Modern Tools for

Reusable Publications and Data Products

Matthew Feickert

(University of Illinois at Urbana-Champaign)

matthew.feickert@cern.ch

LPSC Grenoble Colloquium

November 26th, 2020

Collaborators

Lukas Heinrich

CERN

Giordon Stark

UCSC SCIPP

Kyle Cranmer

NYU

Mark Neubauer

Illinois

Sabine Kraml

LPSC, Grenoble

Wolfgang Waltenberger

HEPHY ÖAW

Gaël Alguero

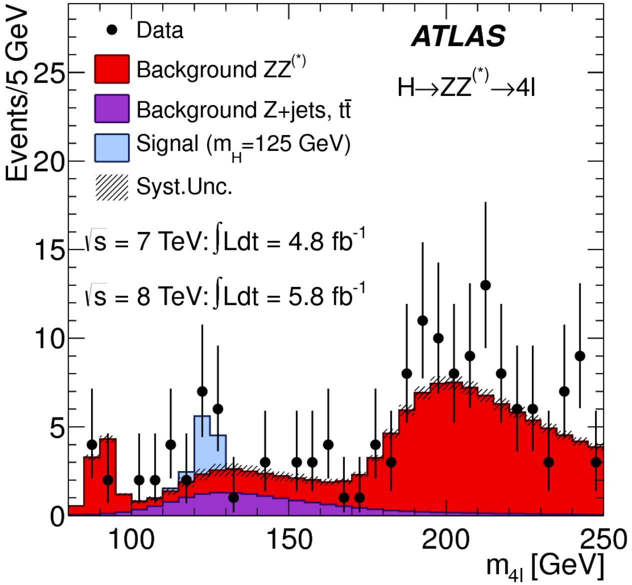
LPSC, Grenoble

Goals for today's discussion

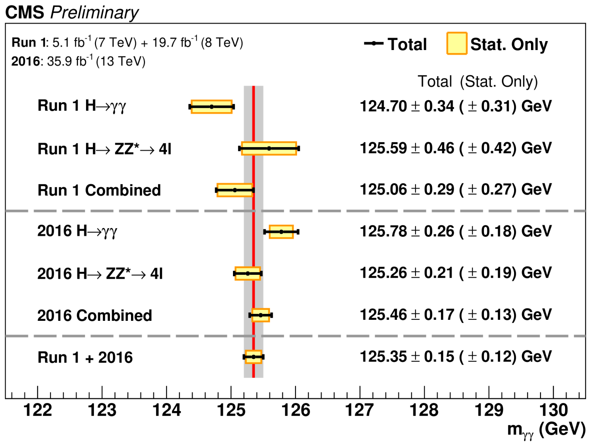
Open and reusable data products are valuable to everyone, especially you!

Show live (toy) example of how easy it can be to get started

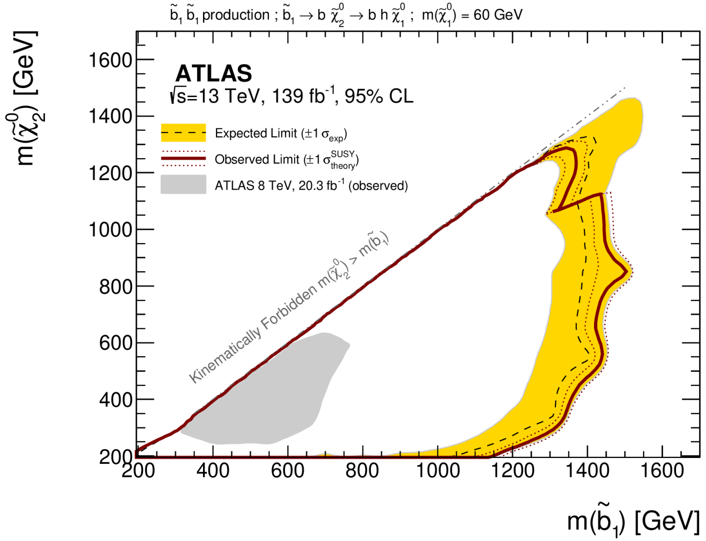
Goals of physics analysis at the LHC



Search for new physics



Make precision measurements

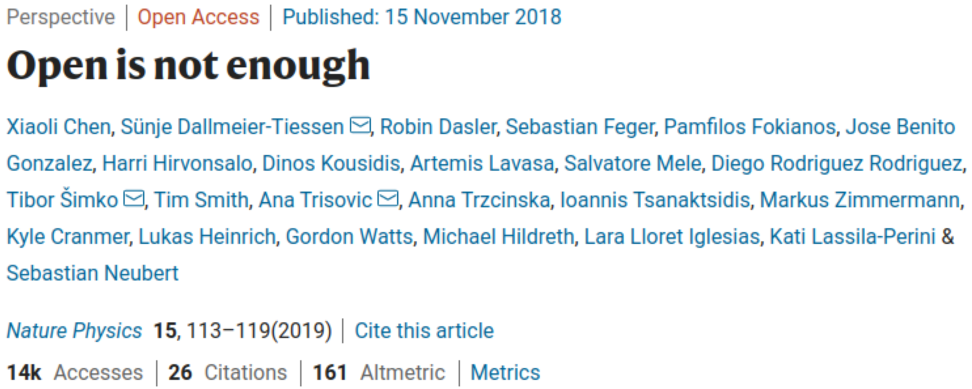


Provide constraints on models through setting best limits

- The results of these analyses are **data products** (plots, tables, likelihoods)
- For these data products to be useful need to be **shared** and **reusable** with experiments and theory
- Requires the data products to be archived, easily findable, and machine readable (and human understandable)

Easy, just make them open...?

- Making data products open is a great first start, but **not enough**!
- If just the code and data were made public, would be **largely useless** to anyone but the authors
  - Millions of lines in codebases, hundreds of TB of data
- Experiments are complex and code and workflows are complicated
- For data products to be useful they need to be able to be **understood** and used by the scientific community
  - Relationship of the data and the analysis needs to be made clear/available as well



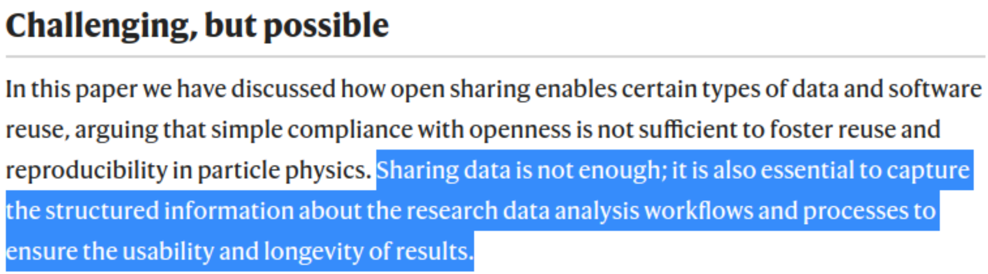
Perspective | Open Access | Published: 15 November 2018

Open is not enough

Xiaoli Chen, Sünje Dallmeier-Tiessen, Robin Dasler, Sebastian Feger, Pamfilos Fokianos, Jose Benito Gonzalez, Harri Hirvonsalo, Dinos Kousidis, Artemis Lavasa, Salvatore Mele, Diego Rodriguez Rodriguez, Tibor Šimko, Tim Smith, Ana Trisovic, Anna Trzcinska, Ioannis Tsanaktsidis, Markus Zimmermann, Kyle Cranmer, Lukas Heinrich, Gordon Watts, Michael Hildreth, Lara Lloret Iglesias, Kati Lassila-Perini & Sebastian Neubert

Nature Physics **15**, 113–119(2019) | Cite this article

14k Accesses | **26** Citations | **161** Altmetric | Metrics



Challenging, but possible

In this paper we have discussed how open sharing enables certain types of data and software reuse, arguing that simple compliance with openness is not sufficient to foster reuse and reproducibility in particle physics. Sharing data is not enough; it is also essential to capture the structured information about the research data analysis workflows and processes to ensure the usability and longevity of results.

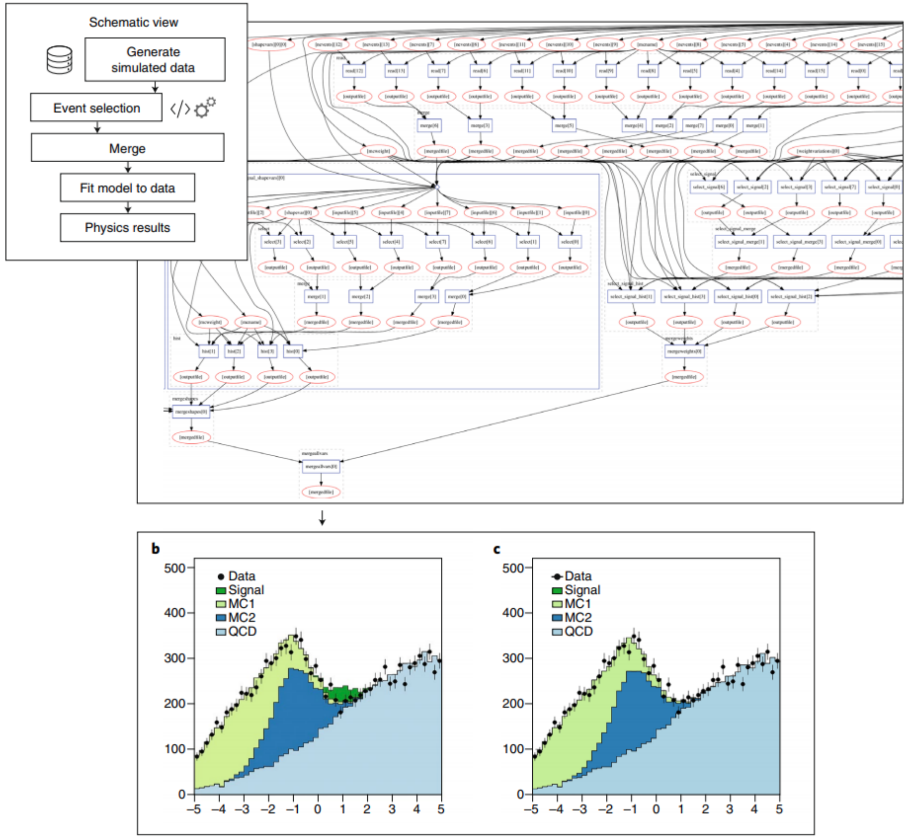
...other extreme: Preserving analyses in full

Ideally preserve the *entire* analysis workflow

- RECAST, REANA: Reproducible research data analysis platform
  - Infrastructure for lossless analysis preservation

  1. Capture software
  2. Capture commands
  3. Capture workflow

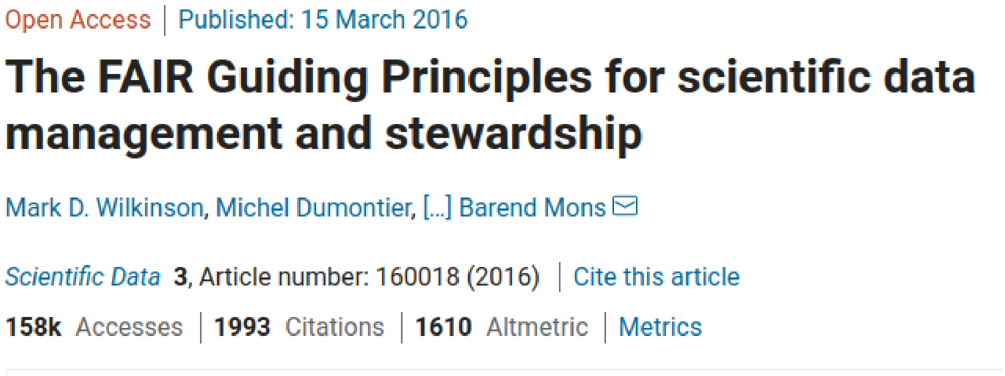
- This is *great* for the LHC experiments (pioneered in ATLAS), but requires huge amounts of resources
  - Hard to scale out for everyone in science
  - Currently still an internal service for CERN
- While the LHC experiments are making this a standard what can be done now?



REANA computational workflow for a beyond the Standard Model (BSM) full analysis

Making data products FAIR

- For data products to be useful they should follow **FAIR** principles
- **F**indable: Metadata and data should be easy to find for both humans and computers.
  - Unique persistent identifier (digital object identifier (DOI)) and rich metadata
- **A**ccessible: Retrievable by identifier using a standardised communications protocol
  - HTTPS, public APIs
- **I**nteroperable: Interoperate with applications or workflows for analysis, storage, and processing
  - Schemas and serialization
  - Formal, shared, broadly applicable
- **R**euseable: Well-described so can be replicated and/or combined in different settings
  - Annotated metadata (Codemeta JSON-LD)



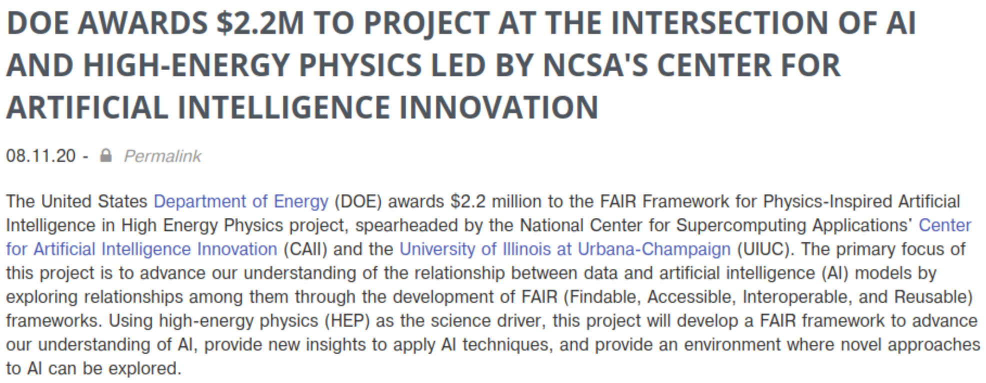
Open Access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons ✉

Scientific Data **3**, Article number: 160018 (2016) | Cite this article

158k Accesses | **1993** Citations | **1610** Altmetric | Metrics



DOE AWARDS $2.2M TO PROJECT AT THE INTERSECTION OF AI AND HIGH-ENERGY PHYSICS LED BY NCSA'S CENTER FOR ARTIFICIAL INTELLIGENCE INNOVATION

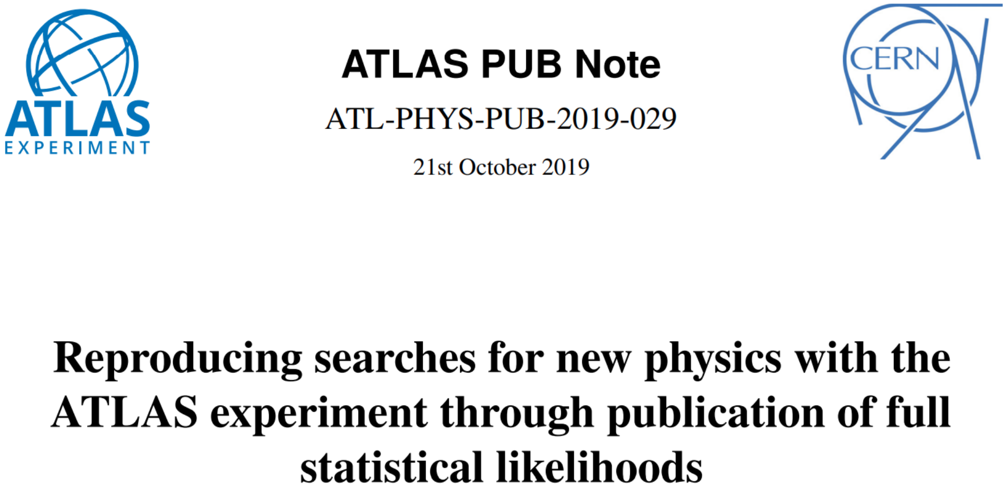
08.11.20 - *Permalink*

The United States Department of Energy (DOE) awards $2.2 million to the FAIR Framework for Physics-Inspired Artificial Intelligence in High Energy Physics project, spearheaded by the National Center for Supercomputing Applications' Center for Artificial Intelligence Innovation (CAII) and the University of Illinois at Urbana-Champaign (UIUC). The primary focus of this project is to advance our understanding of the relationship between data and artificial intelligence (AI) models by exploring relationships among them through the development of FAIR (Findable, Accessible, Interoperable, and Reusable) frameworks. Using high-energy physics (HEP) as the science driver, this project will develop a FAIR framework to advance our understanding of AI, provide new insights to apply AI techniques, and provide an environment where novel approaches to AI can be explored.

