Implementation of full and simplified likelihoods in CheckMATE

Iñaki Lara and Krzysztof Rolbiecki

Faculty of Physics, University of Warsaw
ul. Pasteura 5, PL-02-093 Warsaw, Poland

Inaki.Lara@fuw.edu.pl, Krzysztof.Rolbiecki@fuw.edu.pl

February 4, 2026

Abstract

We present the implementation of simplified and full likelihood models for multibin signal regions in CHECKMATE. A total of 13 searches are included from ATLAS and CMS, and several methods are presented for the implementation and evaluation of likelihood functions. Statistical combinations increase the sensitivity of searches and open up the possibility of combining orthogonal search channels in the CHECKMATE framework.

1 Introduction

The Run 3 at the Large Hadron Collider (LHC) is now in full swing, and by the end of the current run it is expected that the LHC would collect several hundreds of inverse femtobarns of data. Apart from the Higgs boson discovery [1, 2], which made the experimental picture of the Standard Model complete, no sign of new physics has been observed. However, the vast data has only been interpreted in the limited number of theoretical models by the LHC experiments. On many occasions, the interpretation is provided in terms of simplified models [3] which do not directly correspond to realistic TeV-scale physics. This situation has prompted several groups [4–9] to provide computer programs that would allow reinterpretation of experimental results in terms of an arbitrary new physics model [10].

Over the years, searches for new physics were becoming increasingly sophisticated, including the full use of complicated statistical models [11] and machine learning methods [12,13]. Instead of reporting results in a handful of bins, often optimized for a small class of models, and comparing observed and expected numbers of events, the experiments nowadays provide binned data in several observables in signal and control regions. On the one hand, this can significantly improve the sensitivity, but on the other hand, it greatly complicates the reinterpretation of the results. In this approach, experiments in high-energy physics test the compatibility of collision data with theoretical predictions, which can be described as a likelihood function. The function gives the probability of the data assuming a theoretical model and a certain set of parameters. The likelihood can therefore be used to constrain the parameters of the model.

In this paper, we discuss the implementation of the full and simplified likelihoods in the CHECK-MATE framework [14–17]. For implementations in other tools, see, e.g. [18,19]. In this paper, we include 9 ATLAS and 4 CMS searches. The CMS searches use the correlated background model (covariance matrix) [20]. For ATLAS searches, we use the simplified [21] and full likelihood [22] frameworks. In addition, some of the implementations of the ATLAS searches include the full set of control regions. Users have several ways of controlling the actual calculation of likelihoods. The default method of evaluation is based on the SPEY package [23]. We provide validation plots for each of the searches. A comparison of three methods for the calculation of limits is presented: the best-expected signal region, the full likelihood model (if available), and the simplified likelihood.

The paper is organized as follows. In Section 2, we briefly discuss statistical models and their technical implementation in CHECKMATE. We provide a list of implemented searches and options available to a user to control the program's behavior. In Section 3 we present validation plots for each of the searches discussed. A brief summary and outlook are provided in the last section.

2 Technical implementation

In this section, we introduce the methods for the implementation of simplified and full likelihoods in CHECKMATE and user switches to control their execution.

2.1 ATLAS

The functionality of combining signal regions for recasting in ATLAS searches can be implemented using either the full likelihood model [22] or following a simplified approach detailed in Ref. [21]. Table 1 lists the ATLAS analyses with likelihood functionality implemented in CHECKMATE. The simplified likelihood method requires background rates and uncertainties that were already available in the implemented searches. The full likelihood requires an appropriate file in the JSON format and these files were released by ATLAS for 6 searches already implemented in CHECKMATE.

For the searches `atlas_2004_14060`, `atlas_2006_05880`, and `atlas_2111_08372` the full model files are not available but using the published data one can still perform a simplified model fitting in multibin signal regions.

The full-likelihood statistical models are encoded in the JSON files by the ATLAS Collaboration. The files are not shipped with CHECKMATE but are automatically downloaded during installation from the HEPDATA website [24, 25], `https://www.hepdata.net/`. The information provided includes the number of background events for all signal and control regions and for each major background category separately. This results in a large number of nuisance parameters and the complexity of the procedure makes the hypothesis testing very CPU-expensive. Additionally, on the recasting side, in order to fully exploit the method, one should also implement control regions (CR), which was not a standard approach in CHECKMATE. Currently, only two searches `atlas_2010_14293` and `atlas_1911_06660` have complete implementation of all CRs. In other searches, it is assumed that the contribution of signal to CRs is negligible. This assumption is not obviously fulfilled in all imaginable new physics models.

On the technical side, after the usual evaluation of events within CHECKMATE, a JSON patchset is created that encapsulates signal contributions to signal regions (SRs) (and CRs if applicable). The patchset is then combined with the background-only model from ATLAS. Since the signal region names in CHECKMATE do not usually exactly match the names in ATLAS conventions, an additional file is required, `pyhf_conf.json`, which provides a dictionary between two conventions. The likelihood is further evaluated using the PYHF [26–28] package, which is a Python implementation of the HISTFACTORY specification for binned statistical models [29, 30]. The signal strength μ is the parameter of interest, where $\mu = 1$ corresponds to the nominal cross section of a tested model. Depending on the user's choices, the output can contain information about the expected and observed upper limit on μ, with 2-σ bounds, along with the observed and expected CL_s for $\mu = 1$, where

$$\mathrm{CL}_s = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} \tag{1}$$

and CL_{s+b}, CL_b are the p-values of the signal and the null hypotheses, respectively. Asymptotic formulae are the default calculation method; see [26].

By default, the above calculation will be executed using the SPEY program [23] and the SPEY-PYHF plugin [31]. SPEY is a Python-based cross-platform package that allows for statistical inference of hypotheses using different likelihood prescriptions.[♮1] In our setup, it gives a somewhat better control over the calculation than the above-mentioned CHECKMATE-PYHF interface, but nevertheless the calculation is still performed in the PYHF framework. However, the main motivation for using SPEY was the possibility of combining different searches (also between experiments), which is planned in the next release of CHECKMATE. In any case, a direct evaluation using PYHF and bypassing SPEY is also available.

Since the evaluation of full likelihoods is normally time-consuming, it is not practical for large scans of the parameter space. Therefore, the alternative approach is based on the concept of simplified likelihood [21]. In this case, the background model is approximated with the total SM background rate obtained in the background-only fit in the full model. A single nuisance parameter correlated over all bins and representing post-fit background uncertainty is constrained by the unit normal distribution. The evaluation is also performed using the PYHF package. As such, a simplified likelihood can usually be constructed from publicly available information for many ATLAS studies.[♮2]

[♮1]Installation of SPEY is straightforward: `pip install spey`. Please refer to SPEY online documentation and SPEY-PYHF for more details [32].

[♮2]Another approach was adopted in MADANALYSIS: the SIMPLIFY tool [33] is used to construct simplified likelihoods from full likelihood models, and technically both approaches should be equivalent.

Name	Description	#SR	N_{bin}	Full	Ref.
atlas_1908_03122	Search for bottom squarks in final states with Higgs bosons, b-jets and E_T^{miss}	2	7	✓	[36]
atlas_1908_08215	Search for electroweak production of charginos and sleptons in final states with 2 leptons and E_T^{miss}	1	52	✓	[37]
atlas_1911_06660	Search for direct stau production in events with two hadronic taus	1	2	✓	[38]
atlas_1911_12606	Search for electroweak production of supersymmetric particles with compressed mass spectra	2	76	✓	[39]
atlas_2004_14060	Search for stops in hadronic final states with E_T^{miss}	3	14	✗	[40]
atlas_2006_05880	Search for top squarks in events with a Higgs or Z boson	3	23	✗	[41]
atlas_2010_14293	Search for squarks and gluinos in final states with jets and E_T^{miss}	3	60	✓	[42]
atlas_2101_01629	Search for squarks and gluinos in final states with one isolated lepton, jets, and E_T^{miss}	1	26	✓	[43]
atlas_2111_08372	Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates	1	22	✗	[44]

Table 1: List of implemented ATLAS analyses which have likelihood-based signal regions (all searches at $\sqrt{s} = 13$ TeV and $\mathcal{L} = 139$ fb^{-1}).

In CHECKMATE this is performed on the fly in a dedicated routine. This feature should be used with caution, and we advise cross-checking with the full likelihood calculation. In certain situations, the simplified likelihood can result in significant over-constraining of models, as it was shown in Ref. [34] (Appendix A2) in the case of invisible Higgs boson decays in an implementation of the ATLAS analysis [35]. By default, CHECKMATE performs full likelihood fits and the simplified method has to be activated using `Model: simple`, see Tab. 3.

2.2 CMS

The simplified likelihood framework was defined in Ref. [20]. This assumes correlation between background contributions that can be modeled using the multivariate Gaussian distribution:

$$\mathcal{L}_S(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{(\mu \cdot s_i + b_i + \theta_i)^{n_i} e^{-(\mu \cdot s_i + b_i + \theta_i)}}{n_i!} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{V}^{-1} \boldsymbol{\theta}\right) \quad (2)$$

where the product runs over all bins and μ is the signal strength (and the Parameter of Interest - POI), n_i the observed number of events, s_i the expected number of signal events, b_i the expected number of background events, θ_i a background nuisance parameter, and $\mathbf{V}$ the covariance matrix. It is implemented using the covariance matrices provided by the CMS Collaboration, which are included in the CHECKMATE distribution in the JSON format. The evaluation of the above model is performed using the SPEY package and the `default_pdf.correlated_background` method.

2.3 CheckMATE parameters

CHECKMATE provides several switches and parameters to control the details of the statistical evaluation. These are summarized in Tab. 3. The switches are divided into two groups: one providing a control of what statistical tests are performed, and the other to control different modes of calculation. By default, no statistical evaluation is performed. For the sake of speed and

Name	Description	N_{bin}	Ref.
cms_1908_04722	Search for supersymmetry in final states with jets and E_T^{miss}	174	[45]
cms_1909_03460	Search for supersymmetry with M_{T2} variable in final states with jets and E_T^{miss}	282	[46]
cms_2107_13021	Search for new particles in events with energetic jets and large E_T^{miss}	66	[47]
cms_2205_09597	Search for production of charginos and neutralinos in final states containing hadronic decays of WW, WZ, or WH and E_T^{miss}	35	[47]

Table 2: List of implemented CMS analyses which have likelihood-based signal regions (all searches at $\sqrt{s} = 13$ TeV and $\mathcal{L} = 139$ fb^{-1}).

stability, one switch, `scan`, provides a quick and reliable way of obtaining an `Allowed/Excluded` result but with limited additional information. Generally, available statistics include CL_s tests and calculation of upper limits on signal strength, both of which can be obtained as observed and/or expected measures. By choosing a `select` switch, users can control which statistics are calculated. If no explicit choice is made, the *observed upper limit* will be calculated. Finally, the `detailed` switch can be used to request calculation of all available statistics, but it should be noted that its execution can be time consuming. The option `-so` can be used to request the calculation of statistics for previous CHECKMATE runs (it requires the presence of the `evaluation/total_results.dat` file in the output directory).

The second group of parameters is used to choose a method of calculation of requested statistics for the ATLAS searches (it does not affect the calculation for the CMS searches as described in the previous Section). For the default method, the `full` switch chooses the calculation using the full likelihood and the CHECKMATE-SPEY interface. With the `simple` switch, the calculation is performed using simplified likelihood and the CHECKMATE-PYHF interface. Finally, the `fullpyhf` switch requests the calculation using the full likelihood and CHECKMATE-PYHF interface (this is somewhat less flexible with regard to the output compared to the previous options). In any case, users should remember that the full likelihood calculation can be time consuming if many searches and signal regions are requested.

