FAIR Universe

FAIR Universe : HiggsML Uncertainty Challenge

Wahid Bhimji, Paolo Calafiura, **Ragansu Chakkappai**, Yuan-Tang Chou, Sascha Diefenbacher, Jordan Dudley, Steven Farrell, Aishik Ghosh, Isabelle Guyon, Chris Harris, Shih-Chieh Hsu, Elham E Khoda, Benjamin Nachman, Peter Nugent, David Rousseau, Benjamin Sluijter, Benjamin Thorne, Ihsan Ullah, Po Wen, Yulei Zhang

BERKELEY LAB

UNIVERSITY of WASHINGTON

NERSC National Energy Research Scientific Computing Center

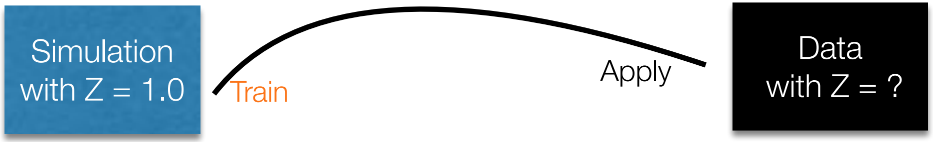
université PARIS-SACLAY

CHA LEARN

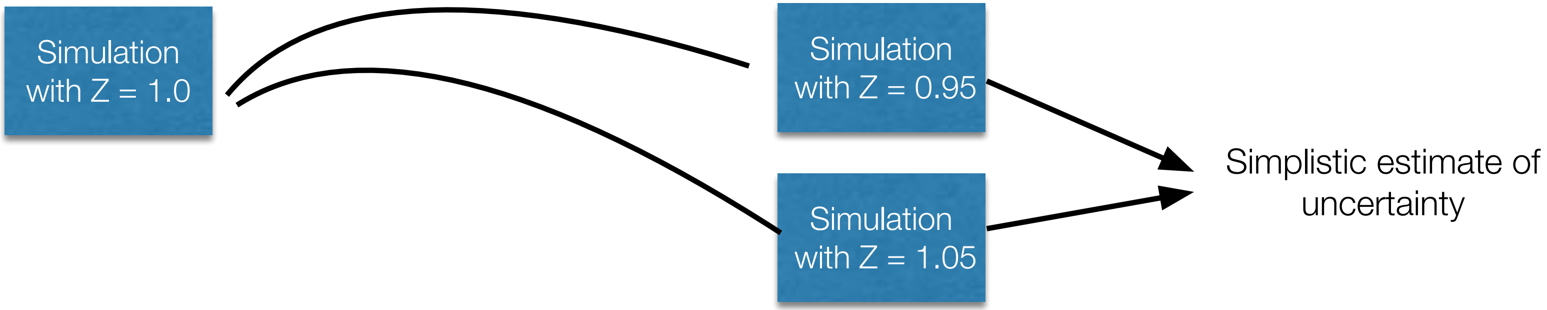
IJCLab Irène Joliot-Curie

Bias and uncertainty in ML in HEP

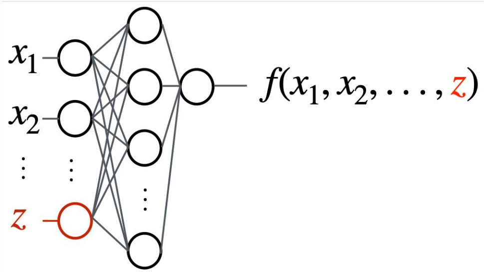
- ML methods in HEP are often trained based on simulation which has estimated systematic uncertainties (“Z”)
- These are then applied in data with the different detector state Z=?



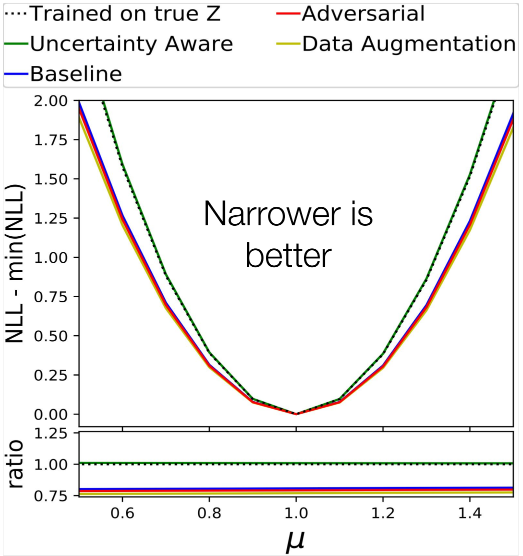
- Common baseline approach: Train classifier on nominal data (e.g. Z=1) and estimate uncertainties with alternate simulations. Shift Z and look at impact or perform full profile likelihood



Increasingly sophisticated approaches

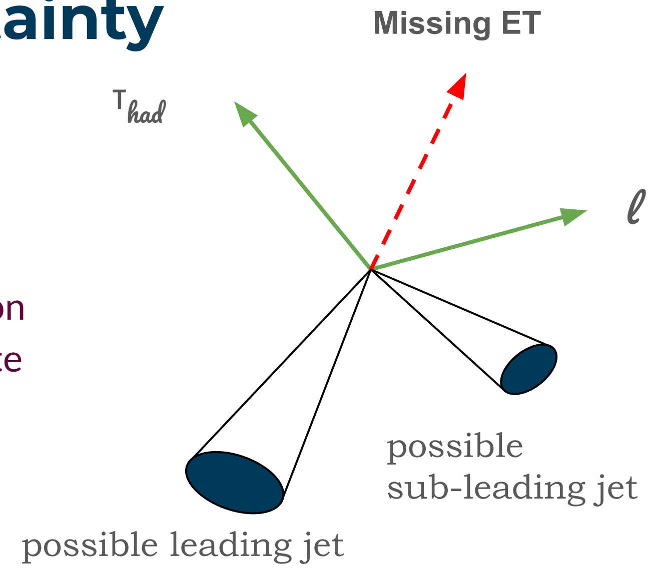


- "pivot" Louppe, Kagan, Cranmer : arXiv:1611.01046
- "Uncertainty-aware" approach of Ghosh, Nachman, Whiteson PhysRevD.104.056026
  - Parameterize classifier using Z
  - Measured on "Toy" 2D Gaussian Dataset and dataset from *HiggsML Challenge* modified to include systematic on tau-energy scale
  - Performs as well as classifier trained on true Z
- Other novel approaches e.g. (not comprehensive)
  - Inferno: arxiv:1806.04743
  - Direct profile-likelihood: e.g. arxiv:2203.13079
  - (Neuro) Simulation Based Inference has to include Z: arXiv:1911.01429