Large interest in FAIR data across all levels

Case study:

publishing LHC experiment full likelihoods



ATLAS PUB Note

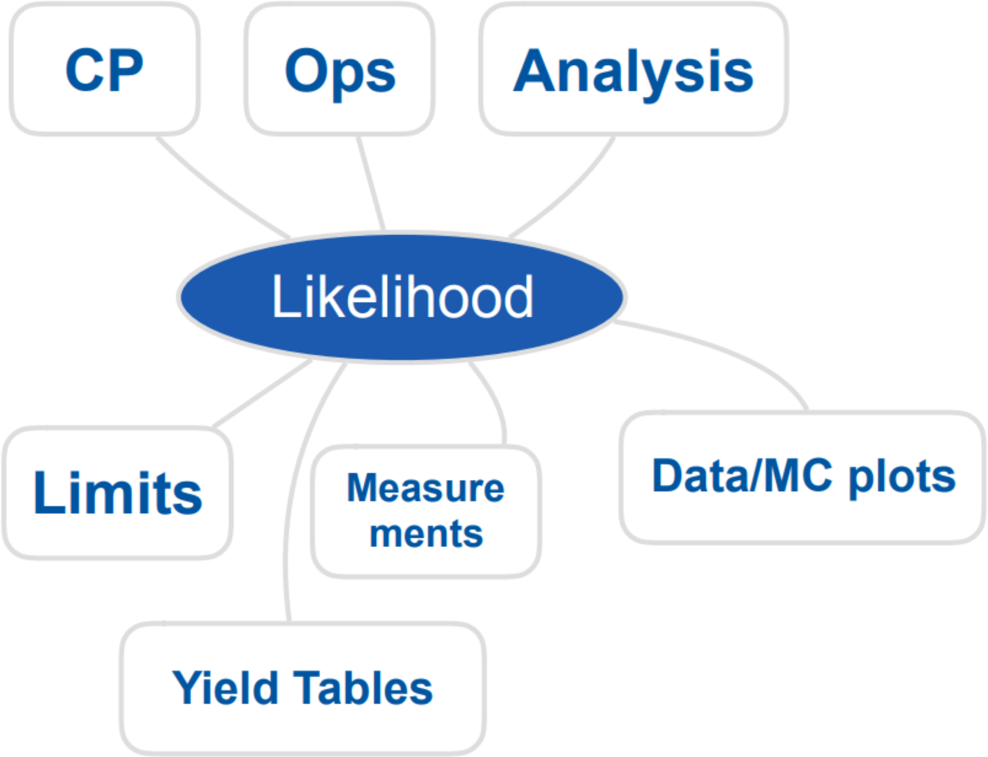
ATL-PHYS-PUB-2019-029

21st October 2019

Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods

Why is the likelihood important?

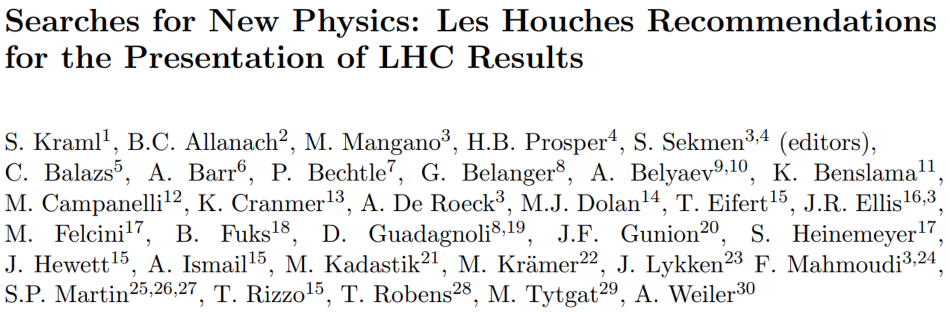
- High information-density summary of analysis
- Almost everything we do in the analysis ultimately affects the likelihood and is encapsulated in it
    - Trigger
    - Detector
    - Combined Performance / Physics Object Groups
    - Systematic Uncertainties
    - Event Selection
- Unique representation of the analysis to reuse and preserve



Importance of public likelihoods

- The statistical model of an experimental analysis provides the **complete mathematical description of that analysis**
  - $p\left(x|\vec{\alpha}\right)$ relating the observed quantities x to the parameters $\vec{\alpha}$
- Given the likelihood, can employ standard statistical approaches for **extracting information** from it
- **Essential information** for any detailed interpretation of experimental results
  - Determining the *compatibility of theoretical predictions with the observations*!

- Requested in Les Houches Recommendations (2012)
- **3b:** When feasible, provide a mathematical description of the final likelihood function in which experimental data and parameters are clearly distinguished, either in the publication or the auxiliary information. Limits of validity should always be clearly specified.
- **3c:** Additionally provide a digitized implementation of the likelihood that is consistent with the mathematical description.

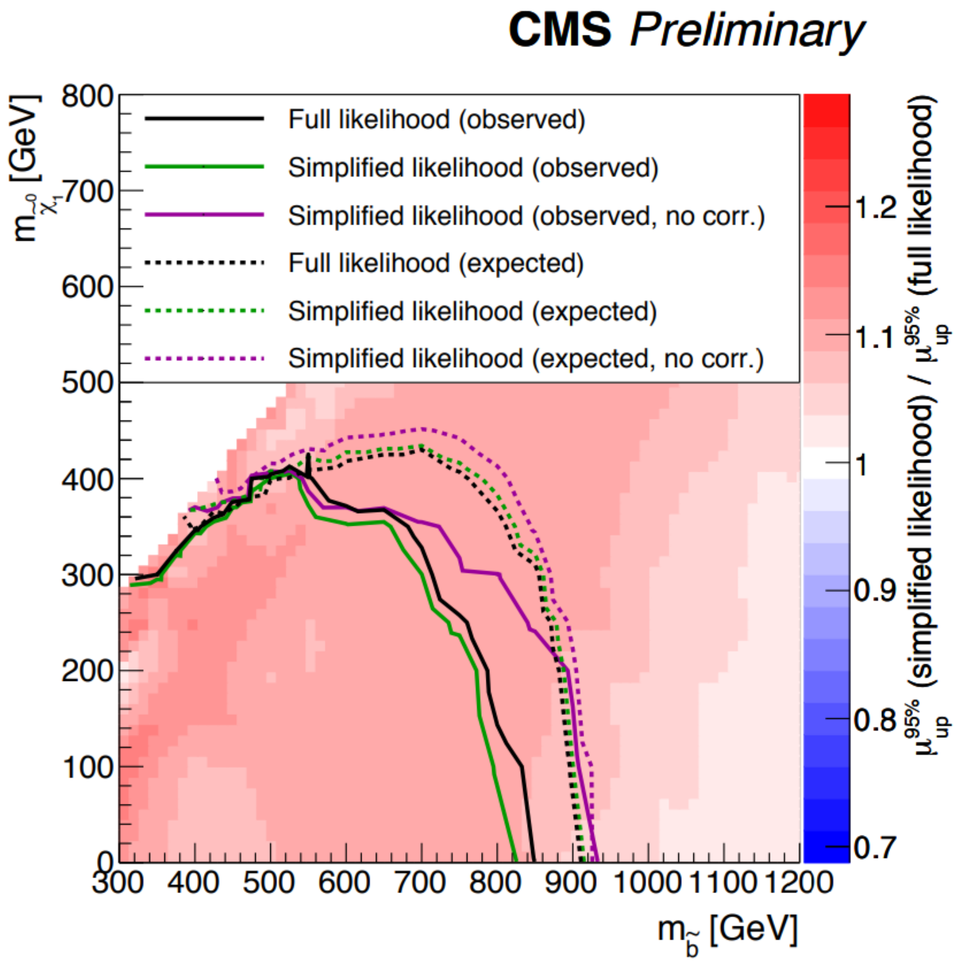


Searches for New Physics: Les Houches Recommendations for the Presentation of LHC Results

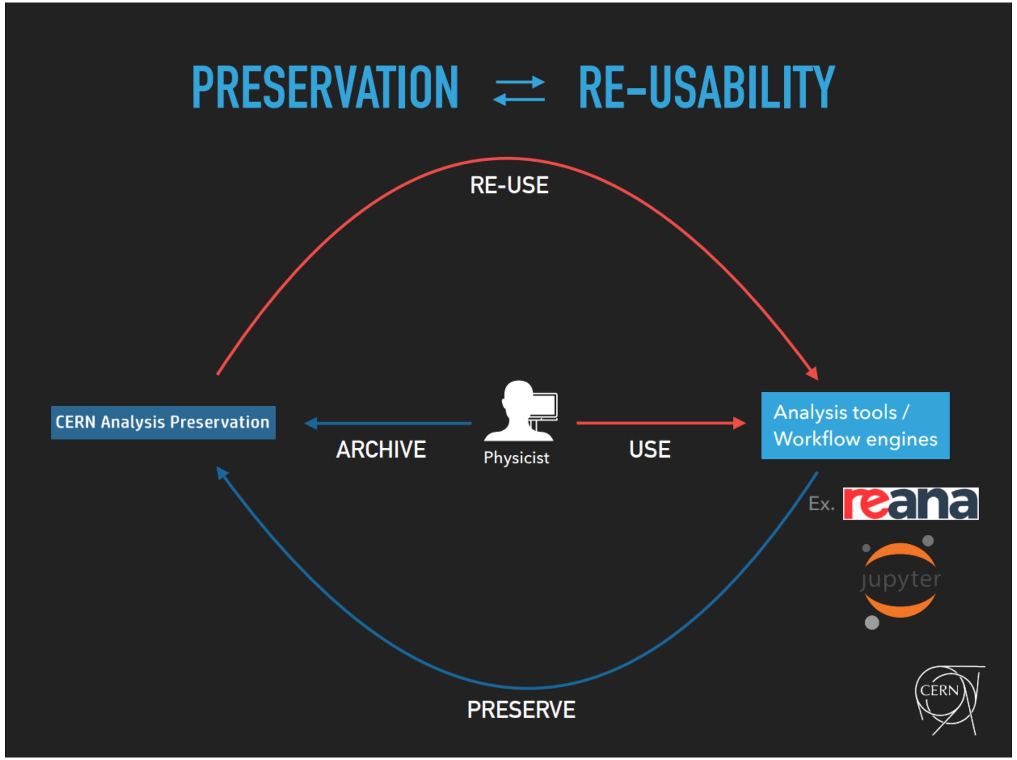
S. Kraml[1], B.C. Allanach[2], M. Mangano[3], H.B. Prosper[4], S. Sekmen[3,4] (editors), C. Balazs[5], A. Barr[6], P. Bechtle[7], G. Belanger[8], A. Belyaev[9,10], K. Benslama[11], M. Campanelli[12], K. Cranmer[13], A. De Roeck[3], M.J. Dolan[14], T. Eifert[15], J.R. Ellis[16,3], M. Felcini[17], B. Fuks[18], D. Guadagnoli[8,19], J.F. Gunion[20], S. Heinemeyer[17], J. Hewett[15], A. Ismail[15], M. Kadastik[21], M. Krämer[22], J. Lykken[23] F. Mahmoudi[3,24], S.P. Martin[25,26,27], T. Rizzo[15], T. Robens[28], M. Tytgat[29], A. Weiler[30]

Slide from Sabine Kraml, Feedback on use of public Likelihoods, 24 September 2020

Partial likelihoods already published/preserved



(CMS, 2017)

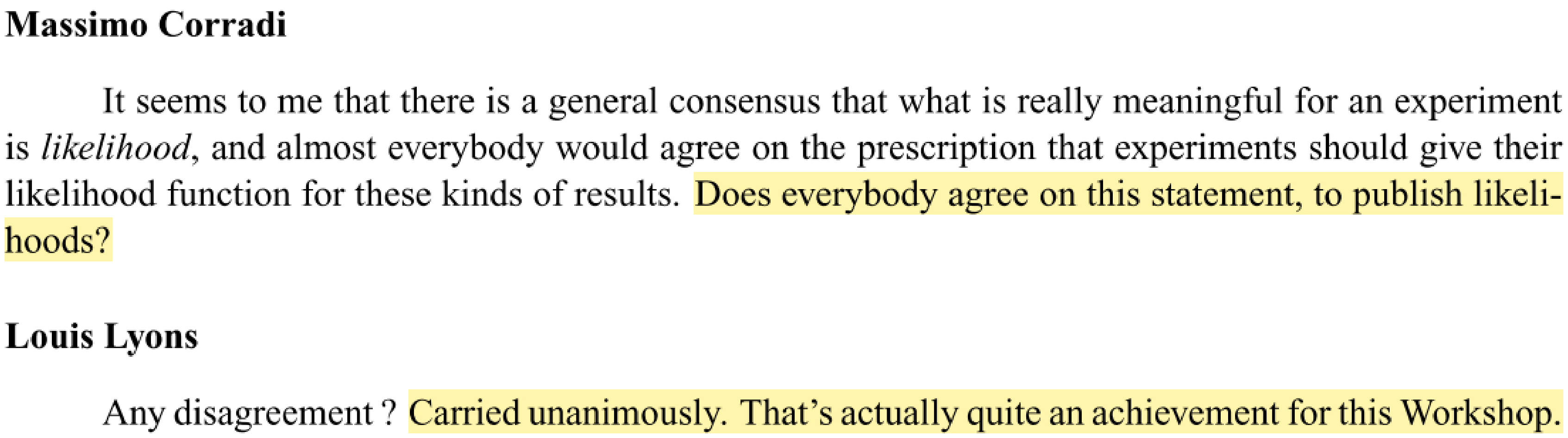

PRESERVATION ⇄ RE-USABILITY
RE-USE
CERN Analysis Preservation
ARCHIVE
Physicist
USE
Analysis tools / Workflow engines
Ex. reana
jupyter
PRESERVE
CERN

CERN Analysis Preservation implements FAIR data

(CERN, CHEP 2019)

Full likelihood serialization...

...making good on 19 year old agreement to publish likelihoods



Massimo Corradi

It seems to me that there is a general consensus that what is really meaningful for an experiment is *likelihood*, and almost everybody would agree on the prescription that experiments should give their likelihood function for these kinds of results. Does everybody agree on this statement, to publish likelihoods?

Louis Lyons

Any disagreement ? Carried unanimously. That's actually quite an achievement for this Workshop.

(1st Workshop on Confidence Limits, CERN, 2000)

This hadn't been done in HEP until 2019

- In an "open world" of statistics this is a difficult problem to solve
- What to preserve and how? All of ROOT?
- Idea: Focus on a single more tractable binned model first

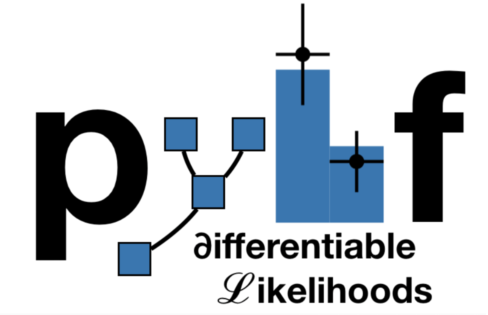
Enter HistFactory and `pyhf`

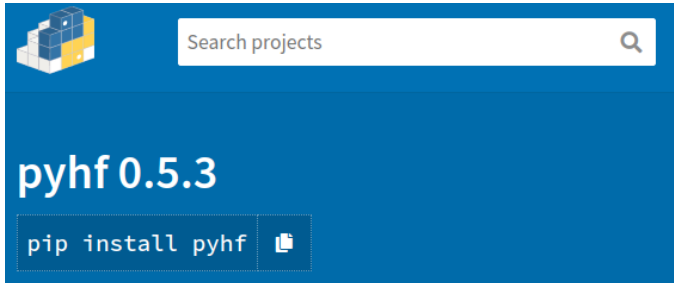
$$f\left(\mathrm{data}|\mathrm{parameters}\right) = f\left(\vec{n}, \vec{a}|\vec{\eta}, \vec{\chi}\right) = \prod_{c\,\in\,\mathrm{channels}}\ \prod_{b\,\in\,\mathrm{bins}_c} \mathrm{Pois}\left(n_{cb}|\nu_{cb}\left(\vec{\eta}, \vec{\chi}\right)\right) \prod_{\chi\,\in\,\vec{\chi}} c_{\chi}\left(a_{\chi}|\chi\right)$$

Use: Multiple disjoint channels (or regions) of binned distributions with multiple samples contributing to each with additional (possibly shared) systematics between sample estimates

HistFactory is used ubiquitously in binned analyses
Focus on this flexible p.d.f. template rather than "open world" of models
This is a mathematical representation! Nowhere is any software spec defined

