



Fair Universe 2024: HiggsML Uncertainty Challenge

David Rousseau,
IJCLab-Orsay

with Paolo Calafiura, Ragansu Chakkappai, Yuan-Tang Chou, Sascha Diefenbacher, Steven Farrell, Aishik Ghosh, Isabelle Guyon, Chris Harris, Shih-Chieh Hsu, Elham Khoda, Benjamin Nachman, Benjamin Thorne, Peter Nugent, Mathis Reymond, David Rousseau, Ihsan Ullah, Daniel Whiteson

EUCaifCon, Amsterdam, May 2024



Background on Fair Universe Project

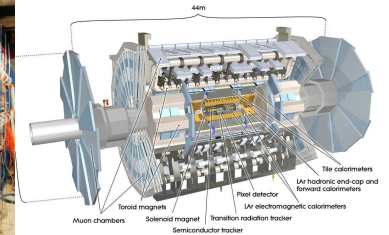
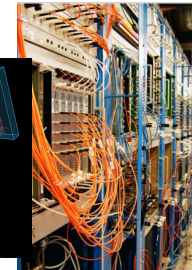
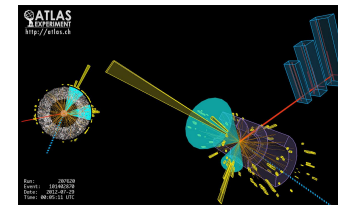
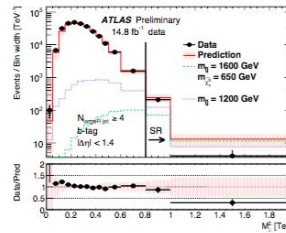
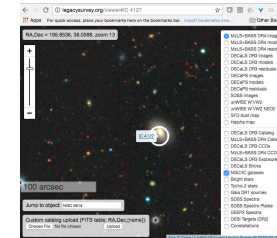
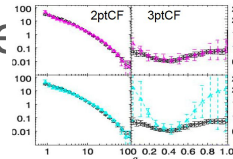
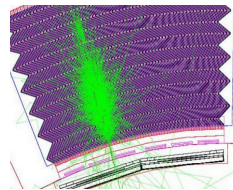
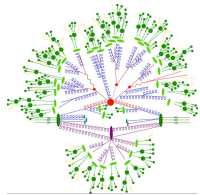
- 3 year US Dept. of Energy, AI for HEP project. Aims to:
 - Provide an open, **large-compute-scale AI ecosystem** for sharing datasets, training large models, fine-tuning those models, and **hosting challenges and benchmarks**.
 - **Organize a challenge series**, progressively rolling in tasks of increasing difficulty, based on novel datasets.
 - Tasks will focus on **measuring and minimizing the effects of systematic uncertainties** in HEP (particle physics and cosmology).
- Broad team in HEP, ML and computing involved in several previous challenges and benchmarks for HEP (e.g. [HiggsML](#) and [TrackML](#), [LHC Olympics](#), [Fast Calorimeter Simulation Challenge](#)) and wider (e.g. [NeurIPS competition track](#), [MLPerf HPC](#)); as well as [Uncertainty aware learning in HEP](#)

Measuring and minimizing the effects of systematic uncertainties in HEP

Bias and uncertainty in the fundamental sciences

Theory into Simulations

- Estimate Systematic Uncertainty



Exp/Obs reconstruction

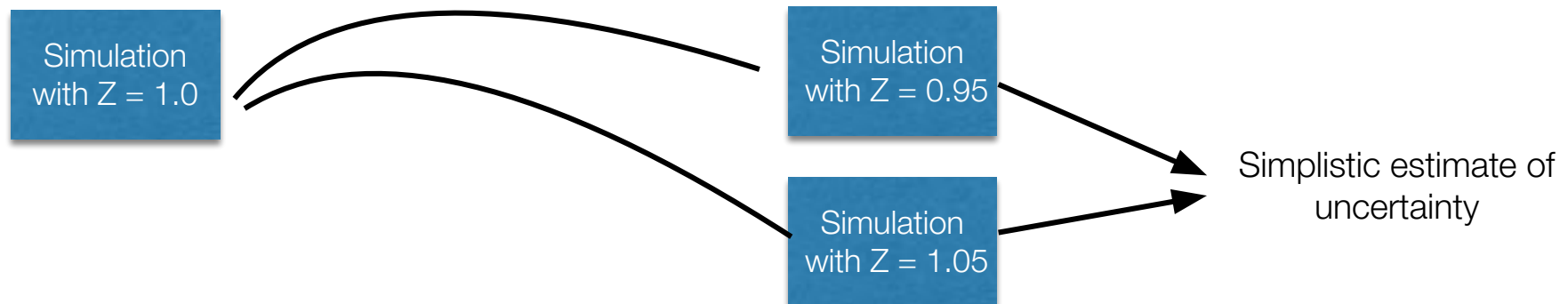
Differences between simulation and data can bias measurements

Bias and uncertainty in ML in the fundamental sciences

- ML methods in HEP are often trained based on simulation which has estimated systematic uncertainties (“Z”)
- These are then applied in data with the different detector state $Z=?$

