ATLAS analysis using PanDA/iDDS based active learning

Christian Weber,
Rui Zhang, Tadashi Maeno, Wen Guan, and Torre Wenaus

BNL Physics 7th joint meeting on AI/ML

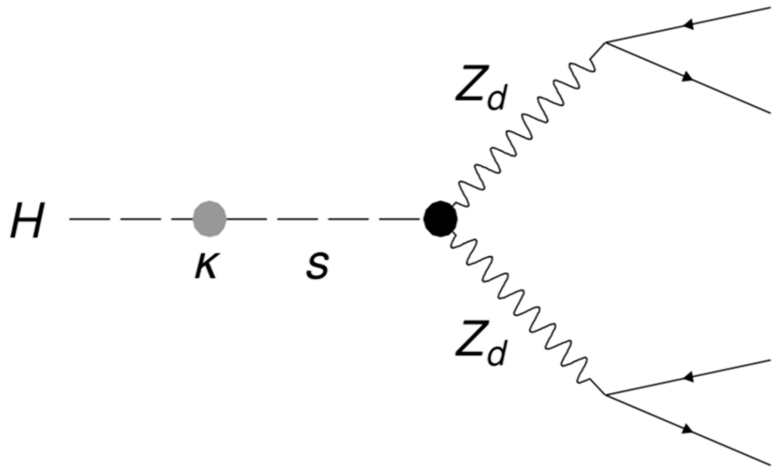
September 27th, 2023

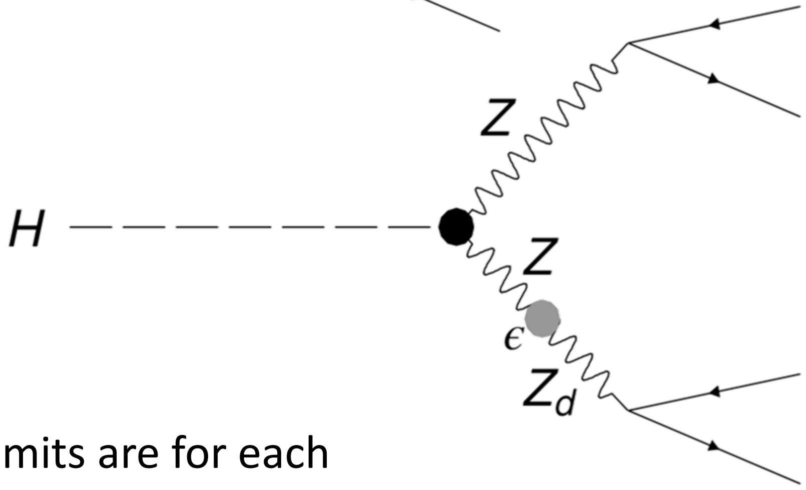
Beyond Standard Model physics parameters

Hidden Abelian Higgs Model
New Scalar S, new vector boson Z_d

Five different parameters:

m_S - new scalar mass
m_{Z_d} - new vector boson mass
Γ_{Z_d} - decay width of the new vector boson
ϵ – mixing between Standard Model Z and Z_d
κ – mixing between Higgs Boson H and S





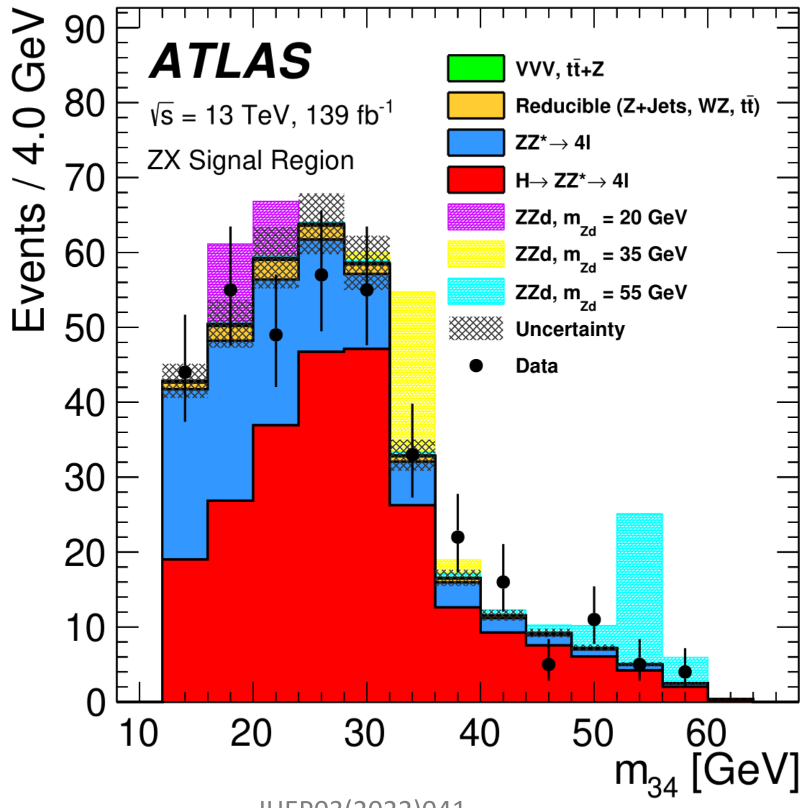
In general, we would like to know what the associated cross section limits are for each combination of parameter values, or which combinations are excluded.

Cross section limit calculation

- Generate Monte Carlo (MC) Events for
  - Standard Model backgrounds
    do this once using Standard Model parameters
  - Signal process under the given signal model
    do once for every point in parameter phase space
- Apply selection for backgrounds and signal
- Profile Likelihoods for each simulated signal to find upper limits on the cross section σ

$$\mathcal{L}(\sigma) = \prod_i \mathrm{Pois}(\mathrm{Data}_i | B_i + S_i \cdot \sigma)$$

- S_i, σ dependent on physics parameters $m_S, m_{Z_d}, \Gamma_{Z_d}, \epsilon$, κ

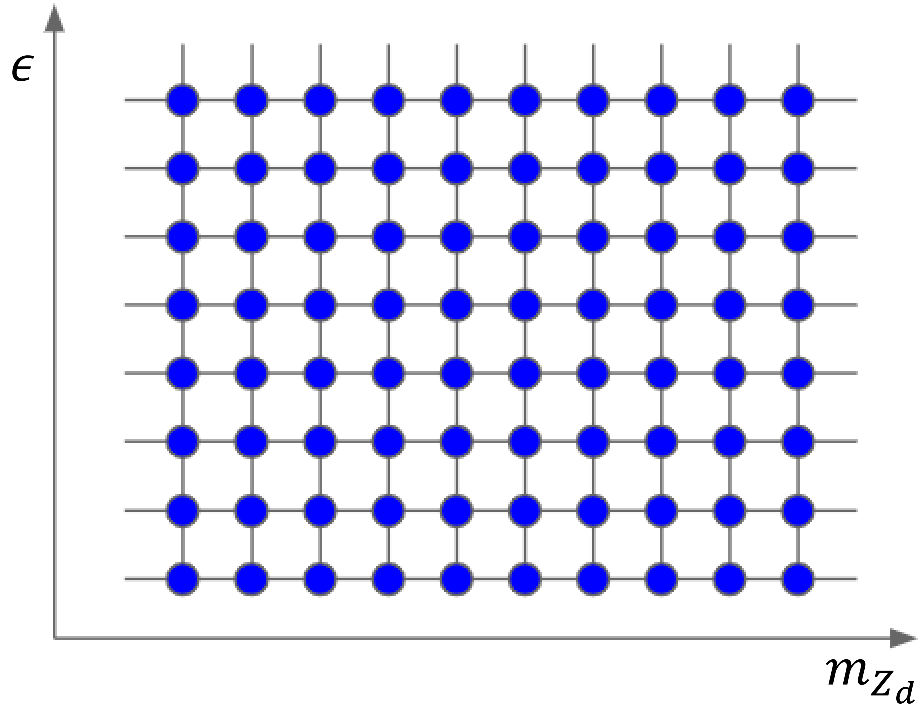


JHEP03(2022)041

Limits in multidimensional parameter space

Process for calculation of limits in phase space conceptionally simple, e.g., do a 'grid search'

- Select boundaries in m_S, m_{Z_d}, Γ_{Z_d}, ϵ, κ
- Generate grid of signal samples
- And calculate cross section limits for each sample
- Bonus: with theory predictions one can infer contours between excluded and not excluded regions of parameter space



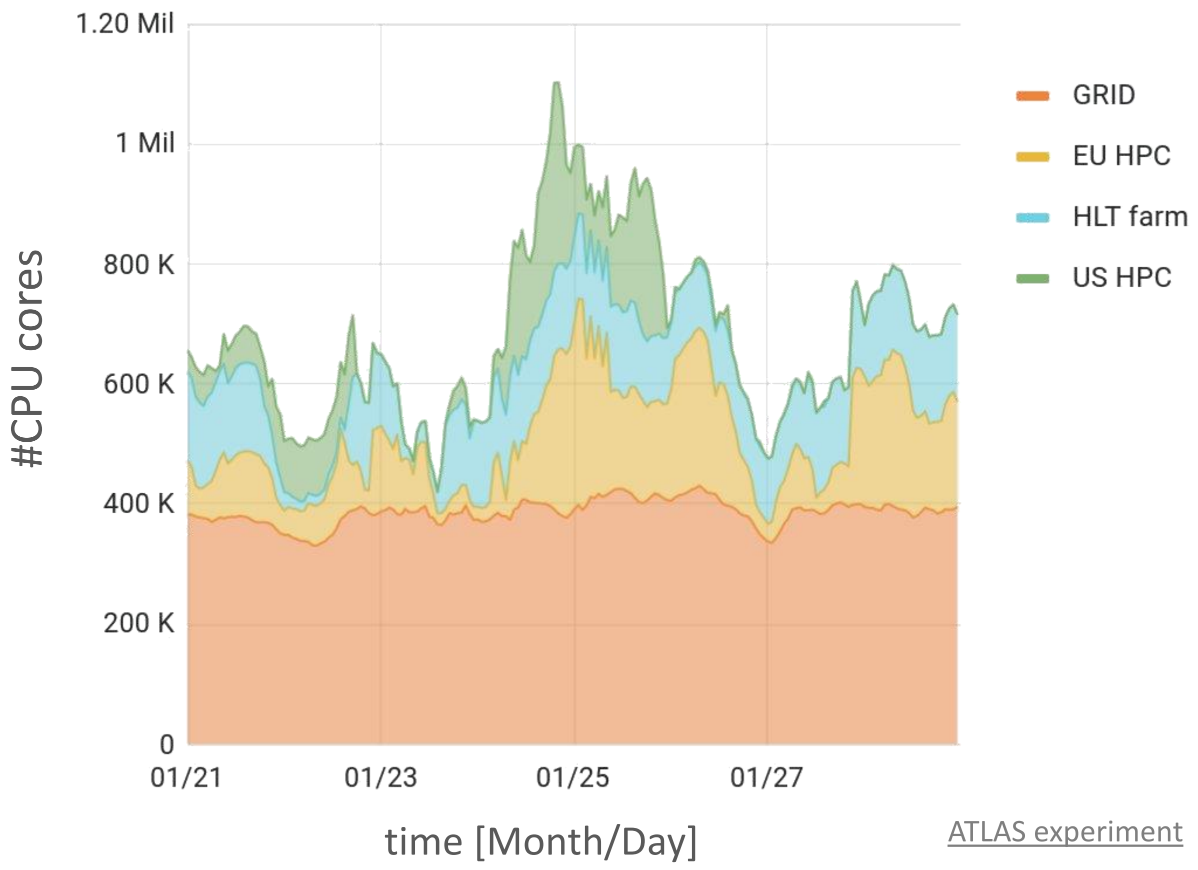
Problem:
- Generating Monte Carlo samples is computationally expensive