`pyhf`: HistFactory in pure Python hardware accelerated with autodiff DOI 10.5281/zenodo.1169739 (how to cite)





pyhf 0.5.3

```
pip install pyhf
```

Lukas Heinrich

CERN

Matthew Feickert

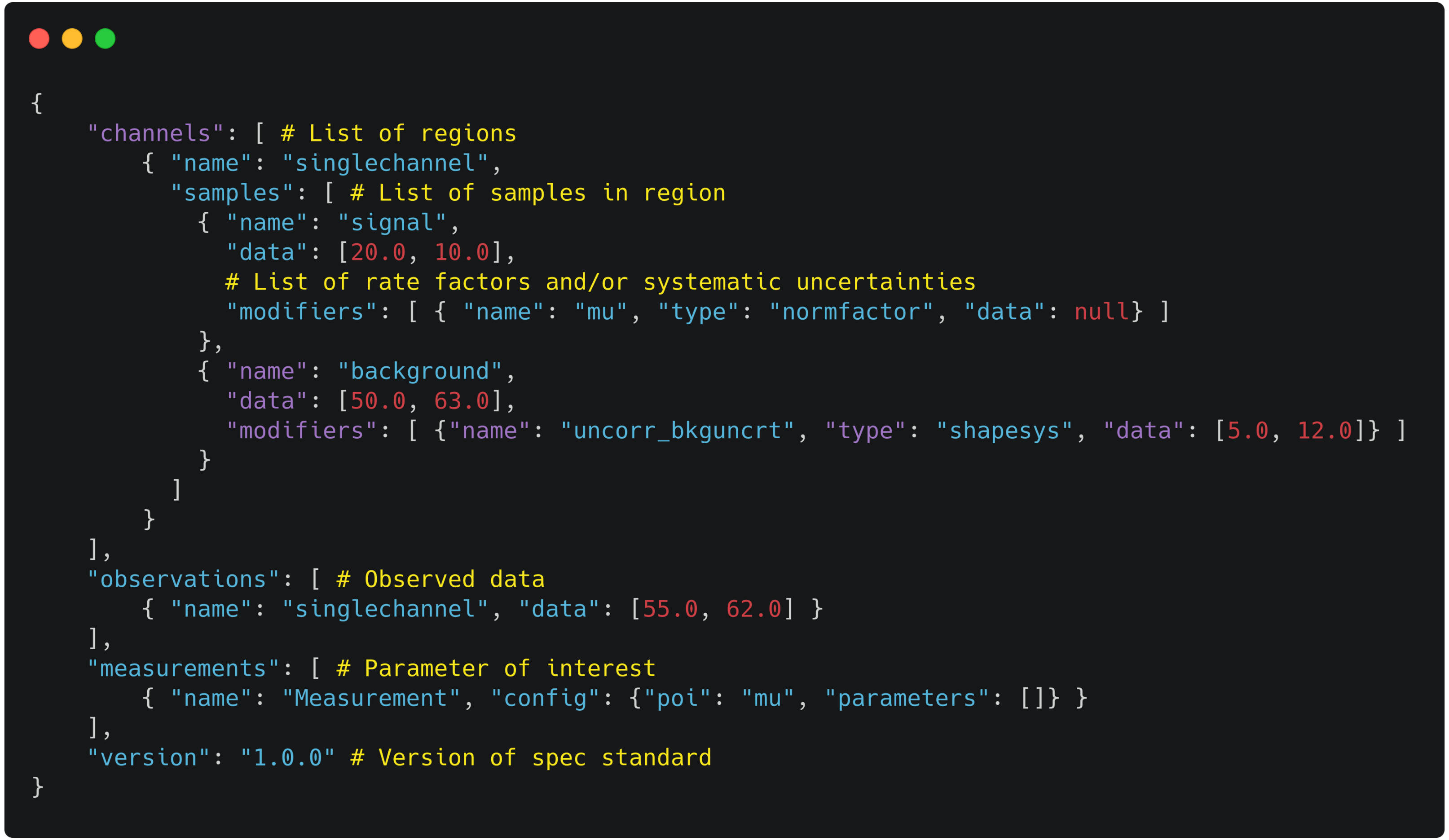
Illinois

Giordon Stark

UCSC SCIPP

JSON spec fully describes the HistFactory model

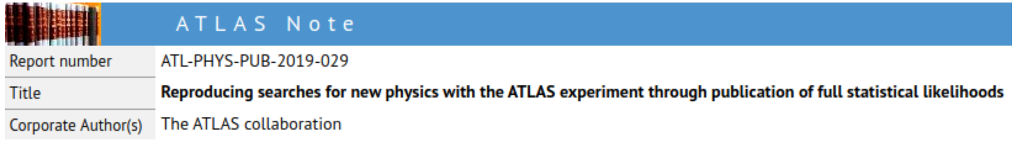
- Human & machine readable **declarative** statistical models
- Industry standard
    - Will be with us forever
- Parsable by every language
    - Highly portable
    - Bidirectional translation with ROOT
- Versionable and easily preserved
    - JSON Schema describing HistFactory specification
    - Attractive for analysis preservation
    - Highly compressible



```
{
    "channels": [ # List of regions
        { "name": "singlechannel",
          "samples": [ # List of samples in region
            { "name": "signal",
              "data": [20.0, 10.0],
              # List of rate factors and/or systematic uncertainties
              "modifiers": [ { "name": "mu", "type": "normfactor", "data": null} ]
            },
            { "name": "background",
              "data": [50.0, 63.0],
              "modifiers": [ {"name": "uncorr_bkguncrt", "type": "shapesys", "data": [5.0, 12.0]} ]
            }
          ]
        }
    ],
    "observations": [ # Observed data
        { "name": "singlechannel", "data": [55.0, 62.0] }
    ],
    "measurements": [ # Parameter of interest
        { "name": "Measurement", "config": {"poi": "mu", "parameters": []} }
    ],
    "version": "1.0.0" # Version of spec standard
}
```

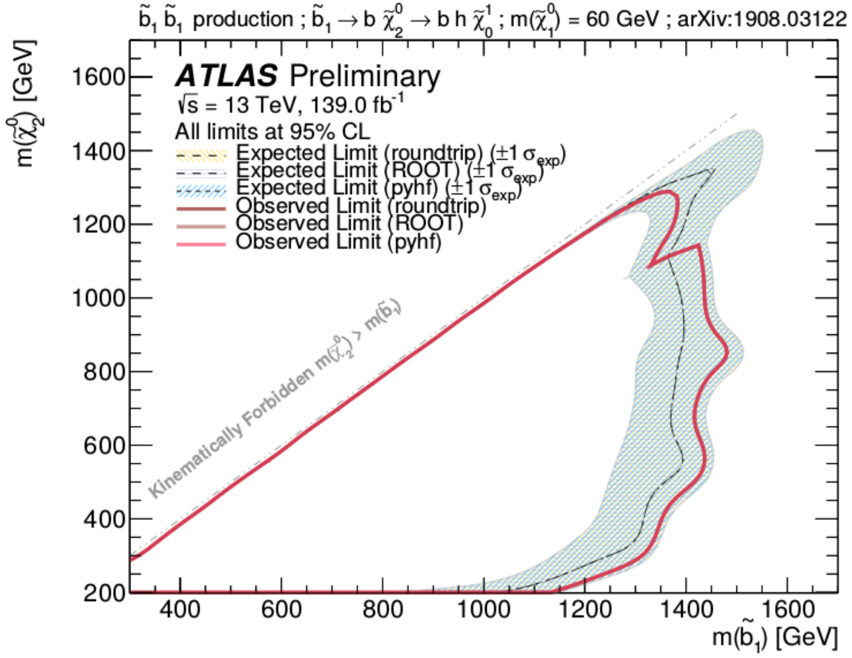
`JSON` defining a single channel, two bin counting experiment with systematics

ATLAS validation and publication of likelihoods



ATLAS Note

Report number	ATL-PHYS-PUB-2019-029
Title	**Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods**
Corporate Author(s)	The ATLAS collaboration



(ATLAS, 2019)

New open release allows theorists to explore LHC data in a new way

The ATLAS collaboration releases full analysis likelihoods, a first for an LHC experiment

9 JANUARY, 2020 | By Katarina Anthony

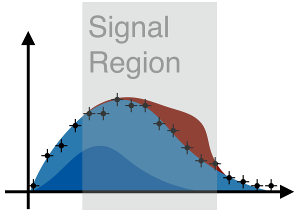
Explore ATLAS open likelihoods on the HEPData platform (Image: CERN)

(CERN, 2020)

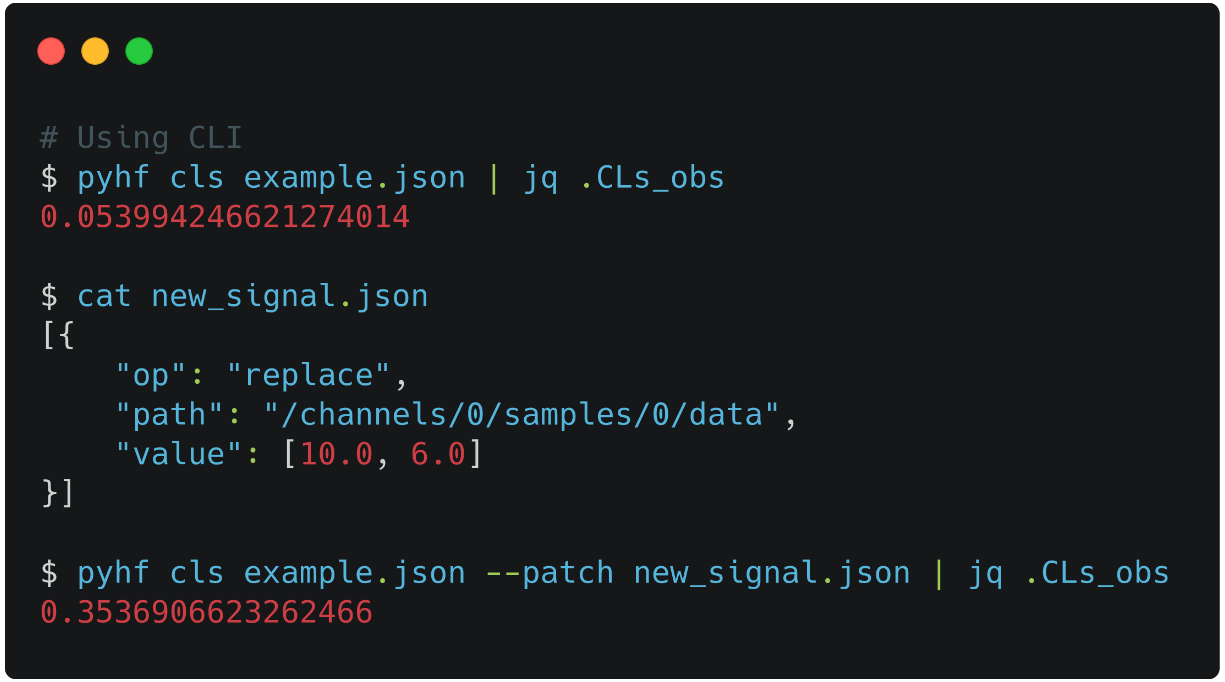
JSON Patch for signal model (reinterpretation)

JSON Patch gives ability to **easily mutate model**
Think: test a **new theory** with a **new patch**!
(c.f. Lukas Heinrich's RECAST talk from Snowmass 2021 Computational Frontier Workshop)

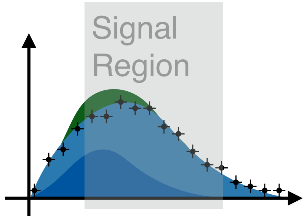
Combined with RECAST gives powerful tool for **reinterpretation studies**



Signal model A



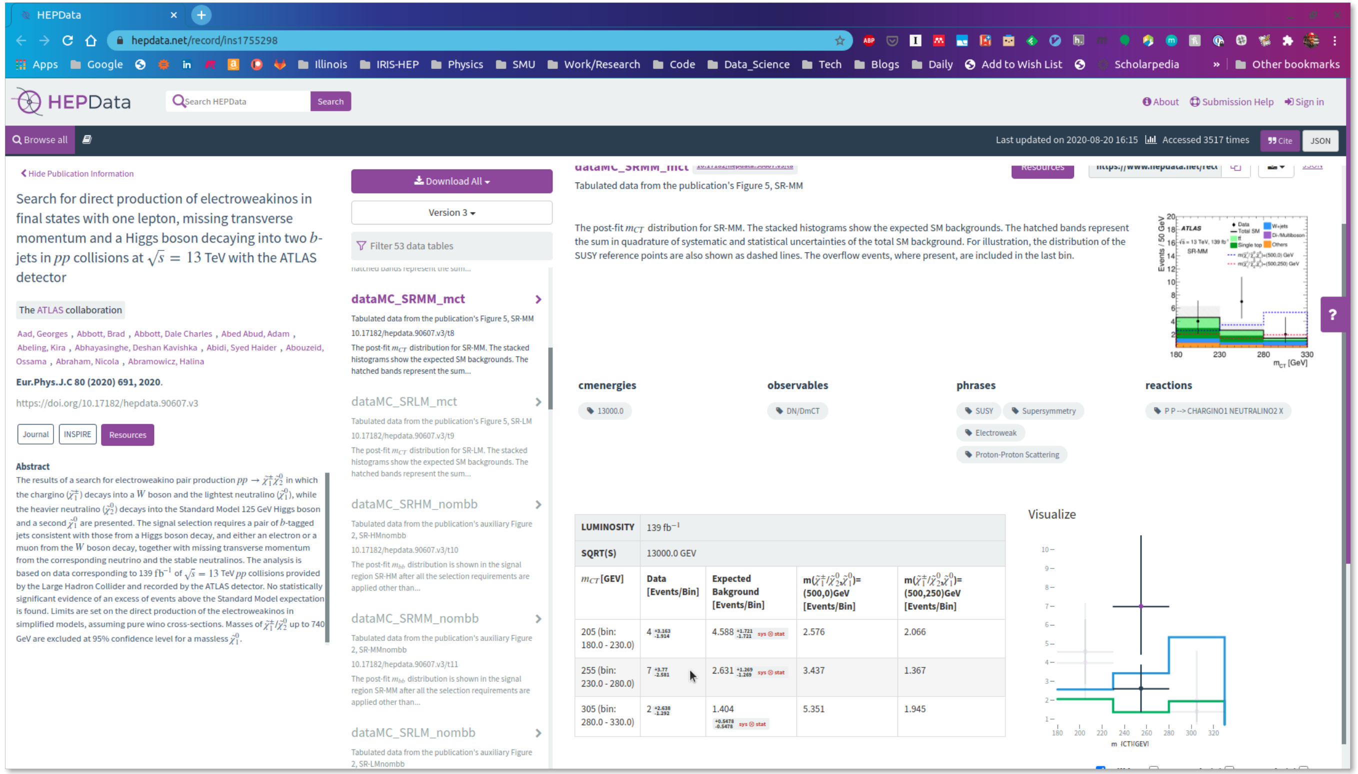
```
# Using CLI
$ pyhf cls example.json | jq .CLs_obs
0.053994246621274014

$ cat new_signal.json
[{
    "op": "replace",
    "path": "/channels/0/samples/0/data",
    "value": [10.0, 6.0]
}]

$ pyhf cls example.json --patch new_signal.json | jq .CLs_obs
0.3536906623262466
```



Signal model B

HEPData

- Repository for publication-related High-Energy Physics data
- Allows for publication of data products that accompany papers
  - Extract values for Figures and Tables
  - Serializations in common formats
  - Associated DOI
- Vital resource in the collaboration cycle between experiment and theory



Last updated on 2020-08-20 16:15 Accessed 3517 times

Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

Aad, Georges , Abbott, Brad , Abbott, Dale Charles , Abed Abud, Adam , Abeling, Kira , Abhayasinghe, Deshan Kavishka , Abidi, Syed Haider , Abouzeid, Ossama , Abraham, Nicola , Abramowicz, Halina

Eur.Phys.J.C 80 (2020) 691, 2020.

https://doi.org/10.17182/hepdata.90607.v3

Abstract

The results of a search for electroweakino pair production $pp \rightarrow \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ in which the chargino ($\tilde{\chi}_1^{\pm}$) decays into a W boson and the lightest neutralino ($\tilde{\chi}_1^0$), while the heavier neutralino ($\tilde{\chi}_2^0$) decays into the Standard Model 125 GeV Higgs boson and a second $\tilde{\chi}_1^0$ are presented. The signal selection requires a pair of b-tagged jets consistent with those from a Higgs boson decay, and either an electron or a muon from the W boson decay, together with missing transverse momentum from the corresponding neutrino and the stable neutralinos. The analysis is based on data corresponding to 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions provided by the Large Hadron Collider and recorded by the ATLAS detector. No statistically significant evidence of an excess of events above the Standard Model expectation is found. Limits are set on the direct production of the electroweakinos in simplified models, assuming pure wino cross-sections. Masses of $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ up to 740 GeV are excluded at 95% confidence level for a massless $\tilde{\chi}_1^0$.

hatched bands represent the sum...

dataMC_SRMM_mct

Tabulated data from the publication's Figure 5, SR-MM

10.17182/hepdata.90607.v3/t8

The post-fit m_{CT} distribution for SR-MM. The stacked histograms show the expected SM backgrounds. The hatched bands represent the sum...

dataMC_SRLM_mct

Tabulated data from the publication's Figure 5, SR-LM

10.17182/hepdata.90607.v3/t9

The post-fit m_{CT} distribution for SR-LM. The stacked histograms show the expected SM backgrounds. The hatched bands represent the sum...

dataMC_SRHM_nombb

Tabulated data from the publication's auxiliary Figure 2, SR-HMnombb

10.17182/hepdata.90607.v3/t10

The post-fit m_{bb} distribution is shown in the signal region SR-HM after all the selection requirements are applied other than...

dataMC_SRMM_nombb

Tabulated data from the publication's auxiliary Figure 2, SR-MMnombb

10.17182/hepdata.90607.v3/t11

The post-fit m_{bb} distribution is shown in the signal region SR-MM after all the selection requirements are applied other than...

dataMC_SRLM_nombb

Tabulated data from the publication's auxiliary Figure 2, SR-LMnombb

dataMC_SRMM_mct

Tabulated data from the publication's Figure 5, SR-MM

The post-fit m_{CT} distribution for SR-MM. The stacked histograms show the expected SM backgrounds. The hatched bands represent the sum in quadrature of systematic and statistical uncertainties of the total SM background. For illustration, the distribution of the SUSY reference points are also shown as dashed lines. The overflow events, where present, are included in the last bin.

