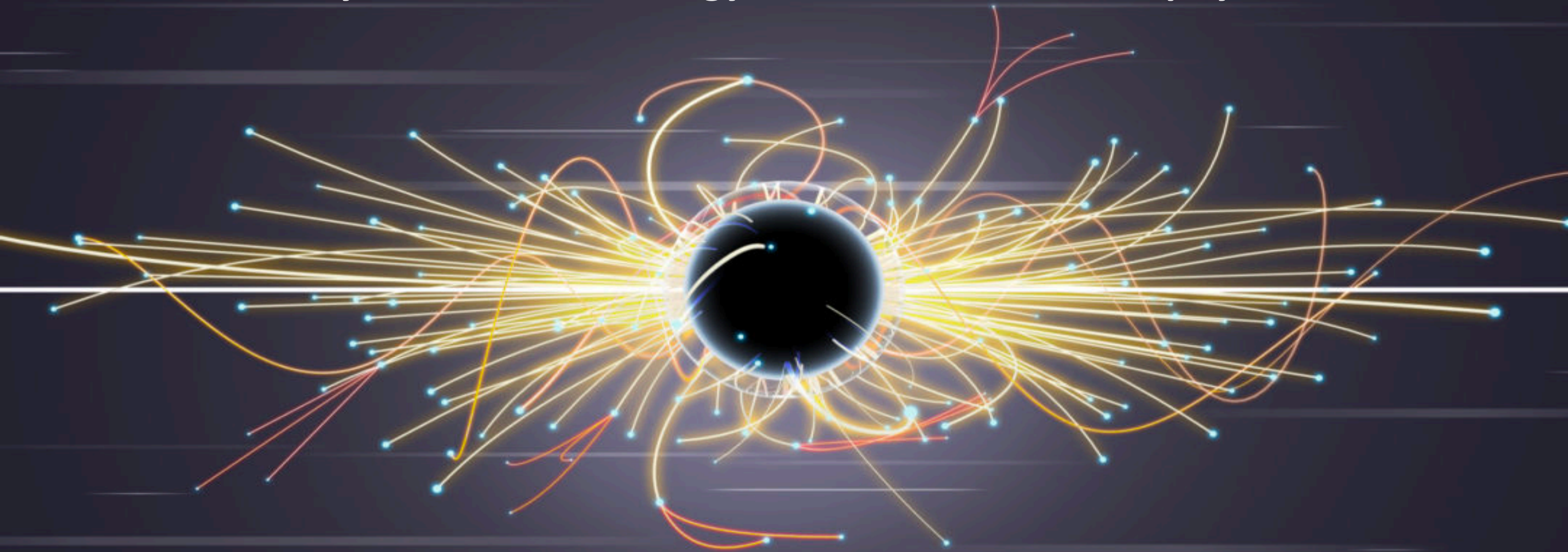
ML approaches for
Particle Physics Phenomenology @ Colliders and Astrophysics