The results of the calculation for each of the multibin signal regions and all requested analyses are stored in the `multibin_limits/results.dat` file. In order to follow the progress of the calculation, the observed limits are also displayed on the screen for each of the signal regions. The final evaluation is decided using the upper limit on the signal strength μ (if calculated):

$$\mu < 1 \implies \texttt{Excluded}$$
$$\mu \geq 1 \implies \texttt{Allowed}.$$

If the results for the upper limit are not available, the decision is made using the observed CL_s statistics at the 95% confidence level:

$$CL_s < 0.05 \implies \texttt{Excluded}$$
$$CL_s \geq 0.05 \implies \texttt{Allowed}.$$

2.4 Evaluation time

Four different backends for the calculation of full likelihood models in the PYHF/SPEY setup are supported: NUMPY [48], TENSORFLOW [49], PYTORCH [50], and JAX [51]. The NUMPY backend does not require additional system components to be installed, but it generally is not recommended

Parameter card	Terminal	X	Description and available choices
`Multibin: X`	`-mb X`	`none`	No signal region combination is performed (default).
		`select`	Calculates user selected statistics.
		`scan`	Calculates observed CL_s; fast and reliable for quick assessment of exclusion.
		`detailed`	Calculates observed and expected upper limits and CL_s.
`Expected: False`	`-exp`		Selects calculation of expected limits.
`CLs: False`	`-mbcls`		Selects calculation of CL_s.
`Uplim: False`	`-uplim`		Selects calculation of upper limits.
`Statonly: False`	`-so`		Calculates statistical combinations without event-level analysis provided the analysis and evaluation steps were already completed.
		`full`	The SPEYinterface to the full likelihood model for ATLAS searches (default).
`Model: X`	`-mod X`	`simple`	The simplified likelihood model for ATLAS searches.
		`fullpyhf`	The full likelihood model for ATLAS searches with PYHF interface.
`Backend: X`		`numpy`	Different backends for calculations in the `full` model; NUMPY is always available.
		`pytorch`	PYTORCH backend (if available)
		`tensorflow`	TENSORFLOW backend (if available)
		`jax`	JAX backend (if available)

Table 3: Summary of options related to multibin signal regions.

	upper limits observed and expected				CL_s
Backend	CPU A	CPU B	GPU B	CPU C	CPU A
NUMPY	11092	12621	N/A	8879	636
JAX	723	526	353	311	55
PYTORCH	583	805	N/A	404	57
TENSORFLOW	2426	2920	1906	1808	156

Table 4: Performance comparison (time in seconds) of different PYHF backends for calculation of the full likelihood in the analysis `atlas_2101_01629`. Systems A and C run on (multithread) CPU, while system B was tested on CPU and GPU. As a reference in the last column the computation time of CL_s observed limit is shown.

due to long computation time. When no backend is explicitly selected, CHECKMATE will check if the JAX backend is available. The JAX backend was also successfully tested with NVIDIA CUDA library and GPU support.

An example calculation was performed for the evaluation of the full likelihood model in the ATLAS search `atlas_2101_01629`, which is a moderately complex one of those implemented. The process considered was $pp \rightarrow \tilde{g}\tilde{g} \rightarrow qqqqWW\tilde{\chi}_1^0\tilde{\chi}_1^0$ with an intermediate chargino; masses $m_{\tilde{g}} = 2200$ GeV, $m_{\tilde{\chi}_1^\pm} = 1100$ GeV, $m_{\tilde{\chi}_1^0} = 1$ GeV; cross section $\sigma = 0.46$ fb. The sample size was 1000 events though the times in Tab. 4 do not include the sample analysis time. The tests were performed on three Ubuntu 22.04.5 systems: A - Intel i7-4770 CPU 3.4 GHz, 4/8 cores/threads;[♮3] B - Intel i7-8565U CPU 1.8 GHz, 4/8 cores/threads and NVIDIA GeForce MX150; C - Intel i7-7700 CPU 3.4 GHz, 4/8 cores/threads. JAX, PYTORCH, and TENSORFLOW take advantage of multithreading even when there are no GPUs present. On system B the CUDA 12.8 toolkit was installed for GPU computation. In Table 4 we show the results. Note that system B was tested in two configurations, with and without GPUs. The best performance is obtained with JAX and PYTORCH. For comparison, we also provide evaluation times for the CL_s observed limit performed in system A. The results agreed between different calculation methods.

As already mentioned, for large scans of parameter space it may not be practical to perform full likelihood evaluation. The simplified models improve the computation and detailed calculation of all signal strength limits and CL_s takes in our example 16 seconds for the JAX backend and 64 seconds for the NUMPY backend. Finally, if the calculation of CL_s is sufficient, it only takes about 1 second.

3 Validation

Multibin signal regions are currently available in 9 ATLAS and 4 CMS analyses, as listed in Tabs. 1 and 2. The searches are based on the full Run 2 luminosity of about 140 fb^{-1} at the center-of-mass energy $\sqrt{s} = 13$ TeV. In this section, we briefly introduce each of the searches and provide validation examples. When possible, we compare the full and simplified likelihood approaches, and provide examples of how the multibin evaluation improves exclusion limits compared to the best-signal-region approach. The best-SR (BSR) is defined as the signal region with the best *expected* sensitivity.

The validation procedure is organized as follows. For SUSY processes, events are generated using MADGRAPH5_AMC@NLO 3.1.0 [52–54] and the `MSSM_SLHA2` [55] UFO [56] model with up to two

---

[♮3]It is really old.

additional partons in the final state. The NNPDF23LO [57–59] parton distribution function (PDF) set is used. The events are then interfaced with PYTHIA 8.3 [60,61] to model decays, hadronization, and showering. The matrix element and parton shower matching for ATLAS searches was performed using the CKKW-L [62] prescription and a matching scale of 1/4 of the SUSY particle mass was set. For CMS searches, the MLM [53,63] matching scheme was used. In the final step, a fast detector simulation is performed using DELPHES 3.5 [64]. Jets are reconstructed using FASTJET [65] and the anti-k_t algorithm [66].

Inclusive signal cross sections for the production of squarks and gluinos are obtained at the approximate next-to-next-to-leading order with soft gluon resummation at the next-to-next-to-leading-logarithm order (approximate NNLO+NNLL) [67–77], following the recommendations of Ref. [78]. The signal cross sections for production of sleptons, charginos and neutralinos are computed at next-to-leading order plus next-to-leading-log precision using RESUMMINO [79–85]. This setup generally follows procedures employed within LHC experiments to obtain exclusion limits used in the validation process.

Events for the simplified scenario of DM production in the CMS search EXO-20-004 [47] were generated at NLO with MADGRAPH5_AMC@NLO 3.1.0 and the DMSIMP model [86]. Up to 2 additional partons were included and the FxFx [87] jet matching scheme was used. The samples for the 2HDM+a model [88] and the validation of HIGG-2018-26 [44] were simulated internally in CHECKMATE interfaced with MADGRAPH5_AMC@NLO 3.4.1 and hadronized with PYTHIA 8.2.

The sampling of parameter space generally follows grids provided by experiments in the supplementary materials on HEPDATA. The number of Monte-Carlo (MC) events per point varies; however, close to exclusion lines at mass limits, we aim at samples 10-20 times larger than the nominal luminosity derived values. The exclusion lines are then interpolated based on the grids of simulated parameter points, with additional sampling points occasionally added. MC errors are consistently included in the PYHF patch files.

For the purpose of validation, we present comparisons between CHECKMATE and the official ATLAS result of both expected and observed exclusion limits. Where available, the results for simplified and full likelihood are provided. In addition, we also include exclusion contours using the best-*expected* signal region. Note that this usually is based on comparing observed signal with the post-fit background-only prediction in a given bin, which typically has reduced uncertainty, and hence can be quite an aggressive approach. Also, a limit obtained with the full likelihood model is not a combination of such bin-by-bin measurements and cannot be compared in a straightforward way. On the other hand, the simplified likelihood will be such a combination in most cases and it is not surprising that it will produce stronger constraints than the full likelihood in some searches; see also Ref. [23] The latter generally shows weaker sensitivity than multibin histogram shape-fits, but can be advantageous in terms of evaluation speed.

We follow the same key for the plots throughout this section. The black solid (dashed) line is for the observed (expected) ATLAS/CMS limit. On some occasions, this is accompanied by theoretical (experimental) $\pm 1\sigma$ uncertainty. The red lines are used for the exclusion obtained with CHECKMATE using the best signal region method, solid and dashed for the observed and expected limits, respectively. Finally, the blue and green lines denote the limits obtained using the simplified and full likelihoods, respectively.

3.1 `atlas_1908_03122` (SUSY-2018-31)

This is a search [36] for bottom squark production in final states containing Higgs bosons, b-jets, and missing transverse momentum. The Higgs boson is reconstructed from two b-tagged jets. The final states contain at least 3 (SRC) or 4 (SRA, SRB) b-jets, no leptons, and large missing transverse
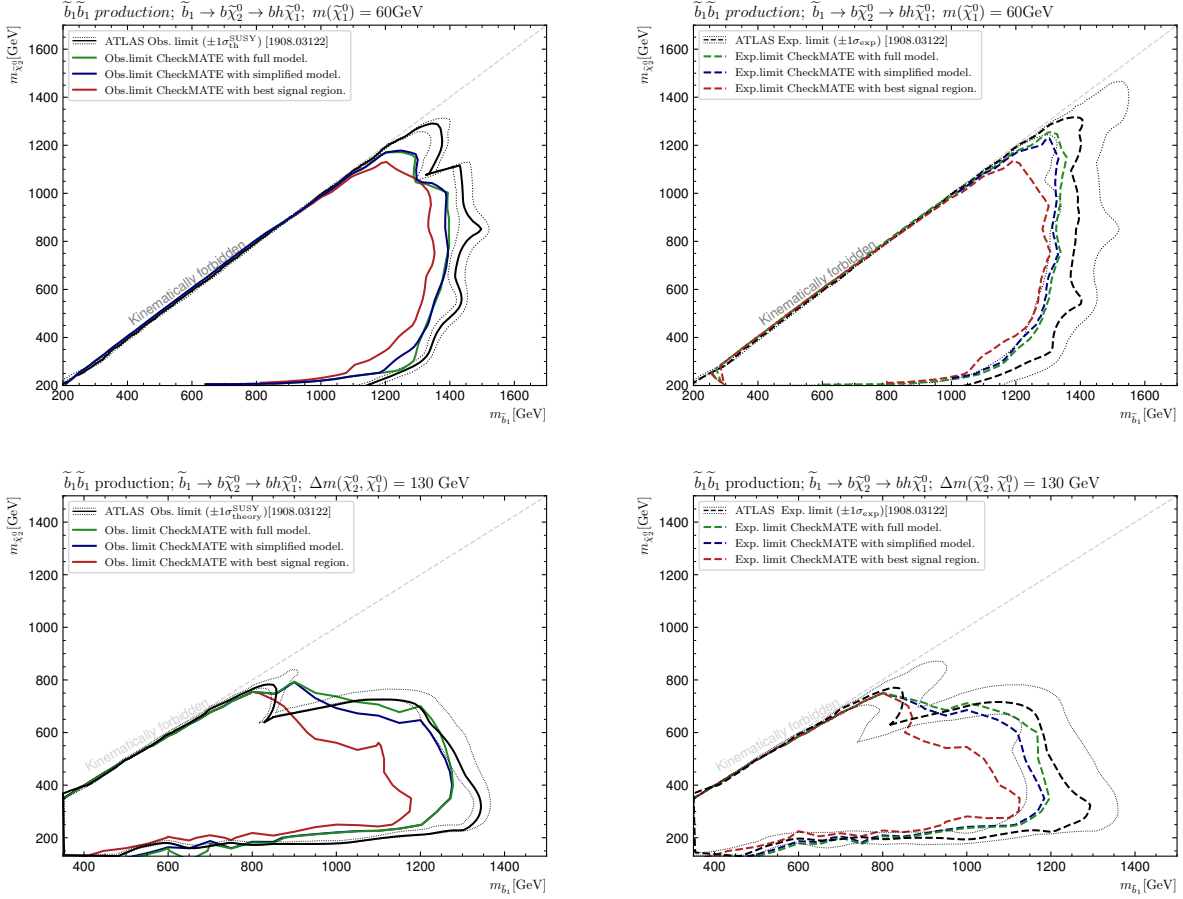
Figure 1: Validation plots for the search `atlas_1908_03122` (SUSY-2018-31). Top row: the model with $m_{\tilde{\chi}_1^0} = 60$ GeV; bottom row: the model with $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = 130$ GeV. Left panels: observed limits; right panels: expected limits.

momentum. The signal region A is divided into 3 bins according to effective mass, m_{eff}, and the signal region C is divided into 4 bins of missing transverse energy significance, $\mathcal{S}$. Thus, both SRA and SRC allow for a shape-fit analysis. Three full likelihood models are provided for the signal regions SRA, SRB, and SRC.