Monte Carlo Simulation

Monte Carlo (MC) Event simulation is very computationally intensive

- ~15h calculation time for 100 simulated events on a single CPU core
- Might need around 10^5 events or more per signal sample
- Using a grid of 10 values in two dimensions would require 10^7 events, about ~2k CPU months
- Estimate is for fast calorimeter simulation, full calorimeter simulation requires even more computational resources

CPU resources available to the ATLAS experiment in late January 2023



ATLAS experiment

Active Learning

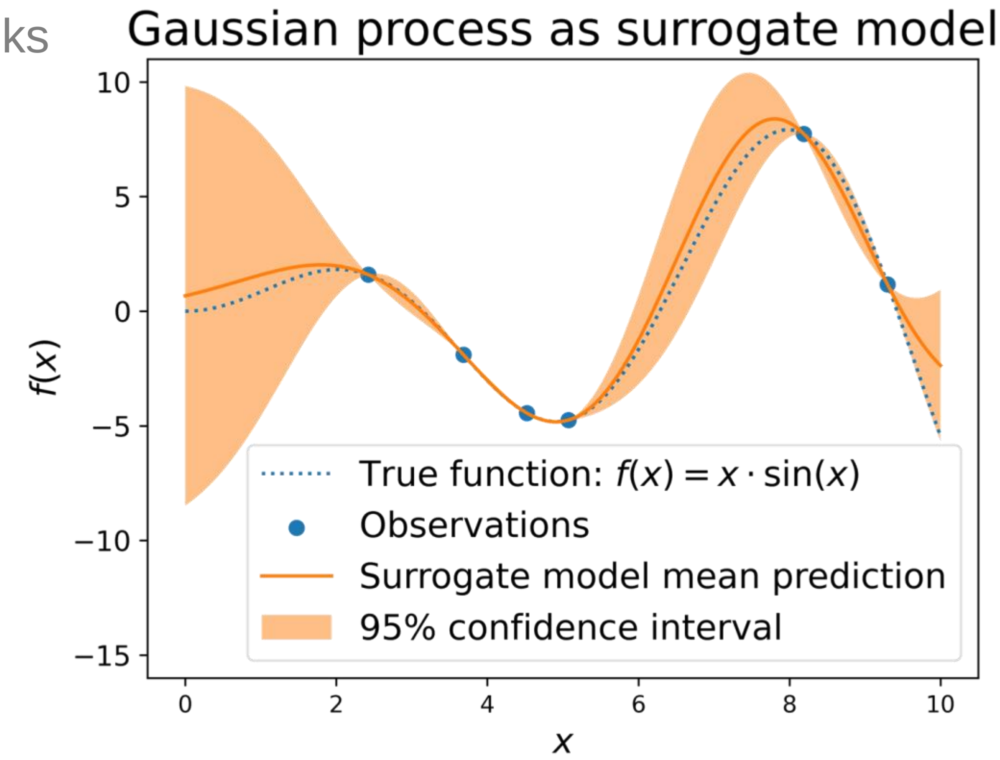
An iterative process to collect new labelled data for optimization tasks

We developed a system to work around the computational constraints by using an Active Learning approach that includes

- Surrogate model that approximates the function mapping physics parameters onto exclusion limits

$$m_S, m_{Z_d}, \Gamma_{Z_d}, \epsilon, \kappa \rightarrow \sigma_{m_S, m_{Z_d}, \Gamma_{Z_d}, \epsilon, \kappa}$$

- An acquisition function that determines which points in the physics parameter space, e.g. $m_S, m_{Z_d}, \Gamma_{Z_d}, \epsilon, \kappa$, to explore next based on the surrogate model



Active Learning Application

Our system builds on ATLAS grid infrastructure to enable the application of the active learning in the exploration of the parameter phase space

It is realized via the intelligent data delivery service (iDDS) that allows a docker container instance to effect the simulation Monte Carlo samples on the ATLAS grid, run an analysis on REANA, and exchange data between them

- The iDDS is a computing service that allows for on-demand data transformation and granular data delivery within the ATLAS grid

- PanDA is the Production and Distributed Analysis workload management system for the ATLAS grid

- REANA is a platform enabling workflow preservation and reuse of data for analysis

iDDS

Surrogate Model

Surrogate Model: Gaussian Process

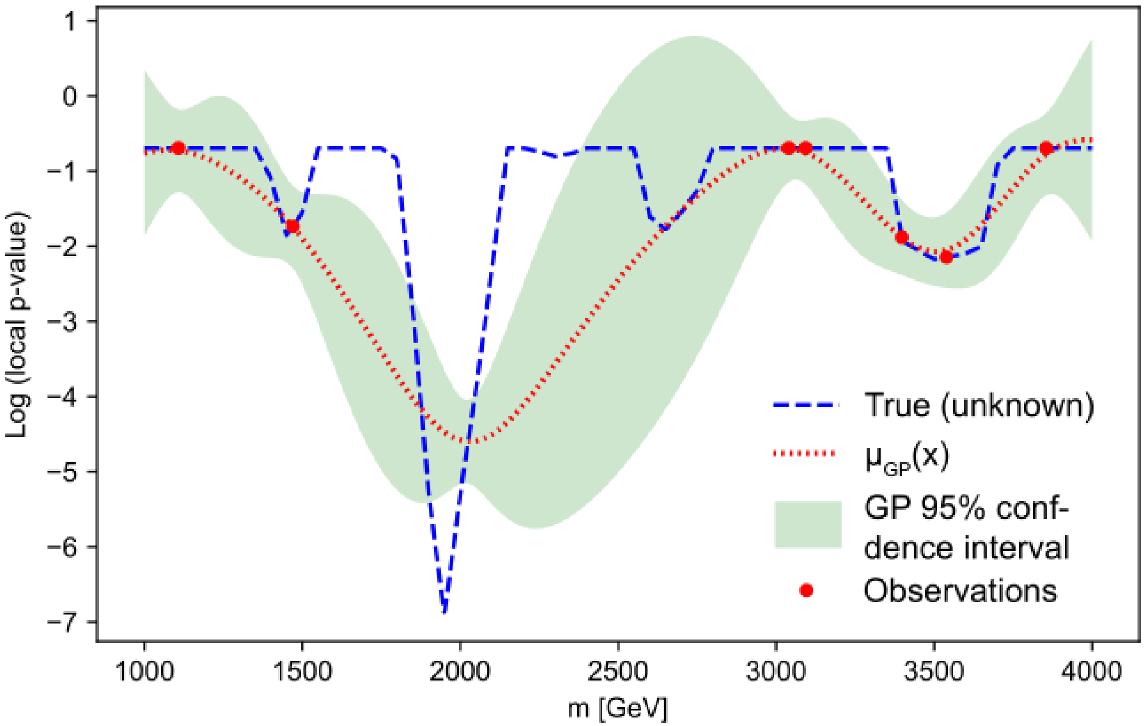
- Non-parametric model, yielding probability distribution over possible functions that fit a set of points.

- Assumes output $f(x)$ is a random variable for each input x and $p(f(x_1), f(x_2), \dots)$ is a multivariate gaussian:

$$p(f(\vec{x})) = G(f(\vec{x}) \mid \vec{\mu}, \boldsymbol{K})$$

mean vector $\mu_i = E(x_i)$
covariance matrix $\boldsymbol{K}_{ij} = k(x_i, x_j)$
kernel k, many options, common choice:
Radial Basis Function kernel