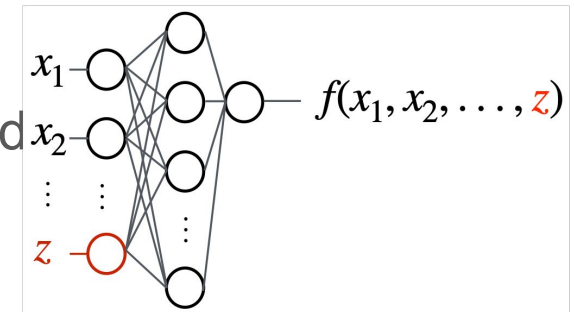


- Common baseline approach: Train classifier on nominal data (e.g. $Z=1$) and estimate uncertainties with alternate simulations. Shift Z and look at impact or perform full profile likelihood



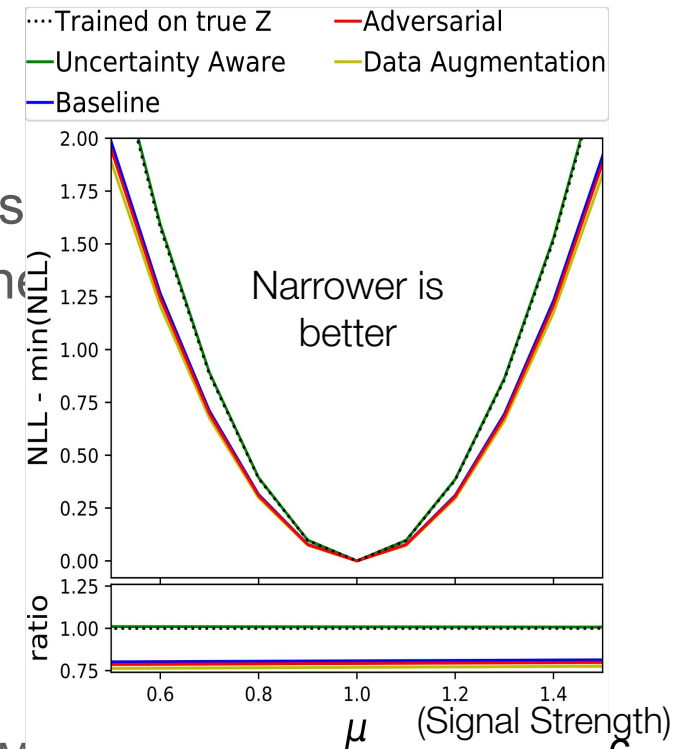
Increasingly sophisticated approaches

- Several focussed on decorrelation, e.g. augmentation; adversarial training; tangent propagation etc.
- “pivot” Louppe, Kagan, Cranmer : [arXiv:1611.01046](https://arxiv.org/abs/1611.01046)
- “Uncertainty-aware” approach of Ghosh, Nachman, Whiteson



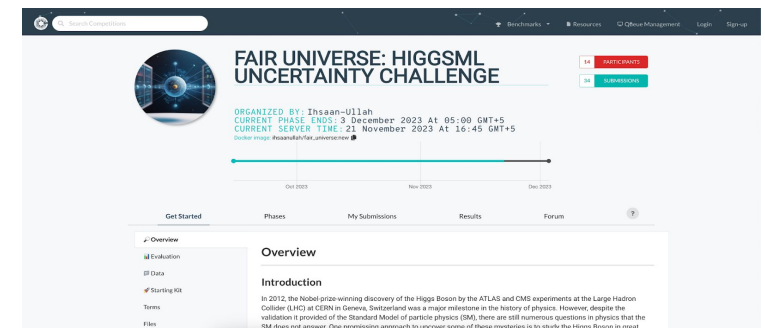
PhysRevD.104.056026

- Parameterize classifier using Z
- Measured on “Toy” 2D Gaussian Dataset and dataset Challenge modified to include systematic on tau-energy
- Performs as well as classifier trained on true Z
- Other novel approaches e.g. (not comprehensive)
 - Inferno: [arxiv:1806.04743](https://arxiv.org/abs/1806.04743)
 - Direct profile-likelihood: e.g. [arxiv:2203.13079](https://arxiv.org/abs/2203.13079)



Progress requires new datasets, metrics, and platform

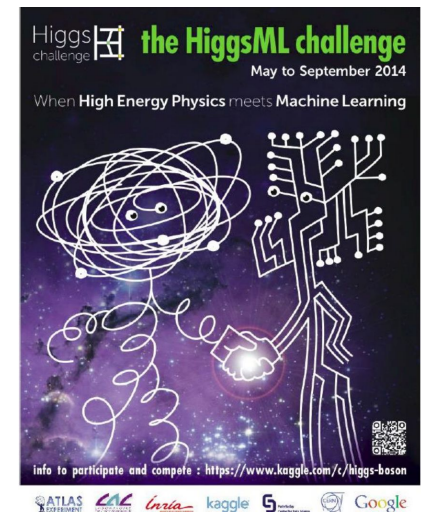
- Uncertainty-aware papers demonstrated on single systematic uncertainty, with limited data
- Original HiggsML dataset too small for ambitious approaches (systematic uncertainty small compared to statistical uncertainty)
- How to scale methods to many values of Z ? (training difficulty increases, profiling approach used is expensive)
- Can faster methods allow for directly evaluating profile likelihood?
- Need for novel metrics to evaluate uncertainty determination for such methods



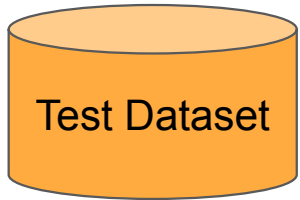
Organize a challenge series,
progressively rolling in tasks of
increasing difficulty, based on novel
datasets