The validation plots for this search are presented in Fig. 1. We compare bottom squark pair production, $pp \to \tilde{b}_1 \tilde{b}_1^*$, and two distinct mass spectra. The first assumes that the mass of the lightest supersymmetric particle (LSP) is $m_{\tilde{\chi}_1^0} = 60$ GeV and the other assumes the mass difference between neutralinos $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = 130$ GeV. Generally a good agreement between CHECKMATE and ATLAS is observed for the shape-fit method with CHECKMATE being slightly weaker. In particular, there is a very good agreement between simplified and full models. The best SR method performs well for the model with fixed LSP mass, however, in the second scenario its exclusion strength can be up to 3 times weaker than shape-fits.

3.2 `atlas_1908_08215` (SUSY-2018-32)

This is a search [37] for the production of electroweakinos and sleptons. The final states with two leptons (opposite sign and same or different flavor SF/DF) and $E_{\text{T}}^{\text{miss}}$ are considered. Events
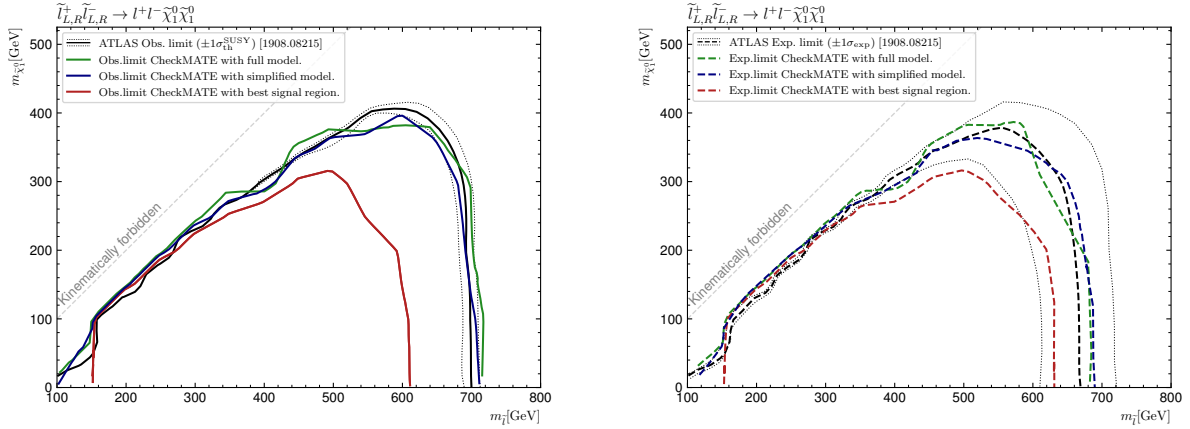
Figure 2: Validation plots for the search `atlas_1908_08215` (SUSY-2018-33), slepton pair production. Left panels: observed limits; right panels: expected limits.

are first separated into SF and DF categories and are further subdivided by the multiplicity of the non-b-tagged jets. A combined multi-bin SR is defined out of the 36 exclusive binned signal regions. The full likelihood statistical model is provided.

The validation plots for the production of slepton pairs are shown in Fig. 2. Both the full- and simplified likelihood models agree very well with the ATLAS result. The BSR method gives visibly weaker constraints. Additional validation material for CHECKMATE, including cut-flows, can be found in [89].

3.3 `atlas_1911_06660` (SUSY-2018-04)

This is a search [38] for the production of staus in final states with two hadronic τ-leptons and missing transverse momentum. Two orthogonal signal regions can be combined in a fit for which the full likelihood model is provided. The implementation includes one of the multijet control regions (CR-A), for which a significant contribution is expected, up to 30%, from signal events.

The validation plots in Fig. 3 show a similar sensitivity of the best signal region, as well as full and simplified likelihoods. In each case, good agreement with the ATLAS result is observed. The observed exclusion limits slightly exceed the ATLAS limit, whereas for the expected limits, the best agreement is obtained for the full-likelihood approach, and comfortably within the experimental 1σ uncertainty. A comparison of observed 95% CL upper limits on cross section for several representative parameter points is shown in Tab. 5. In the bottom plot of Fig. 3 we additionally compare the expected limits obtained with and without the inclusion of control regions. Note that for this search, the signal contribution to the control regions is not negligible. Without control regions, the search over-constrains the parameter space, although the discrepancy remains within the 1-σ expected uncertainty.

3.4 `atlas_1911_12606` (SUSY-2018-16)

This is a search [39] for the production of electroweakinos and sleptons in scenarios with compressed mass spectra. The final states contain two low p_T leptons (opposite sign and same or different flavor). Sensitivity of the search relies on additional initial state radiation (ISR) jets which give transverse boost to the final-state particles and add missing transverse momentum. There are
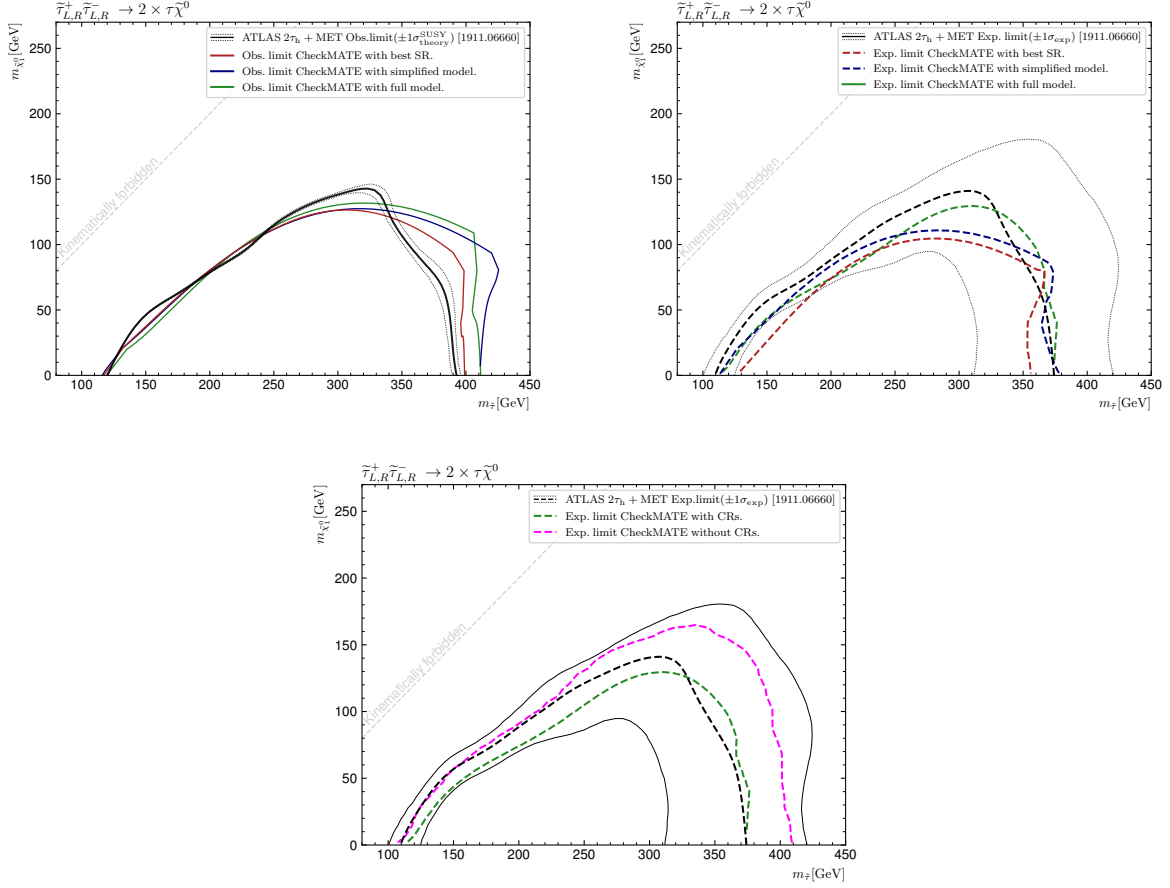
Figure 3: Validation plots for the search `atlas_1911_06660` (SUSY-2018-04). Left panels: observed limits; right panels: expected limits. In the bottom row: a comparison between the observed limits derived with and without control regions.

$(m_{\tilde{\tau}_1}, m_{\tilde{\chi}_1^0})$ [GeV]	full	simple	ATLAS
$(320, 120)$	3.7	4.0	4.6
$(320, 160)$	5.2	5.8	5.0
$(400, 80)$	1.9	1.6	2.0
$(440, 80)$	1.5	1.4	1.7

Table 5: Comparison of observed 95% CL upper limits on cross section in fb for stau pair production in the search `atlas_1911_06660` (SUSY-2018-04); full and simple columns correspond to the limits obtained with CHECKMATE using the full and simplified likelihood, respectively.

two multibin signal regions implemented in CHECKMATE:[♮4] `SR-EWK` targeting electroweakinos production and divided into 44 bins according to the lepton pair invariant mass, E_T^{miss}, and lepton flavor; `SR-S` targets the production of sleptons and is divided into 32 bins. The full likelihood model is provided.

In Figure 4, we show validation plots for slepton search signal regions. The electron and muon channels are shown separately. Very good agreement between ATLAS and CHECKMATE is obtained for the shape-fit analysis. Additionally, by comparing red exclusion curves obtained using the best-SR-method, we find a clear benefit of using multibin signal regions. Since there is a negligible difference between expected and observed results (both for ATLAS and CHECKMATE), we only show the observed limits.

In Figure 5, we show validation plots for the production of wino-like chargino-neutralino pairs (EW search region). The searches for different parities of LSP and NLSP are shown separately: upper row $m(\tilde{\chi}_2^0) \times m(\tilde{\chi}_1^0) > 0$ (same parity) and $m(\tilde{\chi}_2^0) \times m(\tilde{\chi}_1^0) < 0$ (opposite parity) in the lower row; see, e.g. Ref. [90] for a discussion of CP properties of the neutralino sector and consequences for the decay kinematics. Clearly, the simplified likelihood has a problem with reproducing the complicated shape of the observed limit due to significant excesses and deficits in several signal regions. Note that a similar conclusion was obtained by ATLAS in Ref. [21]. Therefore, for this search, we advise against using the simplified likelihood in routine runs. The agreement between the full likelihood and the ATLAS observed limit for the case of same parity neutralinos (top row) is not ideal as well, even though the expected limits comparison is rather encouraging and within 1-σ uncertainty, see the bottom plot in Fig. 5. The search is also included in HACKANALYSIS, cf. Ref. [91], SMODELS [92], and MADANALYSIS [93]. Additional validation material for CHECK-MATE, including cut-flows, can be found in [94].

3.5 `atlas_2004_14060` (SUSY-2018-12)

This is a search [40] for hadronically decaying supersymmetric partners of top quark, and up-type, 3rd generation scalar leptoquark. The final states consist of several jets, 0 leptons, and large E_T^{miss}. The requirements for b jets vary between signal regions. There are 3 multibin signal regions: `SRA-B` that targets scenarios with highly boosted top quarks in the final state and is divided into 6 bins according to an invariant mass of a large-R jet; `SRC`, for scenarios with 3-body decays of stops, that is divided into 5 bins according to a recursive jigsaw reconstruction technique variable R_{ISR} [95]; `SRD` for scenarios with compressed spectra and 4-body decays of stops which is divided into 3 bins according to the number of identified b-jets. No likelihood model is provided, and only simplified likelihood fitting is available. The combinations are performed separately for each signal category (corresponding to three likelihood models), A-B, C, and D, as described in the ATLAS paper. This avoids overlaps and accidental correlations.

The validation plots for this search are shown in Figure 6. We include the full range of stop masses and decay modes. Good agreement across the parameter plane is observed. For the high-stop-mass region the CHECKMATE result is slightly weaker than ATLAS, clearly extending the exclusion reach compared to single SR exclusion. In more compressed scenarios, the simplified fit gives a slightly too strong exclusion in some regions of the parameter space; however, this is below 30% difference in the exclusion strength. Additional validation material can be found in [96].