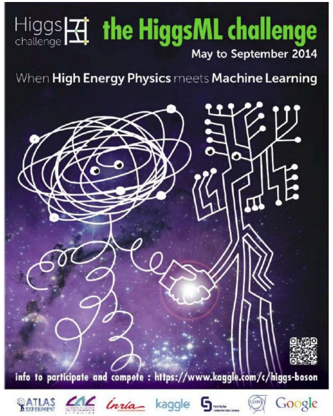


(Signal Strength)

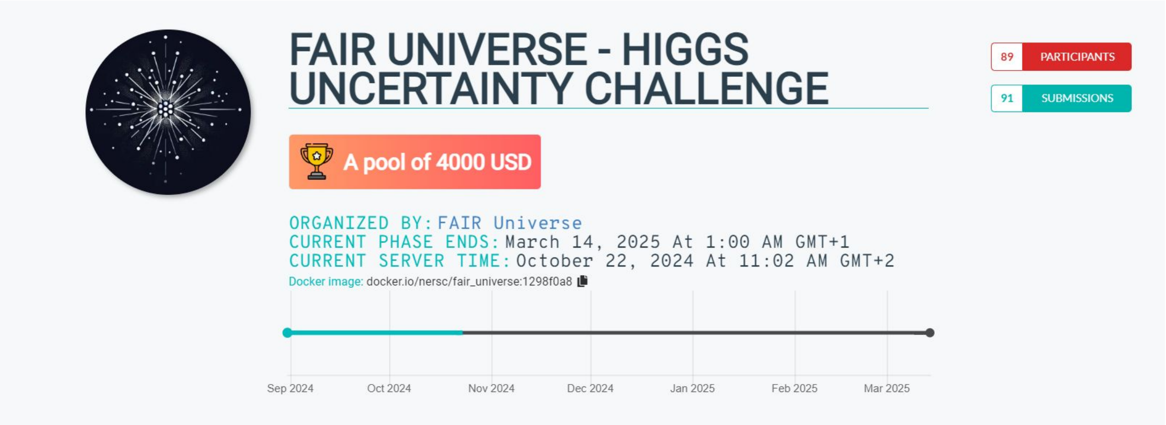
Fair Universe: HiggsML Uncertainty Challenge



- 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):
- New Fair Universe dataset, with following improvements
- Use (much) faster simulation
- Numbers of events 800.000 ⇒ ~280Millions
- Parameterized systematics
- Task : given a **pseudo-experiment** with given signal strength, provide a **Confidence Interval** on signal strength taking into account **statistics** and **systematics** uncertainties



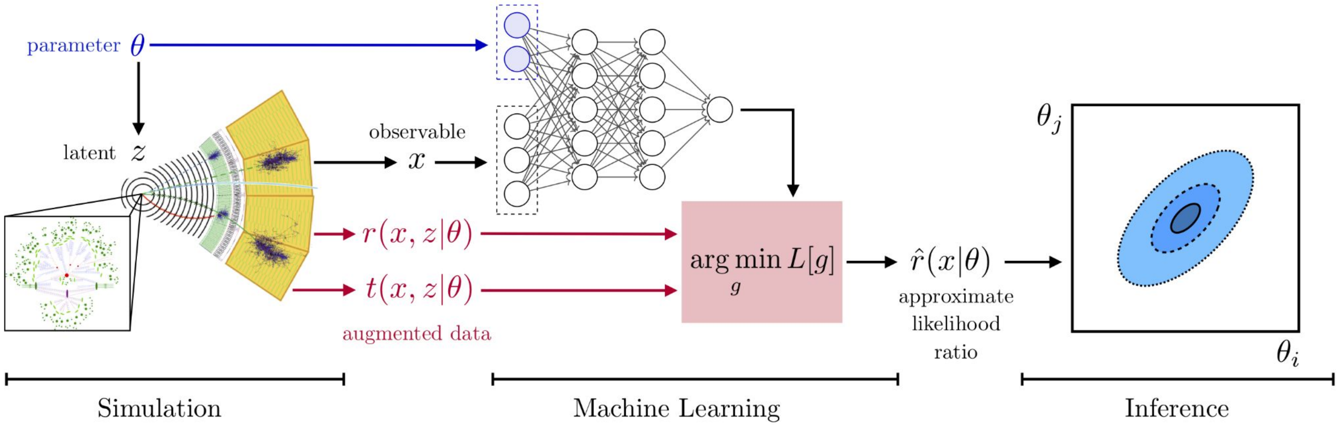
Fair Universe: HiggsML Uncertainty Challenge



- Full HiggsML Uncertainty Challenge Running from September 12 to March 14th
- Accepted as NeurIPS competition 2024
- Dedicated workshop at NeurIPS - 2024 at December 14th, Saturday morning

Challenge Objective

- Train a AI model to improve cross section measurement significance
- The model will be tested with datasets with unknown systematics and signal strength μ. (μ=1 if Standard Model)
- For each pseudo-experiment participants must predict best mu estimate:
  - μ_{hat} : best mu estimate
  - $[\mu_{16}, \mu_{84}]$: 68% Confidence Interval



Challenge Datasets

- We generated data with Pythia 8.2 and Delphes 3.5
- Using the updated Delphes ATLAS card
- Less accurate than Madgraph/Sherpa + Geant4, but much faster
- Generated ~**280 Million** Events after initial cuts equivalent to **220** X **10**fb-1
- Data generated using NERSC supercomputer.
- Data Organised into tabular form with **28** feature per event.