$$k(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$$



Surrogate Model & Acquisition Function

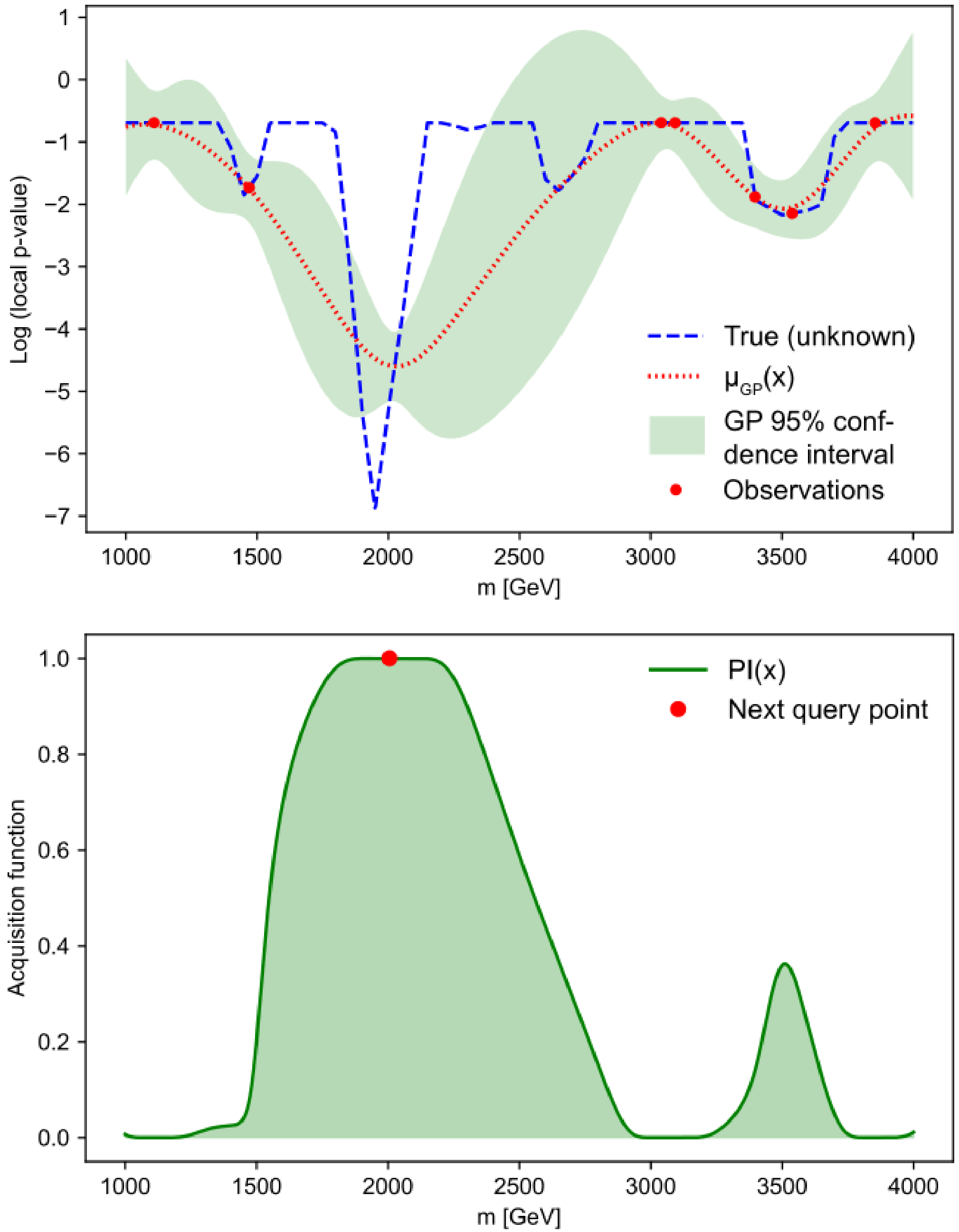
Acquisition Function: Probability of improvement

- Assume we want to find the minimum of an underlying true function, via optimizing surrogate model

- $f(x^*)$ is current minimum, then choose next value to sample for surrogate model, such that
$$\mathrm{PI}(x) = \mathrm{Prob}(f(x) - f(x^*) < 0)$$
is maximal

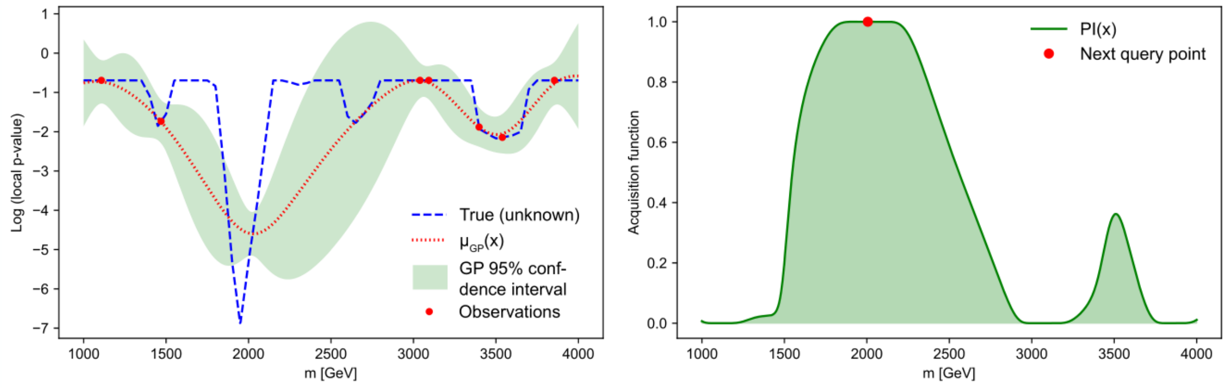
- $\mathrm{PI}(x)$ calculable as $f(x)$ is a gaussian distribution with mean $\mu(x)$, and standard deviation $\sigma(x)$
$$PI(x) = \Phi\left(\frac{f(x^*) - \mu(x)}{\sigma(x)}\right)$$
Φ – Gaussian cumulative distribution function

- See 'Active Learning for Excursion Set Estimation' by Kyle Cranmer et al. for an entropy-based acquisition function