cmenergies: 13000.0

observables: DN/DmCT

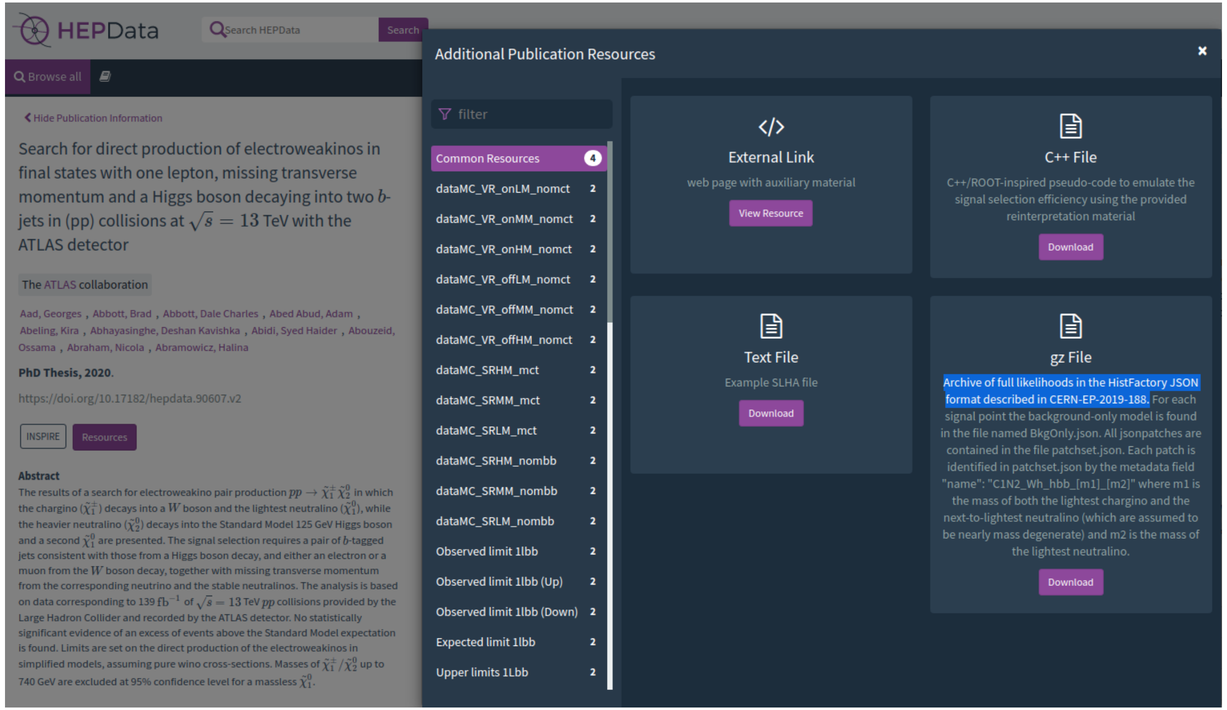
phrases: SUSY, Supersymmetry, Electroweak, Proton-Proton Scattering

reactions: P P --> CHARGINO1 NEUTRALINO2 X

LUMINOSITY	139 fb^{-1}			
SQRT(S)	13000.0 GEV			
m_{CT} [GEV]	Data [Events/Bin]	Expected Bakground [Events/Bin]	m($\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0,\tilde{\chi}_1^0$)=(500,0)GeV [Events/Bin]	m($\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0,\tilde{\chi}_1^0$)=(500,250)GeV [Events/Bin]
205 (bin: 180.0 - 230.0)	$4^{+3.163}_{-1.914}$	$4.588^{+1.721}_{-1.721}$ sys ⊗ stat	2.576	2.066
255 (bin: 230.0 - 280.0)	$7^{+3.77}_{-2.581}$	$2.631^{+1.269}_{-1.269}$ sys ⊗ stat	3.437	1.367
305 (bin: 280.0 - 330.0)	$2^{+2.638}_{-1.292}$	$1.404^{+0.5478}_{-0.5478}$ sys ⊗ stat	5.351	1.945

Likelihoods preserved on HEPData

- `pyhf` pallet:
  - Background-only model JSON stored
  - Hundreds of signal model JSON Patches stored together as a `pyhf` "patch set" file
- Fully preserve and publish the full statistical model and observations to give likelihood
  - with own DOI! DOI 10.17182/hepdata.90607.v3/r3



Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in (pp) collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

The ATLAS collaboration

Aad, Georges , Abbott, Brad , Abbott, Dale Charles , Abed Abud, Adam , Abeling, Kira , Abhayasinghe, Deshan Kavishka , Abidi, Syed Haider , Abouzeid, Ossama , Abraham, Nicola , Abramowicz, Halina

PhD Thesis, 2020.

https://doi.org/10.17182/hepdata.90607.v2

Abstract

The results of a search for electroweakino pair production $pp \rightarrow \tilde{\chi}_1^{\pm} \tilde{\chi}_2^0$ in which the chargino ($\tilde{\chi}_1^{\pm}$) decays into a W boson and the lightest neutralino ($\tilde{\chi}_1^0$), while the heavier neutralino ($\tilde{\chi}_2^0$) decays into the Standard Model 125 GeV Higgs boson and a second $\tilde{\chi}_1^0$ are presented. The signal selection requires a pair of b-tagged jets consistent with those from a Higgs boson decay, and either an electron or a muon from the W boson decay, together with missing transverse momentum from the corresponding neutrino and the stable neutralinos. The analysis is based on data corresponding to 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions provided by the Large Hadron Collider and recorded by the ATLAS detector. No statistically significant evidence of an excess of events above the Standard Model expectation is found. Limits are set on the direct production of the electroweakinos in simplified models, assuming pure wino cross-sections. Masses of $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ up to 740 GeV are excluded at 95% confidence level for a massless $\tilde{\chi}_1^0$.

Additional Publication Resources

Common Resources	4
dataMC_VR_onLM_nomct	2
dataMC_VR_onMM_nomct	2
dataMC_VR_onHM_nomct	2
dataMC_VR_offLM_nomct	2
dataMC_VR_offMM_nomct	2
dataMC_VR_offHM_nomct	2
dataMC_SRHM_mct	2
dataMC_SRMM_mct	2
dataMC_SRLM_mct	2
dataMC_SRHM_nombb	2
dataMC_SRMM_nombb	2
dataMC_SRLM_nombb	2
Observed limit 1lbb	2
Observed limit 1lbb (Up)	2
Observed limit 1lbb (Down)	2
Expected limit 1lbb	2
Upper limits 1Lbb	2

External Link

web page with auxiliary material

C++ File

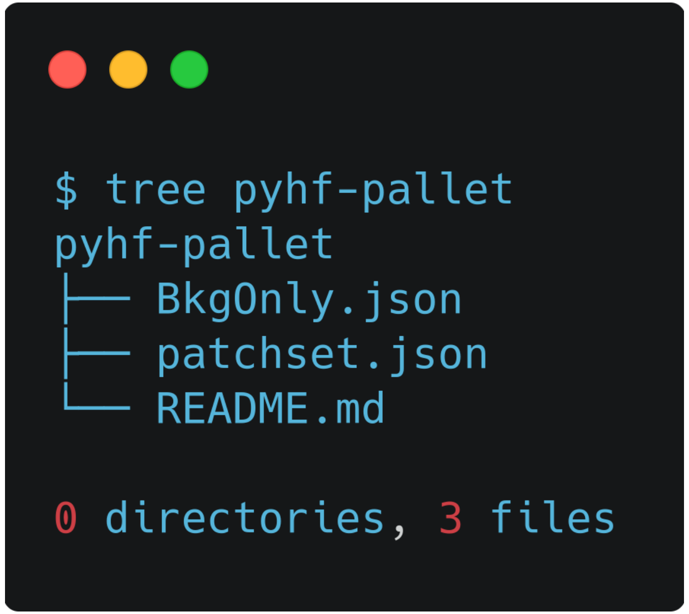
C++/ROOT-inspired pseudo-code to emulate the signal selection efficiency using the provided reinterpretation material

Text File

Example SLHA file

gz File

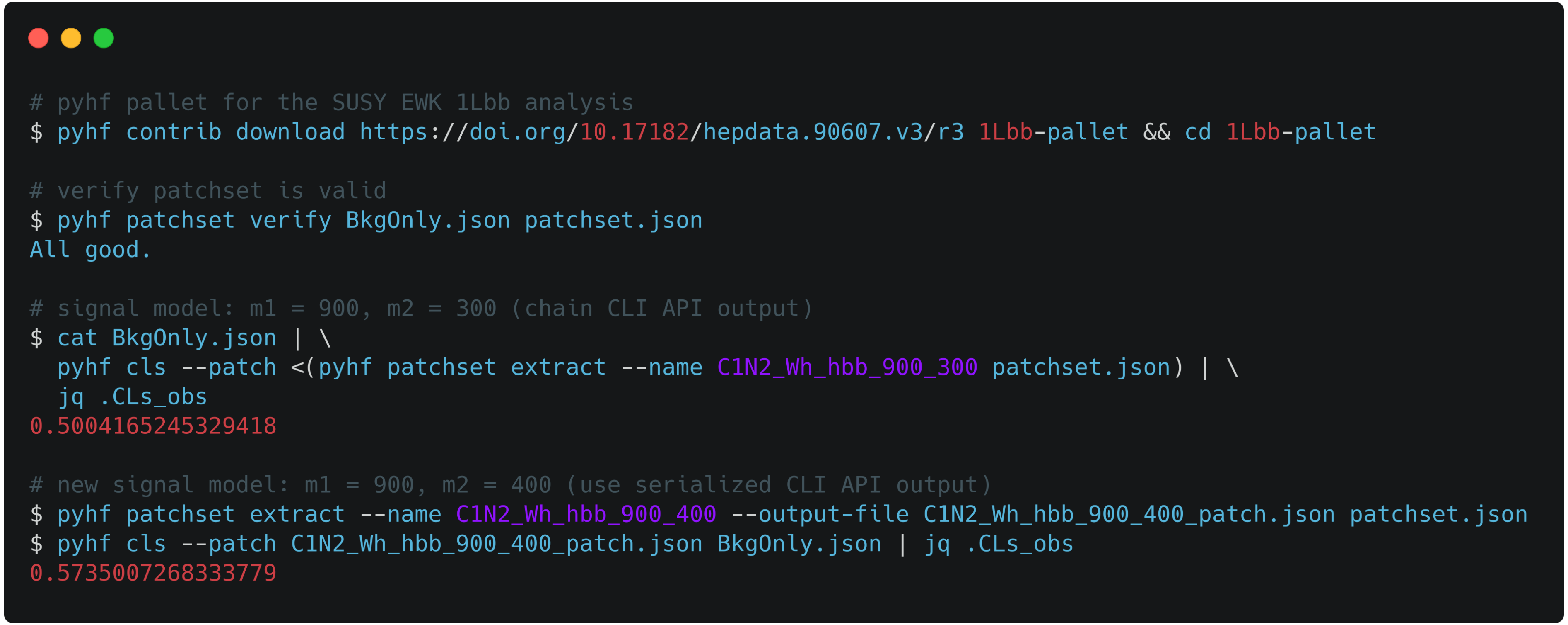
Archive of full likelihoods in the HistFactory JSON format described in CERN-EP-2019-188. For each signal point the background-only model is found in the file named BkgOnly.json. All jsonpatches are contained in the file patchset.json. Each patch is identified in patchset.json by the metadata field "name": "C1N2_Wh_hbb_[m1]_[m2]" where m1 is the mass of both the lightest chargino and the next-to-lightest neutralino (which are assumed to be nearly mass degenerate) and m2 is the mass of the lightest neutralino.



```
$ tree pyhf-pallet
pyhf-pallet
├── BkgOnly.json
├── patchset.json
└── README.md

0 directories, 3 files
```

...can be used from HEPData

- pyhf pallet:
  - Background-only model JSON stored
  - Hundreds of signal model JSON Patches stored together as a pyhf "patch set" file
- Fully preserve and publish the full statistical model and observations to give likelihood
  - with own DOI! DOI 10.17182/hepdata.90607.v3/r3

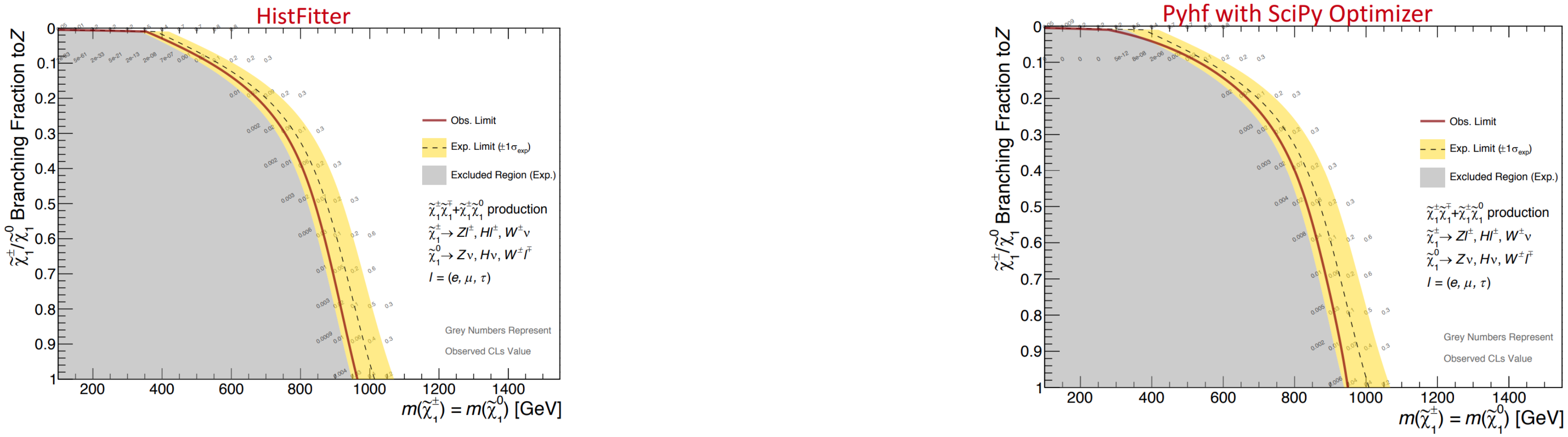


```
# pyhf pallet for the SUSY EWK 1Lbb analysis
$ pyhf contrib download https://doi.org/10.17182/hepdata.90607.v3/r3 1Lbb-pallet && cd 1Lbb-pallet

# verify patchset is valid
$ pyhf patchset verify BkgOnly.json patchset.json
All good.

# signal model: m1 = 900, m2 = 300 (chain CLI API output)
$ cat BkgOnly.json | \
  pyhf cls --patch <(pyhf patchset extract --name C1N2_Wh_hbb_900_300 patchset.json) | \
  jq .CLs_obs
0.5004165245329418

# new signal model: m1 = 900, m2 = 400 (use serialized CLI API output)
$ pyhf patchset extract --name C1N2_Wh_hbb_900_400 --output-file C1N2_Wh_hbb_900_400_patch.json patchset.json
$ pyhf cls --patch C1N2_Wh_hbb_900_400_patch.json BkgOnly.json | jq .CLs_obs
0.5735007268333779
```

Rapid adoption in ATLAS...

- Five ATLAS analyses with full likelihoods published to HEPData
- ATLAS SUSY will be continuing to publish full Run 2 likelihoods
- More FAIR than traditional ROOT workspaces

- direct staus, doi:10.17182/hepdata.89408 (2019)
- sbottom multi-b, doi:10.17182/hepdata.91127 (2019)
- 1Lbb, doi:10.17182/hepdata.92006 (2019)
- 3L eRJR, doi:10.17182/hepdata.90607 (2020)
- ss3L search, doi:10.17182/hepdata.91214 (2020)

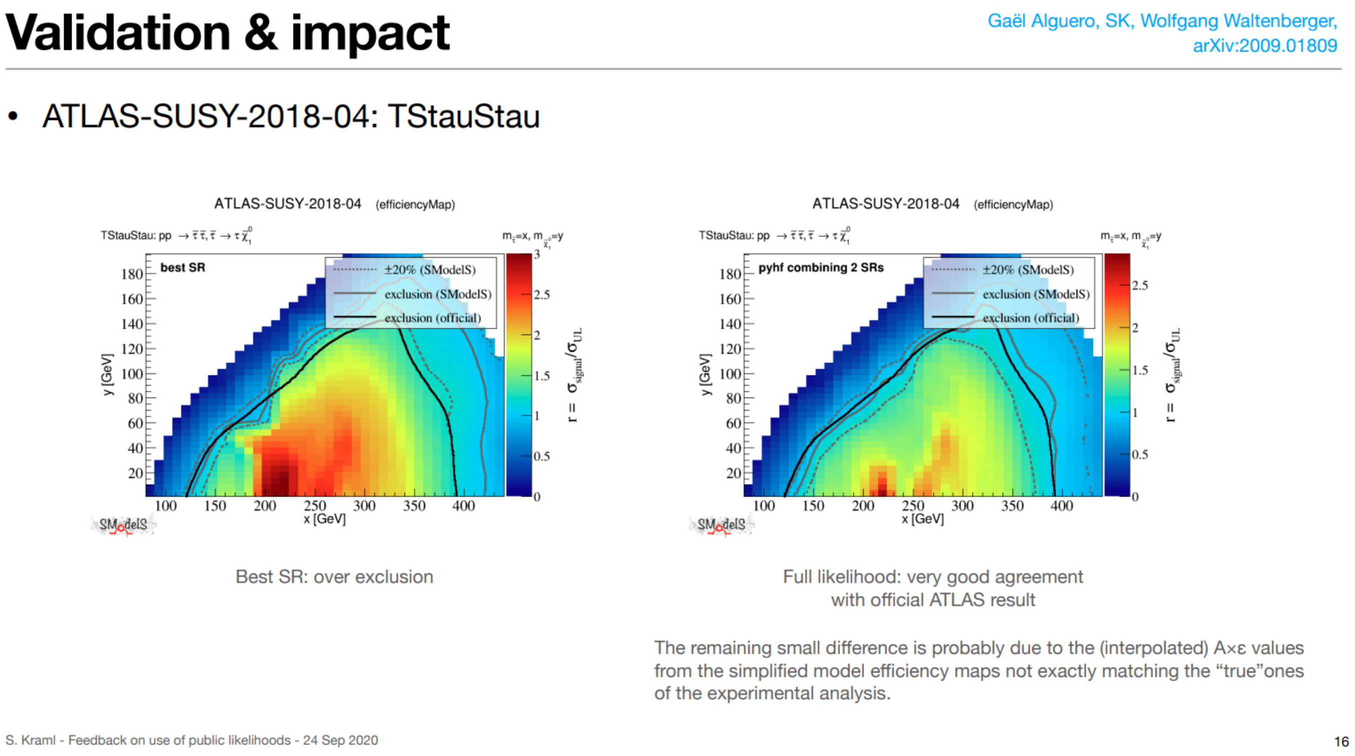

HistFitter

Pyhf with SciPy Optimizer

SUSY EWK 3L RPV analysis (ATLAS-CONF-2020-009): Exclusion curves as a function of mass and branching fraction to Z bosons

...and by theory

- `pyhf` likelihoods discussed in
  - Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report
  - Higgs boson potential at colliders: status and perspectives
- SModelS team has implemented a `SModelS/pyhf` interface [arXiv:2009.01809]
  - tool for interpreting simplified-model results from the LHC
  - designed to be used by theorists
  - `SModelS` authors gave tutorial at TOOLS 2020 workshop



Validation & impact

Gaël Alguero, SK, Wolfgang Waltenberger, arXiv:2009.01809

- ATLAS-SUSY-2018-04: TStauStau

Best SR: over exclusion

Full likelihood: very good agreement with official ATLAS result

The remaining small difference is probably due to the (interpolated) A×ε values from the simplified model efficiency maps not exactly matching the "true"ones of the experimental analysis.