Fair Universe: HiggsML Uncertainty Challenge

- Extension of previous HiggsML challenge from 2014 (a classification problem for Higgs decaying to Tau leptons based on final state 3-momenta and derived quantities):
 - 30 features : l,h,MissingET,up to 2 jets, and high level quantities like transverse masses
- Dataset : HiggsML 2014 data set on CERN Open Data portal [doi:10.7483/OPENDATA.ATLAS.ZBP2.M5T8](https://doi.org/10.7483/OPENDATA.ATLAS.ZBP2.M5T8)
- ⇒new Fair Universe dataset, with following improvements
 - Instead of ATLAS G4 simulation, use Pythia LO + Delphes
 - Numbers of events 800.000⇒ >50 millions
 - Parametrised systematics (Nuisance Parameters) :
 - Tau Energy Scale : on had Tau Pt (and correlated MET)
 - Jet Energy Scale (and correlated MET impact)
 - additional randomised Soft MET
 - background normalisation
 - W background normalisation (a subdominant poorly constrained BKG)
- Task : given a pseudo-experiment with given signal strength (==amount of signal), provide a Confidence Interval on signal strength



Pseudo-experiments



Test Dataset

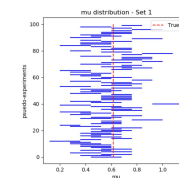
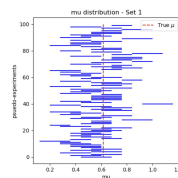
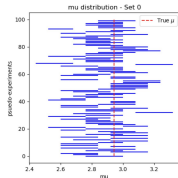
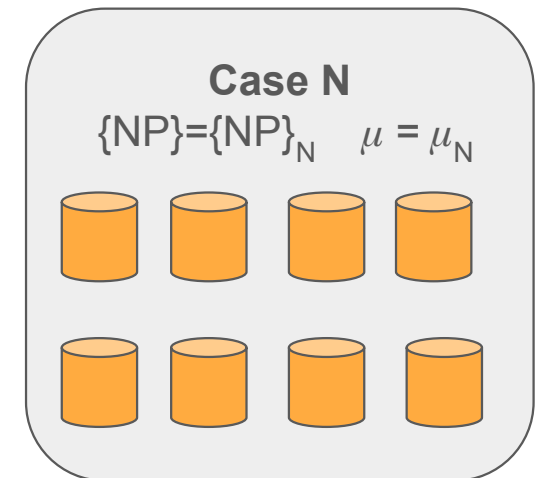
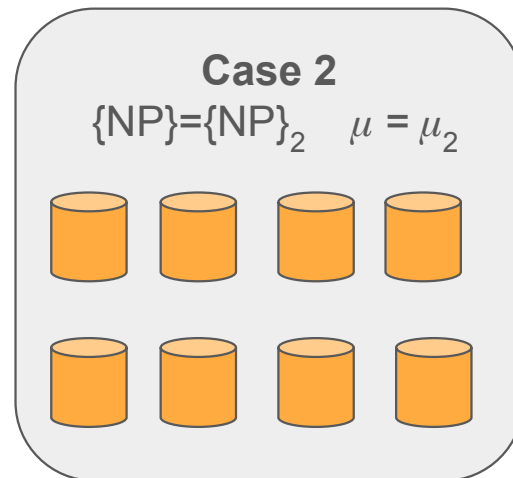
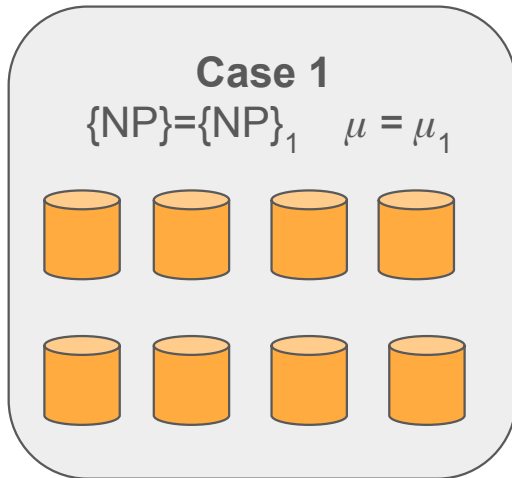
$\{\mathbf{x}, \mathbf{w}\}$

