **C40**, 100001 (2016).

[46] B. C. Allanach, A. Djouadi, J. L. Kneur, W. Porod, and P. Slavich, JHEP **09**, 044 (2004), arXiv:hep-ph/0406166 [hep-ph].

[47] "LEPSUSYWG, ALEPH, DELPHI, L3 and OPAL experiments, note LEPSUSYWG/01-03.1," http://lepsusy.web.cern.ch/lepsusy/Welcome.html, online; accessed 25 October 2016.

[48] F. Feroz and M. Hobson, Mon.Not.Roy.Astron.Soc. **384**, 449 (2008), arXiv:0704.3704 [astro-ph].

[49] F. Feroz, M. P. Hobson, and M. Bridges, (2008), arXiv:0809.3437 [astro-ph].

[50] R. Brun and F. Rademakers, *New computing techniques in physics research V. Proceedings, 5th International Workshop, AIHENP '96, Lausanne, Switzerland, September 2-6, 1996*, Nucl. Instrum. Meth. **A389**, 81 (1997).

[51] W. Verkerke and D. P. Kirkby, *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, September 12-15, 2005*, eConf **C0303241**, MOLT007 (2003), [,186(2003)], arXiv:physics/0306116 [physics].

[52] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, *Proceedings, 13th International Workshop on Advanced computing and analysis techniques in physics research: Jaipur, India, February 22-27*, PoS **ACAT2010**, 057 (2010), arXiv:1009.1003 [physics.data-an].

[53] R. R. de Austri, R. Trotta, and L. Roszkowski, JHEP **0605**, 002 (2006), arXiv:hep-ph/0602028 [hep-ph].

[54] L. Roszkowski, R. Ruiz de Austri, and R. Trotta, JHEP **0707**, 075 (2007), arXiv:0705.2012 [hep-ph].

[55] R. Trotta, F. Feroz, M. P. Hobson, L. Roszkowski, and R. Ruiz de Austri, JHEP **0812**, 024 (2008), arXiv:0809.3792 [hep-ph].

[56] A. Shah, A. G. Wilson, and Z. Ghahramani, in *AISTATS* (2014) pp. 877–885.

[57] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, Proceedings of the IEEE **104**, 148 (2016).

[58] S. Seo, M. Wallat, T. Graepel, and K. Obermayer, in *Mustererkennung 2000* (Springer, 2000) pp. 27–34.

[59] B. Settles, *Active Learning Literature Survey*, Computer Science Technical Report 1648 (University of Wisconsin–Madison, 2009).