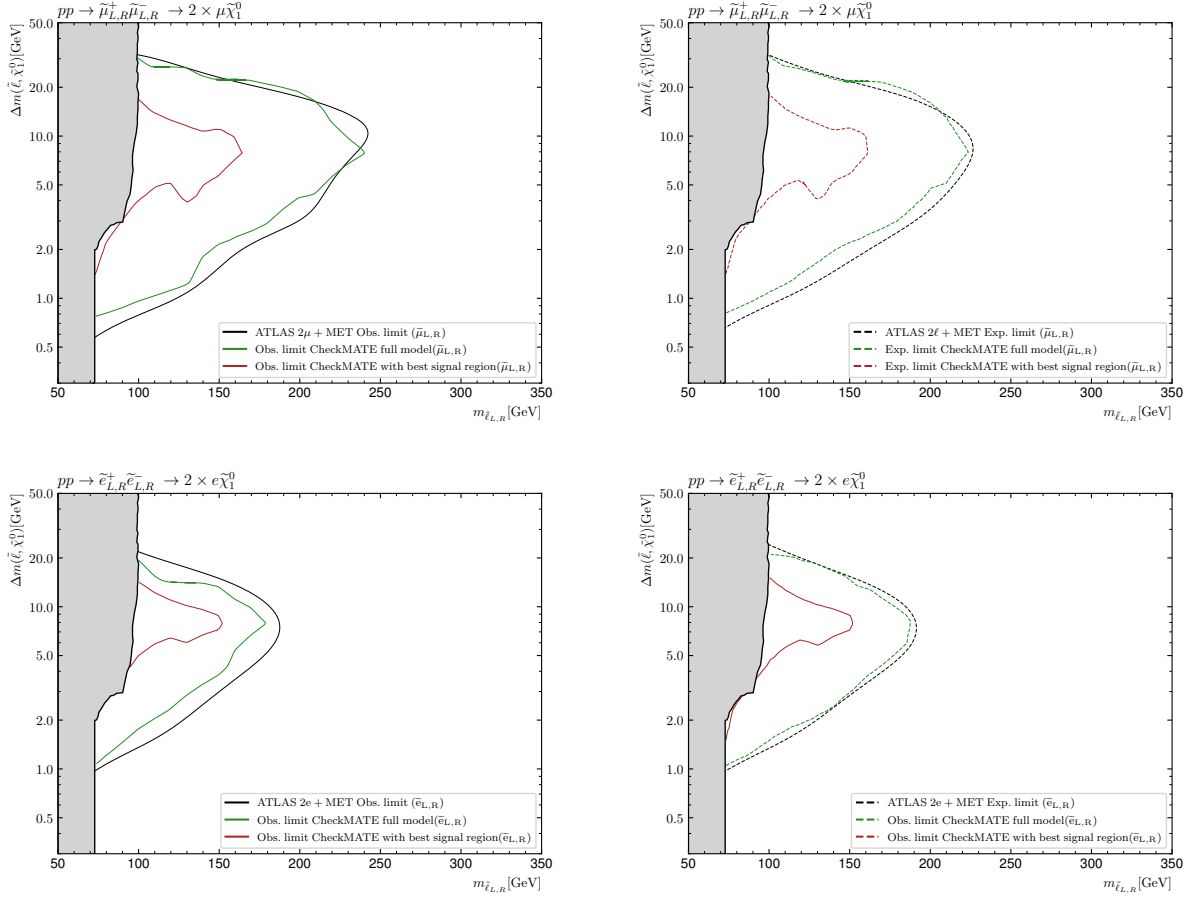[♮4]vector boson fusion (VBF) SRs are currently not available.

Figure 4: Validation plots (left: observed limits; right: expected limits) for the search `atlas_1911_12606` (SUSY-2018-16), slepton pair production. Top row: smuon search; bottom row: selectron search. Left panel: observed limits; right panel: expected limits.
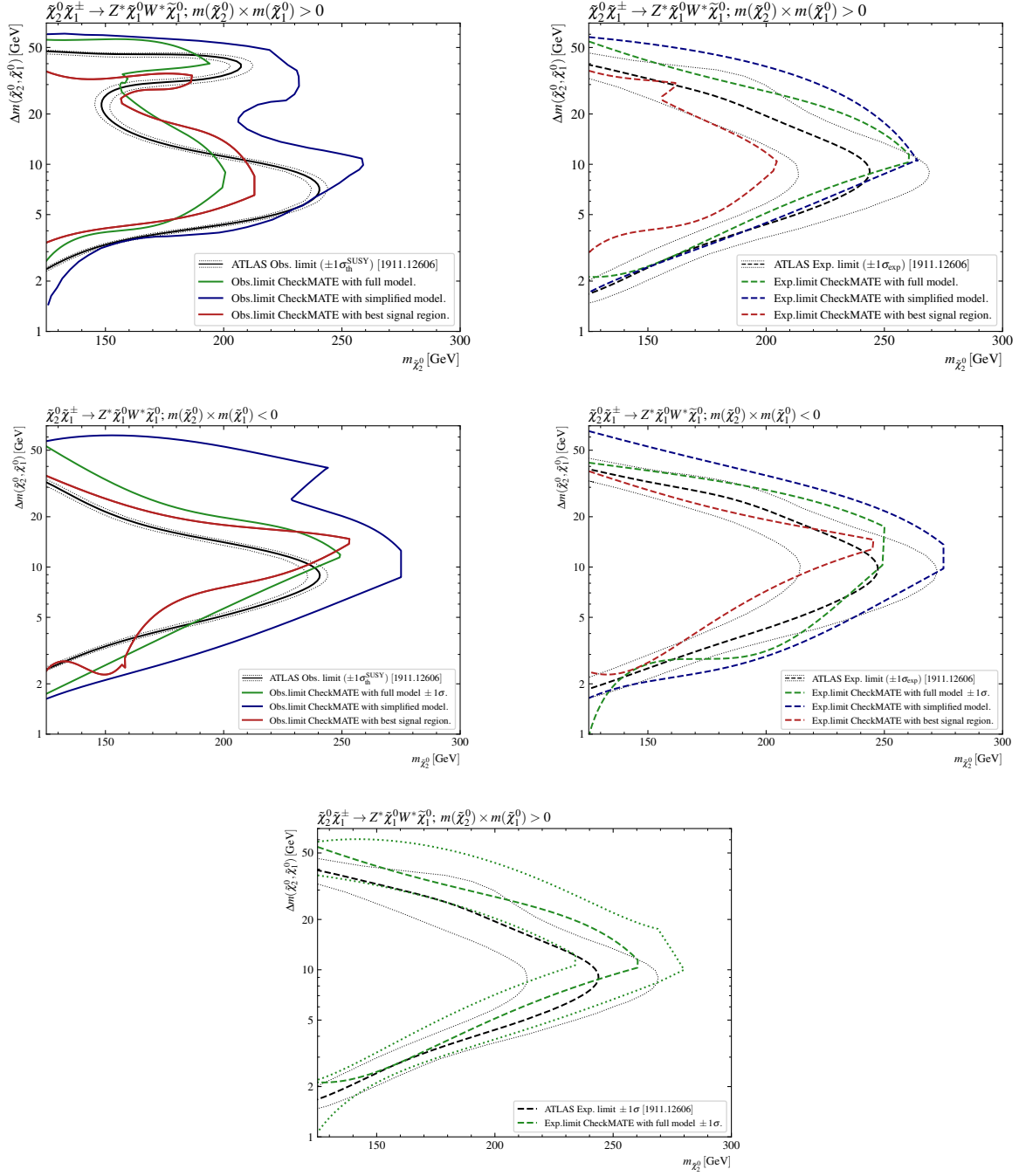
Figure 5: Validation plots for the search `atlas_1911_12606` (SUSY-2018-16). Top row: $m(\tilde{\chi}_2^0) \times m(\tilde{\chi}_1^0) > 0$; middle row: $m(\tilde{\chi}_2^0) \times m(\tilde{\chi}_1^0) < 0$. Left panels: observed limits; right panels: expected limits. In the bottom row comparison of expected 1-σ bands of ATLAS and CHECKMATE full likelihood model for the same-parity case.
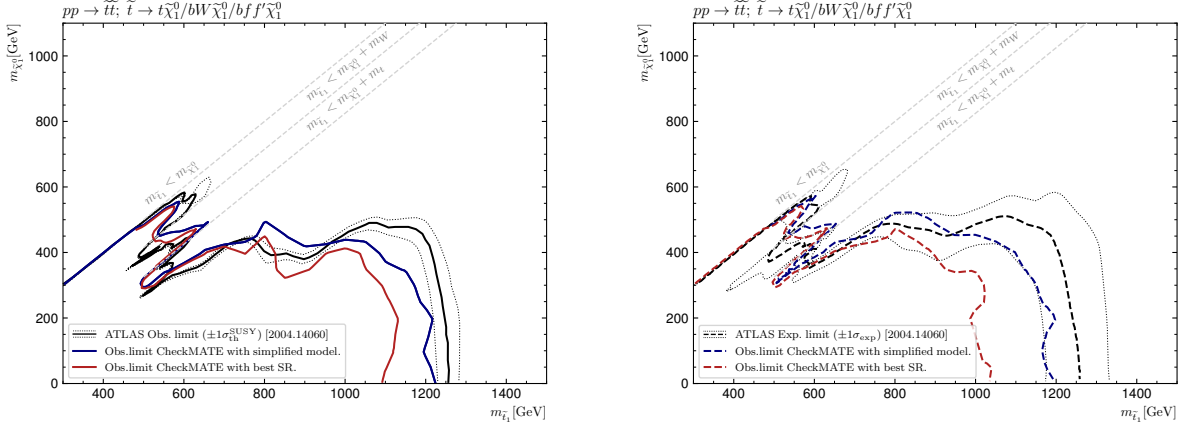
Figure 6: Validation plots for the search `atlas_2004_14060` (SUSY-2018-12). Left panels: observed limits; right panels: expected limits.
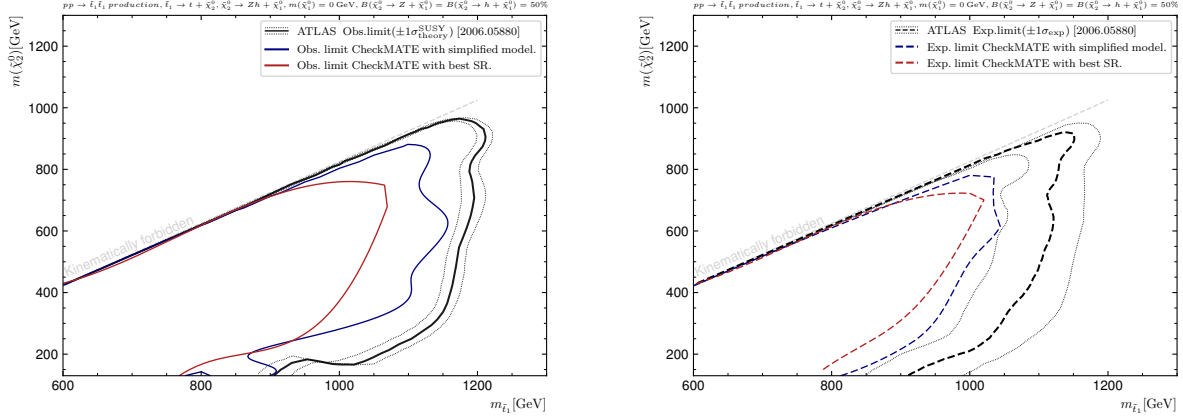

Figure 7: Validation plots for the search `atlas_2006_05880` (SUSY-2018-21). Left panels: observed limits; right panels: expected limits.

3.6 `atlas_2006_05880` (SUSY-2018-21)

This is a search [41] for supersymmetric partners of top quark decaying to a Higgs or Z boson. The final state Higgs boson is reconstructed from a pair of b-jets, whereas the Z boson is reconstructed from a same-flavor opposite-sign dilepton pair. There are 3 multibin signal regions that are shape-fits in E_T^{miss}, E_T^{miss}-significance, and p_T of the Z candidate. No likelihood model is provided, and only a simplified likelihood model is available. The combinations are performed separately for three signal categories: SR_{1A}^{hZ} (combining signal regions SR_{1A}^{Z}, SR_{1A}^{h}, SR_{1AB}^{h}), SR_{1B}^{hZ} (combining signal regions SR_{1B}^{Z}, SR_{1B}^{h}, SR_{1AB}^{h}), SR_{2}^{Z} (combining signal regions SR_{2A}^{Z} and SR_{2B}^{Z}). This ensures statistical independence of signal regions in each combination and follows ATLAS definitions of likelihood models.