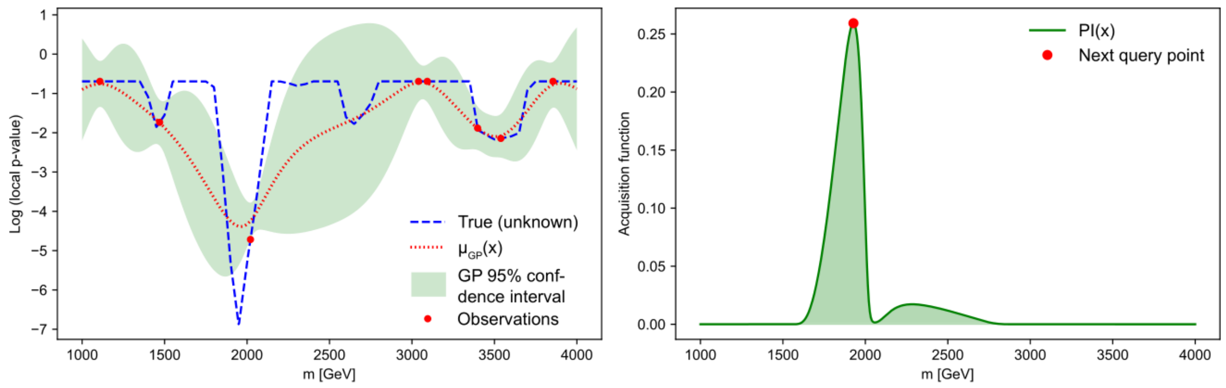


Iterative feedback & learning

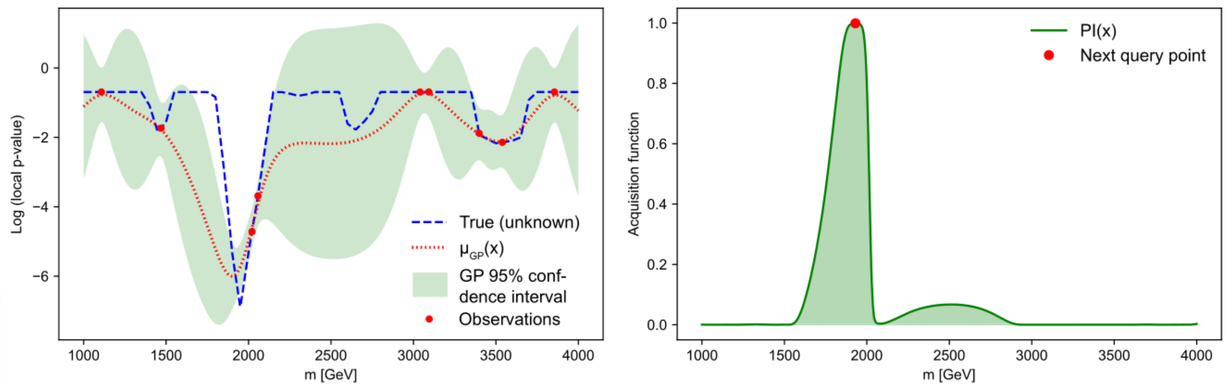
start with 7 random initial points



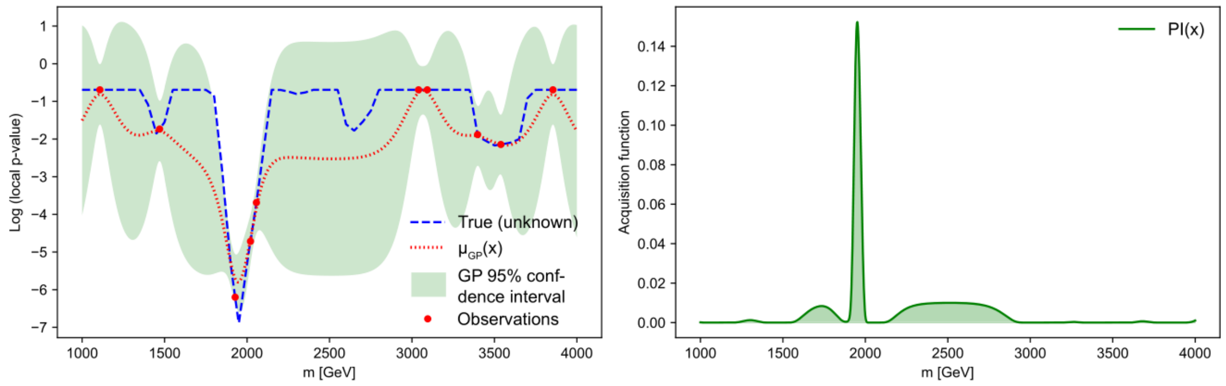
7 initial points + 1 predicted point



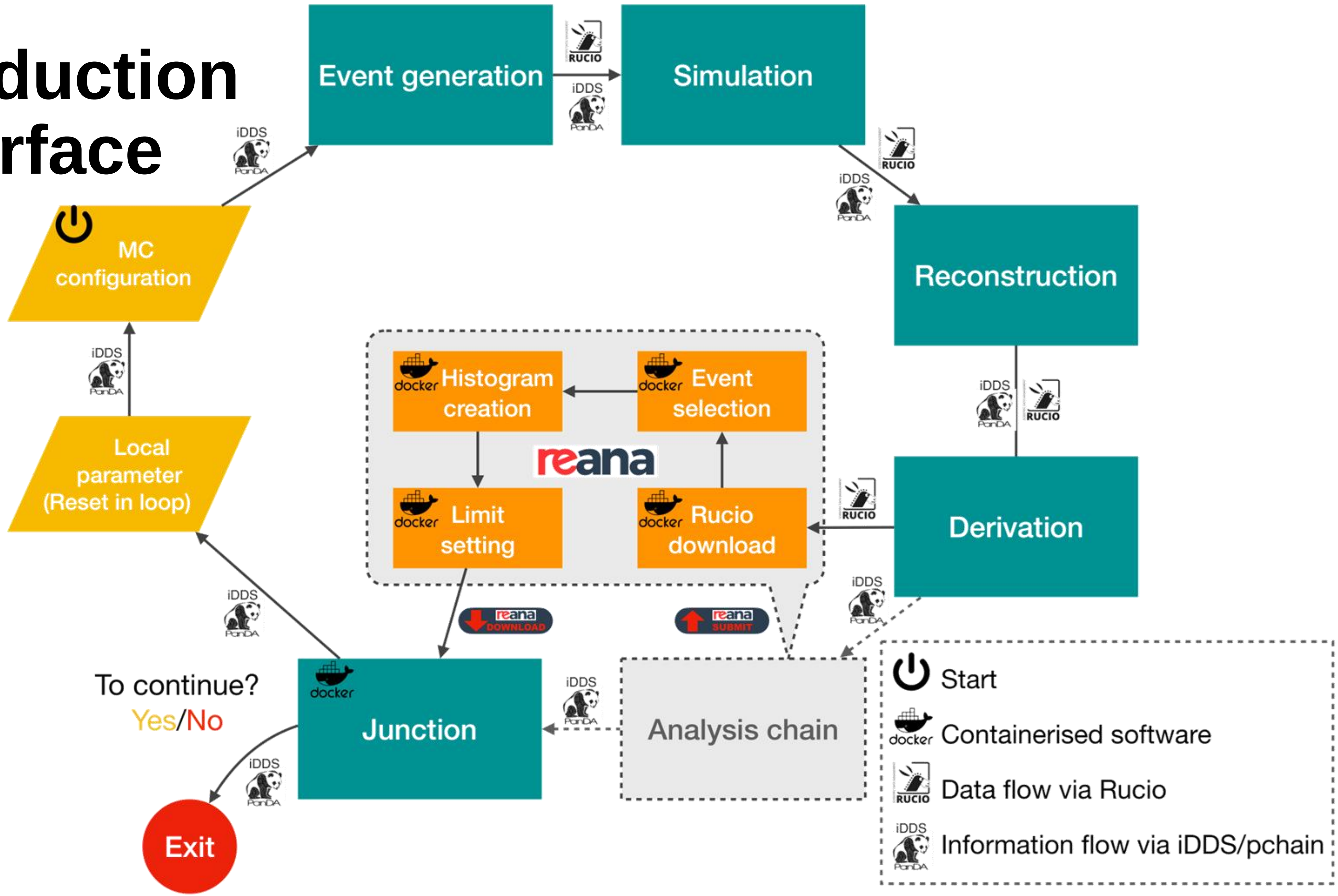
7 initial points + 2 predicted points



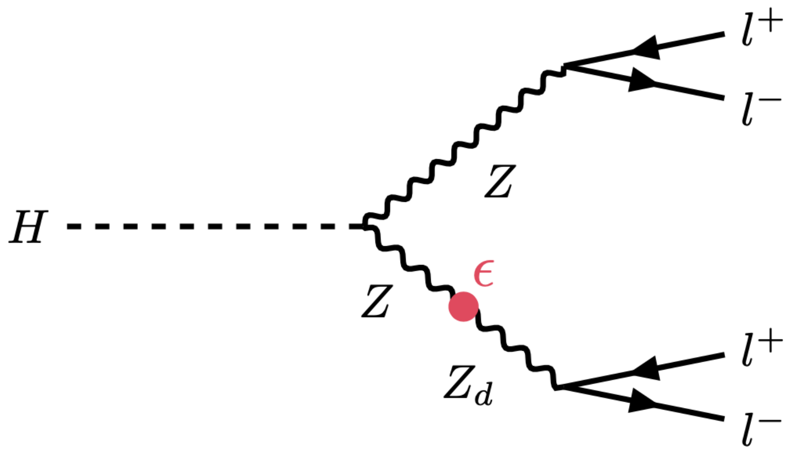
7 initial points + 3 predicted point



Production Interface



Dark Sector Search



Demonstrate active learning pipeline on one of the analysis channels from JHEP03(2022)041:

Search for Higgs bosons decaying into new spin-0 or spin-1 particles in four-lepton final states with the ATLAS detector with 139 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV

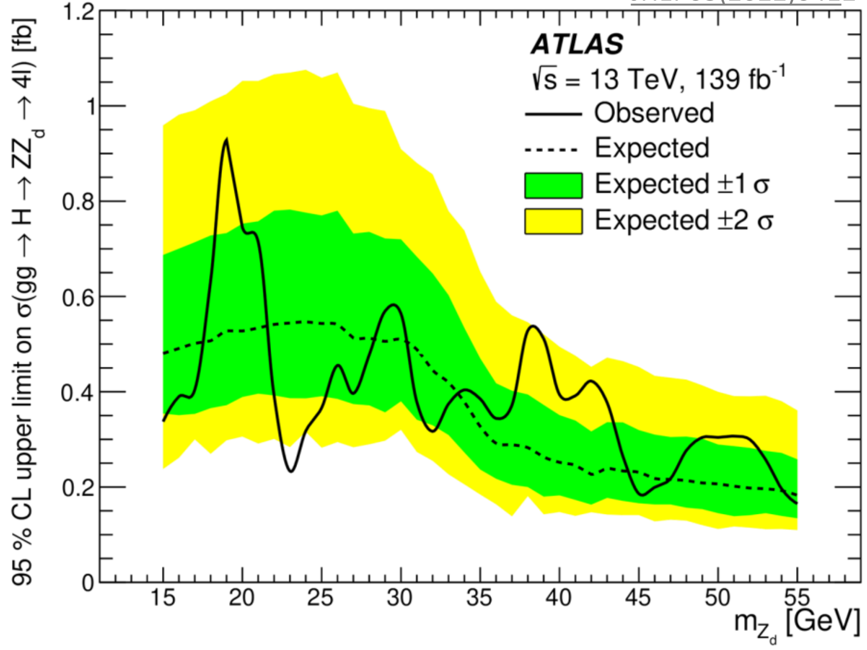
Low-mass (LM)	$H \to Z_d Z_d \to 4\mu$	$1\text{ GeV} < m_{Z_d} < 15\text{ GeV}$	targeting Higgs portal
High-mass (HM)	$H \to Z_d Z_d \to 4\ell$	$15\text{ GeV} < m_{Z_d} < 60\text{ GeV}$	targeting Higgs portal
Single Z boson (ZX)	$H \to Z Z_d \to 4\ell$	$15\text{ GeV} < m_{Z_d} < 55\text{ GeV}$	targeting hypercharge portal

Focusing on ZX channel

- Original analysis performed 1D scan in m_{Z_d}
- Extend to 2D scan in m_{Z_d} and ϵ

JHEP03(2022)041

Verify MC Production & Analysis Chain

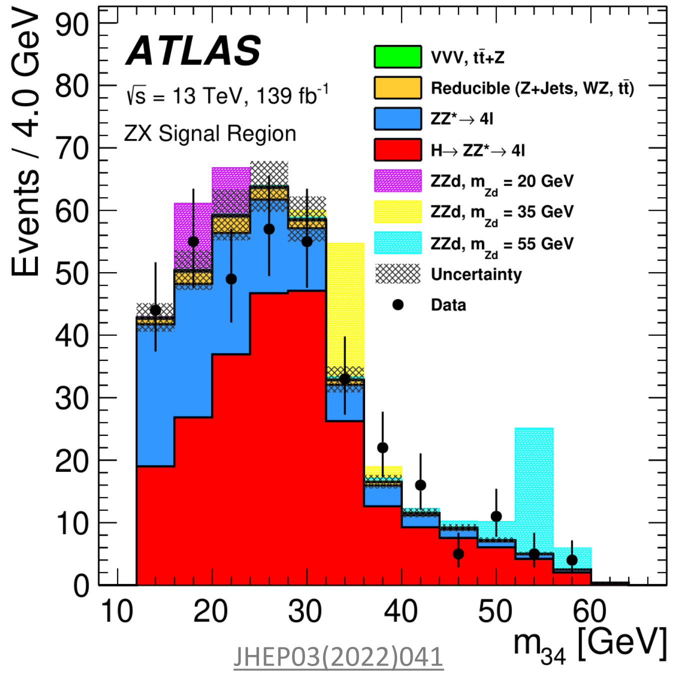


Slight difference in MC Production and Analysis setup with respect to published analysis at: JHEP03(2022)041

- Slightly different MC production setup
- Smaller amount of MC statistics 10k vs ~200k

Check against published cross section limit values on HEPData

m_{Z_d} [GeV]	HepData Obs. [fb]	HepData Exp. [fb]	This work Obs. [fb]	This work Exp. [fb]	Ratio of this work to HepData Obs. [fb]	Ratio of this work to HepData Exp. [fb]
15	0.34	0.48	0.32	0.48	0.96	1.00
20	0.74	0.53	0.73	0.51	0.97	0.96
25	0.37	0.54	0.37	0.55	1.01	1.01
30	0.56	0.51	0.57	0.52	1.01	1.02
35	0.39	0.33	0.39	0.34	1.01	1.03
40	0.39	0.25	0.39	0.26	1.00	1.03
50	0.30	0.21	0.30	0.20	0.99	0.97