Create pseudo experiments for each test cases

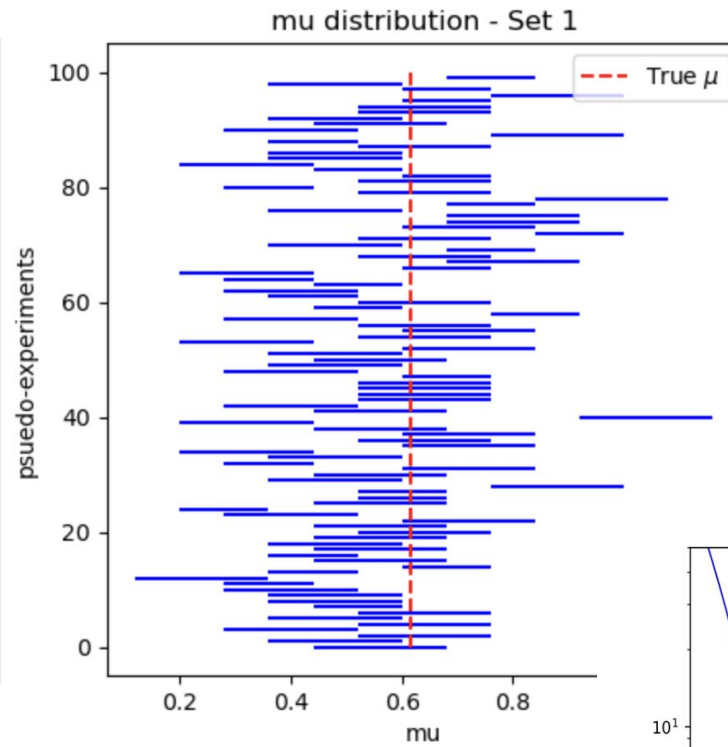
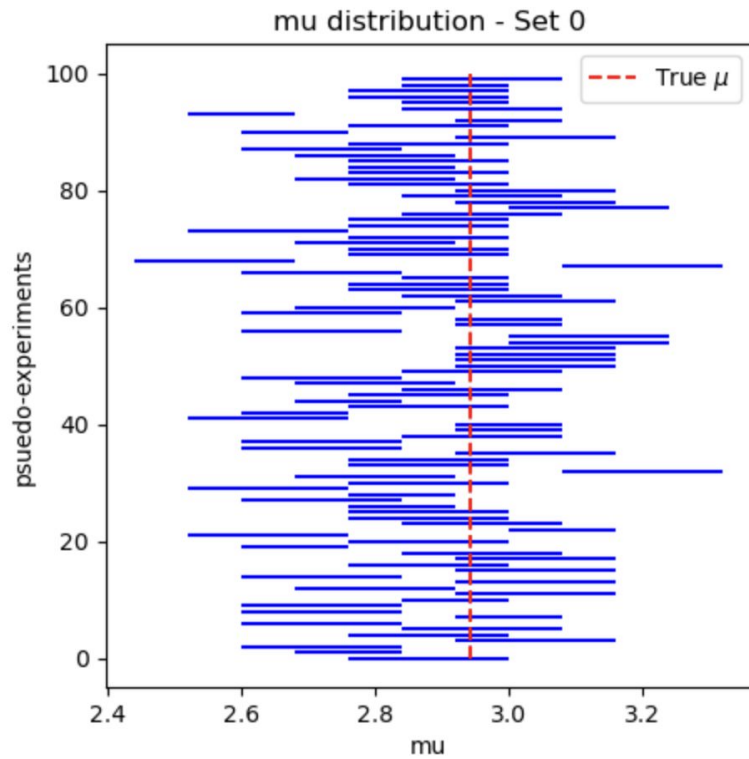
$\{\mathbf{x}, \mathbf{w}_1\}, w_1 = \text{Pois}(w)$

$\{\mathbf{x}, \mathbf{w}_2\}, w_2 = \text{Pois}(w)$

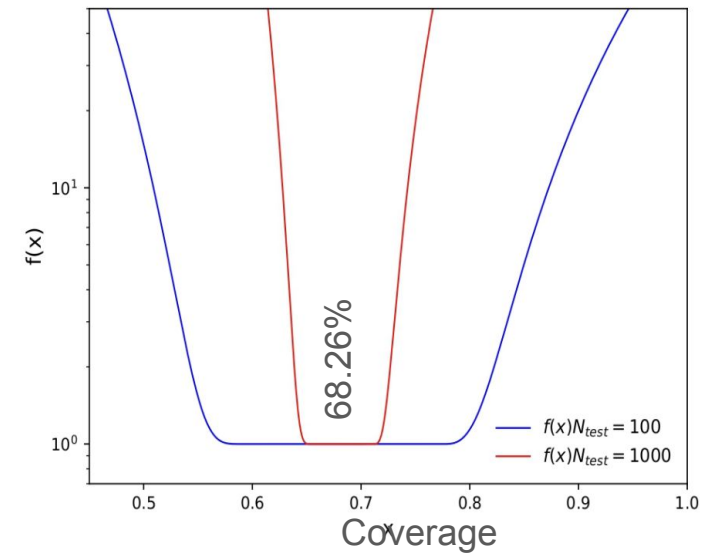
$\{\mathbf{x}, \mathbf{w}_N\}, w_N = \text{Pois}(w)$