The validation plots in Figure 7 again show a good agreement, with CHECKMATE somewhat weaker than ATLAS in the high stop-mass region. The observed limits clearly extend the reach compared to the best SR, increasing sensitivity by up to factor 2.
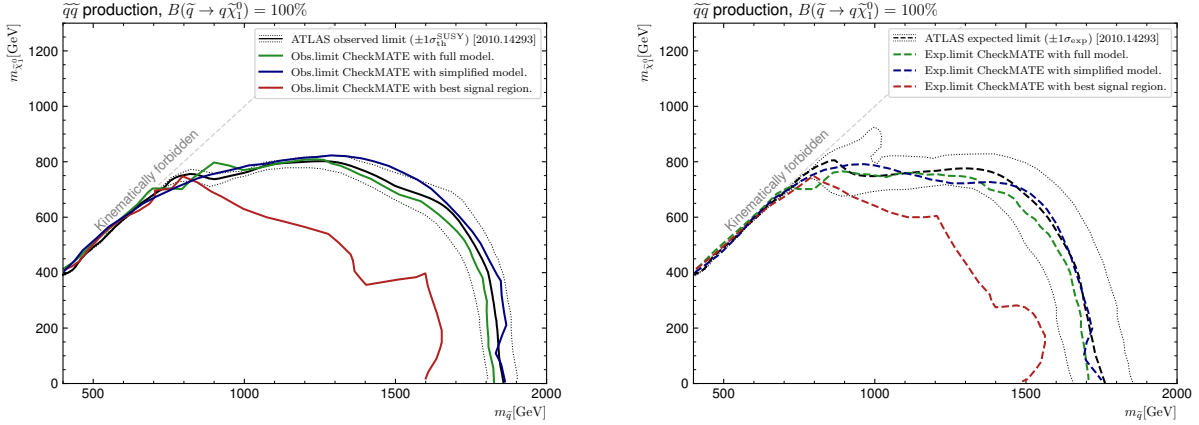
Figure 8: Validation plots for the search `atlas_2010_14293` (SUSY-2018-22), squark pair production - 8-fold degeneracy. Left panels: observed limits; right panels: expected limits.

3.7 `atlas_2010_14293` (SUSY-2018-22)

This is a search [42] for squarks (1st and 2nd generation) and gluinos in final states with 2-6 jets, 0 leptons, and missing transverse momentum. There are three multibin signal regions in this search: `MB-SSd` which targets squark pair production and is divided into 24 bins according to m_{eff}, $E_{\text{T}}^{\text{miss}}/\sqrt{H_{\text{T}}}$ and the number of jets; `MB-GGd` which targets gluino production and is divided into 18 bins according to m_{eff}, $E_{\text{T}}^{\text{miss}}/\sqrt{H_{\text{T}}}$; `MB-C` which targets compressed spectra and is divided into 18 bins according to m_{eff}, $E_{\text{T}}^{\text{miss}}/\sqrt{H_{\text{T}}}$ and the number of jets. The control regions are also implemented, which gives an opportunity to fully exploit the full likelihood model provided by the ATLAS Collaboration.

Figure 8 shows the validation plots for the squark pair-production process. The simplified shape-fit has excellent agreement with the official result, while the full likelihood model gives a somewhat weaker exclusion. In either case, there is a clearly visible advantage over the best-SR method. In Appendix we provide detailed signal yields for the multibin SRs of this search. Additional validation material (cut flows) can be found in [97].

3.8 `atlas_2101_01629` (SUSY-2018-10)

This is a search [43] for squarks and gluinos in final states with an isolated lepton, jets, and missing transverse momentum. Benchmark models assume long decay chains for squarks and gluinos with charginos, neutralinos and gauge bosons in intermediate states that give rise to the final state lepton. There is a multibin signal region that combines 26 bins defined according to the number of jets, the number of identified b-jets, and m_{eff}. The full likelihood model is provided.

The validation plots for gluino pair production, followed by the decay to intermediate mass chargino, are shown in Fig. 9. The shape of the ATLAS exclusion line is well reproduced for the likelihood models, whereas the BSR exclusion is significantly weaker.

3.9 `atlas_2111_08372` (HIGG-2018-26)

This is a search [44] for invisible decays of a Higgs boson or dark matter (DM) particles, χ, produced in association with a Z boson. The final state Z boson is reconstructed from a same-flavor opposite-sign dilepton pair. A multibin signal region, optimized for the 2HDM+a model [88],
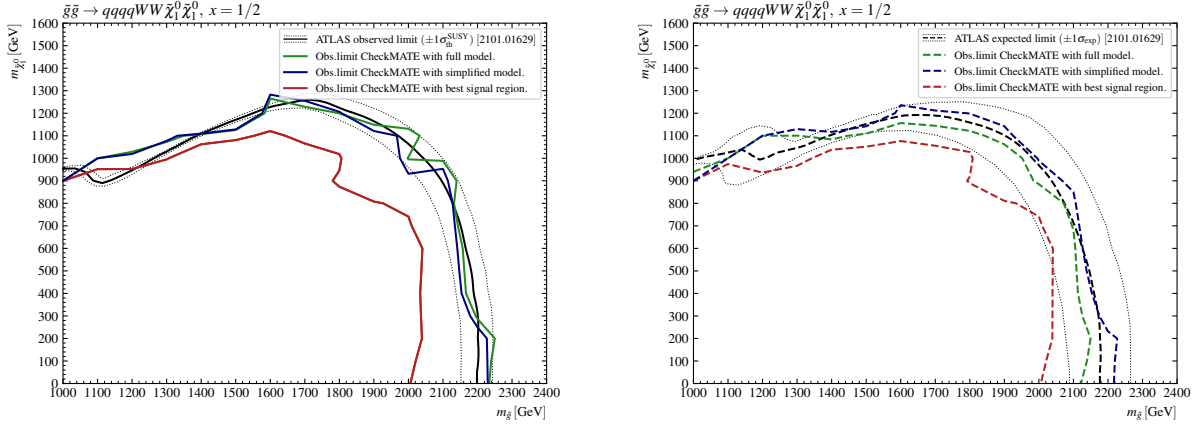
Figure 9: Validation plots for the search `atlas_2101_01629`, gluino pair production and the decay chain: $\tilde{g} \to q\bar{q}'\tilde{\chi}_1^\pm \to q\bar{q}'W\tilde{\chi}_1^0$ with $x = (m(\tilde{\chi}_1^\pm) - m(\tilde{\chi}_1^0))/(m(\tilde{g}) - m(\tilde{\chi}_1^0)) = 1/2$.
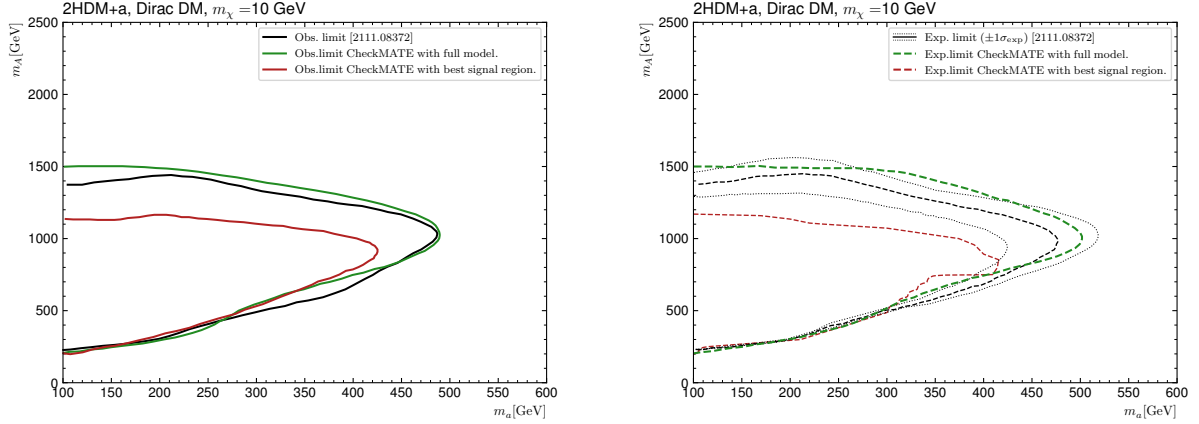

Figure 10: Validation plots for the search `atlas_2111_08372` (HIGG-2018-26) in the 2HDM+a model, $pp \to Z(\to \ell^+\ell^-)a(\to \chi\chi)$. Left panel: observed limits; right panel: expected limits.

is implemented in CHECKMATE. It is divided into 22 bins according to the transverse mass, m_T. The implementation is based on the simplified likelihood model.

As can be seen in Figure 10 we find excellent agreement between the simplified shape fit result and the official result. The best-SR exclusion is significantly weaker. It should be noted that in the parts of the plot with large m_a and/or m_A the exclusion is driven by the shape fit in the range of large m_T, which increases the sensitivity by factor ~ 2.

3.10 `cms_1908_04722` (SUS-19-006)

This is a search [45] for supersymmetric particles in final states with jets (≥ 2) and missing transverse momentum. The main discriminating variables are H_T and H_T^{miss}, the scalar sum of jet transverse momenta, and the magnitude of the vector sum of jet transverse momenta, respectively. Only the signal regions for prompt production are currently implemented in CHECKMATE. The search combines 174 bins in the simplified likelihood framework with the covariance matrix provided by the CMS Collaboration. Individual bins are defined according to the number of jets and
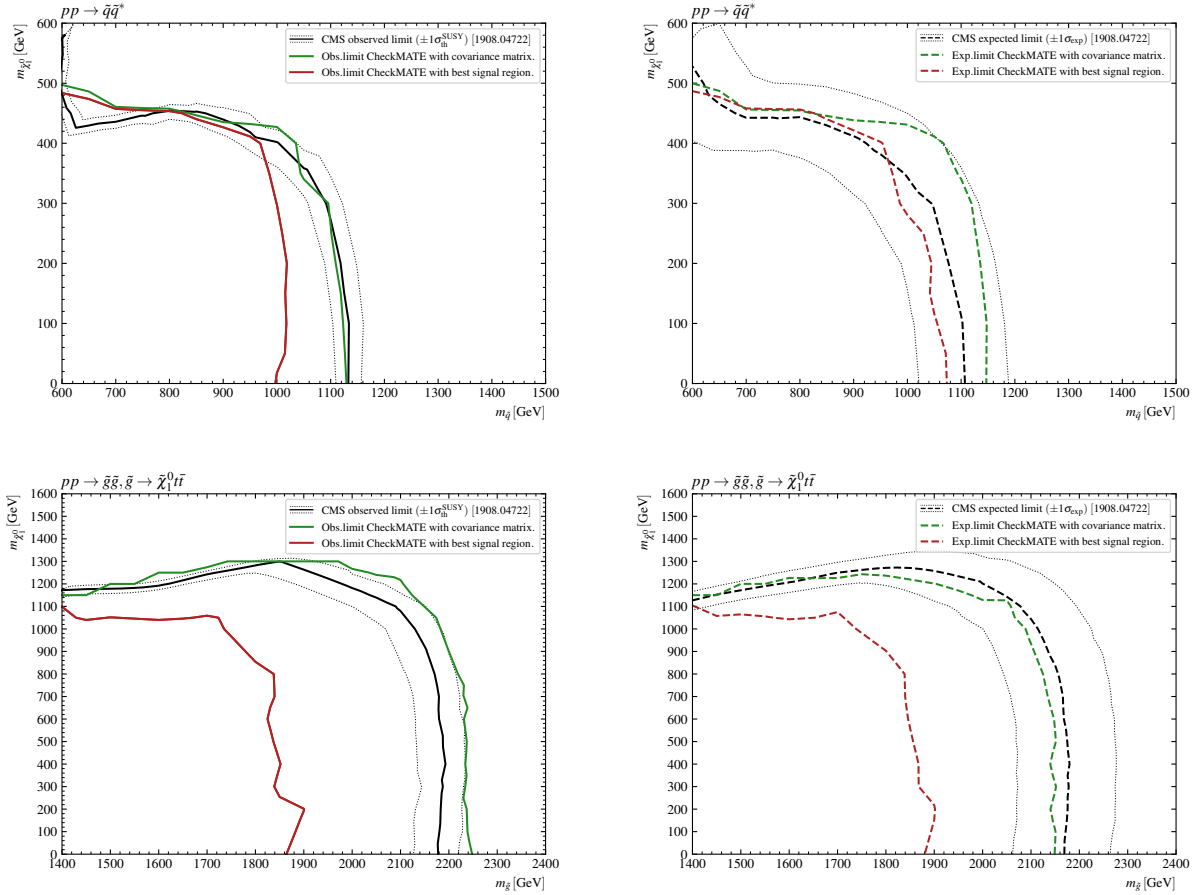
Figure 11: Validation plots for the search `cms_1908_04722`, squark pair production (1-fold degeneracy; upper row), gluino pair production with $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$ (lower row). The dotted lines around ATLAS limit denote 1-sigma uncertainty: theoretical for the observed limit and experimental for the expected limit. Left panels: observed limits; right panels: expected limits.

b-jets, H_{T}, and $H_{\mathrm{T}}^{\mathrm{miss}}$. In addition, there are 12 aggregate signal regions defined.