Process	Number Generated	LHC Events @10fb-1	Label
Higgs	52101127	1015	signal
Z Boson	221724480	1002395	background
Di-Boson	2105415	3783	background
$t\bar{t}$	12073068	44190	background

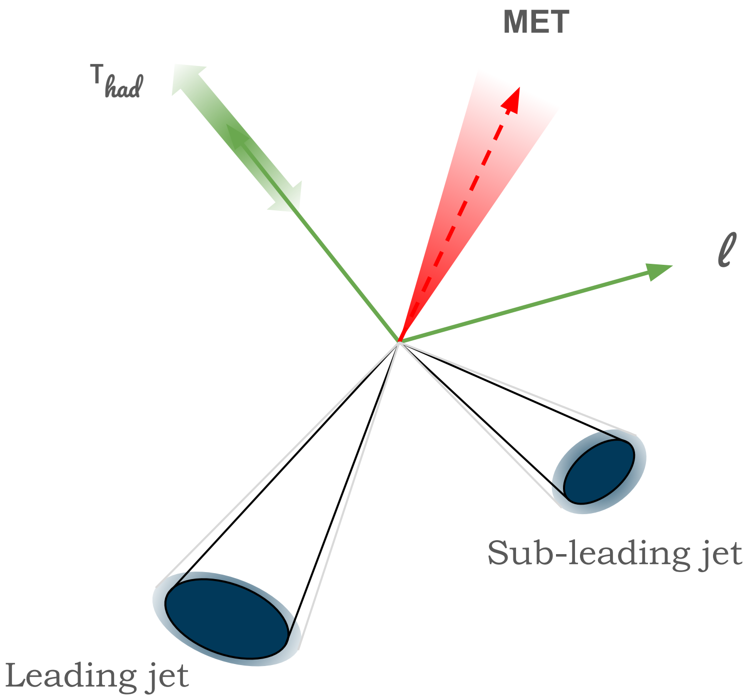
DELPHES
fast simulation

Challenge Datasets - Systematics

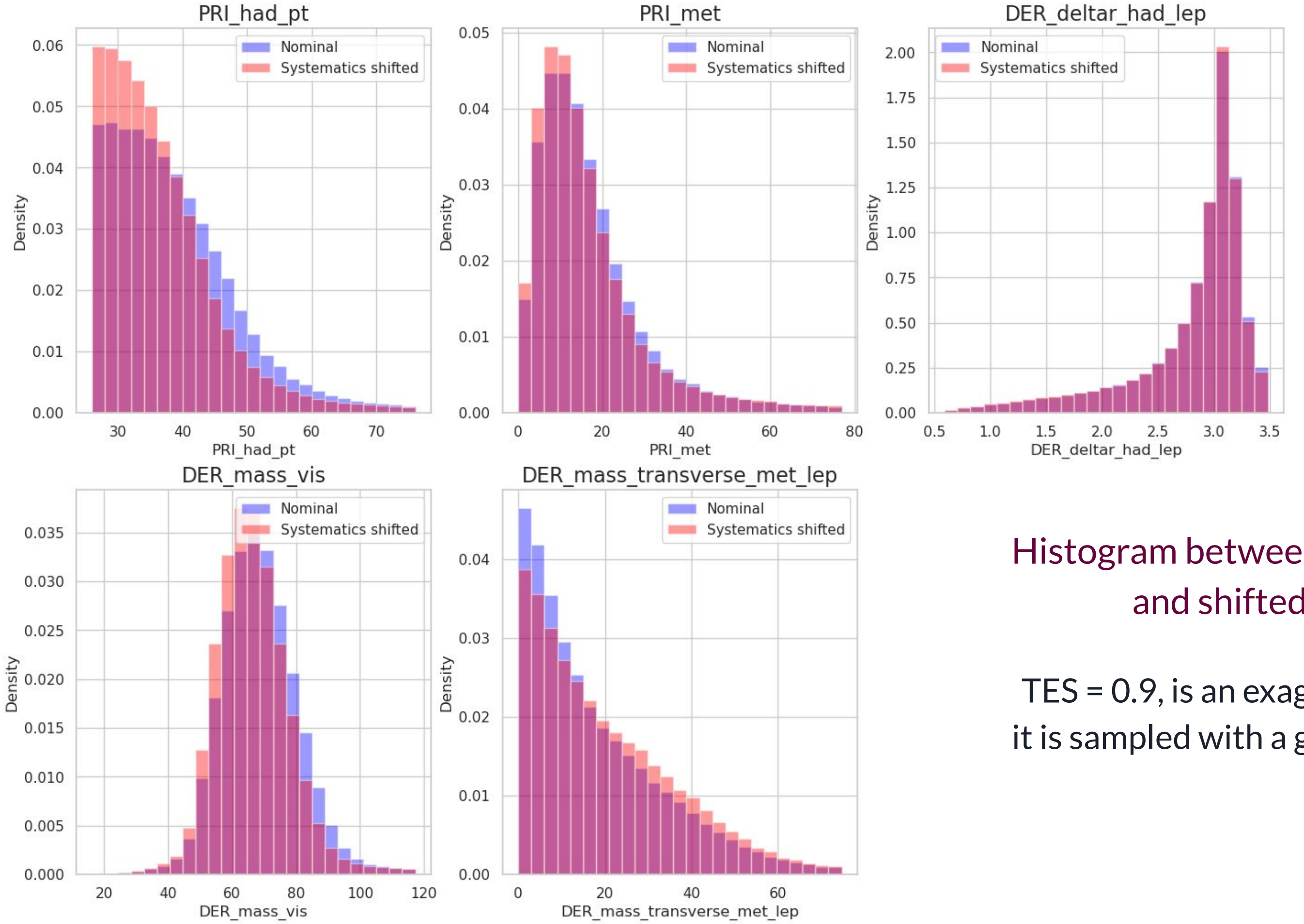
Apply parameterized systematics (Nuisance Parameters) :

- Feature distortions:
  - Tau Energy Scale (and correlated MET)
  - Jet Energy Scale (and correlated MET impact)
  - Additional randomised Soft MET

- Event category normalisation
  - Background overall normalisation
  - Di-boson background normalisation
  - ttbar background normalisation