Coverage evaluation



Coverage penalisation function






















score : $\langle \text{CI length} \rangle \times \text{coverage penalisation}$

Fair Universe, David Rousseau, EUCaifCon, Amste

sd

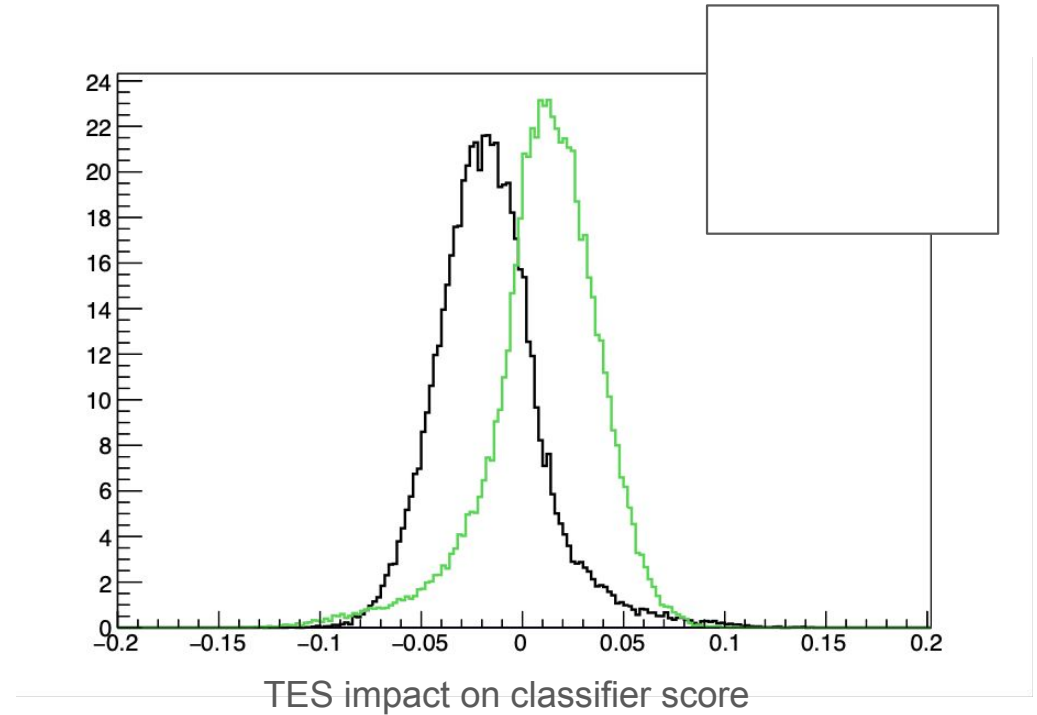
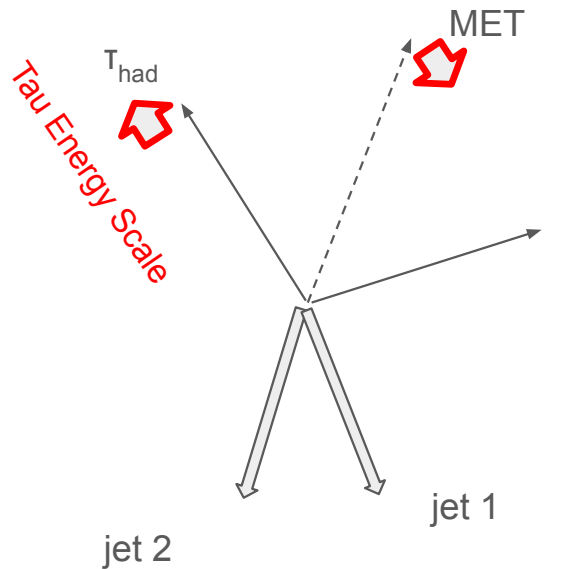
sd

Task:				Fact Sheet Answers	Higgs Uncertainty Challenge			
#	Participant	Entries	Date of last entry	Method Name	Quantile Score	Interval	Coverage	Detailed Results
	ragansu	30	2024-01-22	Histogram_10	1.45	0.226	0.57	
	ragansu	30	2024-01-22	One_bin NLL	1.07	0.333	0.57	
	laurensslu	20	2023-12-01	cheat7	0.68	0.504	0.63	
	laurensslu	20	2023-12-01	cheat7	0.61	0.544	0.68	
	laurensslu	20	2023-12-01	cheat4	0.31	0.732	0.61	
6	laurensslu	20	2023-12-01	cheat4	0.16	0.852	0.71	
7	laurensslu	20	2023-12-01	Cheat2	-0.44	1.55	0.62	
8	laurensslu	20	2023-12-01	Cheat2	-0.74	1.375	0.55	
9	ragansu	30	2024-01-22	tes_finder	-0.95	1.124	0.54	
10	laurensslu	20	2023-12-01	Cheat2	-1.59	1.325	0.53	
11	Ihsan Ullah	4	2024-01-18	Sascha sys aware 8	-2.69	0.329	0.47	
12	Rafał Masełek	10	2023-12-01	1binNLL	-2.9	1.233	0.5	
13	ihsanchalearn	16	2023-12-18	1 bin NLL	-2.9	1.233	0.5	
14	Rafał Masełek	10	2023-12-01	1binNLL	-2.9	1.233	0.5	
15	ihsanchalearn	16	2023-12-18	Sascha sys aware 8	-3.01	0.33	0.46	

Systematic parameterisation

Systematic parameterisation

- provide to participant a systematic parameterisation script
- recompute all features and weight consistently
- e.g. for Tau Energy Scale