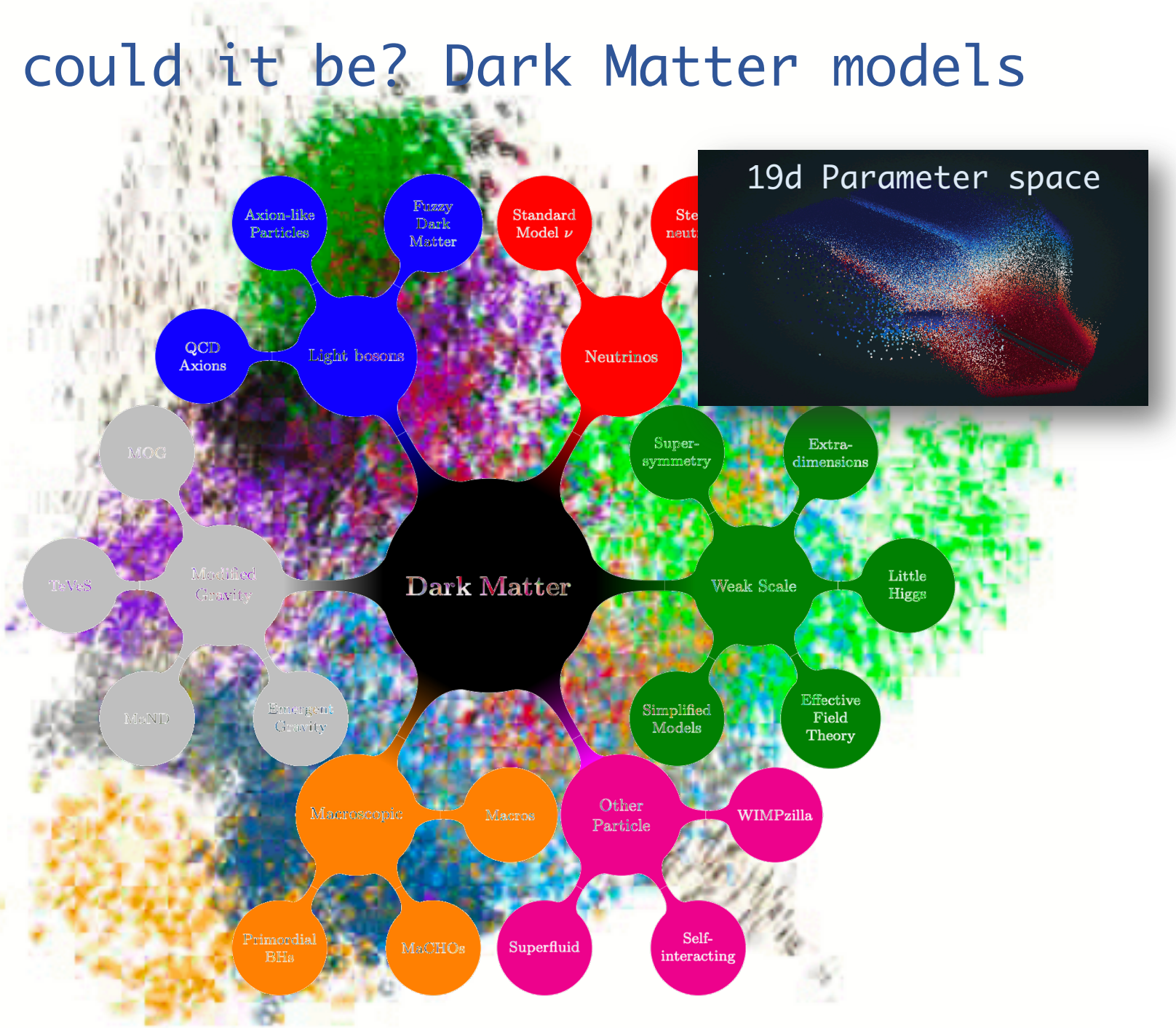
Sascha Caron and many others

Presented works are mainly within or in connection with darkmachines.org

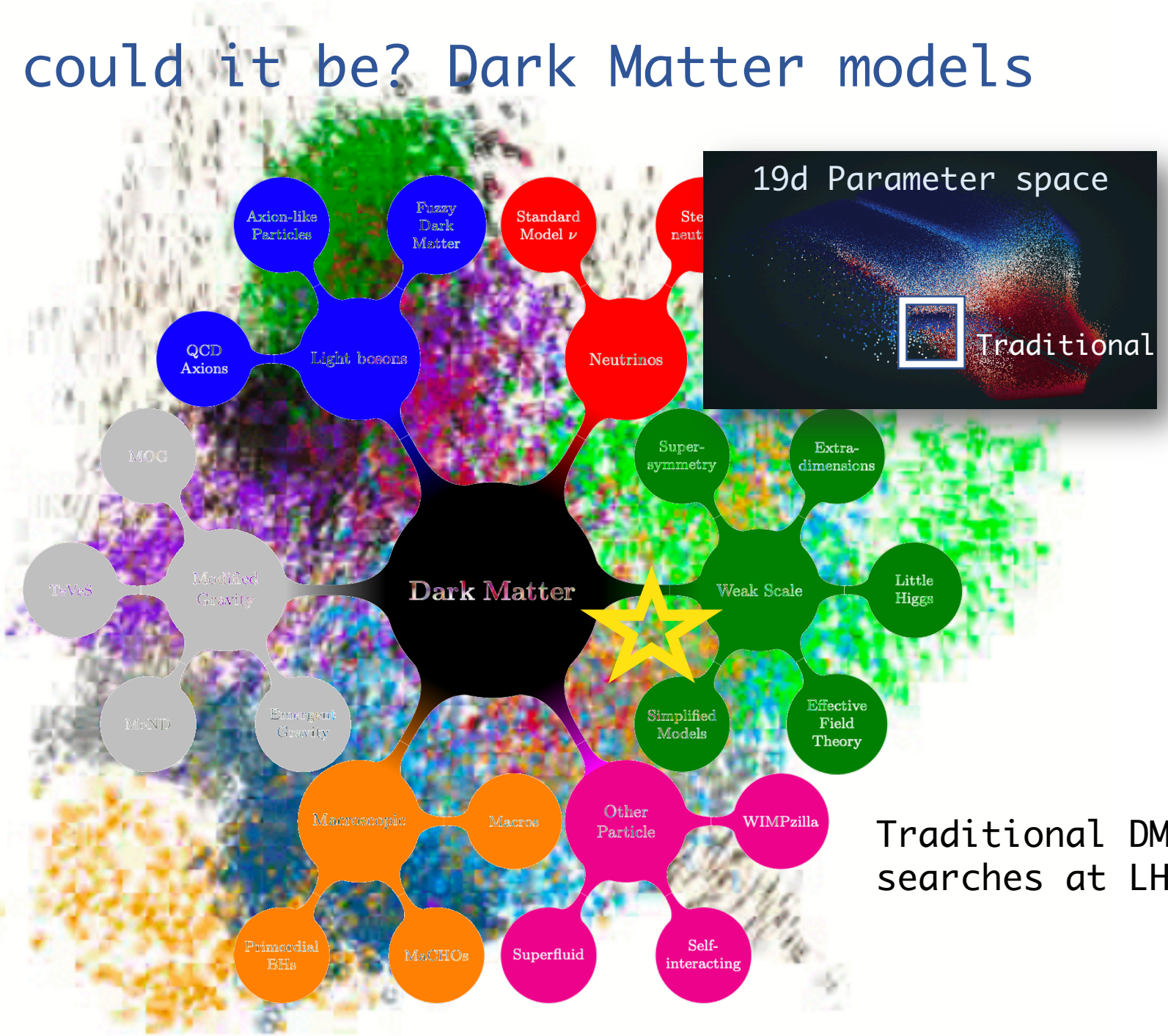
Picture taken from CERN

What could it be? Dark Matter models

What could it be? Dark Matter models



What could it be? Dark Matter models



Traditional DM/SUSY etc. searches at LHC

Some opportunities to address the wide range of physical opportunities ...

1. Learning physical models better without (2d) simplifications ?
2. Finding solutions in multimodal high-dimensional likelihood landscapes
3. Generators for HEP with generative ML models
4. New physics as an anomaly detection
5. Organising Challenges / Comparisons for the community
6. Computer Vision to determine physical parameters
7. … I stop when time is up…

Learning physical models without (2d) simplifications

Example 1

Practical Machine Learning for regression and classification and applications in HEP phenomenology

S. Caron, A. Coccaro, S. Ganguly, S. Kraml, A. Lessa, S. Otten, R. Ruiz, H. Reyes-González, R. Ruiz de Austri, B. Stienen, R. Torre

https://arxiv.org/abs/2002.12220

Aim: simplify the creation and reuse of ML models made for HEP phenomenology. For this we like to encourage to save the ML model, the HEP data set and the code to train the ML models.