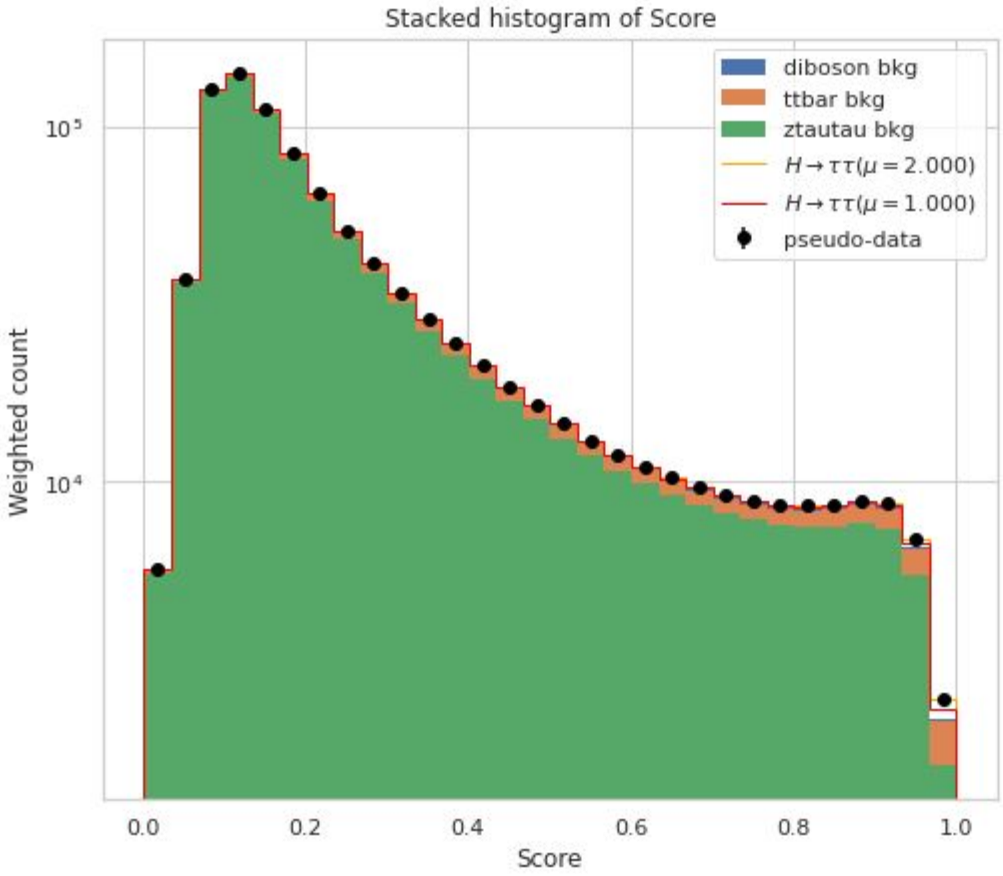
Tau Energy Scale Systematics Applied



Histogram between nominal (TES = 1) and shifted (TES = 0.9)

TES = 0.9, is an exaggeration, in practice it is sampled with a gaussian of 1 +/- 0.01

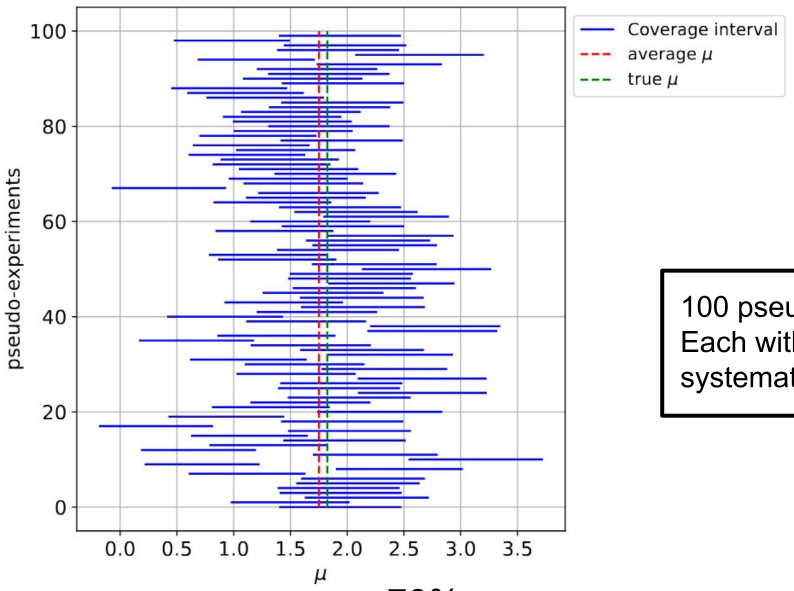
Simple Classifier



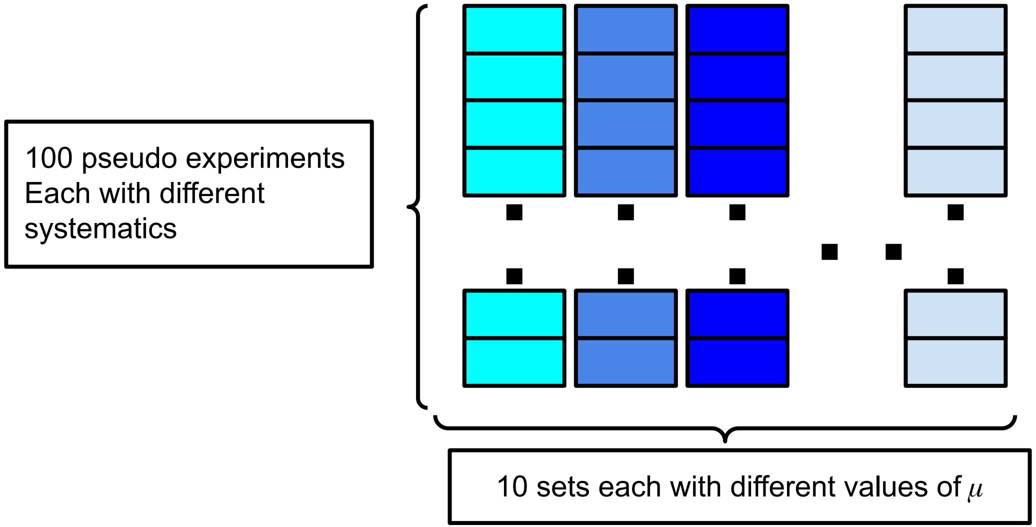
- Simple **BDT** Classifier
- **No** special training for systematics
- Histogram is what is obtained on one pseudo experiment