In Figure 11 we provide validation plots for two cases: squark pair production (single generation) and gluino pair production (followed by $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$ decay). In both cases, we observe good agreement, within 1-sigma uncertainty) between the CHECKMATE likelihood model and the CMS results. In case of gluinos there is a significant improvement with respect to the best-SR result, especially in the large gluino-mass limit. It is worth noting the small production cross section in this region: $\sigma \sim 0.6$ fb at $m_{\tilde{g}} \sim 2.2$ TeV.

3.11   `cms_1909_03460` (SUS-19-005)

This is a search [45] for supersymmetric particles (squarks and gluinos) in final states with jets (≥ 2) and a large missing transverse momentum. The events are required to have at least one energetic jet. For the inclusive search, events are required to have at least 2 jets and signal regions are based on the number of jets, b-jets, H_T, and the M_{T2} variable calculated using two pseudojets. Only the signal regions for prompt production are currently implemented in CHECKMATE. There are 21 inclusive aggregate signal regions and 282 exclusive bins, which are combined into two multi-
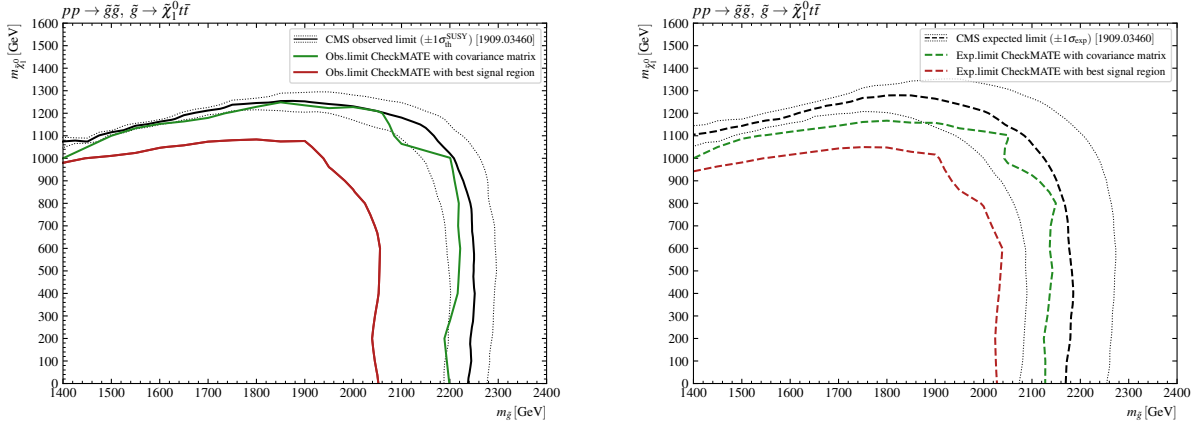
Figure 12: Validation plots for the search `cms_1909_03460`. The dotted lines around ATLAS limit denote 1-sigma uncertainty: theoretical for the observed limit and experimental for the expected limit. Left panel: observed limits; right panel: expected limits.

bin signal regions using a covariance matrix. Due to numerical problems with the entire matrix 282×282, we split the calculation into two multibin signal regions (corresponding to final states with low and high-H_T, respectively) that are marginally correlated [98]. However, these two signal regions can be further combined by a user with the `SRCombination` input parameter, and the `UnCorrStatisticsCombiner` function of the SPEY package.

In Figure 12 we show the validation plot for gluino pair production followed by decay $\tilde{g} \to t\bar{t}\tilde{\chi}_1^0$. The result is based on the above-mentioned combination of multibin signal regions. There is good agreement between CHECKMATE (green lines) and ATLAS (black lines). There is also a clear advantage of the multibin fir over the best-SR (here the aggregate SRs were used).

3.12   `cms_2107_13021` (EXO-20-004)

This is a search [47] for new particles in the final states with at least one jet, no leptons, and missing transverse momentum. A main focus of the analysis are invisible particles that can be dark-matter candidates and that are produced with at least one ISR jet. The simplified likelihood fit is performed on 66 bins: 3 sets for different data-taking periods and divided according to missing transverse energy.

For validation, we compare exclusion limits for the vector mediator model in the $m_{\mathrm{mediator}}-m_{\mathrm{DM}}$ plane. The mediator-quark and mediator-DM couplings are fixed to $g_q = 0.25$ and $g_\chi = 1$, respectively. The validation plot, Figure 13, again shows good agreement between the CHECKMATE and CMS limits.

3.13   `cms_2205_09597` (SUS-21-002)

This is a search [99] for charginos and neutralinos that looks for final states with large missing transverse momentum and pairs of hadronically decaying bosons WW, WZ and WH. This search makes use of specific algorithms (taggers) defined to identify W boson/Higgs boson candidates out of the identified large-radius signal jets. Individual bins are defined orthogonally based on the number of b-tagged jets, the number of identified W, Z, and Higgs boson candidates, and missing
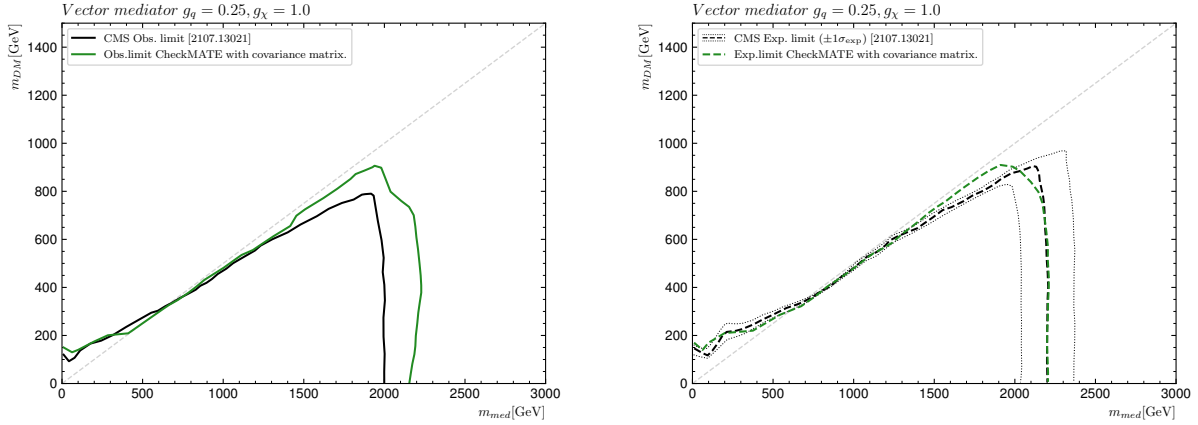
Figure 13: Validation plots for the search `cms_2107_13021` and the vector mediator model in the m_{mediator}–m_{DM} plane. The mediator-quarks and mediator-DM couplings are fixed to $g_q = 0.25$ and $g_\chi = 1$, respectively. The diagonal line indicates $m_{\text{med}} = 2m_{\text{DM}}$.
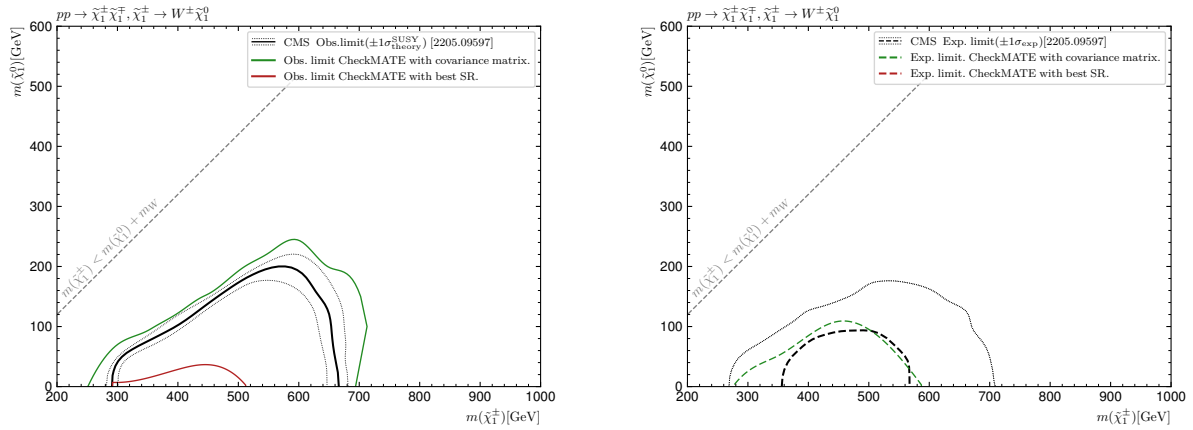

Figure 14: Validation plots for the search `cms_2205_09597` and chargino pair production followed by $\tilde{\chi}_1^\pm \to W^\pm \tilde{\chi}_1^0$.

transverse momentum. The simplified likelihood is built out of 35 signal regions. Additionally, there are 4 aggregate signal regions defined.

Currently, the implementation is validated for the production of charginos with W bosons in the final state. As can be seen in Fig. 14 there is good agreement between CHECKMATE and the CMS exclusion contours. The multibin fit clearly improves sensitivity, the red vs. green line in the left panel of Fig. 14. The remaining signal regions for Z and h channels are implemented, however, due to missing mistag rates for different types of final states, they should be used with caution.

4   Conclusions and outlook

In this paper, we present a version of CHECKMATE, which significantly updates the database of available experimental searches. All included searches are based on the full data set of LHC Run 2. Here, we describe the implementation of 9 ATLAS searches and 4 CMS searches. Another update

comes with the inclusion of statistical models, which in a number of examples significantly improve sensitivity and expand exclusion limits. This implementation includes the full and simplified likelihoods provided by the ATLAS Collaboration and the simplified correlated background models provided by CMS. With a number of options available, it allows users to fully control the numerical evaluation. CHECKMATE can be downloaded from the GitHub repository:

https://github.com/CheckMATE2/checkmate2/ .

More information, including previous versions and expanded validation notes, can be found at:

http://checkmate.hepforge.org/ .

In future releases, we plan to further expand the CHECKMATEcapabilities by allowing a statistical combination of different searches and orthogonal signal regions. In a separate note, we will cover another new development to CHECKMATE, which is an interface to machine learning methods which are becoming increasingly common tools in experimental analyses at the LHC.

Acknowledgements

Appendix

Event yields for atlas_2010_14293 (SUSY-2018-22)

In Tables 6–8 we provide a detailed comparison of the event yields in the signal regions of the atlas_2010_14293 (SUSY-2018-22) search where the MC sample size in CHECKMATE was 100k. A good agreement (within statistical uncertainties) between ATLAS and CHECKMATE is observed except for the 4-jet bins in the squark search region MB-SSd. However, the test sample, the squark pair production with direct decays $\tilde{q} \to q\tilde{\chi}_1^0$, at the lowest parton order does not contribute to these bins and requires 2 additional jets from ISR/FSR. It is therefore very sensitive to details in MC modeling and is difficult to reproduce in recasts. These discrepancies do not affect the fit for this signal process due to the small yields compared to expected backgrounds.

References

[1] **ATLAS** Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," Phys. Lett. B **716** (2012) 1–29 [arXiv:1207.7214].

[2] **CMS** Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," Phys. Lett. B **716** (2012) 30–61 [arXiv:1207.7235].

[3] **LHC New Physics Working Group** Collaboration, "Simplified Models for LHC New Physics Searches," J. Phys. G **39** (2012) 105005 [arXiv:1105.2838].