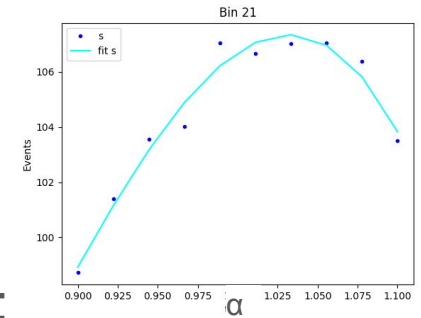


A simple Baseline Model

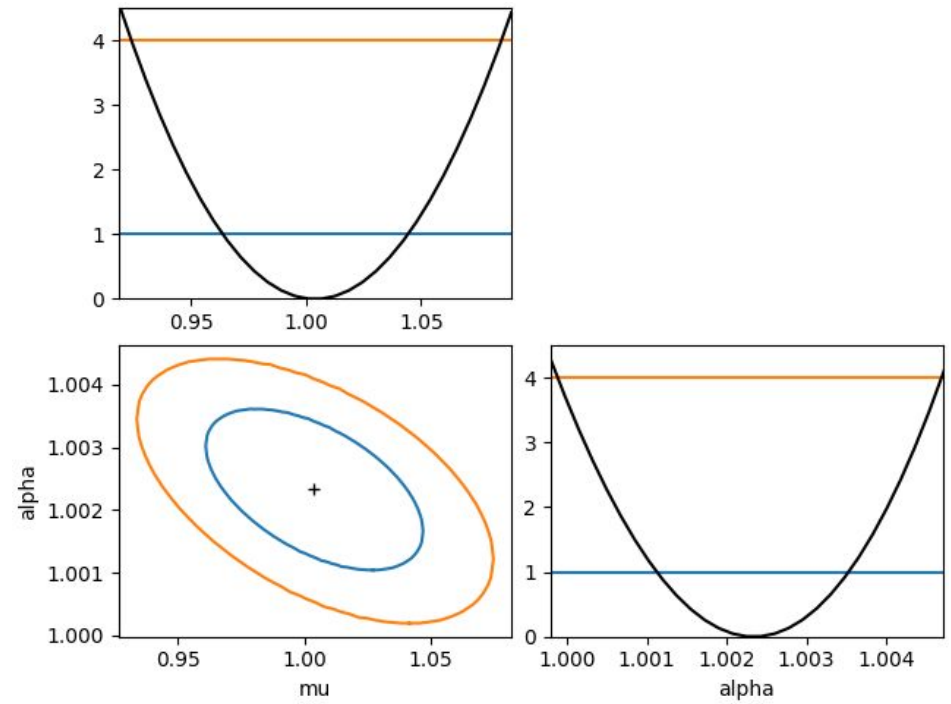
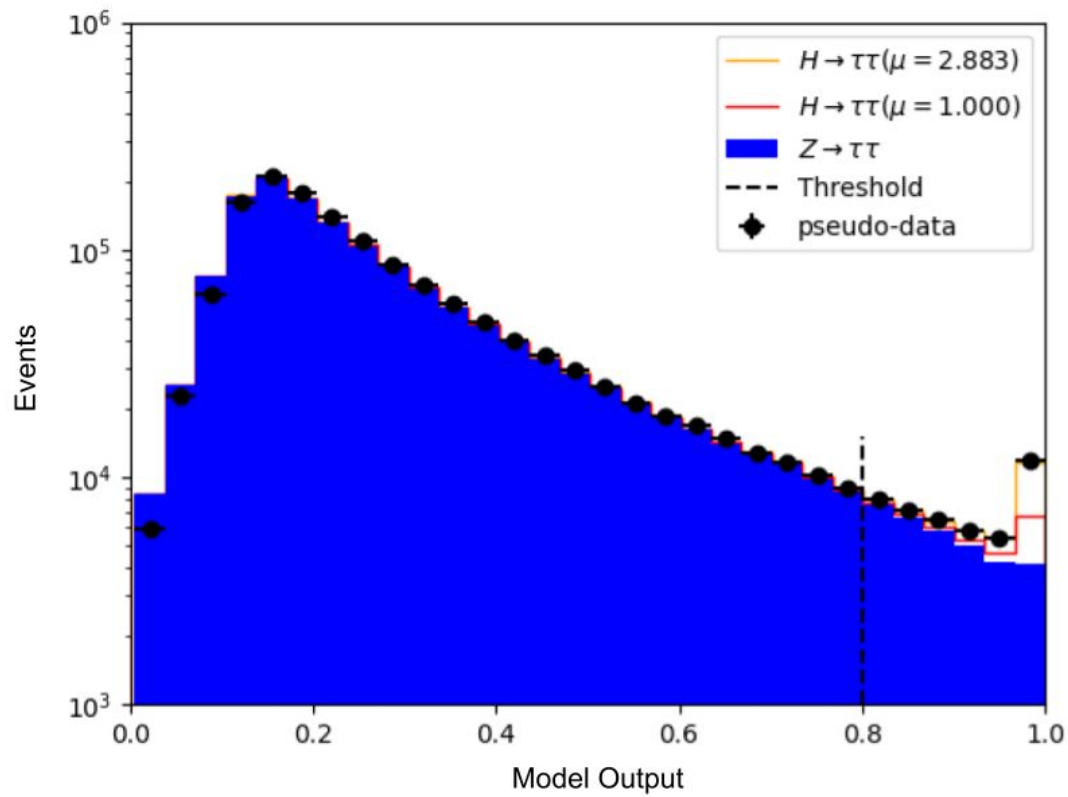
...to get people started

Baseline Model

- Train a simple classifier (dense NN, BDT,...)
- Bin the score for Background and Signal : B_i, S_i
- Parameterise as a function of nuisance parameter α $B_i(\alpha), S_i(\alpha)$:
- Binned Negative Log Likelihood function as function of NP and μ \Rightarrow
- For Each pseudo experiment
 - a. compute score distribution
 - b. NLL regress μ (and α but throw it away)
 - c. Returns μ 68% confidence interval
- works nicely for 1 NP, breaks down for more
- ...people can take it from there with the more sophisticated uncertainty-aware approaches mentioned earlier



Fit on one pseudo experiment



Large-compute-scale AI ecosystem ...
hosting challenges and benchmarks.