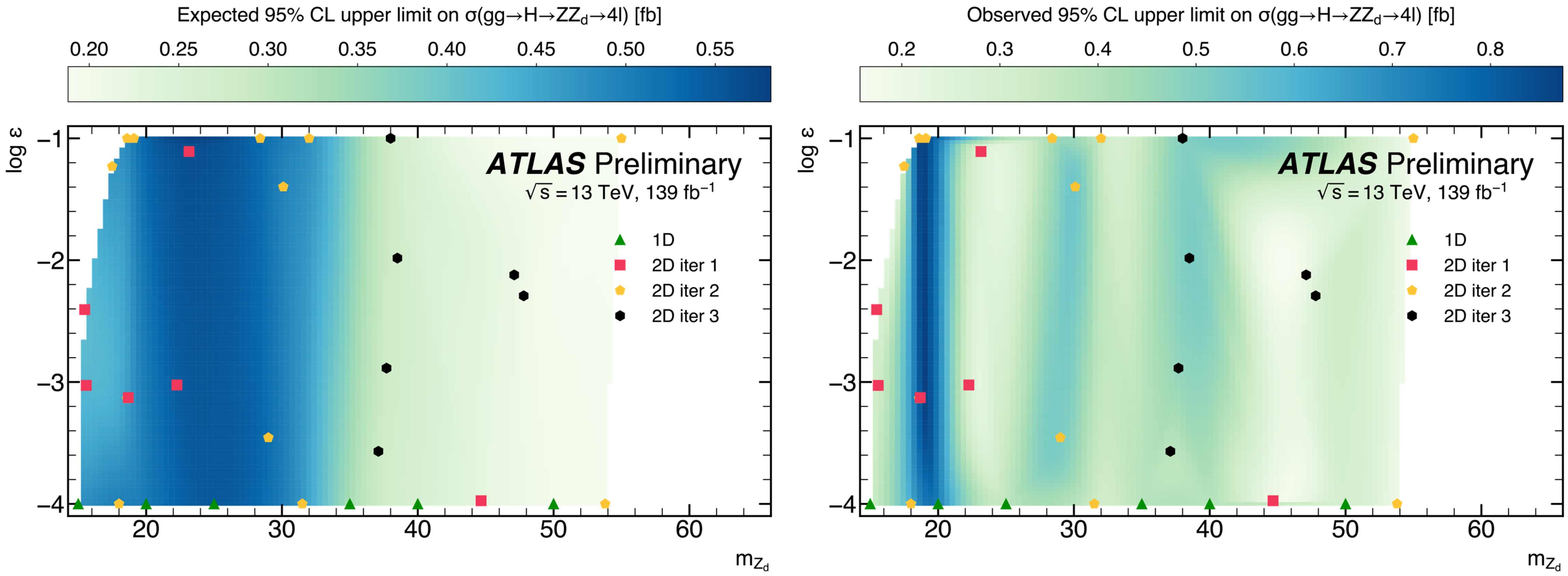
JHEP03(2022)041

Agreement within 4% - good!

2D Demonstration

Parameter phase space bounded by
- $15 \leq m_{Z_d} \leq 55$ GeV
- $10^{-4} \leq \epsilon \leq 10^{-1}$

Run in 4 iterations with 30 points total



ATL-PHYS-PUB-2023-010

Heavy Higgs H to WW

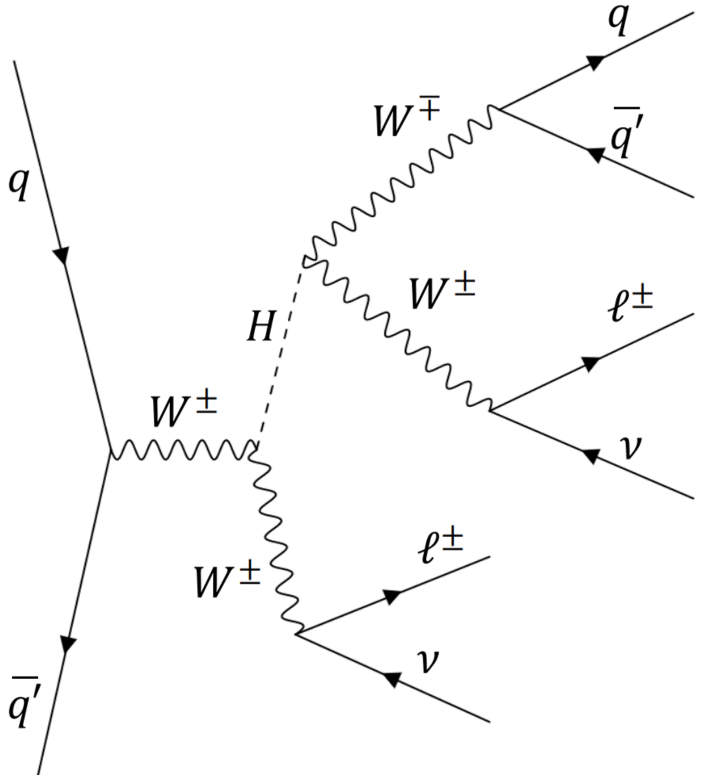
Published here: arXiv:2211.02617

Search for process $W^{\pm}H \rightarrow W^{\pm}W^{\pm}W^{\mp} \rightarrow \ell^{\pm}\nu\; \ell^{\pm}\nu\; jj$
Assume no specific model, instead consider effective Lagrangian up to dimension 6

$$\mathcal{L}^{(4)}_{hWW} = \rho_h g m_W h W^{\mu} W_{\mu},$$
$$\mathcal{L}^{(4)}_{hZZ} = \rho_h \frac{g m_W}{2c^2_W} h Z^{\mu} Z_{\mu},$$
$$\mathcal{L}^{(4)}_{HWW} = \rho_H g m_W H W^{\mu} W_{\mu},$$
$$\mathcal{L}^{(4)}_{HZZ} = \rho_H \frac{g m_W}{2c^2_W} H Z^{\mu} Z_{\mu},$$
$$\mathcal{L}^{(6)}_{HWW} = \rho_H g m_W \frac{f_W}{2\Lambda^2}\left(W^{+}_{\mu\nu} W^{-\mu}\partial^{\nu} H + h.c.\right) - \rho_H g m_W \frac{f_{WW}}{\Lambda^2} W^{+}_{\mu\nu} W^{-\mu\nu} H,$$
$$\mathcal{L}^{(6)}_{HZZ} = \rho_H g m_W \frac{c^2_W f_W + s^2_W f_B}{2c^2_W \Lambda^2} Z_{\mu\nu} Z^{\mu}\partial^{\nu} H - \rho_H g m_W \frac{c^4_W f_{WW} + s^4_W f_{BB}}{2c^2_W \Lambda^2} Z_{\mu\nu} Z^{\mu\nu} H,$$