S. Kraml - Feedback on use of public likelihoods - 24 Sep 2020 16

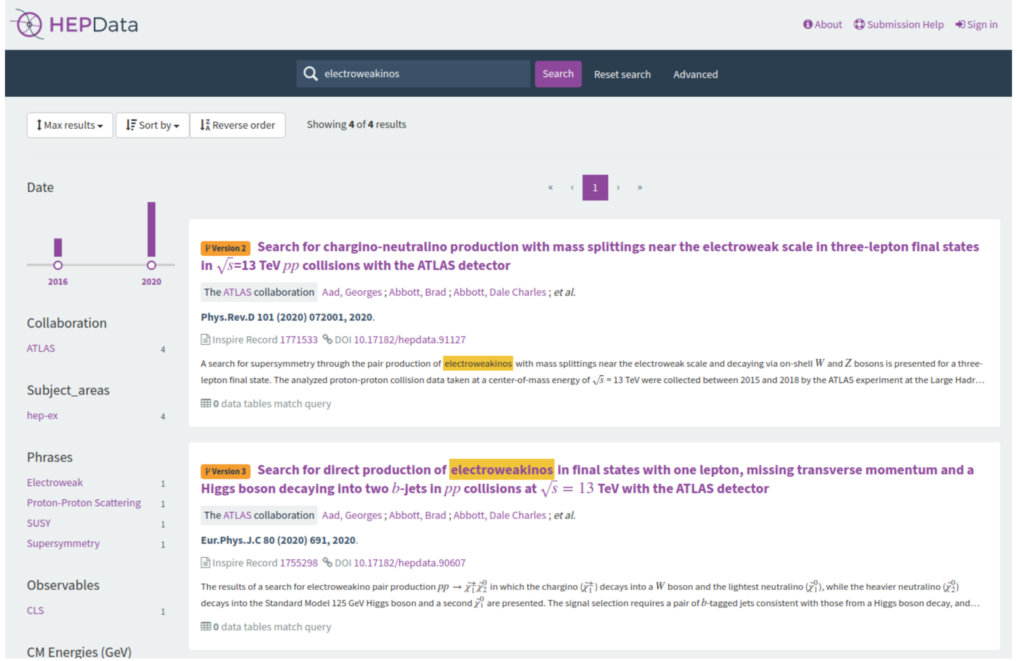
Feedback on use of public Likelihoods, Sabine Kraml
(ATLAS Exotics + SUSY Reinterpretations Workshop)

- Have produced three comparisons to published ATLAS likelihoods: ATLAS-SUSY-2018-04, ATLAS-SUSY-2018-31, ATLAS-SUSY-2019-08
  - Compare simplified likelihood (bestSR) to full likelihood (`pyhf`) using `SModelS`

FAIR summary for likelihoods: Findable

Metadata and data should be easy to find for both humans and computers

Searchable database



electroweakinos

Showing **4** of **4** results

Date

Collaboration
ATLAS 4

Subject_areas
hep-ex 4

Phrases
Electroweak 1
Proton-Proton Scattering 1
SUSY 1
Supersymmetry 1

Observables
CLS 1

CM Energies (GeV)

Version 2 **Search for chargino-neutralino production with mass splittings near the electroweak scale in three-lepton final states in $\sqrt{s}$=13 TeV pp collisions with the ATLAS detector**

The ATLAS collaboration Aad, Georges ; Abbott, Brad ; Abbott, Dale Charles ; *et al.*

Phys.Rev.D 101 (2020) 072001, 2020.

Inspire Record 1771533 DOI 10.17182/hepdata.91127

A search for supersymmetry through the pair production of electroweakinos with mass splittings near the electroweak scale and decaying via on-shell W and Z bosons is presented for a three-lepton final state. The analyzed proton-proton collision data taken at a center-of-mass energy of $\sqrt{s}$ = 13 TeV were collected between 2015 and 2018 by the ATLAS experiment at the Large Hadr…

0 data tables match query

Version 3 **Search for direct production of electroweakinos in final states with one lepton, missing transverse momentum and a Higgs boson decaying into two b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector**

The ATLAS collaboration Aad, Georges ; Abbott, Brad ; Abbott, Dale Charles ; *et al.*

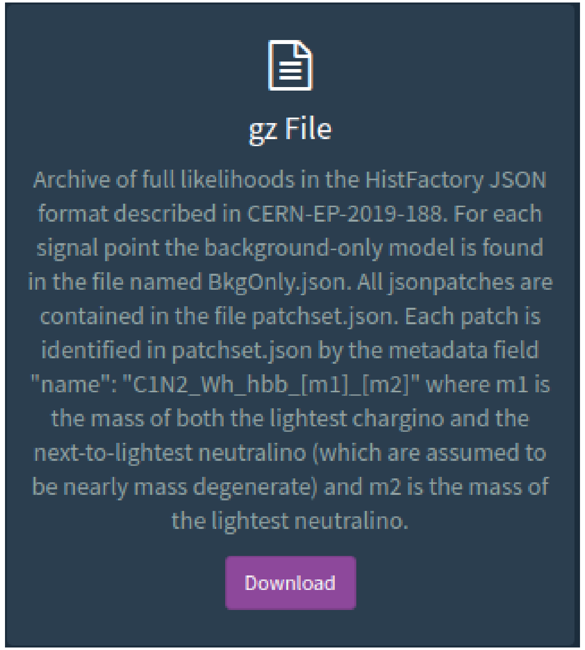
Eur.Phys.J.C 80 (2020) 691, 2020.

Inspire Record 1755298 DOI 10.17182/hepdata.90607

The results of a search for electroweakino pair production $pp \to \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ in which the chargino ($\tilde{\chi}_1^{\pm}$) decays into a W boson and the lightest neutralino ($\tilde{\chi}_1^0$), while the heavier neutralino ($\tilde{\chi}_2^0$) decays into the Standard Model 125 GeV Higgs boson and a second $\tilde{\chi}_1^0$ are presented. The signal selection requires a pair of b-tagged jets consistent with those from a Higgs boson decay, and…

0 data tables match query

Transparent data products



gz File

Archive of full likelihoods in the HistFactory JSON format described in CERN-EP-2019-188. For each signal point the background-only model is found in the file named BkgOnly.json. All jsonpatches are contained in the file patchset.json. Each patch is identified in patchset.json by the metadata field "name": "C1N2_Wh_hbb_[m1]_[m2]" where m1 is the mass of both the lightest chargino and the next-to-lightest neutralino (which are assumed to be nearly mass degenerate) and m2 is the mass of the lightest neutralino.

DOI 10.17182/hepdata.90607

DOI 10.17182/hepdata.90607.v3/r3

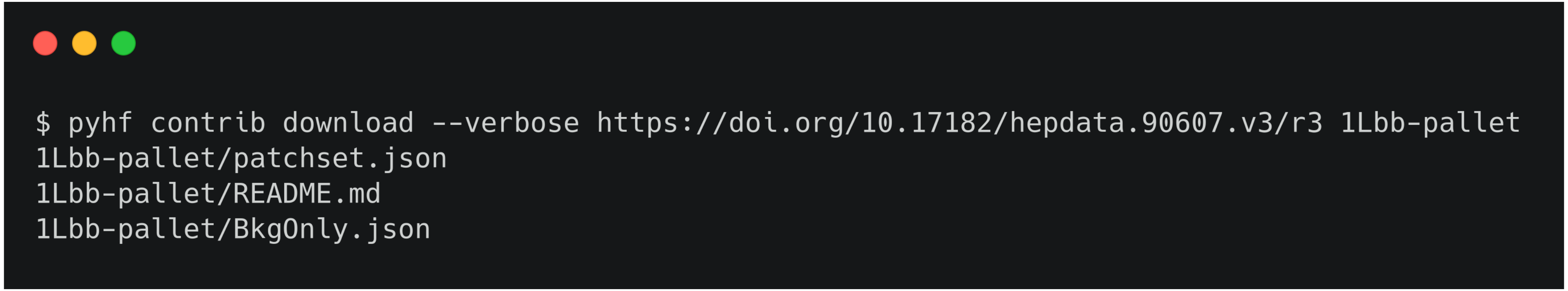
FAIR summary for likelihoods: Accessible

Once found metadata and data need to be easily accessed

Clickable buttons to download

DOI 10.17182/hepdata.90607.v3/r3

CLI tools and Python API

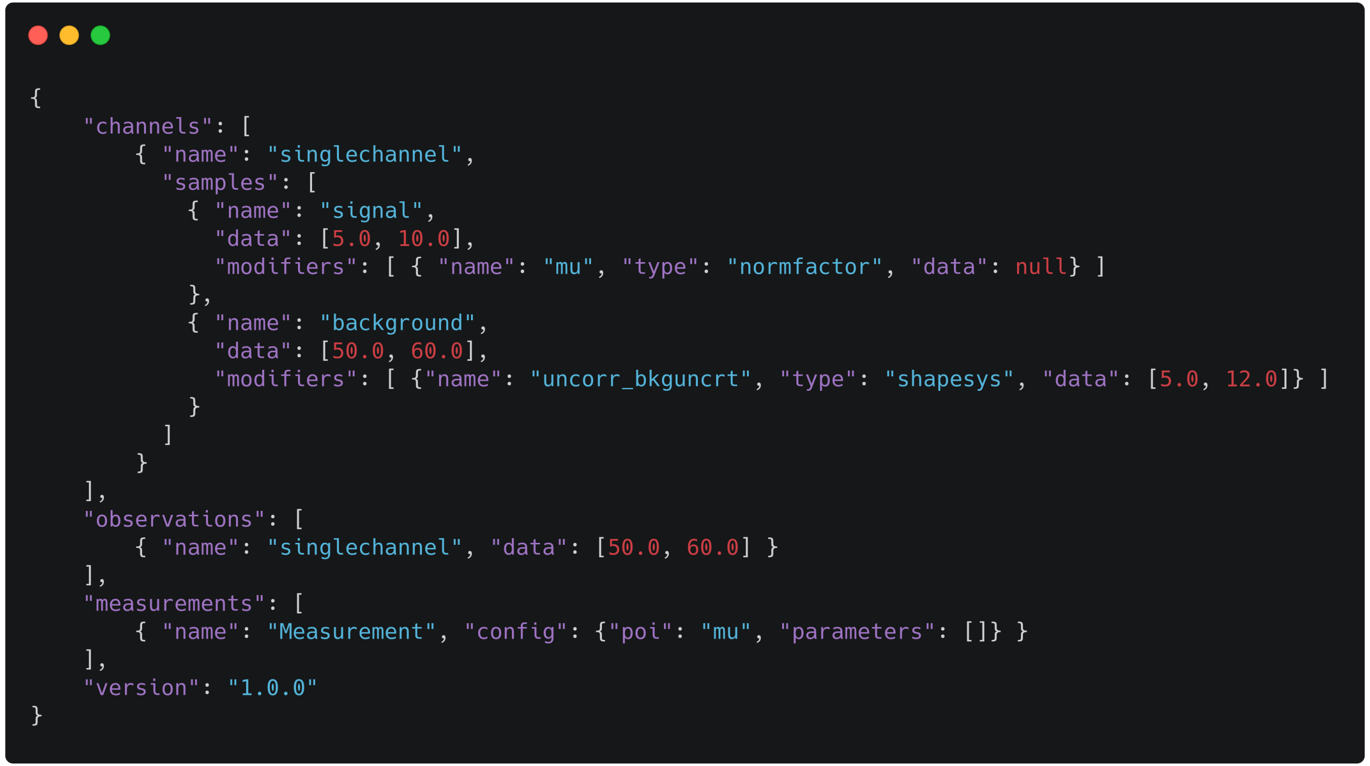


```
$ pyhf contrib download --verbose https://doi.org/10.17182/hepdata.90607.v3/r3 1Lbb-pallet
1Lbb-pallet/patchset.json
1Lbb-pallet/README.md
1Lbb-pallet/BkgOnly.json
```

FAIR summary for likelihoods: Interoperable

Interoperate with applications for analysis, storage, and processing

JSON spec



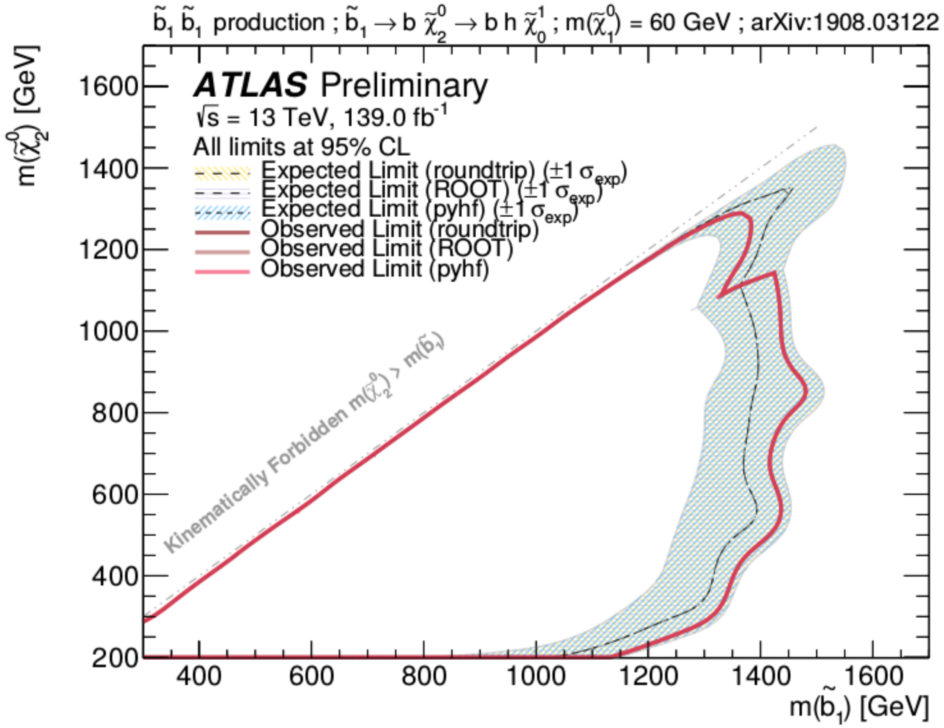
```json
{
    "channels": [
        { "name": "singlechannel",
          "samples": [
            { "name": "signal",
              "data": [5.0, 10.0],
              "modifiers": [ { "name": "mu", "type": "normfactor", "data": null} ]
            },
            { "name": "background",
              "data": [50.0, 60.0],
              "modifiers": [ {"name": "uncorr_bkguncrt", "type": "shapesys", "data": [5.0, 12.0]} ]
            }
          ]
        }
    ],
    "observations": [
        { "name": "singlechannel", "data": [50.0, 60.0] }
    ],
    "measurements": [
        { "name": "Measurement", "config": {"poi": "mu", "parameters": []} }
    ],
    "version": "1.0.0"
}
```

FAIR summary for likelihoods: Reusable

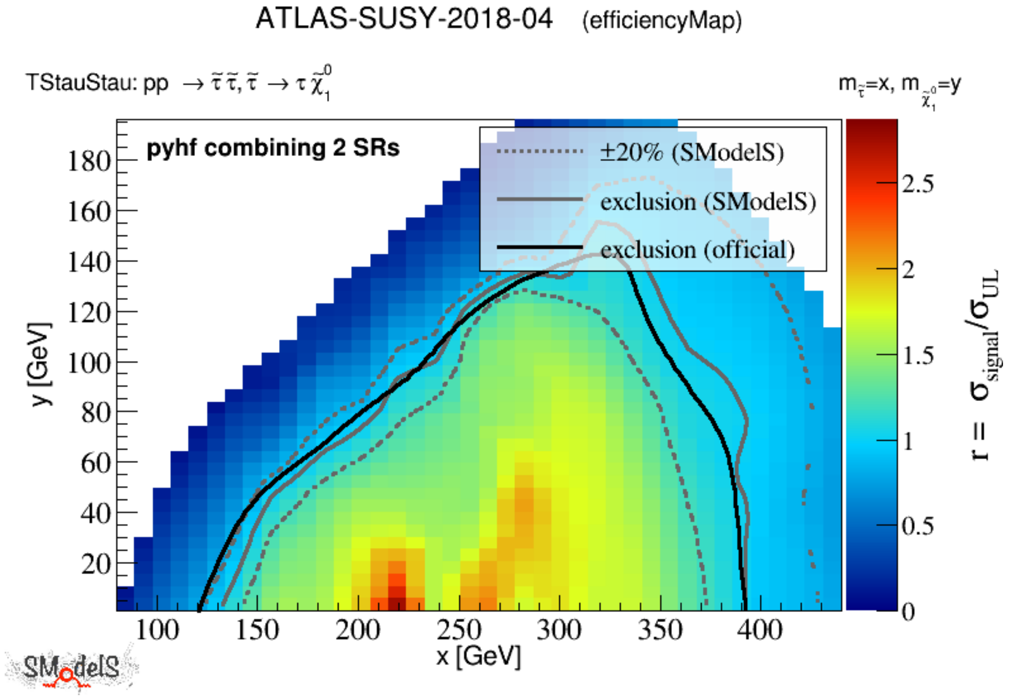
Well-described so can be replicated and/or combined in different settings