We'll not be on Kaggle

The screenshot shows the Codabench website interface. At the top, there is a search bar for competitions and a navigation menu with links for Benchmarks, Resources, Queue Management, Login, and Sign-up. The main content area features an announcement box, two columns of benchmark cards (Popular and Featured), and a 'Get Started' section with three icons: Participate, Organize, and Contribute.

Announcement

Welcome to Codabench!
Join the [Google group](#) to connect with the community!

Popular Benchmarks

Track 1: Pedestrian Attribute Recognition - WACV'24 As a part of the WACV'2024 Pedestrian Attribute Recognition and Person Retrieval Challenge... <i>Organized by: julioj</i>	September 9, 2023 50 Participants
AutoML Cup Phase 1 AutoML Cup Phase 1 <i>Organized by: automlcup</i>	June 6, 2023 49 Participants
(ended) Auto-Survey Challenge'23 Auto-Survey Challenge'23 <i>Organized by: fnachalearn</i>	July 7, 2023 36 Participants
SNAKE #1 SaNitization Algorithms under attack <i>Organized by: louisbeziaud</i>	May 25, 2023 22 Participants

Featured Benchmarks

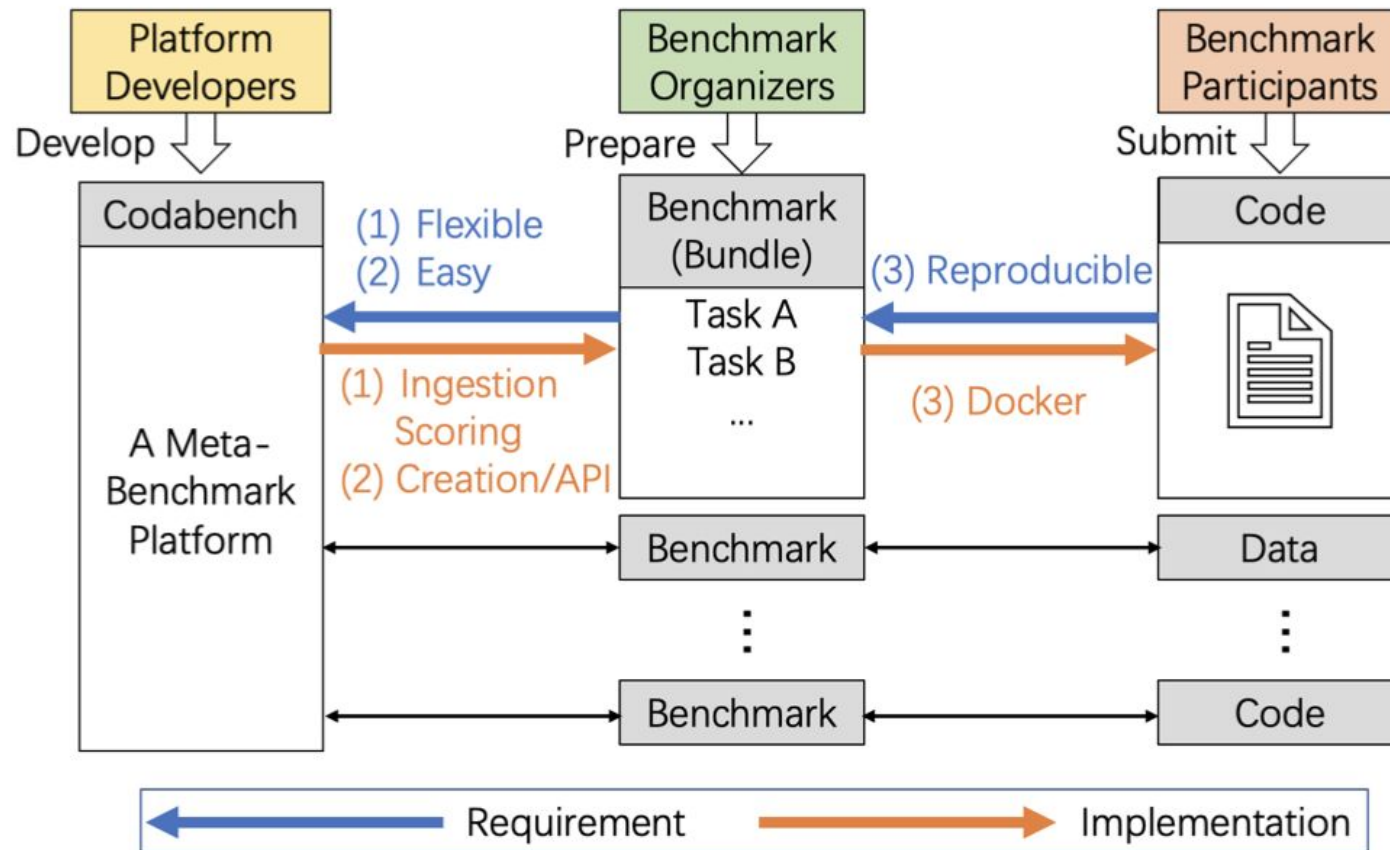
WACV 2024 - Grand Challenge - AV-Sync Error Measurement Regression prediction of temporal offset between audio and video <i>Organized by: highamdh</i>	October 11, 2023 7 Participants
AutoML Cup Phase 2 AutoML Cup Phase 2 <i>Organized by: spencr</i>	July 15, 2023 20 Participants
Cross-Domain MetaDL Any-Way Any-Shot Learning Competition with Novel Datasets from Practical Domains <i>Organized by: pavao</i>	November 15, 2022 9 Participants
RescueNet - Semi-Supervised Semantic Segmentation Semi-Supervised Semantic Segmentation of RescueNet dataset into 10 defined classes <i>Organized by: binalab</i>	October 2, 2023 6 Participants