Coverage Evaluation

- Form multiple pseudo-experiment test sets: different signal strengths ($\boldsymbol{\mu}$) and systematics
  - 10$\boldsymbol{\mu}$ times 100 pseudo-experiments
- Task: predict uncertainty interval $[\boldsymbol{\mu}_{16}, \boldsymbol{\mu}_{84}]$
  - E.g. 68% quantile of likelihood or assume 1σ



coverage=70%



Uncertainty Quantification Metric

- **Interval width (w)** averaged over N test sets
- **Coverage (c)**: fraction of time μ is contained
- Combined using a **coverage function f(x):**

$$x \geq 0.68 - 2\sigma_{68} \text{ and } x \leq 0.68 + 2\sigma_{68} \ : \ 1.$$

$$x < 0.68 - 2\sigma_{68} \ : \ 1 + \left|\frac{x-(0.68-2\sigma_{68})}{\sigma_{68}}\right|^4$$

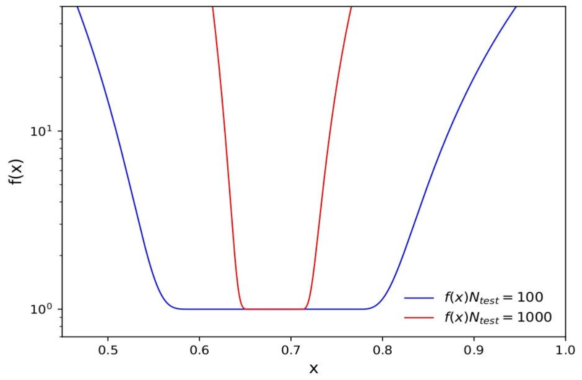
$$x > 0.68 + 2\sigma_{68} \ : \ 1 + \left|\frac{x-(0.68+2\sigma_{68})}{\sigma_{68}}\right|^3$$

$$\text{with } \sigma_{68} = \frac{\sqrt{(1-0.68)0.68N)}}{N}$$

- N dependance for equivalent ideal coverage
- Penalizes undercoverage more
- Final score (s) designed to avoid large values or gaming

$$w = \frac{1}{N}\sum_{i=0}^{N} |\mu_{84,i} - \mu_{16,i}|$$

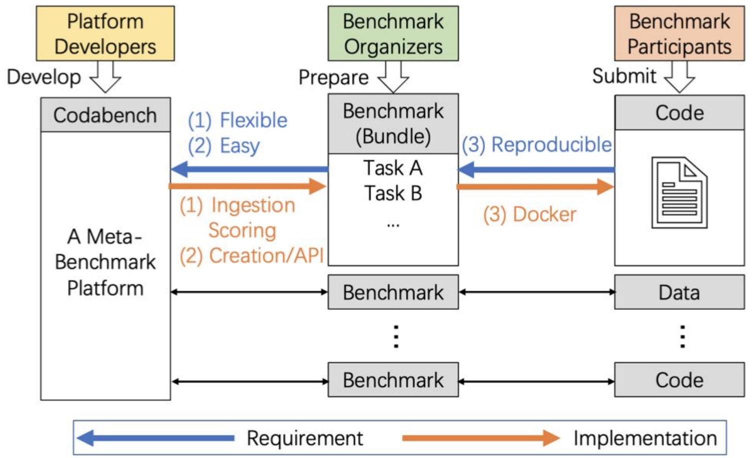
$$c = \frac{1}{N}\sum_{i=0}^{N} 1 \text{ if}(\mu_{true,i} \in [\mu_{84,i} - \mu_{16,i}])$$



$$\boxed{s = -\ln((w+\epsilon)f(c))}$$

See also [Sascha Diefenbacher's AISSAI Workshop presentation]

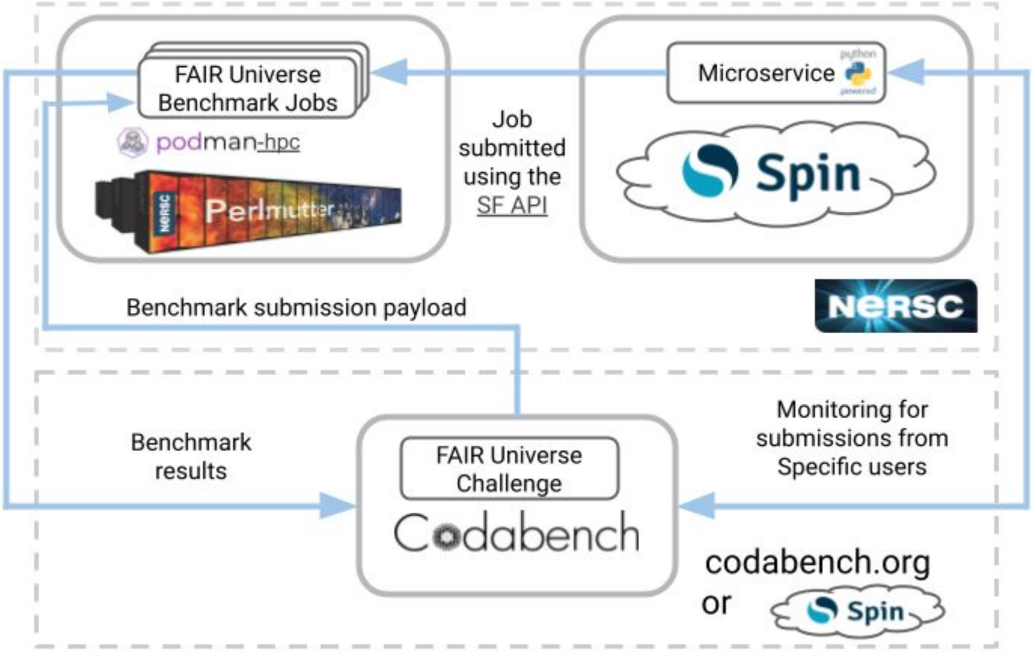
Codabench Platform



Codabench - open source platform for AI benchmarks and challenges

- Originally (CodaLab) Microsoft/Stanford now a Paris-Saclay/LISN led community
- > 600 challenges since 2013
- Completely open-ended competition design.
- Allows code submission as well as results e.g. for evaluation timing or reproducibility
- Also data-centric AI "inverted competitions"
- Queues for evaluation can run on diverse compute resources
- Platform itself can be deployed on different compute resources
- Ranked best challenge platform for ML by ML contests