HEPData platform and rich metadata make possible

Replication of published results



New applications and reinterpretation



Validation example

Tools for making data products FAIR

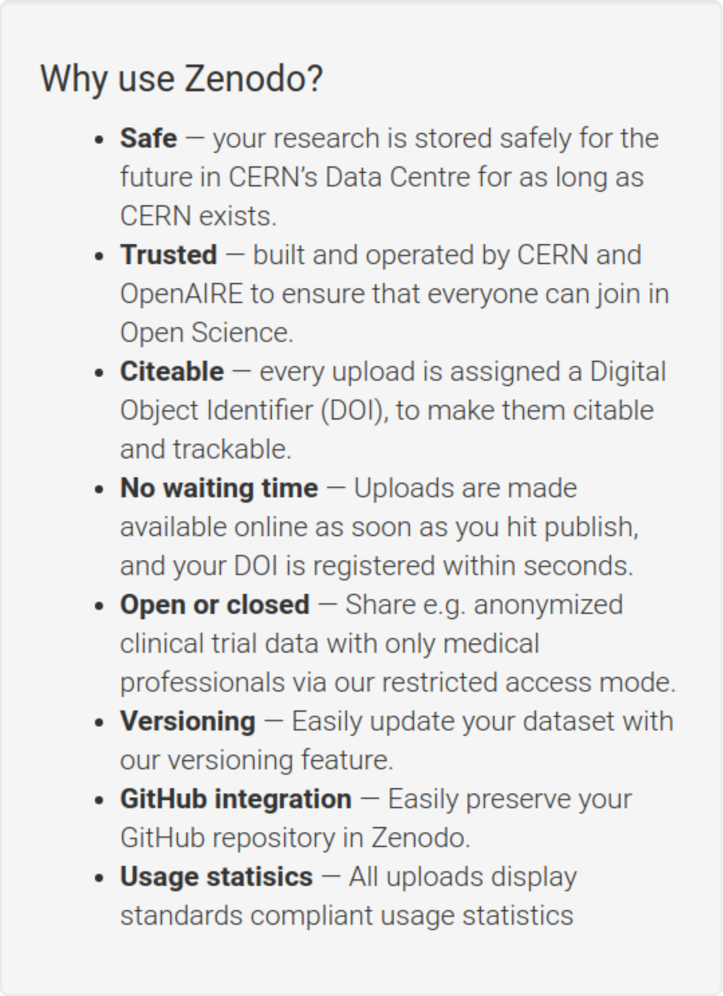
HEPData

- HEP centric by design (FAIR data core tool)
- Open source
- HEP data products:
  - table source data
  - figure source data
  - likelihoods

zenodo

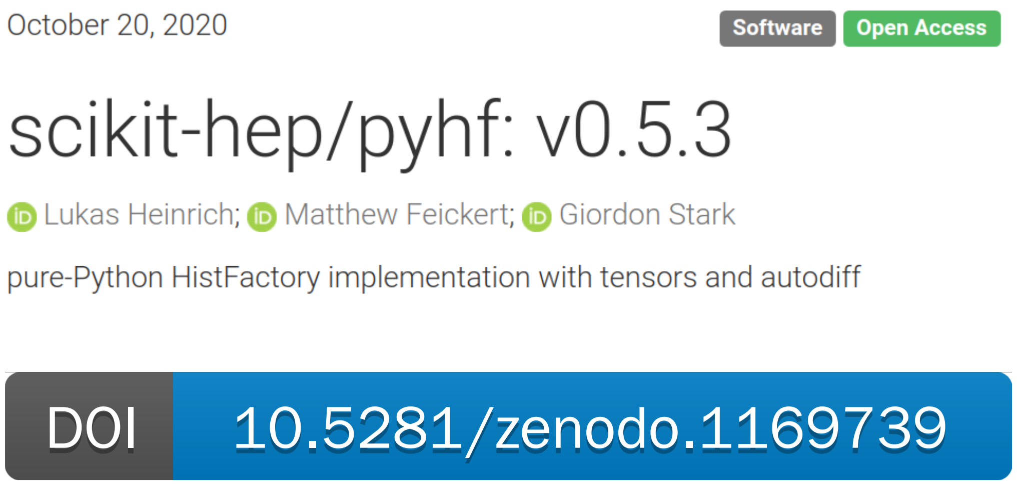
- Open science resource for all fields
- Open source (but your files can be closed access)
- Versioned archival of everything:
  - code
  - documents
  - data products
  - data sets

Zenodo



Why use Zenodo?

- **Safe** — your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** — built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** — every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** — Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** — Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** — Easily update your dataset with our versioning feature.
- **GitHub integration** — Easily preserve your GitHub repository in Zenodo.
- **Usage statisics** — All uploads display standards compliant usage statistics



October 20, 2020

Software Open Access

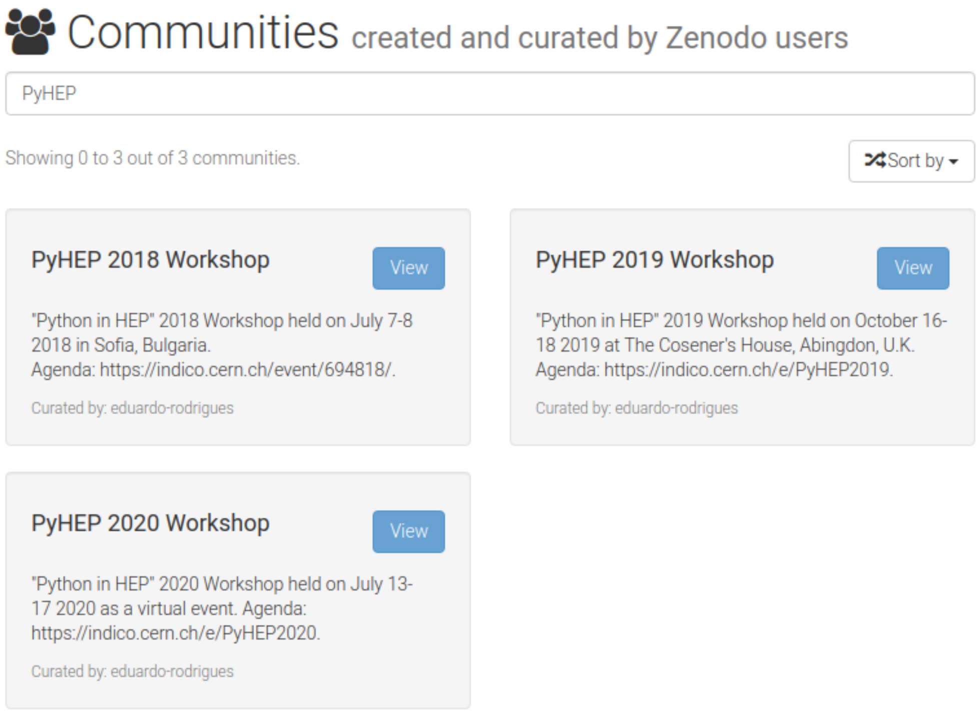
scikit-hep/pyhf: v0.5.3

Lukas Heinrich; Matthew Feickert; Giordon Stark

pure-Python HistFactory implementation with tensors and autodiff

DOI 10.5281/zenodo.1169739

Communities allow for archival of collections



Communities created and curated by Zenodo users

PyHEP

Showing 0 to 3 out of 3 communities.

PyHEP 2018 Workshop

"Python in HEP" 2018 Workshop held on July 7-8 2018 in Sofia, Bulgaria.
Agenda: https://indico.cern.ch/event/694818/.

Curated by: eduardo-rodrigues

PyHEP 2019 Workshop

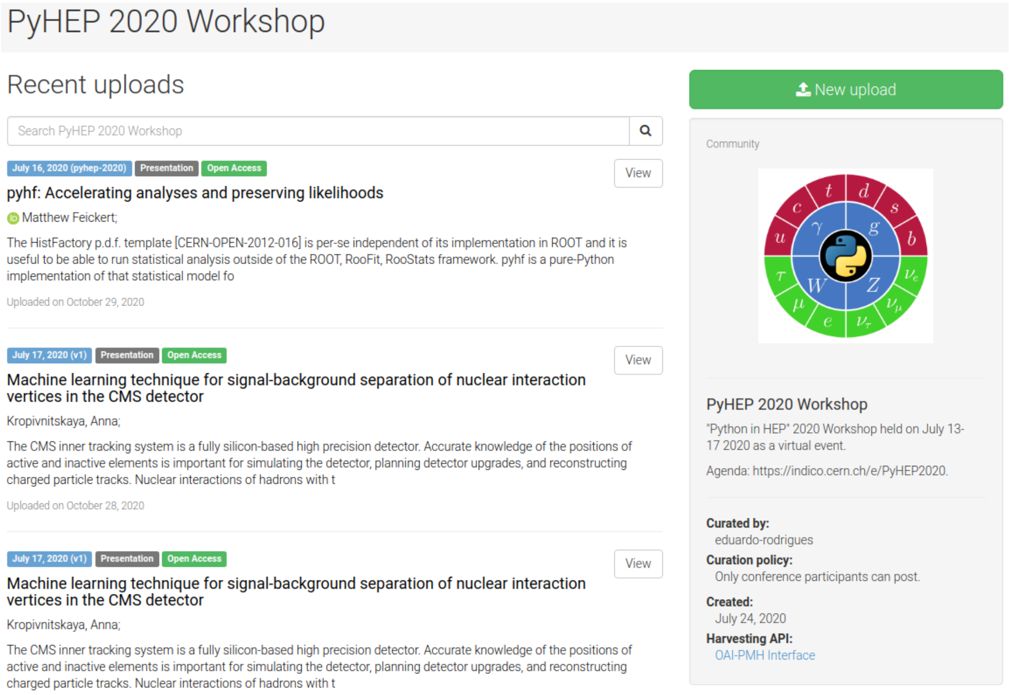
"Python in HEP" 2019 Workshop held on October 16-18 2019 at The Cosener's House, Abingdon, U.K.
Agenda: https://indico.cern.ch/e/PyHEP2019.

Curated by: eduardo-rodrigues

PyHEP 2020 Workshop

"Python in HEP" 2020 Workshop held on July 13-17 2020 as a virtual event. Agenda: https://indico.cern.ch/e/PyHEP2020.

Curated by: eduardo-rodrigues



PyHEP 2020 Workshop

Recent uploads

Search PyHEP 2020 Workshop

July 16, 2020 (pyhep-2020) | Presentation | Open Access

pyhf: Accelerating analyses and preserving likelihoods

Matthew Feickert;

The HistFactory p.d.f. template [CERN-OPEN-2012-016] is per-se independent of its implementation in ROOT and it is useful to be able to run statistical analysis outside of the ROOT, RooFit, RooStats framework. pyhf is a pure-Python implementation of that statistical model fo

Uploaded on October 29, 2020

July 17, 2020 (v1) | Presentation | Open Access

Machine learning technique for signal-background separation of nuclear interaction vertices in the CMS detector

Kropivnitskaya, Anna;

The CMS inner tracking system is a fully silicon-based high precision detector. Accurate knowledge of the positions of active and inactive elements is important for simulating the detector, planning detector upgrades, and reconstructing charged particle tracks. Nuclear interactions of hadrons with t

Uploaded on October 28, 2020

July 17, 2020 (v1) | Presentation | Open Access

Machine learning technique for signal-background separation of nuclear interaction vertices in the CMS detector

Kropivnitskaya, Anna;

The CMS inner tracking system is a fully silicon-based high precision detector. Accurate knowledge of the positions of active and inactive elements is important for simulating the detector, planning detector upgrades, and reconstructing charged particle tracks. Nuclear interactions of hadrons with t

Community

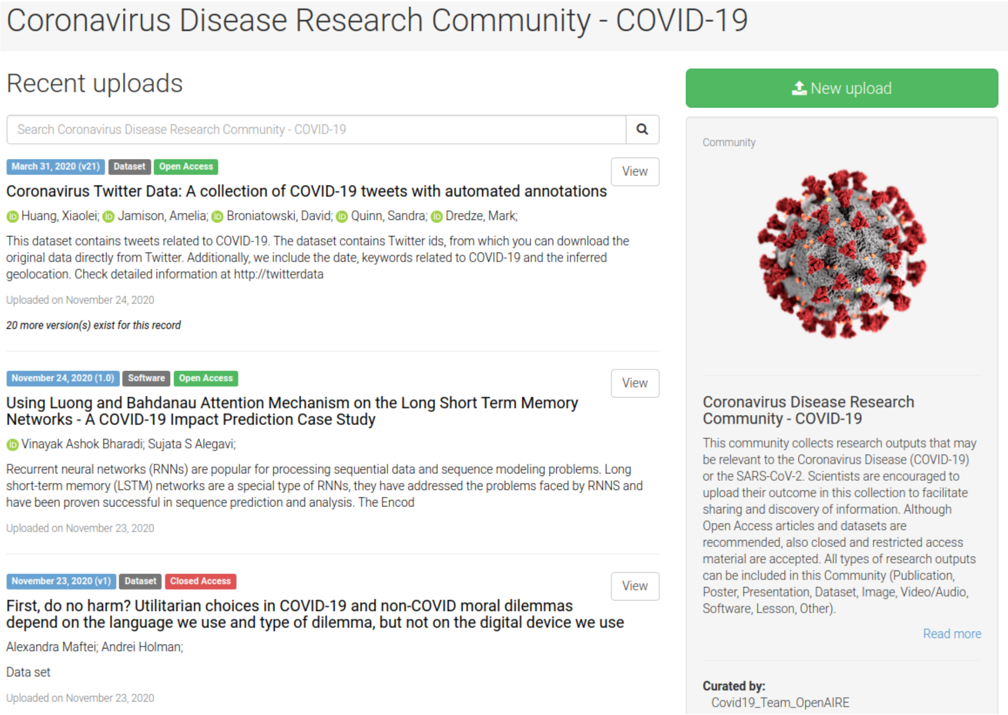
PyHEP 2020 Workshop

"Python in HEP" 2020 Workshop held on July 13-17 2020 as a virtual event.

Agenda: https://indico.cern.ch/e/PyHEP2020.

Curated by:
eduardo-rodrigues
Curation policy:
Only conference participants can post.
Created:
July 24, 2020
Harvesting API:
OAI-PMH Interface

...and finding topical data sets



Coronavirus Disease Research Community - COVID-19

Recent uploads

March 31, 2020 (v21) Dataset Open Access

Coronavirus Twitter Data: A collection of COVID-19 tweets with automated annotations

Huang, Xiaolei; Jamison, Amelia; Broniatowski, David; Quinn, Sandra; Dredze, Mark;

This dataset contains tweets related to COVID-19. The dataset contains Twitter ids, from which you can download the original data directly from Twitter. Additionally, we include the date, keywords related to COVID-19 and the inferred geolocation. Check detailed information at http://twitterdata

Uploaded on November 24, 2020

20 more version(s) exist for this record

November 24, 2020 (1.0) Software Open Access

Using Luong and Bahdanau Attention Mechanism on the Long Short Term Memory Networks - A COVID-19 Impact Prediction Case Study

Vinayak Ashok Bharadi; Sujata S Alegavi;

Recurrent neural networks (RNNs) are popular for processing sequential data and sequence modeling problems. Long short-term memory (LSTM) networks are a special type of RNNs, they have addressed the problems faced by RNNS and have been proven successful in sequence prediction and analysis. The Encod

Uploaded on November 23, 2020

November 23, 2020 (v1) Dataset Closed Access

First, do no harm? Utilitarian choices in COVID-19 and non-COVID moral dilemmas depend on the language we use and type of dilemma, but not on the digital device we use

Alexandra Maftei; Andrei Holman;

Data set

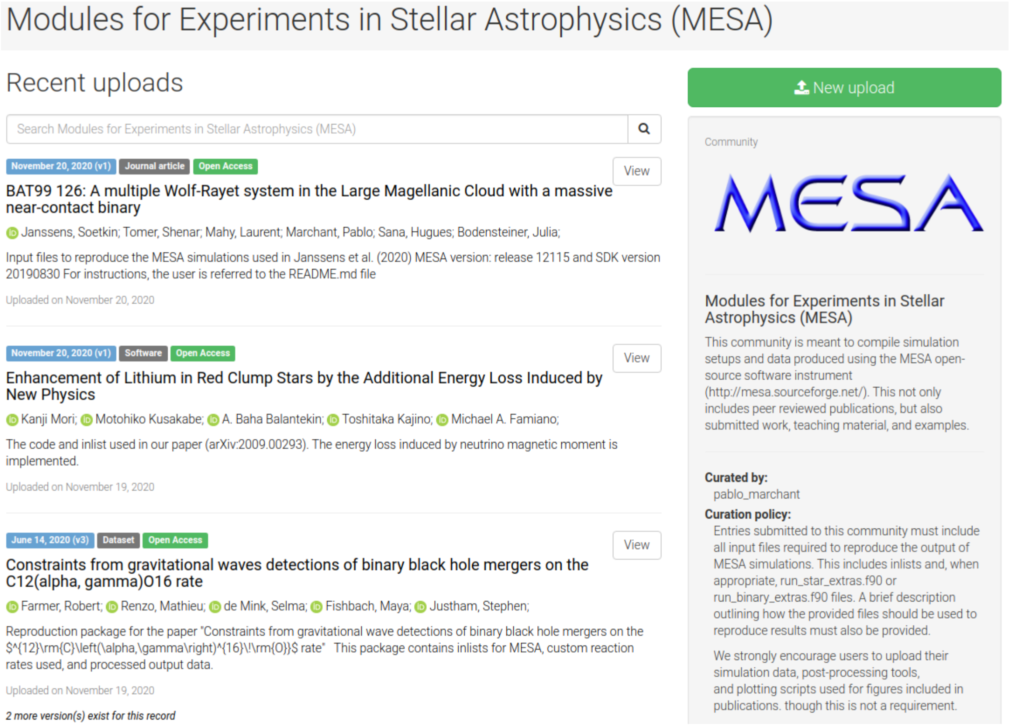
Uploaded on November 23, 2020

Community

Coronavirus Disease Research Community - COVID-19

This community collects research outputs that may be relevant to the Coronavirus Disease (COVID-19) or the SARS-CoV-2. Scientists are encouraged to upload their outcome in this collection to facilitate sharing and discovery of information. Although Open Access articles and datasets are recommended, also closed and restricted access material are accepted. All types of research outputs can be included in this Community (Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, Other).