Get Started

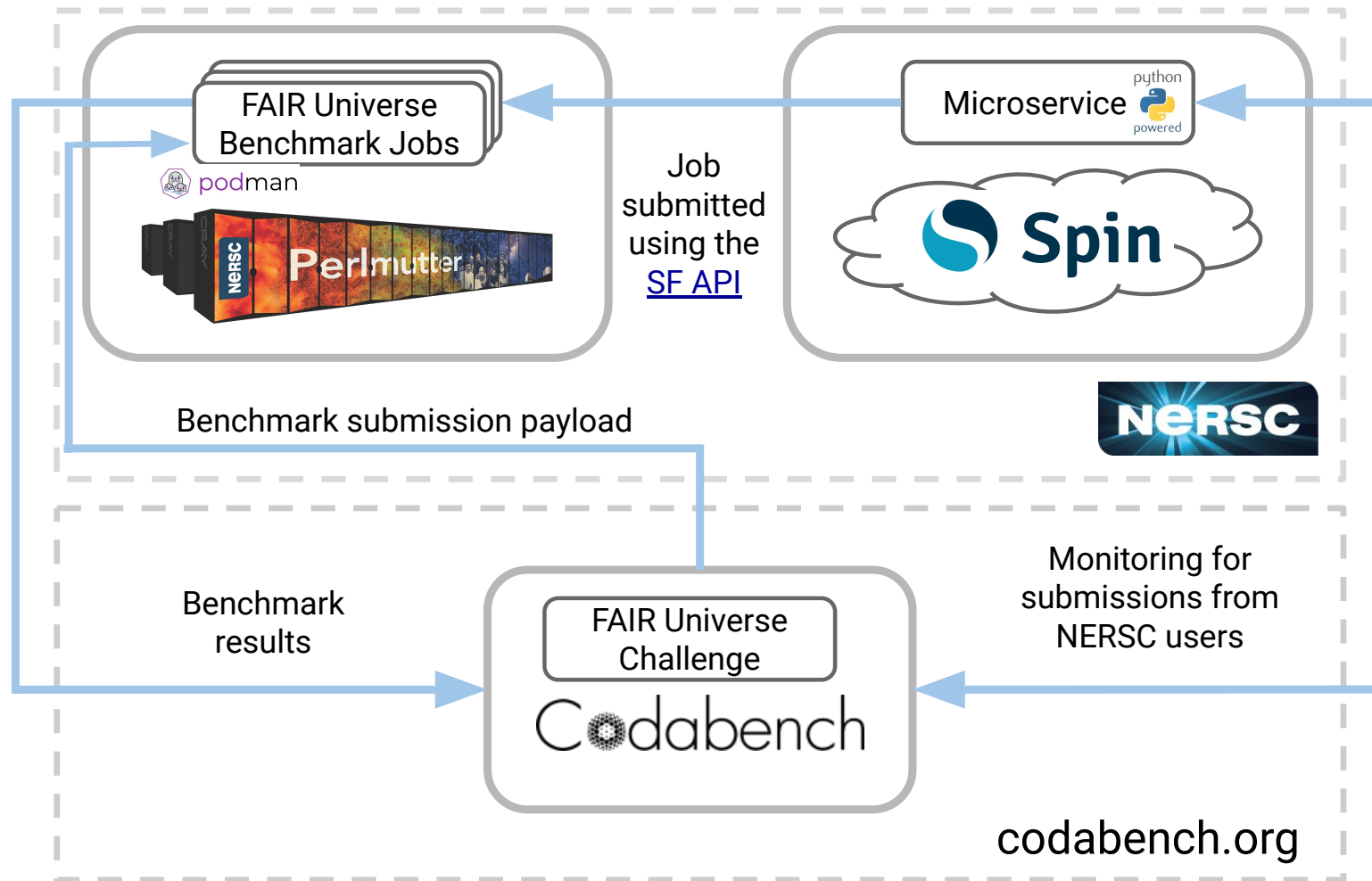
- Participate**
Find benchmarks that pique your interest! A benchmark allows you to test new algorithms against reference datasets OR (inverted benchmark) submit challenging data to reference algorithms.
- Organize**
Organize a benchmark on Codabench. Start with our [tutorial](#).
- Contribute**
Interested in joining the development team? Join us on [Github](#) or [contact us directly](#).

Codabench/“Fair Universe” Platform

Evolution of Codalab
<https://www.codabench.org/>



Fair Universe Platform: Current Codabench/NERSC integration



Conclusion

- a major new scientific competition on measuring Higgs cross-section,
 - taking into account/minimizing impact from modelisation systematics
 - winner to provide a narrow confidence interval with good coverage
- on Codabench platform with NERSC back-end for precise evaluation of submissions
- early prototype run as part of [Paris AI uncertainties workshop](#) in Nov 2023 and ACAT 2024
- to run June-Sep 2024
- submitted to NeurIPS 2024 competition track
- will be announced on : lhc-machinelearning-wg@cern.ch

- A cosmology challenge (weak-lensing) is also in the pipeline