Fair Universe Platform: Codabench-NERSC integration



System Specifications

Partition	# of nodes	CPU	GPU
GPU	1536	1x AMD EPYC 7763	4x NVIDIA A100 (40GB)
	256	1x AMD EPYC 7763	4x NVIDIA A100 (80GB)

Conclusion

- AI challenge which addresses Systematic Uncertainty in HEP problem.
- Large Data Set with ~280M Events (signal + background)
- New Scoring to take Coverage and Confidence interval into account.
- Custom ingestion algorithm to test multiple pseudo-experiments in parallel.
- Large Computing Infrastructure as back_end
- You can enter the **HiggsML Uncertainty Challenge** now!
- You can win an invitation to **NeurIPS** workshop if meaningful participation by **15th Nov**
    - https://www.codabench.org/competitions/2977/

Help and feedback: #higgsml-uncertainty-challenge channel on the Fair Universe Slack
Ongoing information Google Group: Fair-Universe-Announcements
Collaborations, questions, comments: fair-universe@lbl.gov
Ragansu Chakkappai, Sascha Diefenbacher and David Rousseau are here the full week, talk to us!

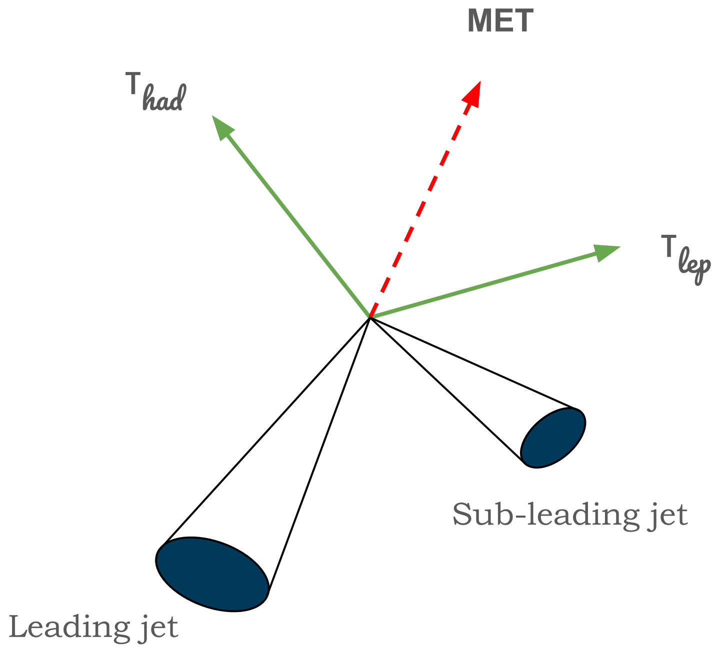
Thank you for your attention!

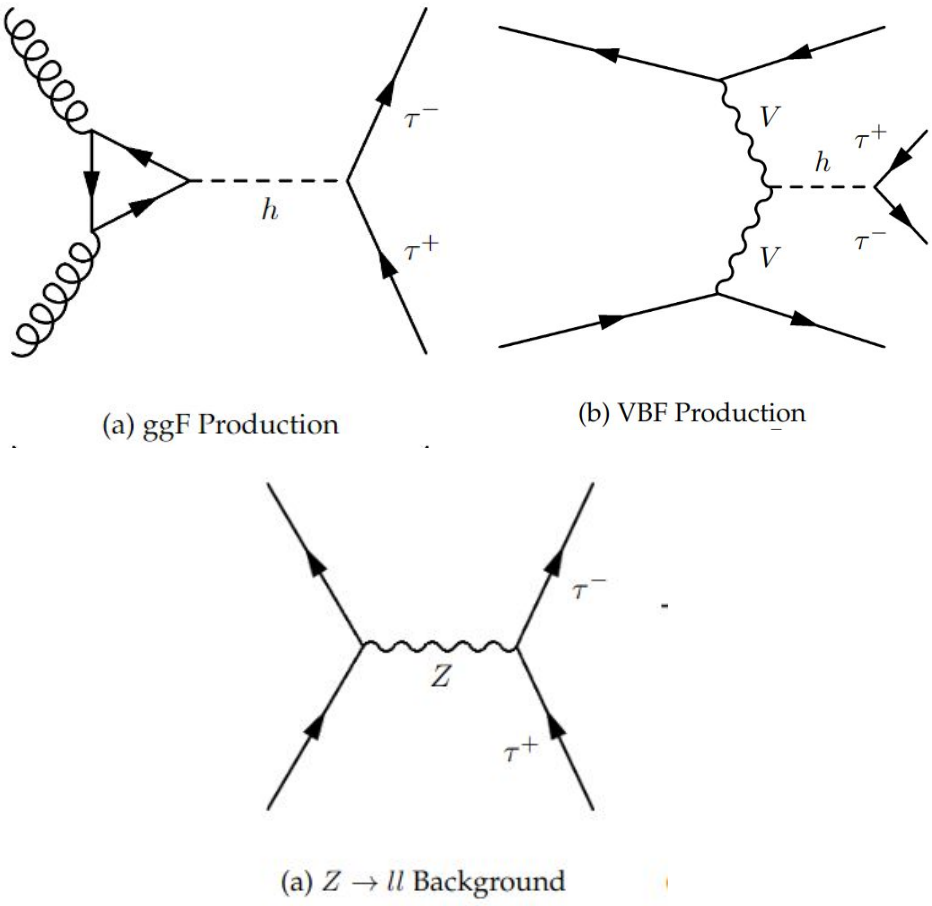
Back-up

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).
- This funding went to LBL, NERSC, U Washington, and Chalearn (Isabelle Guyon's Non-Profit US Organisation).