Curated by:
Covid19_Team_OpenAIRE



Modules for Experiments in Stellar Astrophysics (MESA)

Recent uploads

New upload
Search Modules for Experiments in Stellar Astrophysics (MESA)

November 20, 2020 (v1) Journal article Open Access

BAT99 126: A multiple Wolf-Rayet system in the Large Magellanic Cloud with a massive near-contact binary

Janssens, Soetkin; Tomer, Shenar; Mahy, Laurent; Marchant, Pablo; Sana, Hugues; Bodensteiner, Julia;

Input files to reproduce the MESA simulations used in Janssens et al. (2020) MESA version: release 12115 and SDK version 20190830 For instructions, the user is referred to the README.md file

Uploaded on November 20, 2020

November 20, 2020 (v1) Software Open Access

Enhancement of Lithium in Red Clump Stars by the Additional Energy Loss Induced by New Physics

Kanji Mori; Motohiko Kusakabe; A. Baha Balantekin; Toshitaka Kajino; Michael A. Famiano;

The code and inlist used in our paper (arXiv:2009.00293). The energy loss induced by neutrino magnetic moment is implemented.

Uploaded on November 19, 2020

June 14, 2020 (v3) Dataset Open Access

Constraints from gravitational waves detections of binary black hole mergers on the C12(alpha, gamma)O16 rate

Farmer, Robert; Renzo, Mathieu; de Mink, Selma; Fishbach, Maya; Justham, Stephen;

Reproduction package for the paper "Constraints from gravitational wave detections of binary black hole mergers on the $^{12}\rm{C}\left(\alpha,\gamma\right)^{16}\!\rm{O}}$ rate" This package contains inlists for MESA, custom reaction rates used, and processed output data.

Uploaded on November 19, 2020

2 more version(s) exist for this record

Community

Modules for Experiments in Stellar Astrophysics (MESA)

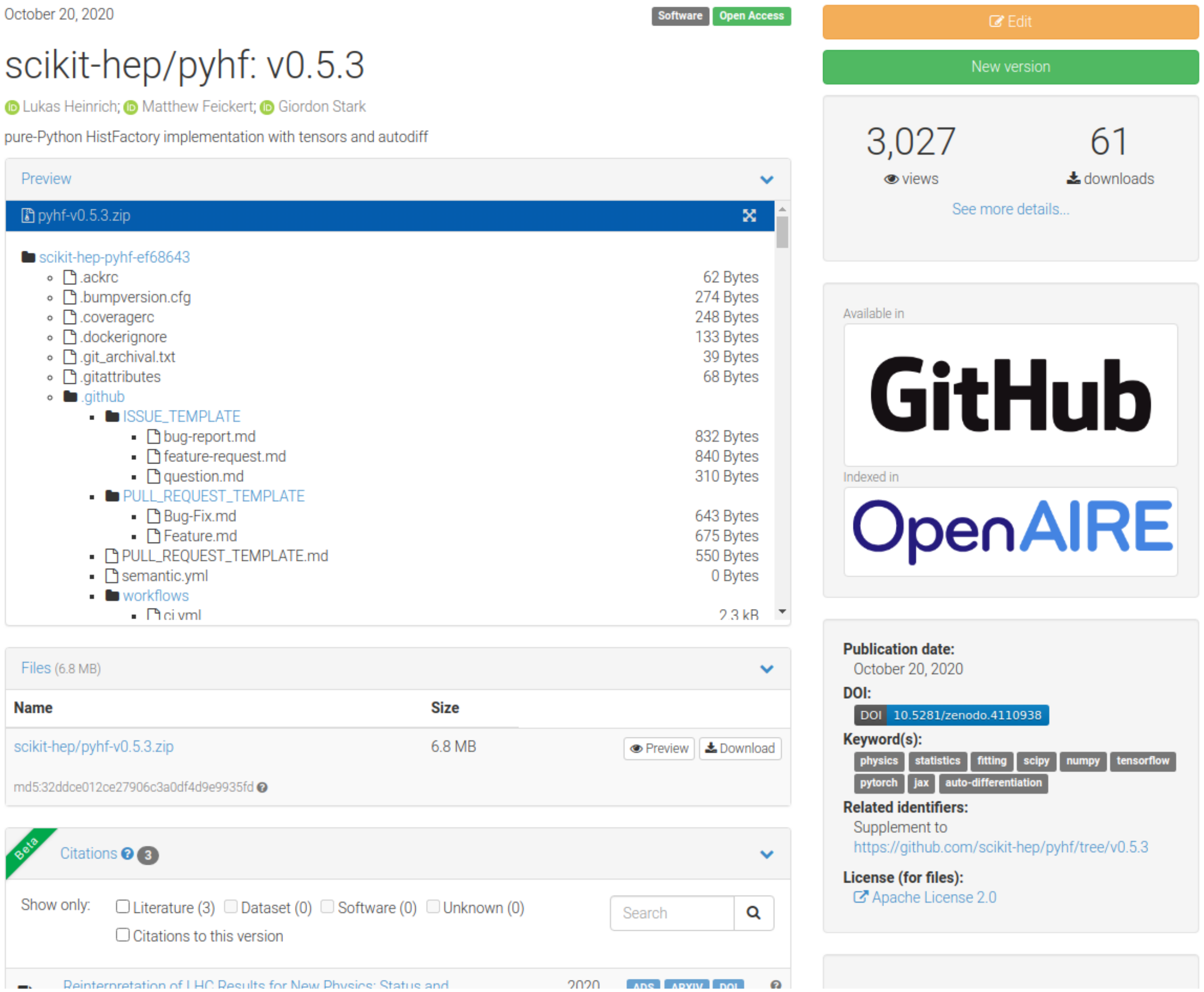
This community is meant to compile simulation setups and data produced using the MESA open-source software instrument (http://mesa.sourceforge.net/). This not only includes peer reviewed publications, but also submitted work, teaching material, and examples.

Curated by:
pablo_marchant

Curation policy:
Entries submitted to this community must include all input files required to reproduce the output of MESA simulations. This includes inlists and, when appropriate, run_star_extras.f90 or run_binary_extras.f90 files. A brief description outlining how the provided files should be used to reproduce results must also be provided.

We strongly encourage users to upload their simulation data, post-processing tools, and plotting scripts used for figures included in publications. though this is not a requirement.

Anatomy of a Zenodo archive



October 20, 2020

Software | Open Access

Edit

New version

scikit-hep/pyhf: v0.5.3

Lukas Heinrich; Matthew Feickert; Giordon Stark

pure-Python HistFactory implementation with tensors and autodiff

3,027 views

61 downloads

See more details...

Preview

pyhf-v0.5.3.zip

```
scikit-hep-pyhf-ef68643
  ∘ .ackrc                                   62 Bytes
  ∘ .bumpversion.cfg                        274 Bytes
  ∘ .coveragerc                             248 Bytes
  ∘ .dockerignore                           133 Bytes
  ∘ .git_archival.txt                        39 Bytes
  ∘ .gitattributes                           68 Bytes
  ∘ .github
      ▪ ISSUE_TEMPLATE
          ▪ bug-report.md                   832 Bytes
          ▪ feature-request.md              840 Bytes
          ▪ question.md                     310 Bytes
      ▪ PULL_REQUEST_TEMPLATE
          ▪ Bug-Fix.md                      643 Bytes
          ▪ Feature.md                      675 Bytes
      ▪ PULL_REQUEST_TEMPLATE.md            550 Bytes
      ▪ semantic.yml                          0 Bytes
      ▪ workflows
          ▪ ci.yml                           2.3 kB
```

Available in

GitHub

Indexed in

OpenAIRE

Files (6.8 MB)

Name	Size	
scikit-hep/pyhf-v0.5.3.zip md5:32ddce012ce27906c3a0df4d9e9935fd	6.8 MB	Preview Download

Publication date:
October 20, 2020

DOI:
DOI 10.5281/zenodo.4110938

Keyword(s):
physics statistics fitting scipy numpy tensorflow pytorch jax auto-differentiation

Related identifiers:
Supplement to
https://github.com/scikit-hep/pyhf/tree/v0.5.3

License (for files):
Apache License 2.0

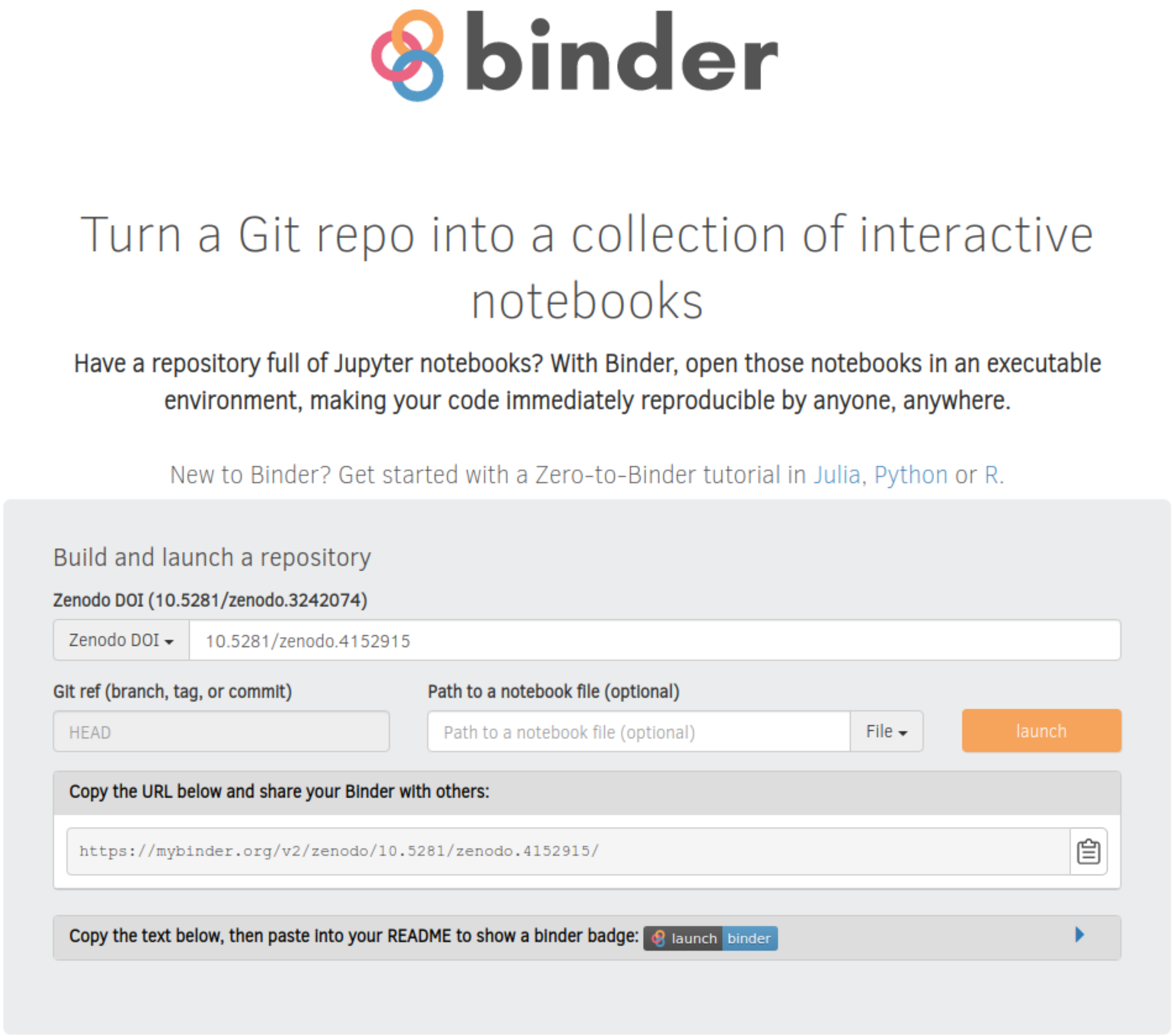
Beta Citations 3

Show only: □ Literature (3) □ Dataset (0) □ Software (0) □ Unknown (0)
□ Citations to this version

Search

Reinterpretation of LHC Results for New Physics: Status and 2020 ADS ARXIV DOI

Zenodo DOI → reproducible examples



binder

Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

New to Binder? Get started with a Zero-to-Binder tutorial in Julia, Python or R.

Build and launch a repository

Zenodo DOI (10.5281/zenodo.3242074)

Zenodo DOI: 10.5281/zenodo.4152915

Git ref (branch, tag, or commit): HEAD

Path to a notebook file (optional): Path to a notebook file (optional)

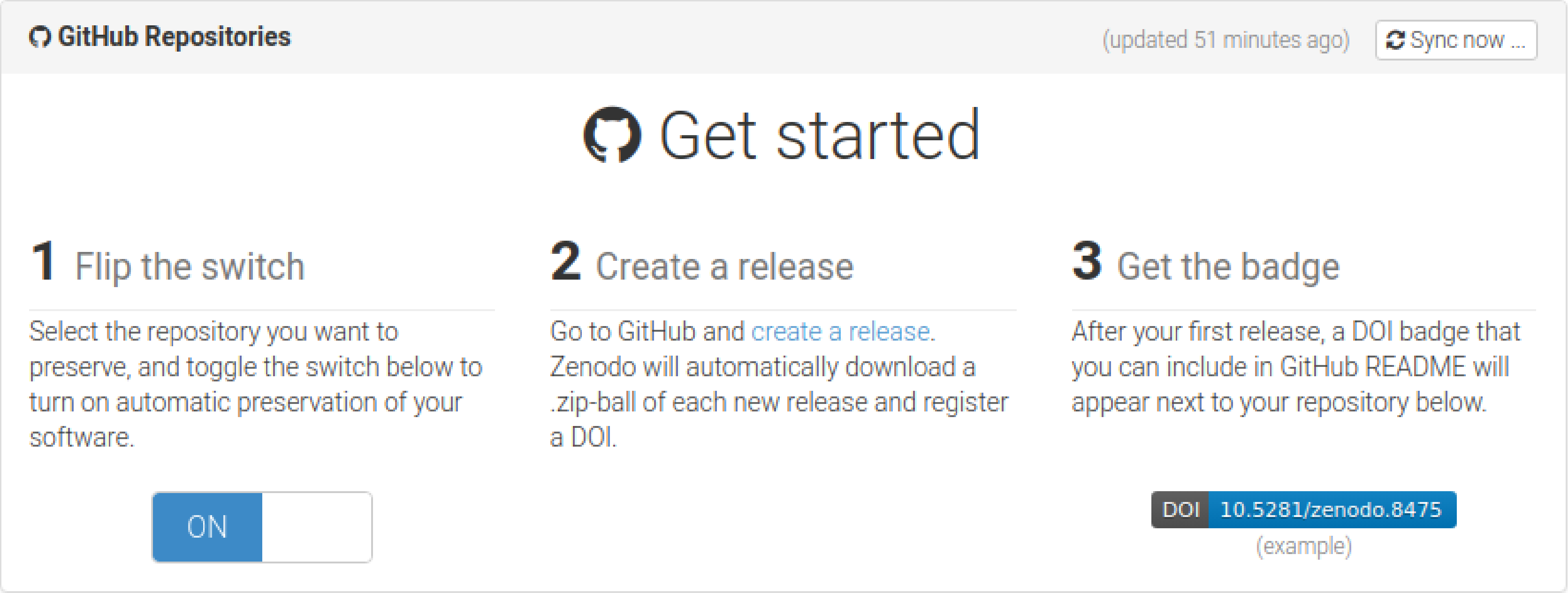
Copy the URL below and share your Binder with others:

```
https://mybinder.org/v2/zenodo/10.5281/zenodo.4152915/
```

Copy the text below, then paste into your README to show a binder badge:

DOI minting made easy

- Everything on Zenodo has a DOI
- You can upload directly to Zenodo or can enable it to automatically preserve work from GitHub
- We're going to do so now



Get started

1 Flip the switch

Select the repository you want to preserve, and toggle the switch below to turn on automatic preservation of your software.

2 Create a release

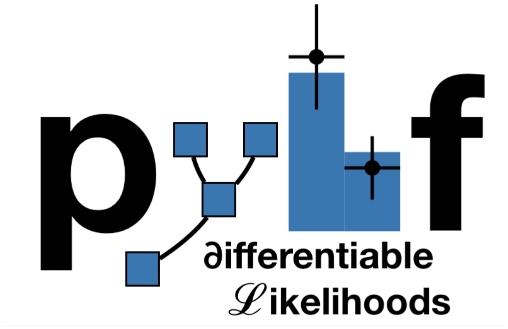
Go to GitHub and create a release. Zenodo will automatically download a .zip-ball of each new release and register a DOI.