Bin	CHECKMATE	ATLAS	ATLAS/CHECKMATE
MB-SSd-2-1000-10	0.94 ± 0.033	1 ± 0.15	1.1 ± 0.16
MB-SSd-2-1000-16	0.74 ± 0.029	0.59 ± 0.1	0.8 ± 0.14
MB-SSd-2-1000-22	0.41 ± 0.022	0.27 ± 0.071	0.66 ± 0.17
MB-SSd-2-1600-10	3.8 ± 0.066	3.7 ± 0.26	0.96 ± 0.069
MB-SSd-2-1600-16	2.2 ± 0.05	2 ± 0.18	0.91 ± 0.085
MB-SSd-2-1600-22	4 ± 0.067	3.5 ± 0.26	0.88 ± 0.064
MB-SSd-2-2200-16	1.8 ± 0.045	2 ± 0.2	1.1 ± 0.11
MB-SSd-2-2200-22	7.8 ± 0.094	8.4 ± 0.39	1.1 ± 0.051
MB-SSd-2-2800-16	1.4 ± 0.04	1.3 ± 0.16	0.95 ± 0.11
MB-SSd-2-2800-22	5.9 ± 0.082	6 ± 0.33	1 ± 0.055
MB-SSd-2-3400-22	1.2 ± 0.038	0.93 ± 0.14	0.77 ± 0.12
MB-SSd-2-3400-28	2.6 ± 0.055	2.2 ± 0.19	0.84 ± 0.075
MB-SSd-2-4000-22	0.47 ± 0.024	0.22 ± 0.063	0.47 ± 0.13
MB-SSd-2-4000-28	0.89 ± 0.033	0.67 ± 0.11	0.76 ± 0.12
MB-SSd-4-1000-10	0.13 ± 0.012	0.63 ± 0.1	4.7 ± 0.78
MB-SSd-4-1000-16	0.067 ± 0.0088	0.39 ± 0.08	5.8 ± 1.2
MB-SSd-4-1000-22	0.015 ± 0.0041	0.17 ± 0.064	11 ± 4.3
MB-SSd-4-1600-10	1.7 ± 0.045	2.4 ± 0.21	1.4 ± 0.13
MB-SSd-4-1600-16	0.52 ± 0.025	1.3 ± 0.16	2.6 ± 0.3
MB-SSd-4-1600-22	0.39 ± 0.021	1.7 ± 0.17	4.3 ± 0.44
MB-SSd-4-2200-16	0.77 ± 0.03	1.4 ± 0.16	1.8 ± 0.21
MB-SSd-4-2200-22	1.6 ± 0.042	3.9 ± 0.26	2.5 ± 0.17
MB-SSd-4-2800-16	1.1 ± 0.036	0.99 ± 0.14	0.92 ± 0.13
MB-SSd-4-2800-22	1.9 ± 0.047	4.6 ± 0.3	2.4 ± 0.16

Table 6: Event yields for the squark pair production, $m_{\tilde{q}} = 1700$ GeV, $m_{\tilde{\chi}_1^0} = 500$ GeV with MC statistical uncertainties, $s \pm \Delta s$, normalized to the nominal cross section (8-fold degeneracy) at the approximate NNLO+NNLL accuracy, $\sigma = 0.74$ fb.

Bin	CHECKMATE	ATLAS	ATLAS/CHECKMATE
MB-GGd-4-1000-10	4 ± 0.13	2.9 ± 0.49	0.74 ± 0.12
MB-GGd-4-1000-16	1.4 ± 0.076	0.64 ± 0.21	0.46 ± 0.15
MB-GGd-4-1000-22	0.18 ± 0.027	0.097 ± 0.097	0.55 ± 0.55
MB-GGd-4-1600-10	17 ± 0.26	17 ± 1.3	1 ± 0.076
MB-GGd-4-1600-16	13 ± 0.23	14 ± 1.1	1.1 ± 0.086
MB-GGd-4-1600-22	5.1 ± 0.15	4.4 ± 0.63	0.85 ± 0.12
MB-GGd-4-2200-10	11 ± 0.21	13 ± 1.1	1.2 ± 0.1
MB-GGd-4-2200-16	14 ± 0.24	17 ± 1.2	1.2 ± 0.088
MB-GGd-4-2200-22	12 ± 0.23	14 ± 1.1	1.1 ± 0.086
MB-GGd-4-2800-10	2.6 ± 0.11	2.4 ± 0.42	0.91 ± 0.16
MB-GGd-4-2800-16	3.6 ± 0.12	3.6 ± 0.54	1 ± 0.15
MB-GGd-4-2800-22	4.3 ± 0.13	5.3 ± 0.65	1.2 ± 0.15
MB-GGd-4-3400-10	0.49 ± 0.046	0.56 ± 0.21	1.1 ± 0.42
MB-GGd-4-3400-16	0.47 ± 0.046	0.3 ± 0.14	0.64 ± 0.3
MB-GGd-4-3400-22	0.71 ± 0.056	1.5 ± 0.36	2.1 ± 0.51
MB-GGd-4-4000-10	0.11 ± 0.023	0.24 ± 0.14	2.2 ± 1.3
MB-GGd-4-4000-16	0.11 ± 0.022	0.11 ± 0.1	0.98 ± 0.95
MB-GGd-4-4000-22	0.21 ± 0.031	0.1 ± 0.1	0.48 ± 0.48

Table 7: Event yields for the gluino pair production, $m_{\tilde{g}} = 1800$ GeV, $m_{\tilde{\chi}_1^0} = 800$ GeV with MC statistical uncertainties, $s \pm \Delta s$, normalized to the nominal cross section at the approximate NNLO+NNLL accuracy, $\sigma = 2.93$ fb.

Bin	CHECKMATE	ATLAS	ATLAS/CHECKMATE
MB-C-2-1600-16	2.2 ± 0.05	1.9 ± 0.18	0.89 ± 0.084
MB-C-2-1600-22	3.9 ± 0.066	3.3 ± 0.24	0.85 ± 0.062
MB-C-2-2200-16	2.1 ± 0.049	2.3 ± 0.22	1.1 ± 0.1
MB-C-2-2200-22	9.1 ± 0.1	9.8 ± 0.42	1.1 ± 0.047
MB-C-2-2800-16	1.7 ± 0.044	1.5 ± 0.17	0.91 ± 0.098
MB-C-2-2800-22	9.5 ± 0.1	9.5 ± 0.41	1 ± 0.044
MB-C-4-1600-16	0.76 ± 0.029	0.71 ± 0.12	0.93 ± 0.16
MB-C-4-1600-22	0.69 ± 0.028	0.63 ± 0.11	0.91 ± 0.16
MB-C-4-2200-16	1 ± 0.034	1.1 ± 0.14	1.1 ± 0.14
MB-C-4-2200-22	2.9 ± 0.058	2.5 ± 0.21	0.86 ± 0.07
MB-C-4-2800-16	1 ± 0.035	0.76 ± 0.13	0.73 ± 0.12
MB-C-4-2800-22	4.7 ± 0.074	4 ± 0.28	0.84 ± 0.06
MB-C-5-1600-16	0.35 ± 0.02	0.28 ± 0.07	0.8 ± 0.2
MB-C-5-1600-22	0.19 ± 0.015	0.24 ± 0.068	1.2 ± 0.35
MB-C-5-2200-16	1 ± 0.034	0.61 ± 0.11	0.61 ± 0.11
MB-C-5-2200-22	1.6 ± 0.043	1.2 ± 0.15	0.77 ± 0.094
MB-C-5-2800-16	1.5 ± 0.043	0.75 ± 0.13	0.5 ± 0.085
MB-C-5-2800-22	4.6 ± 0.074	3.3 ± 0.25	0.71 ± 0.055

Table 8: Event yields for the squark pair production, $m_{\tilde{q}} = 1700$ GeV, $m_{\tilde{\chi}_1^0} = 500$ GeV with MC statistical uncertainties, $s \pm \Delta s$, normalized to the nominal cross section (8-fold degeneracy) at the approximate NNLO+NNLL accuracy, $\sigma = 0.74$ fb.

[4] S. Kraml, S. Kulkarni, U. Laa, A. Lessa, *et al.*, "SModelS: a tool for interpreting simplified-model results from the LHC and its application to supersymmetry," Eur. Phys. J. C **74** (2014) 2868 [`arXiv:1312.4175`].

[5] G. Alguero, J. Heisig, C. K. Khosa, S. Kraml, *et al.*, "Constraining new physics with SModelS version 2," JHEP **08** (2022) 068 [`arXiv:2112.00769`].

[6] E. Conte and B. Fuks, "Confronting new physics theories to LHC data with MADANALYSIS 5," Int. J. Mod. Phys. A **33** (2018) 1830027 [`arXiv:1808.00480`].

[7] C. Bierlich *et al.*, "Robust Independent Validation of Experiment and Theory: Rivet version 3," SciPost Phys. **8** (2020) 026 [`arXiv:1912.05451`].

[8] **GAMBIT** Collaboration, "ColliderBit: a GAMBIT module for the calculation of high-energy collider observables and likelihoods," Eur. Phys. J. C **77** (2017) 795 [`arXiv:1705.07919`].

[9] G. Unel, S. Sekmen, A. M. Toon, B. Gokturk, *et al.*, "CutLang v2: Advances in a Runtime-Interpreted Analysis Description Language for HEP Data," Front. Big Data **4** (2021) 659986 [`arXiv:2101.09031`].

[10] **LHC Reinterpretation Forum** Collaboration, "Reinterpretation of LHC Results for New Physics: Status and Recommendations after Run 2," SciPost Phys. **9** (2020) 022 [`arXiv:2003.07868`].

[11] K. Cranmer *et al.*, "Publishing statistical models: Getting the most out of particle physics experiments," SciPost Phys. **12** (2022) 037 [`arXiv:2109.04981`].

[12] K. Albertsson *et al.*, "Machine Learning in High Energy Physics Community White Paper," J. Phys. Conf. Ser. **1085** (2018) 022008 [`arXiv:1807.02876`].

[13] J. Y. Araz *et al.*, "Les Houches guide to reusable ML models in LHC analyses," SciPost Phys. Comm. Rep. (2024) 3 [`arXiv:2312.14575`].

[14] M. Drees, H. Dreiner, D. Schmeier, J. Tattersall, and J. S. Kim, "CheckMATE: Confronting your Favourite New Physics Model with LHC Data," Comput. Phys. Commun. **187** (2015) 227–265 [`arXiv:1312.2591`].

[15] J. S. Kim, D. Schmeier, J. Tattersall, and K. Rolbiecki, "A framework to create customised LHC analyses within CheckMATE," Comput. Phys. Commun. **196** (2015) 535–562 [`arXiv:1503.01123`].

[16] D. Dercks, N. Desai, J. S. Kim, K. Rolbiecki, *et al.*, "CheckMATE 2: From the model to the limit," Comput. Phys. Commun. **221** (2017) 383–418 [`arXiv:1611.09856`].

[17] **CheckMATE** Collaboration, "Constraining electroweak and strongly charged long-lived particles with CheckMATE," Eur. Phys. J. C **81** (2021) 968 [`arXiv:2104.04542`].

[18] G. Alguero, S. Kraml, and W. Waltenberger, "A SModelS interface for pyhf likelihoods," Comput. Phys. Commun. **264** (2021) 107909 [`arXiv:2009.01809`].

[19] G. Alguero, J. Y. Araz, B. Fuks, and S. Kraml, "Signal region combination with full and simplified likelihoods in MadAnalysis 5," SciPost Phys. **14** (2023) 009 [`arXiv:2206.14870`].

[20] **CMS** Collaboration, "Simplified likelihood for the re-interpretation of public CMS results," CMS–NOTE–2017–001, CERN, Geneva, 2017.

[21] **ATLAS** Collaboration, "Implementation of simplified likelihoods in HistFactory for searches for supersymmetry," ATL–PHYS–PUB–2021–038, CERN, Geneva, 2021.

[22] **ATLAS** Collaboration, "Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods," ATL–PHYS–PUB–2019–029, CERN, Geneva, 2019.

[23] J. Y. Araz, "Spey: smooth inference for reinterpretation studies," SciPost Phys. **16** (2024) 032 [`arXiv:2307.06996`].

[24] T. Speer *et al.*, eds., "HepData reloaded: Reinventing the HEP data archive," PoS **ACAT2010** (2010) 067 [`arXiv:1006.0517`].

[25] R. Mount and C. Tull, eds., "HEPData: a repository for high energy physics data," J. Phys. Conf. Ser. **898** (2017) 102006 [`arXiv:1704.05473`].

[26] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics," Eur. Phys. J. C **71** (2011) 1554 [`arXiv:1007.1727`]. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[27] L. Heinrich, M. Feickert, G. Stark, and K. Cranmer, "pyhf: pure-Python implementation of HistFactory statistical models," Journal of Open Source Software **6** (2021) 2823.

[28] L. Heinrich, M. Feickert, and G. Stark, "pyhf: v0.7.2." `doi:10.5281/zenodo.1169739`. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.2.