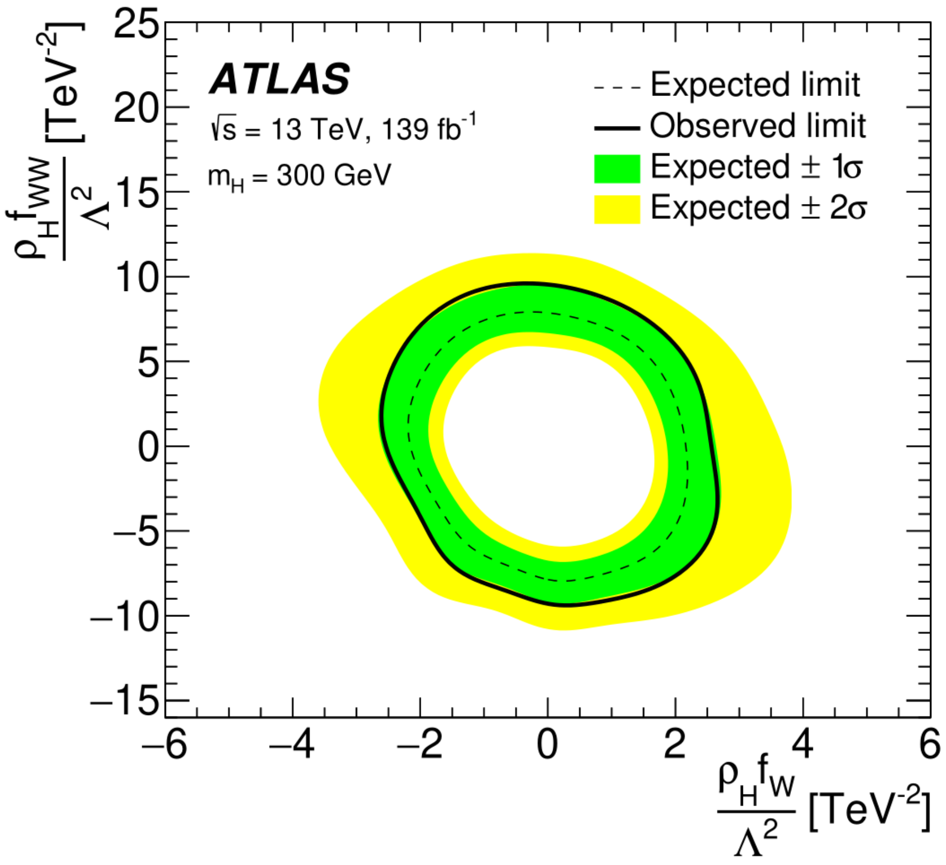
HWW couplings parametrized by couplings

$$\rho_H, f_W, f_{WW}$$



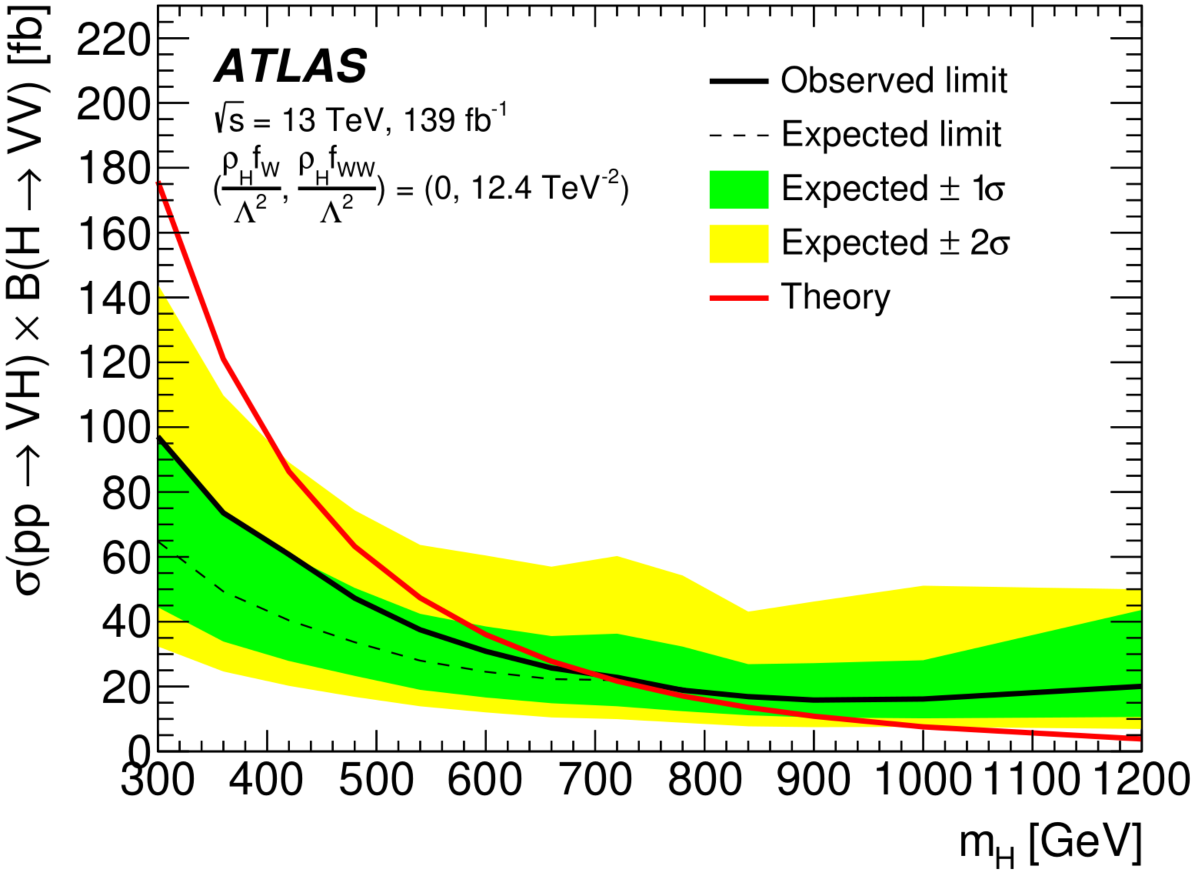
Heavy Higgs Results

Two-dimensional results in the f_w-f_{ww} plane for $m_H = 300, 600, 900$ GeV



One-dimensional results as a function of m_H for two combinations of f_w and f_{ww}:

$$\left(\frac{\rho_H f_w}{\Lambda^2}, \frac{\rho_H f_{ww}}{\Lambda^2}\right) = (0, 12.4/\mathrm{TeV}^2), (2.7/\mathrm{TeV}^2, 0)$$



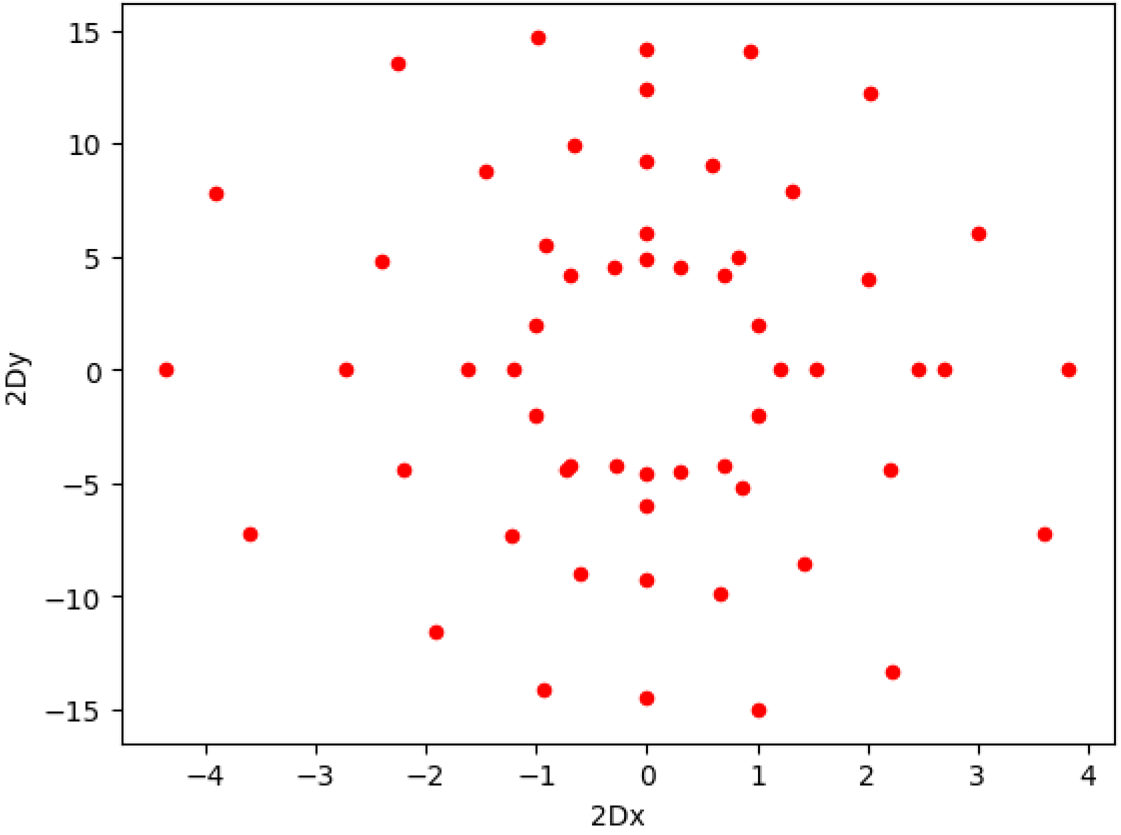
ρ_H set to 0.05, Λ set to 5 TeV in this study

Obtain cross section theory predictions from heavy SM like Higgs

Signal samples

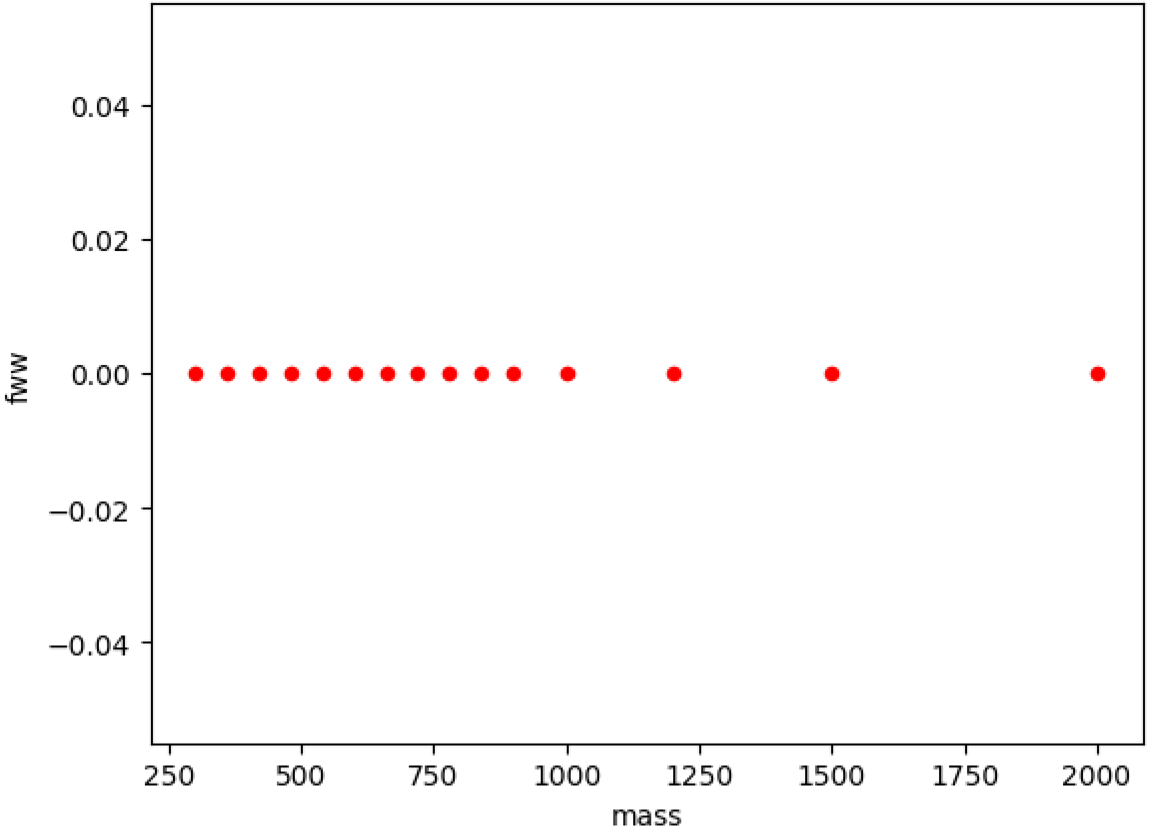
Generated signal samples in 16 directions for f_w-f_{ww} plane limits