Introduction

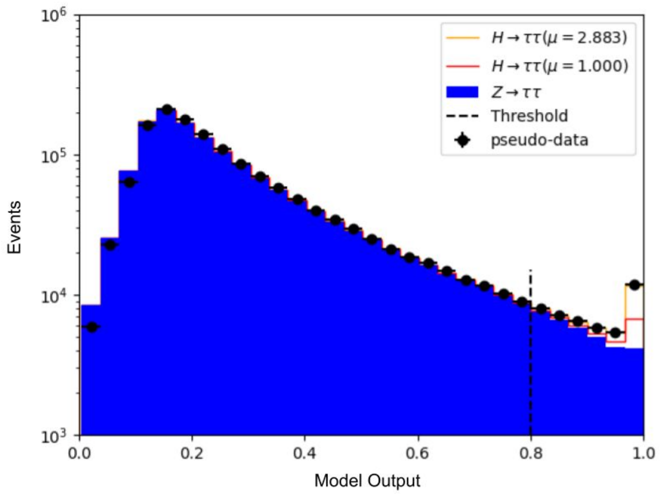




Basic Algorithm

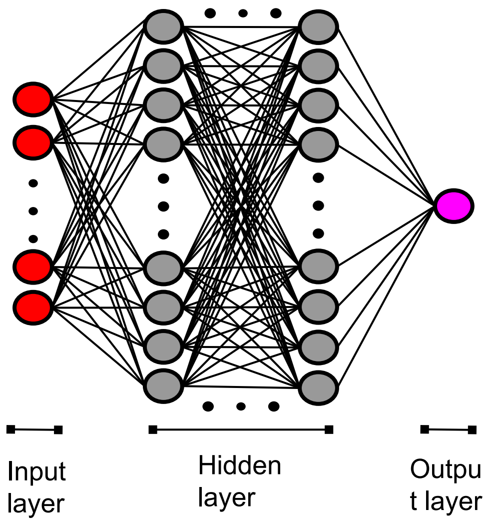
1. Divide data into ***train_set*** and ***holdout_set***
2. Use ***train_set*** to Train the simple dense NN
3. Define S_i and B_i : predicted score bin content
4. Construct for S_i (α) and B_i(α) functions from ***holdout_set***
5. Combine define Binned Negative Log Likelihood function as function of NPs and μ
6. For Each pseudo experiment

    a. Predict score for pseudo experiment
    b. Use Minuit to find value of mu, sigma_mu and NP
    c. Returns
        - ***mu_hat***
        - ***p16 = mu - sigma_mu***
        - ***p84 = mu + sigma_mu***



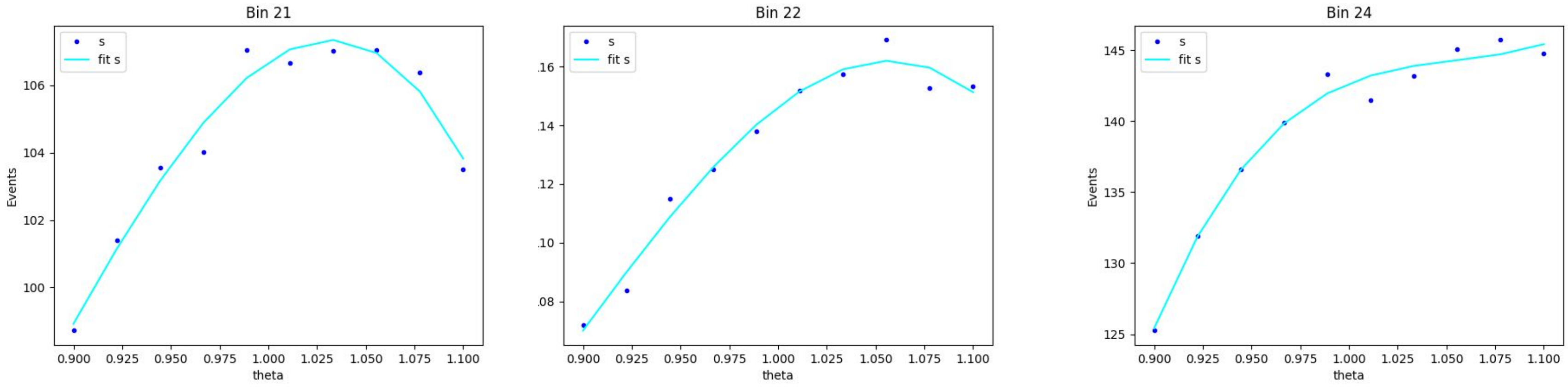
NN with L2 regularization using PyTorch

- PyTorch NN Classifier is Trained to distinguish Signal (Higgs) from Background (Z)
- 32 features,
- Architecture
  - 3 Hidden layers with 100 nodes
  - 1 Output node
  - ReLU Activation between layers
  - L2 Regularization during training
- Model return score between 0 (background) and 1(signal),

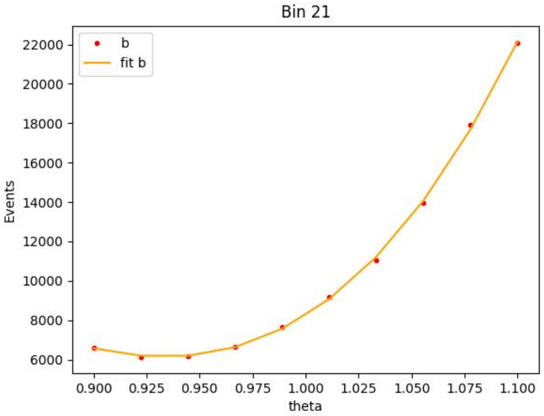


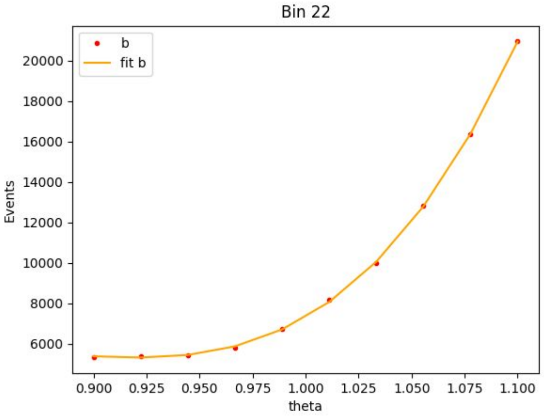
Parameterisation of S(α)