[29] **ROOT** Collaboration, "HistFactory: A tool for creating statistical models for use with RooFit and RooStats," CERN–OPEN–2012–016, CERN, Geneva, 2012.

[30] M. Baak, G. J. Besjes, D. Côte, A. Koutsman, *et al.*, "HistFitter software framework for statistical data analysis," Eur. Phys. J. C **75** (2015) 153 [`arXiv:1410.1280`].

[31] J. Araz, "spey-pyhf: v0.2.0." `doi:10.5281/zenodo.14945825`.

[32] `https://speysidehep.github.io/spey/quick_start.html`.

[33] E. Schanet, "Simplify." `https://https://pypi.org/project/simplify-hep/`. https://https://pypi.org/project/simplify-hep/.

[34] J. Lahiri, T. Robens, and K. Rolbiecki, "Constraining the Inert Doublet Model at the LHC." `arXiv:2511.23133`.

[35] **ATLAS** Collaboration, "Search for invisible Higgs-boson decays in events with vector-boson fusion signatures using 139 fb^{-1} of proton-proton data recorded by the ATLAS experiment," JHEP **08** (2022) 104 [`arXiv:2202.07953`].

[36] **ATLAS** Collaboration, "Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, *b*-jets and missing transverse momentum," JHEP **12** (2019) 060 [`arXiv:1908.03122`].

[37] **ATLAS** Collaboration, "Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector," Eur. Phys. J. C **80** (2020) 123 [arXiv:1908.08215].

[38] **ATLAS** Collaboration, "Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector," Phys. Rev. D **101** (2020) 032009 [arXiv:1911.06660].

[39] **ATLAS** Collaboration, "Searches for electroweak production of supersymmetric particles with compressed mass spectra in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector," Phys. Rev. D **101** (2020) 052005 [arXiv:1911.12606].

[40] **ATLAS** Collaboration, "Search for a scalar partner of the top quark in the all-hadronic $t\bar{t}$ plus missing transverse momentum final state at $\sqrt{s} = 13$ TeV with the ATLAS detector," Eur. Phys. J. C **80** (2020) 737 [arXiv:2004.14060].

[41] **ATLAS** Collaboration, "Search for top squarks in events with a Higgs or Z boson using 139 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector," Eur. Phys. J. C **80** (2020) 1080 [arXiv:2006.05880].

[42] **ATLAS** Collaboration, "Search for squarks and gluinos in final states with jets and missing transverse momentum using 139 fb^{-1} of $\sqrt{s} =13$ TeV pp collision data with the ATLAS detector," JHEP **02** (2021) 143 [arXiv:2010.14293].

[43] **ATLAS** Collaboration, "Search for squarks and gluinos in final states with one isolated lepton, jets, and missing transverse momentum at $\sqrt{s} = 13$ with the ATLAS detector," Eur. Phys. J. C **81** (2021) 600 [arXiv:2101.01629]. [Erratum: Eur.Phys.J.C 81, 956 (2021)].

[44] **ATLAS** Collaboration, "Search for associated production of a Z boson with an invisibly decaying Higgs boson or dark matter candidates at $\sqrt{s} = 13$ TeV with the ATLAS detector," Phys. Lett. B **829** (2022) 137066 [arXiv:2111.08372].

[45] **CMS** Collaboration, "Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum," JHEP **10** (2019) 244 [arXiv:1908.04722].

[46] **CMS** Collaboration, "Searches for physics beyond the standard model with the M_{T2} variable in hadronic final states with and without disappearing tracks in proton-proton collisions at $\sqrt{s} = 13$ TeV," Eur. Phys. J. C **80** (2020) 3 [arXiv:1909.03460].

[47] **CMS** Collaboration, "Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV," JHEP **11** (2021) 153 [arXiv:2107.13021].

[48] "NumPy." https://numpy.org/.

[49] "TensorFlow." https://www.tensorflow.org/.

[50] "PyTorch." https://pytorch.org/.

[51] "JAX." https://docs.jax.dev/en/latest/quickstart.html.

[52] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, *et al.*, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," JHEP **07** (2014) 079 [arXiv:1405.0301].

[53] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," Eur. Phys. J. C **53** (2008) 473–500 [arXiv:0706.2569].

[54] J. Alwall, S. de Visscher, and F. Maltoni, "QCD radiation in the production of heavy colored particles at the LHC," JHEP **02** (2009) 017 [arXiv:0810.5350].

[55] C. Duhr and B. Fuks, "A superspace module for the FeynRules package," Comput. Phys. Commun. **182** (2011) 2404–2426 [arXiv:1102.4191].

[56] L. Darmé *et al.*, "UFO 2.0: the 'Universal Feynman Output' format," Eur. Phys. J. C **83** (2023) 631 [arXiv:2304.09883].

[57] R. D. Ball *et al.*, "Parton distributions with LHC data," Nucl. Phys. B **867** (2013) 244–289 [arXiv:1207.1303].

[58] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, *et al.*, "LHAPDF6: parton density access in the LHC precision era," Eur. Phys. J. C **75** (2015) 132 [arXiv:1412.7420].

[59] **NNPDF** Collaboration, "Parton distributions for the LHC Run II," JHEP **04** (2015) 040 [arXiv:1410.8849].

[60] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, *et al.*, "An introduction to PYTHIA 8.2," Comput. Phys. Commun. **191** (2015) 159–177 [arXiv:1410.3012].

[61] C. Bierlich *et al.*, "A comprehensive guide to the physics and usage of PYTHIA 8.3," SciPost Phys. Codeb. **2022** (2022) 8 [arXiv:2203.11601].

[62] L. Lonnblad and S. Prestel, "Matching Tree-Level Matrix Elements with Interleaved Showers," JHEP **03** (2012) 019 [arXiv:1109.4829].

[63] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," JHEP **01** (2007) 013 [hep-ph/0611129].

[64] **DELPHES 3** Collaboration, "DELPHES 3, A modular framework for fast simulation of a generic collider experiment," JHEP **02** (2014) 057 [arXiv:1307.6346].

[65] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," Eur. Phys. J. C **72** (2012) 1896 [arXiv:1111.6097].

[66] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," JHEP **04** (2008) 063 [arXiv:0802.1189].

[67] W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, and E. Laenen, "NNLL-fast: predictions for coloured supersymmetric particle production at the LHC with threshold and Coulomb resummation," JHEP **12** (2016) 133 [arXiv:1607.07741].

[68] W. Beenakker, R. Hopker, M. Spira, and P. M. Zerwas, "Squark and gluino production at hadron colliders," Nucl. Phys. B **492** (1997) 51–103 [hep-ph/9610490].

[69] A. Kulesza and L. Motyka, "Threshold resummation for squark-antisquark and gluino-pair production at the LHC," Phys. Rev. Lett. **102** (2009) 111802 [`arXiv:0807.2405`].

[70] A. Kulesza and L. Motyka, "Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC," Phys. Rev. D **80** (2009) 095004 [`arXiv:0905.4749`].

[71] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, *et al.*, "Soft-gluon resummation for squark and gluino hadroproduction," JHEP **12** (2009) 041 [`arXiv:0909.4418`].

[72] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, *et al.*, "NNLL resummation for squark-antisquark pair production at the LHC," JHEP **01** (2012) 076 [`arXiv:1110.2446`].

[73] W. Beenakker, T. Janssen, S. Lepoeter, M. Krämer, *et al.*, "Towards NNLL resummation: hard matching coefficients for squark and gluino hadroproduction," JHEP **10** (2013) 120 [`arXiv:1304.6354`].

[74] W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, *et al.*, "NNLL resummation for squark and gluino production at the LHC," JHEP **12** (2014) 023 [`arXiv:1404.3134`].

[75] W. Beenakker, M. Kramer, T. Plehn, M. Spira, and P. M. Zerwas, "Stop production at hadron colliders," Nucl. Phys. B **515** (1998) 3–14 [`hep-ph/9710451`].

[76] W. Beenakker, S. Brensing, M. Kramer, A. Kulesza, *et al.*, "Supersymmetric top and bottom squark production at hadron colliders," JHEP **08** (2010) 098 [`arXiv:1006.4771`].

[77] W. Beenakker, C. Borschensky, R. Heger, M. Krämer, *et al.*, "NNLL resummation for stop pair-production at the LHC," JHEP **05** (2016) 153 [`arXiv:1601.02954`].

[78] J. Butterworth *et al.*, "PDF4LHC recommendations for LHC Run II," J. Phys. G **43** (2016) 023001 [`arXiv:1510.03865`].

[79] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering, "Precision predictions for electroweak superpartner production at hadron colliders with Resummino," Eur. Phys. J. C **73** (2013) 2480 [`arXiv:1304.0790`].

[80] J. Fiaschi and M. Klasen, "Slepton pair production at the LHC in NLO+NLL with resummation-improved parton densities," JHEP **03** (2018) 094 [`arXiv:1801.10357`].

[81] G. Bozzi, B. Fuks, and M. Klasen, "Joint resummation for slepton pair production at hadron colliders," Nucl. Phys. B **794** (2008) 46–60 [`arXiv:0709.3057`].

[82] G. Bozzi, B. Fuks, and M. Klasen, "Threshold Resummation for Slepton-Pair Production at Hadron Colliders," Nucl. Phys. B **777** (2007) 157–181 [`hep-ph/0701202`].

[83] J. Fiaschi and M. Klasen, "Neutralino-chargino pair production at NLO+NLL with resummation-improved parton density functions for LHC Run II," Phys. Rev. D **98** (2018) 055014 [`arXiv:1805.11322`].

[84] B. Fuks, M. Klasen, D. R. Lamprea, and M. Rothering, "Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV," JHEP **10** (2012) 081 [`arXiv:1207.2159`].

[85] J. Debove, B. Fuks, and M. Klasen, "Joint Resummation for Gaugino Pair Production at Hadron Colliders," Nucl. Phys. B **849** (2011) 64–79 [`arXiv:1102.4422`].

[86] M. Backović, M. Krämer, F. Maltoni, A. Martini, *et al.*, "Higher-order QCD predictions for dark matter production at the LHC in simplified models with s-channel mediators," Eur. Phys. J. C **75** (2015) 482 [`arXiv:1508.05327`].

[87] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO," JHEP **12** (2012) 061 [`arXiv:1209.6215`].

[88] M. Bauer, U. Haisch, and F. Kahlhoefer, "Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators," JHEP **05** (2017) 138 [`arXiv:1701.07427`].

[89] "SUSY-2018-31 validation notes." `http://checkmate.hepforge.org/validationNotes/validation_1908_08215_v2.pdf`.

[90] S. Y. Choi, B. C. Chung, J. Kalinowski, Y. G. Kim, and K. Rolbiecki, "Analysis of the neutralino system in three-body leptonic decays of neutralinos," Eur. Phys. J. C **46** (2006) 511–520 [`hep-ph/0504122`].

[91] D. Agin, B. Fuks, M. D. Goodsell, and T. Murphy, "Seeking a coherent explanation of LHC excesses for compressed spectra," Eur. Phys. J. C **84** (2024) 1218 [`arXiv:2404.12423`].

[92] M. M. Altakach, S. Kraml, A. Lessa, S. Narasimha, *et al.*, "SModelS v3: going beyond $\mathcal{Z}_2$ topologies," JHEP **11** (2024) 074 [`arXiv:2409.12942`].

[93] J. Y. Araz, B. Fuks, M. D. Goodsell, and T. Murphy, "Deciphering compressed electroweakino excesses with MadAnalysis 5." `arXiv:2507.08927`.

[94] "SUSY-2018-16 validation notes." `https://checkmate.hepforge.org/validationNotes/validation_1911_12606.pdf`.

[95] P. Jackson, C. Rogan, and M. Santoni, "Sparticles in motion: Analyzing compressed SUSY scenarios with a new method of event reconstruction," Phys. Rev. D **95** (2017) 035031 [`arXiv:1607.08307`].

[96] "SUSY-2018-12 validation notes." `https://checkmate.hepforge.org/validationNotes/validation_2004_14060.pdf`.

[97] "SUSY-2018-22 validation notes." `https://checkmate.hepforge.org/validationNotes/validation_atlas_2010_14293.pdf`.

[98] M. Masciovecchio. Communication with CMS SUSY convenors.

[99] **CMS** Collaboration, "Search for electroweak production of charginos and neutralinos at $\sqrt{s} =$ 13 TeV in final states containing hadronic decays of WW, WZ, or WH and missing transverse momentum," Phys. Lett. B **842** (2023) 137460 [`arXiv:2205.09597`].