$\times 3$

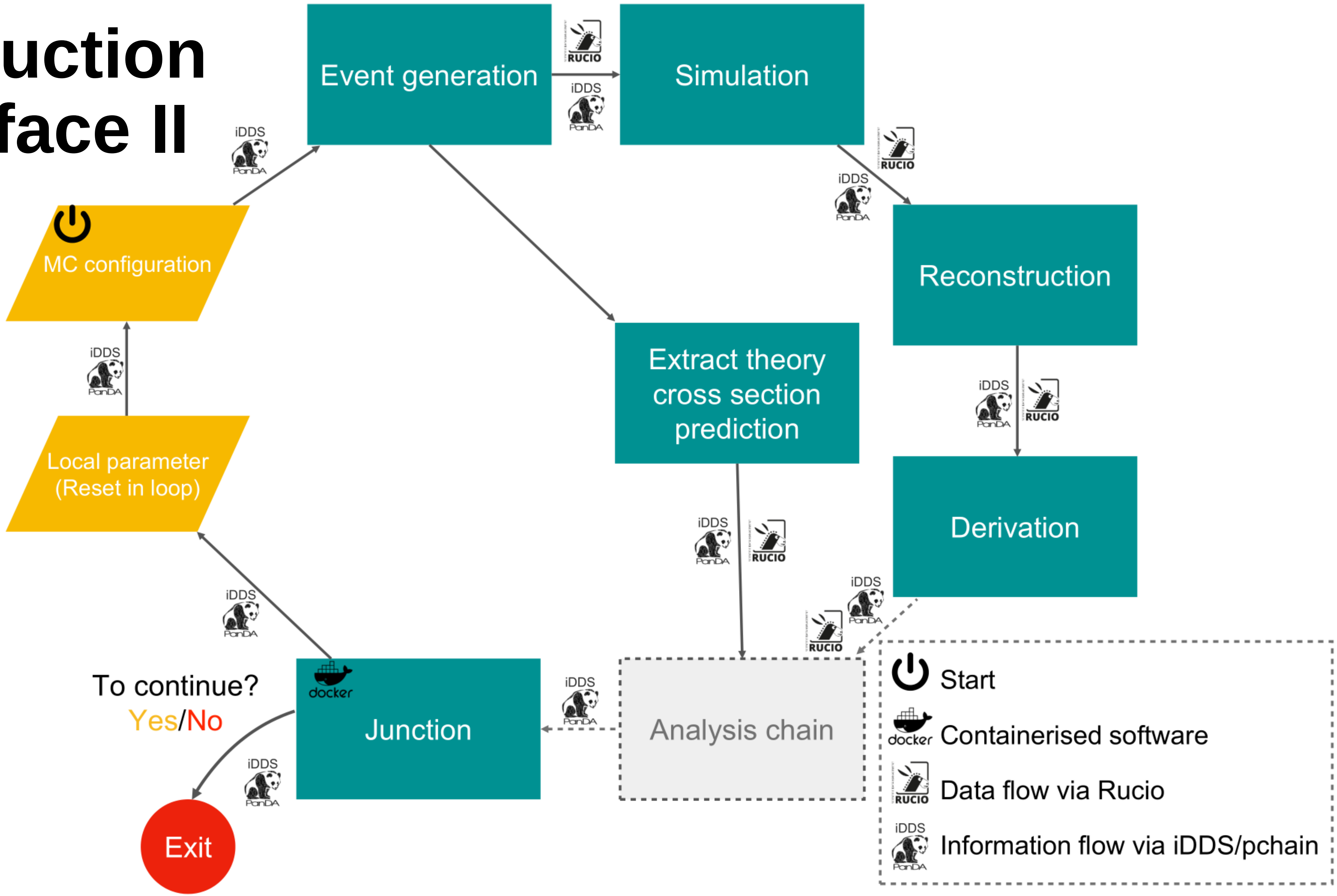


Study relied on 196 different signal samples in total

$\times 2$



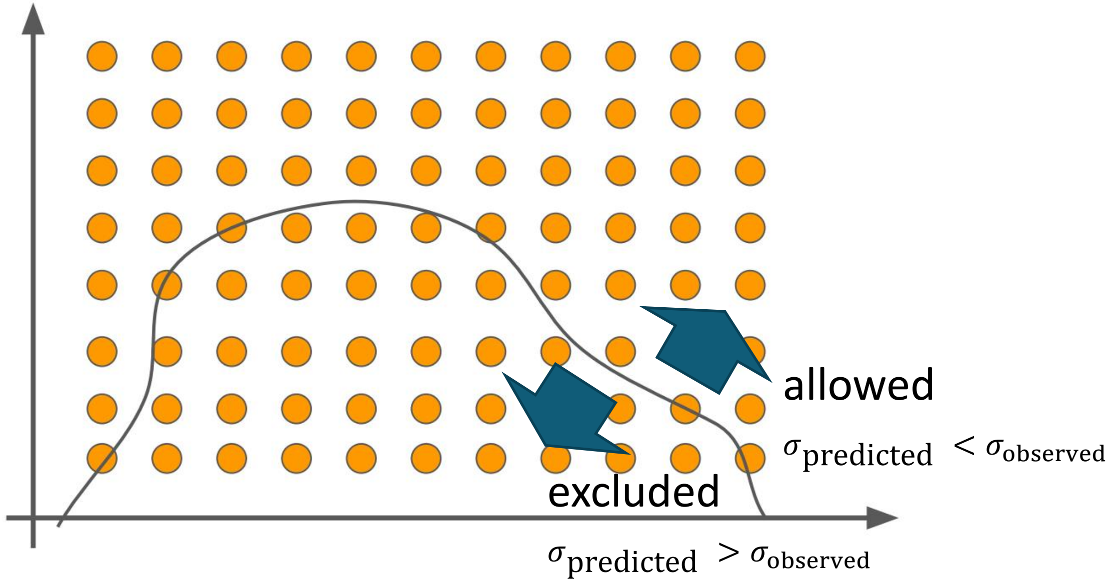
Production Interface II



Active Learning for Excursion Set Estimation

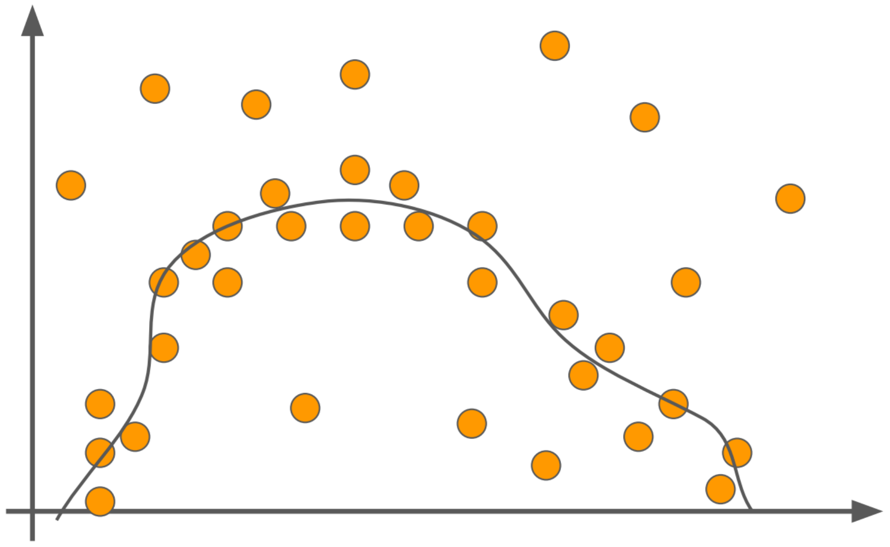
Kyle Cranmer, Lukas Heinrich, Gilles Louppe

Presume we conduct a grid search with the goal to find an exclusion contour



Active Learning for Excursion Set Estimation

Ideally, we would like to have an active learning method that optimizes the exclusion contour, predicting phase space points that help in that task.



Active Learning for Excursion Set Estimation

Cranmer et. al have developed such an active learning method: https://github.com/diana-hep/excursion

Entropy based acquisition function

Physics parameter θ, $\sigma_{obs}^{upperLimit} \equiv \sigma(\theta)_{\text{obs}}^{\text{upperLimit}}$, $\sigma_{\text{theory}} \equiv \sigma(\theta)_{\text{theory}}$

Entropy based acquisition function

Physics parameter θ, $\sigma_{obs}^{upperLimit} \equiv \sigma(\theta)_{\mathrm{obs}}^{\mathrm{upperLimit}}$, $\sigma_{\mathrm{theory}} \equiv \sigma(\theta)_{\mathrm{theory}}$

θ is excluded if $\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}} \leq \sigma_{\mathrm{theory}}$ or $y = \frac{\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}}}{\sigma_{\mathrm{theory}}} \leq 1$

So find contour where $\frac{\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}}}{\sigma_{\mathrm{theory}}} = 1$

Entropy based acquisition function

Physics parameter θ, $\sigma_{obs}^{upperLimit} \equiv \sigma(\theta)_{\mathrm{obs}}^{\mathrm{upperLimit}}$, $\sigma_{\mathrm{theory}} \equiv \sigma(\theta)_{\mathrm{theory}}$

θ is excluded if $\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}} \leq \sigma_{\mathrm{theory}}$ or $y = \frac{\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}}}{\sigma_{\mathrm{theory}}} \leq 1$

So find contour where $\frac{\sigma_{\mathrm{obs}}^{\mathrm{upperLimit}}}{\sigma_{\mathrm{theory}}} = 1$

Gaussian Process maps every θ onto a Gaussian: $y(\theta)$ is distributed as $G(y(\theta)|\,\vec{\mu}, \boldsymbol{K})$

$$P(\theta \text{ is excluded}) = P_{\mathrm{excl}}(\theta) = \int_{-\infty}^{1} dy\, G(y(\theta)|\,\vec{\mu}, \boldsymbol{K})$$

Entropy based acquisition function

Physics parameter θ, $\sigma_{obs}^{upperLimit} \equiv \sigma(\theta)_{\text{obs}}^{\text{upperLimit}}$, $\sigma_{\text{theory}} \equiv \sigma(\theta)_{\text{theory}}$

θ is excluded if $\sigma_{\text{obs}}^{\text{upperLimit}} \leq \sigma_{\text{theory}}$ or $y = \frac{\sigma_{\text{obs}}^{\text{upperLimit}}}{\sigma_{\text{theory}}} \leq 1$

So find contour where $\frac{\sigma_{\text{obs}}^{\text{upperLimit}}}{\sigma_{\text{theory}}} = 1$

Gaussian Process maps every θ onto a Gaussian: $y(\theta)$ is distributed as $G(y(\theta)|\,\vec{\mu}, \boldsymbol{K})$

$$P(\theta \text{ is excluded}) = P_{\text{excl}}(\theta) = \int_{-\infty}^{1} dy\, G(y(\theta)|\,\vec{\mu}, \boldsymbol{K})$$