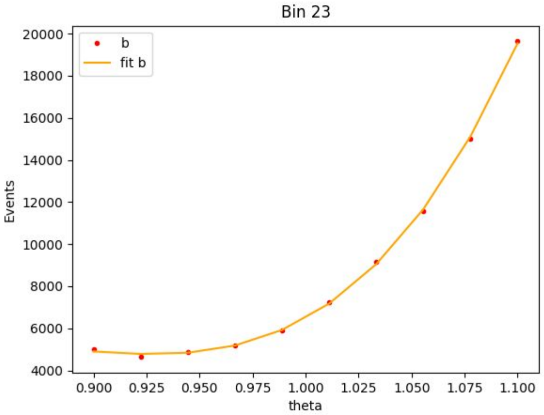
With the help of the *holdout_set* for we get values of S and B for each NP in each bin.
A polynomial function is used to fit them. This function is later used in the NLL formalism

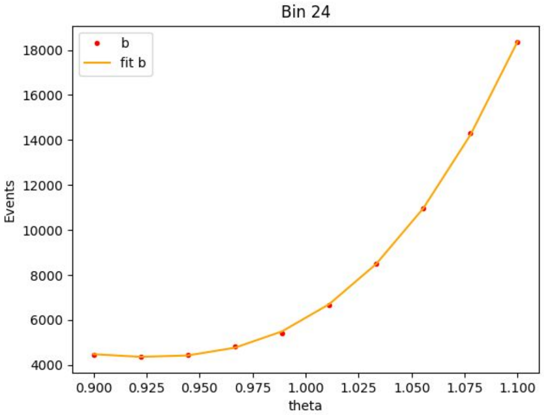


Parameterisation of B(alpha)









Profile μ and α simultaneously

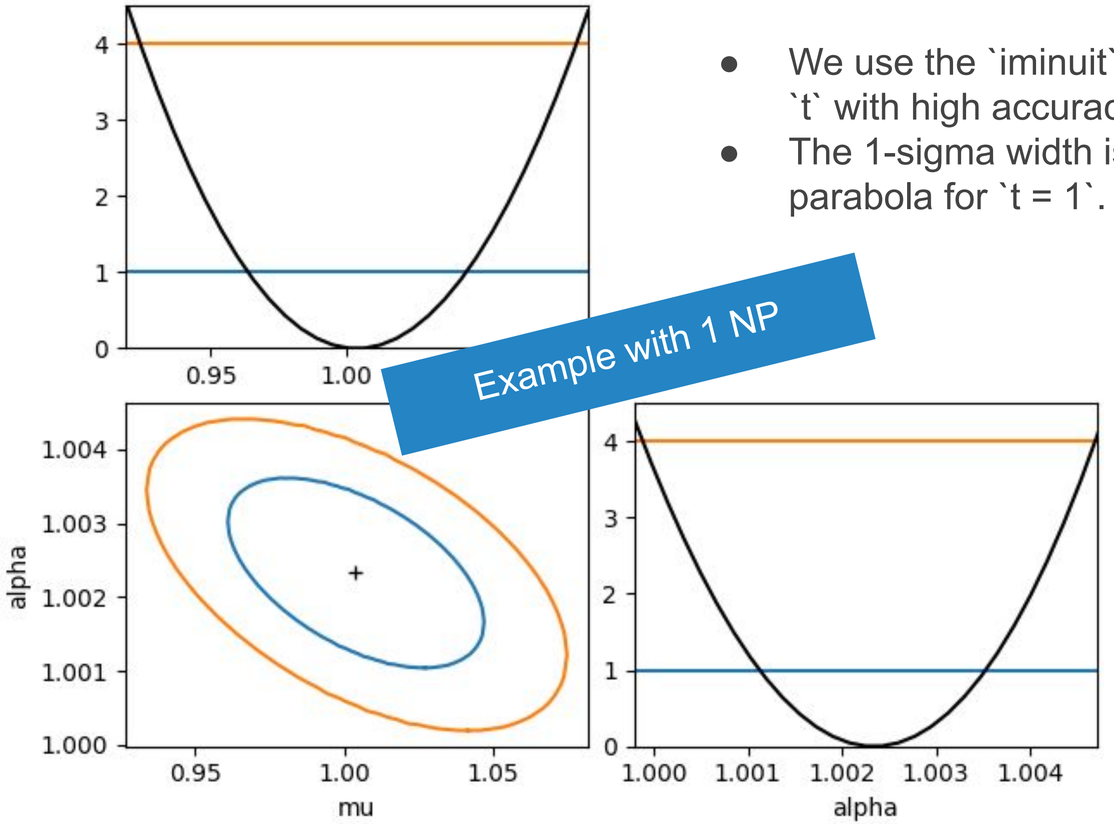
$$L(\mu, \vec{\alpha}|\mathcal{D}) = \prod_{i=1}^{N_{\text{bins}}} \frac{(\mu S_i(\vec{\alpha}) + B_i(\vec{\alpha}))^{n_i} e^{-(\mu S_i(\vec{\alpha}) + B_i(\vec{\alpha}))}}{n_i!}$$

$$\Rightarrow t_{\mu,\vec{\alpha}} = -2 \log \left(L(\mu, \vec{\alpha} \mid \mathcal{D}) \right)$$

$$= -2 \sum_{i}^{N_{\text{bins}}} n_i \log(\mu S_i(\vec{\alpha}) + B_i(\vec{\alpha})) + (\mu S_i(\vec{\alpha}) + B_i(\vec{\alpha}))$$

L here is the likelihood estimator which depends on μ and α, where α is a vector of 5 NP thus the μ at which L is maximum or *t* is minimum is the predicted μ,

NLL curve and contour



- We use the `iminuit` package to find the minimum of `t` with high accuracy.
- The 1-sigma width is the width between points on the parabola for `t = 1`.

Leaderboard so far

Results										
Task:					Fact Sheet Answers	Higgs NeurIPS Task 10x100				
#	Participant	Entries	Date	ID	Method Name	Quantile Score	Interval	Coverage	RMSE	Run Time (mins
1	jdudley	2	2024-10-15 23:23	91180	game	**-0.68**	1.97	0.68	2.42	7.0
2	benevedes	2	2024-10-07 19:30	87099	Test submission	**-5.67**	1.67	0.6	n/a	7.0
3	avencast	7	2024-10-17 13:21	92509	xgboost correct 2.0	**-8.19**	4.02	0.85	15.66	8.0