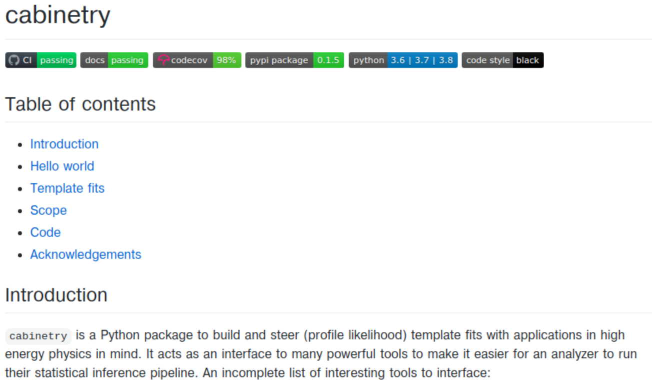
3 Get the badge

After your first release, a DOI badge that you can include in GitHub README will appear next to your repository below.

Open data creates space for open ecosystems

All these tools are able to make use of public likelihoods





cabinetry

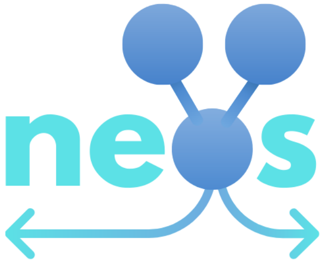
CI passing | docs passing | codecov 98% | pypi package 0.1.5 | python 3.6 | 3.7 | 3.8 | code style black

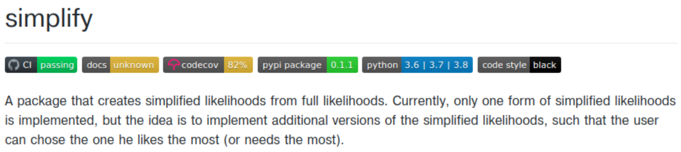
Table of contents

- Introduction
- Hello world
- Template fits
- Scope
- Code
- Acknowledgements

Introduction

`cabinetry` is a Python package to build and steer (profile likelihood) template fits with applications in high energy physics in mind. It acts as an interface to many powerful tools to make it easier for an analyzer to run their statistical inference pipeline. An incomplete list of interesting tools to interface:





simplify

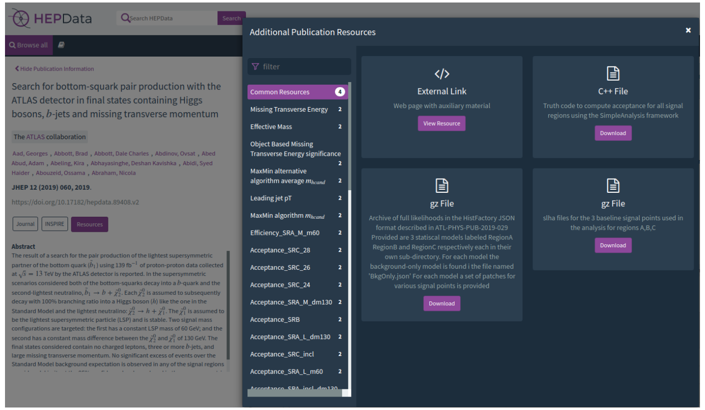
CI passing | docs unknown | codecov 82% | pypi package 0.1.1 | python 3.6 | 3.7 | 3.8 | code style black

A package that creates simplified likelihoods from full likelihoods. Currently, only one form of simplified likelihoods is implemented, but the idea is to implement additional versions of the simplified likelihoods, such that the user can chose the one he likes the most (or needs the most).

FAIR data products are able to expand and accelerate research and tooling

Summary

- Application of **F**indable, **A**ccessible, **I**nteroperable, and **R**euseable data principles extends the usefulness of data far beyond the publication
  - Case study of FAIR open likelihoods increasing the communication and sharing of data products between HEP experiment and theory
  - Expanding scope of existing work and creating new
- Democratizing process
  - Data products in open common formats ensures accessible to everyone, regardless of resources
  - Breeds best practices and open ecosystem growth
- Existing infrastructure makes FAIR data practices easy and smart to do
  - Archival, storage hosting, DOI/citation



Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum

The ATLAS collaboration

Aad, Georges , Abbott, Brad , Abbott, Dale Charles , Abdinov, Ovsat , Abed Abud, Adam , Abeling, Kira , Abhayasinghe, Deshan Kavishka , Abidi, Syed Haider , Abouzeid, Ossama , Abraham, Nicola

JHEP 12 (2019) 060, 2019.

https://doi.org/10.17182/hepdata.89408.v2

Abstract

The result of a search for the pair production of the lightest supersymmetric partner of the bottom quark ($\tilde{b}_1$) using 139 fb^{-1} of proton-proton data collected at $\sqrt{s} = 13$ TeV by the ATLAS detector is reported. In the supersymmetric scenarios considered both of the bottom-squarks decay into a b-quark and the second-lightest neutralino, $\tilde{b}_1 \to b + \tilde{\chi}_2^0$. Each $\tilde{\chi}_2^0$ is assumed to subsequently decay with 100% branching ratio into a Higgs boson (h) like the one in the Standard Model and the lightest neutralino: $\tilde{\chi}_2^0 \to h + \tilde{\chi}_1^0$. The $\tilde{\chi}_1^0$ is assumed to be the lightest supersymmetric particle (LSP) and is stable. Two signal mass configurations are targeted: the first has a constant LSP mass of 60 GeV; and the second has a constant mass difference between the $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^0$ of 130 GeV. The final states considered contain no charged leptons, three or more b-jets, and large missing transverse momentum. No significant excess of events over the Standard Model background expectation is observed in any of the signal regions

External Link

Web page with auxiliary material

C++ File

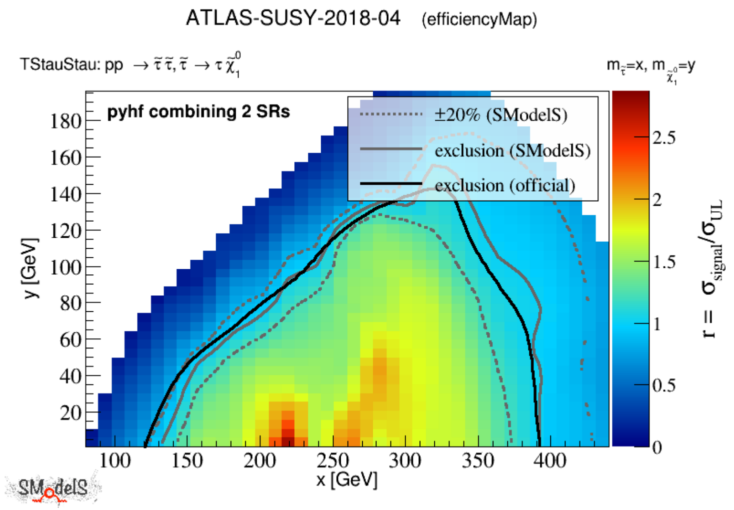
Truth code to compute acceptance for all signal regions using the SimpleAnalysis framework

gz File

Archive of full likelihoods in the HistFactory JSON format described in ATL-PHYS-PUB-2019-029 Provided are 3 statiscal models labeled RegionA RegionB and RegionC respectively each in their own sub-directory. For each model the background-only model is found i the file named 'BkgOnly.json' For each model a set of patches for various signal points is provided

gz File

slha files for the 3 baseline signal points used in the analysis for regions A,B,C



Backup

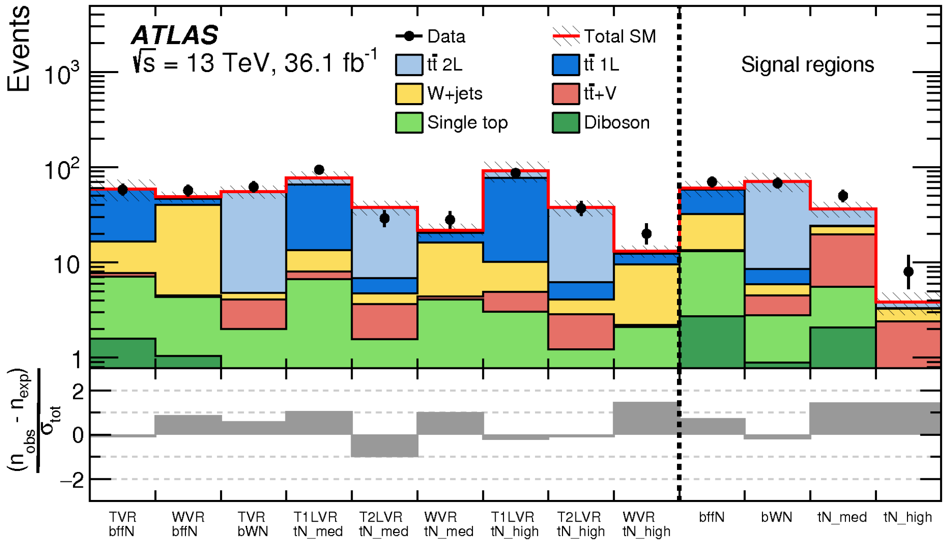
HistFactory Template

$$f\left(\mathrm{data}|\mathrm{parameters}\right) = f\left(\vec{n}, \vec{a}|\vec{\eta}, \vec{\chi}\right) = \prod_{c\,\in\,\mathrm{channels}} \prod_{b\,\in\,\mathrm{bins}_c} \mathrm{Pois}\left(n_{cb}|\nu_{cb}\left(\vec{\eta}, \vec{\chi}\right)\right) \prod_{\chi\,\in\,\vec{\chi}} c_{\chi}\left(a_{\chi}|\chi\right)$$

Use: Multiple disjoint channels (or regions) of binned distributions with multiple samples contributing to each with additional (possibly shared) systematics between sample estimates

Main pieces:

- Main Poisson p.d.f. for simultaneous measurement of multiple channels
- Event rates ν_{cb} (nominal rate ν_{scb}^{0} with rate modifiers)
- Constraint p.d.f. (+ data) for "auxiliary measurements"
  - encode systematic uncertainties (e.g. normalization, shape)
- $\vec{n}$: events, $\vec{a}$: auxiliary data, $\vec{\eta}$: unconstrained pars, $\vec{\chi}$: constrained pars



Example: **Each bin** is separate (1-bin) channel, each **histogram** (color) is a sample and share a **normalization systematic** uncertainty

HistFactory Template

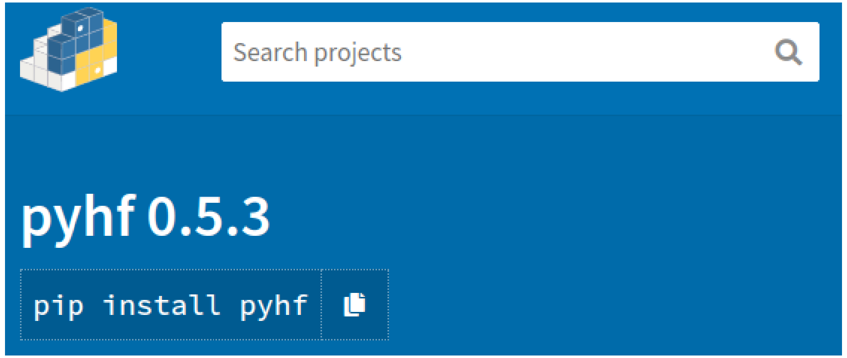
$$f\left(\vec{n},\vec{a}|\vec{\eta},\vec{\chi}\right) = \prod_{c\,\in\,\text{channels}} \prod_{b\,\in\,\text{bins}_c} \text{Pois}\left(n_{cb}|\nu_{cb}\left(\vec{\eta},\vec{\chi}\right)\right) \prod_{\chi\,\in\,\vec{\chi}} c_{\chi}\left(a_{\chi}|\chi\right)$$

Mathematical grammar for a simultaneous fit with

- multiple "channels" (analysis regions, (stacks of) histograms)
- each region can have multiple bins
- coupled to a set of constraint terms

This is a mathematical representation! Nowhere is any software spec defined
Until now (2018), the only implementation of HistFactory has been in `ROOT`

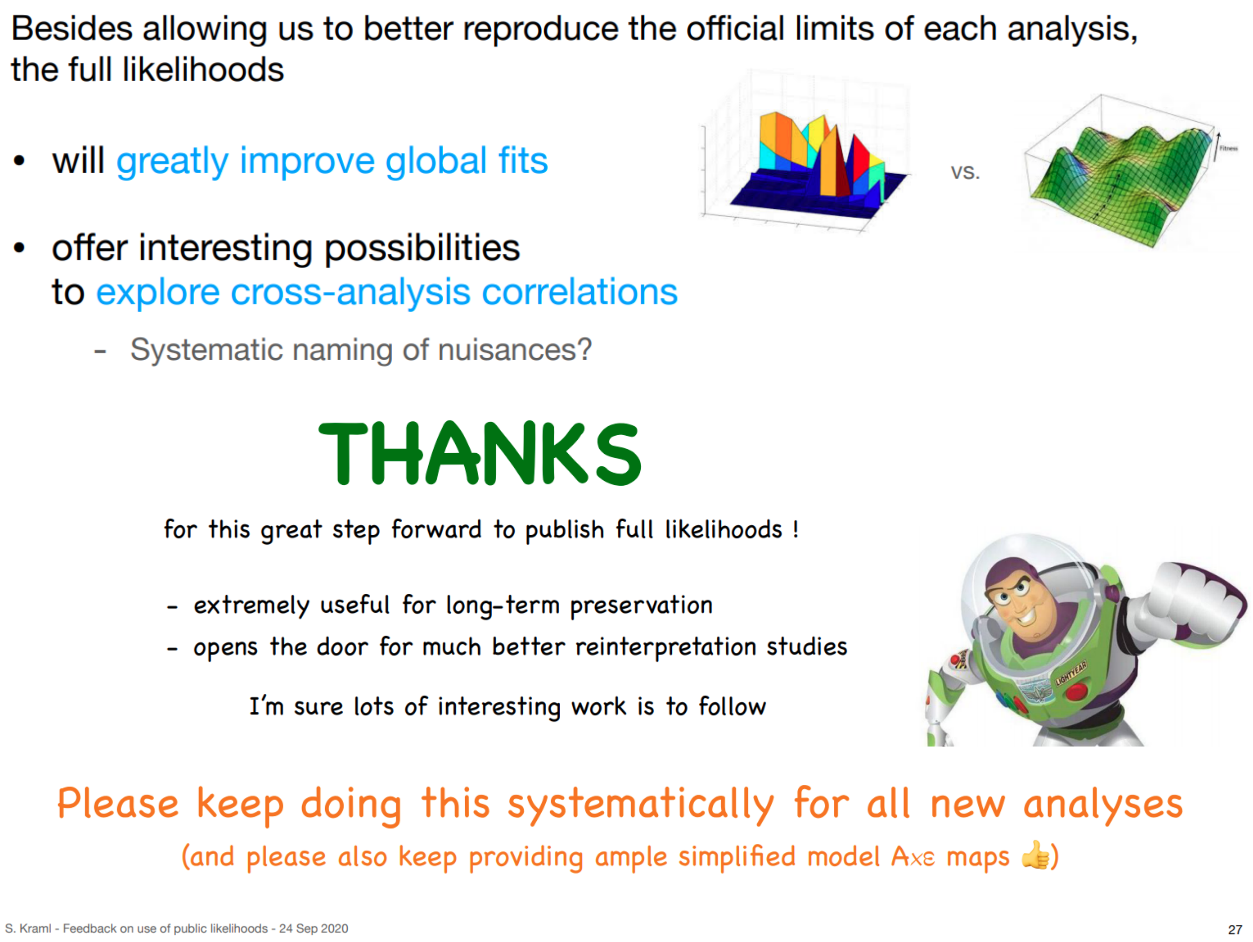
`pyhf`: HistFactory in pure Python



pyhf 0.5.3

```
pip install pyhf
```

Full likelihood plans and ideas



Besides allowing us to better reproduce the official limits of each analysis, the full likelihoods

- will greatly improve global fits
- offer interesting possibilities to explore cross-analysis correlations
  - Systematic naming of nuisances?

THANKS

for this great step forward to publish full likelihoods !

- extremely useful for long-term preservation
- opens the door for much better reinterpretation studies

I'm sure lots of interesting work is to follow

Please keep doing this systematically for all new analyses

(and please also keep providing ample simplified model A×ε maps 👍)

S. Kraml - Feedback on use of public likelihoods - 24 Sep 2020 27

Feedback on use of public Likelihoods, Sabine Kraml

What is pyhf?

Please checkout the many resources we have starting with the website and the SciPy 2020 talk!

SciPy 2020

Scientific Computing with Python
Virtual Conference • July 6-12, 2020

Matthew Feickert
A Pure Python Statistical Fitting Library with Tensors and Autograd

References

1. F. James, Y. Perrin, L. Lyons, *Workshop on confidence limits: Proceedings*, 2000.
2. ROOT collaboration, K. Cranmer, G. Lewis, L. Moneta, A. Shibata and W. Verkerke, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, 2012.
3. L. Heinrich, H. Schulz, J. Turner and Y. Zhou, *Constraining A_4 Leptonic Flavour Model Parameters at Colliders and Beyond*, 2018.
4. A. Read, *Modified frequentist analysis of search results (the CL_s method)*, 2000.
5. K. Cranmer, *CERN Latin-American School of High-Energy Physics: Statistics for Particle Physicists*, 2013.
6. ATLAS collaboration, *Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum*, 2019
7. ATLAS collaboration, *Reproducing searches for new physics with the ATLAS experiment through publication of full statistical likelihoods*, 2019
8. ATLAS collaboration, *Search for bottom-squark pair production with the ATLAS detector in final states containing Higgs bosons, b-jets and missing transverse momentum: HEPData entry*, 2019

The end.