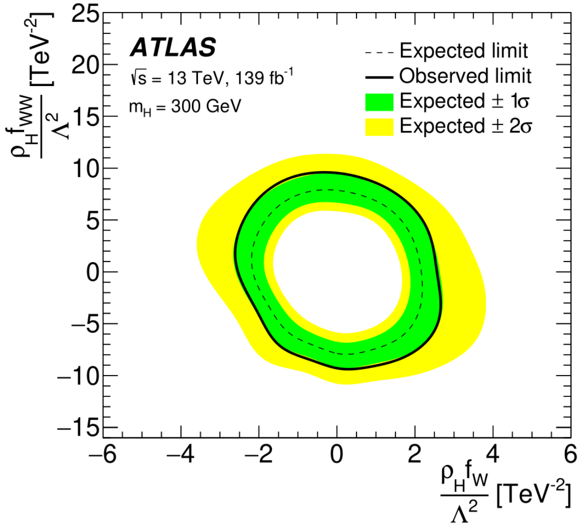
Entropy

$$H(\theta) = -\sum_{x \in X} p(x) \log\big(p(x)\big)$$

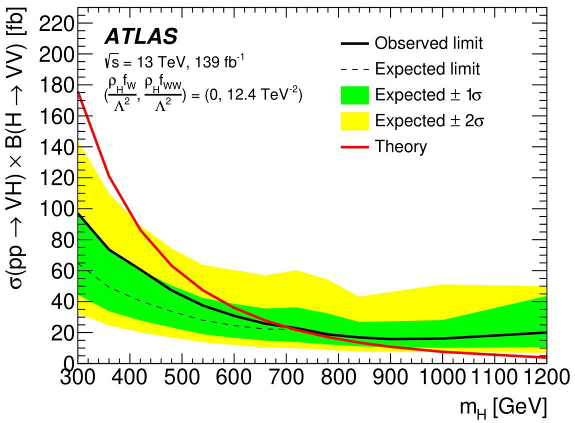
$$H(\theta) = -P_{\text{excl}}(\theta) \log\big(P_{\text{excl}}(\theta)\big) - \big(1 - P_{\text{excl}}(\theta)\big) \log\big(1 - P_{\text{excl}}(\theta)\big)$$

Select new points to explore where $H(\theta)$ is maximal

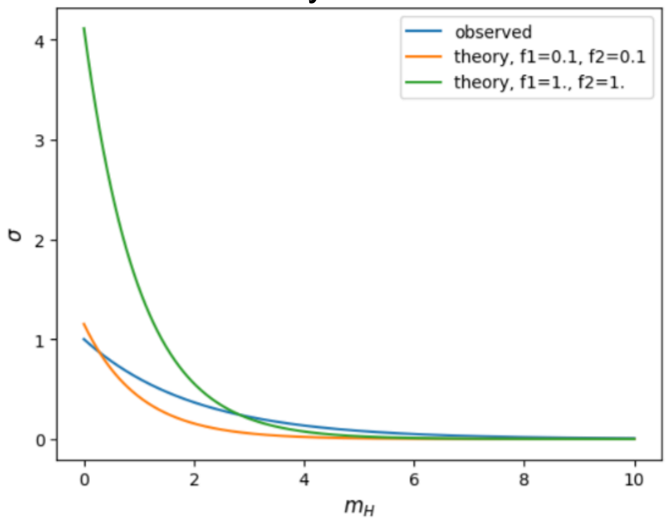
Prototype updated Active Learning step



Approximate topology of observed and theory limits in ρ_H, f_w, f_{ww} space with dummy functions

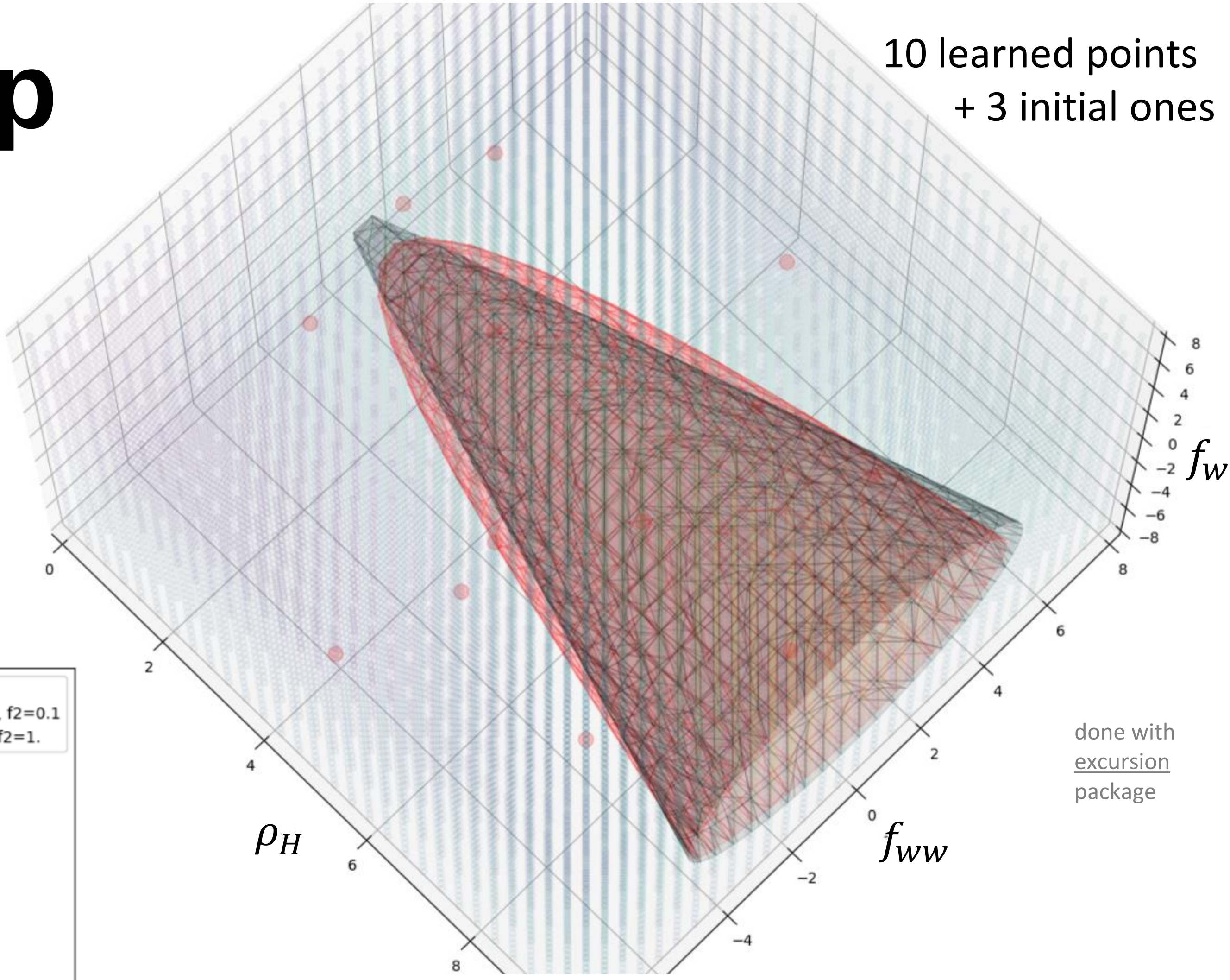


‘dummy functions’



Grey Mesh: true dummy exclusion contour

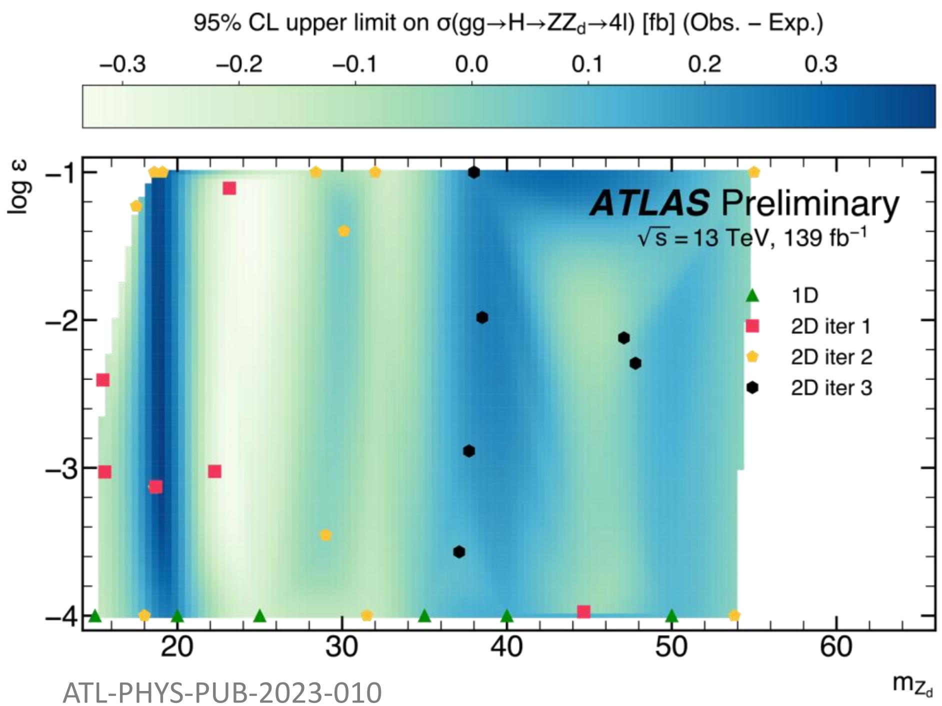
Red Mesh: surrogate exclusion contour



Use dummy function to test the ability of the active learning process to approximate the topology of exclusion contour

Conclusion

- Demonstrated active learning driven re-analysis for published dark sector analysis – see ATLAS PubNote ATL-PHYS-PUB-2023-010
- Extended ZX analysis channel from 1D to $2D$ parameter space in m_{Z_d} and ϵ
- Many other applications and possibilities for this tool and workflow
- Documentation here
- Another ATLAS application of active learning using a different tool: *Active Learning reinterpretation of an ATLAS Dark Matter search constraining a model of a dark Higgs boson decaying to two b-quarks* ATL-PHYS-PUB-2022-045
- Preparing second demonstrator based on heavy Higgs Boson search
  - 3D parameter space
  - theory predictions to prioritize exclusion contours



ATL-PHYS-PUB-2023-010


- Demonstrated active learning driven re-analysis for published dark sector analysis – see ATLAS PubNote ATL-PHYS-PUB-2023-010
- Extended ZX analysis channel from 1D to $2D$ parameter space in m_{Z_d} and ϵ
- Many other applications and possibilities for this tool and workflow
- Documentation here
- Another ATLAS application of active learning using a different tool: *Active Learning reinterpretation of an ATLAS Dark Matter search constraining a model of a dark Higgs boson decaying to two b-quarks* ATL-PHYS-PUB-2022-045
- Preparing second demonstrator based on heavy Higgs Boson search
  - 3D parameter space

A search for heavy Higgs bosons decaying into vector bosons in same-sign two-lepton final states in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector
arXiv:2211.02617

The End.

Thank you!

Backup.