Questions:

- **How good can we regress time-consuming cross sections for HEP models ?**
- **Learning of exclusion boundaries for high-dim models** (e.g. www.susy-ai.com)
- **Learning of likelihood for high-dim models**

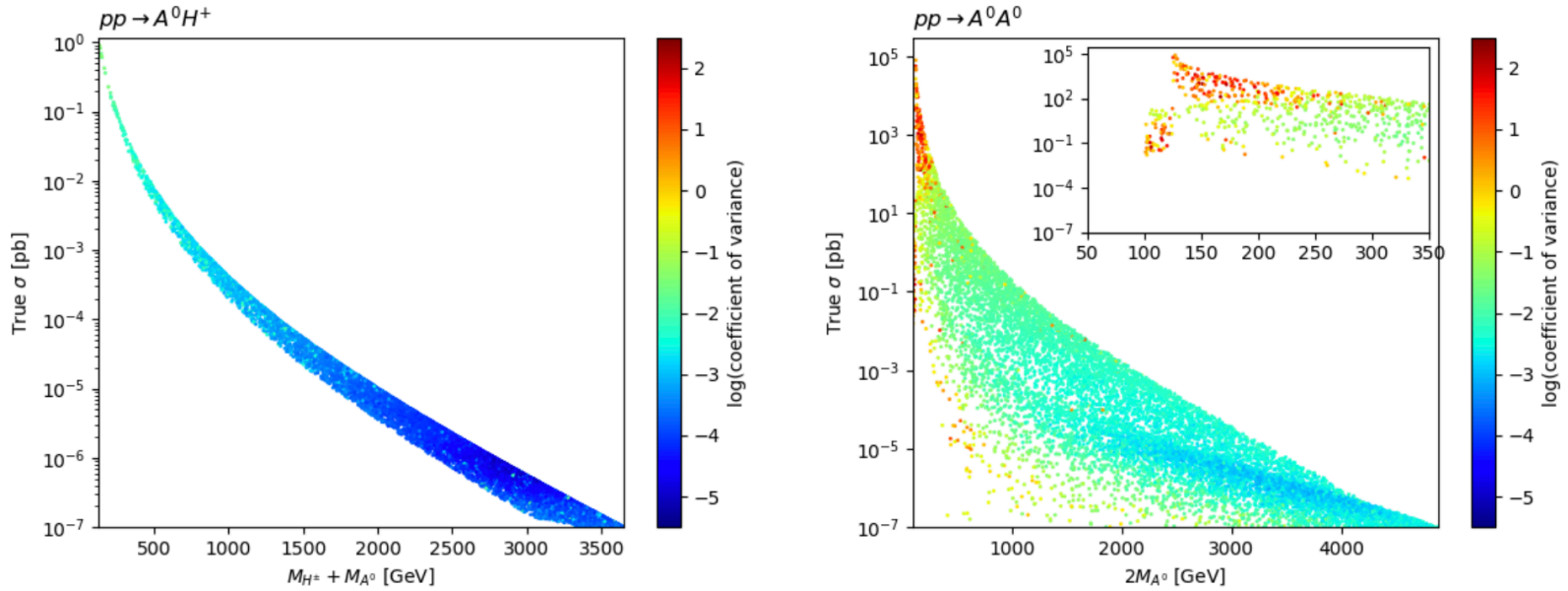
➔ Needs some change of publication strategy for HEP experiments and pheno/theory community !

Learning physical models without (2d) simplifications

https://arxiv.org/abs/2002.12220

Example cross sections…

Various examples , here best and worst case for heavy Higgs



(a) $pp \to A^0 H^+$ (b) $pp \to A^0 A^0$

Figure 4: True cross section vs. effective mass for the processes with (a) the best and (b) the worst performance. In color, the logarithm of the coefficient of variance is shown.

Learning physical models without (2d) simplifications

https://arxiv.org/abs/2002.12220

What is published

Put in here your favoured 2d likelihood / exclusion etc. plot

Learning physical models without (2d) simplifications

https://arxiv.org/abs/2002.12220

What we could publish

Model parameter set 1, likelihood
Model parameter set 2, likelihood…
… 1000s of more lines…

Learning physical models without (2d) simplifications

https://arxiv.org/abs/2002.12220

What can we do with this data ?

Train a ML model to predict

Likelihood (20d Model-parameters)

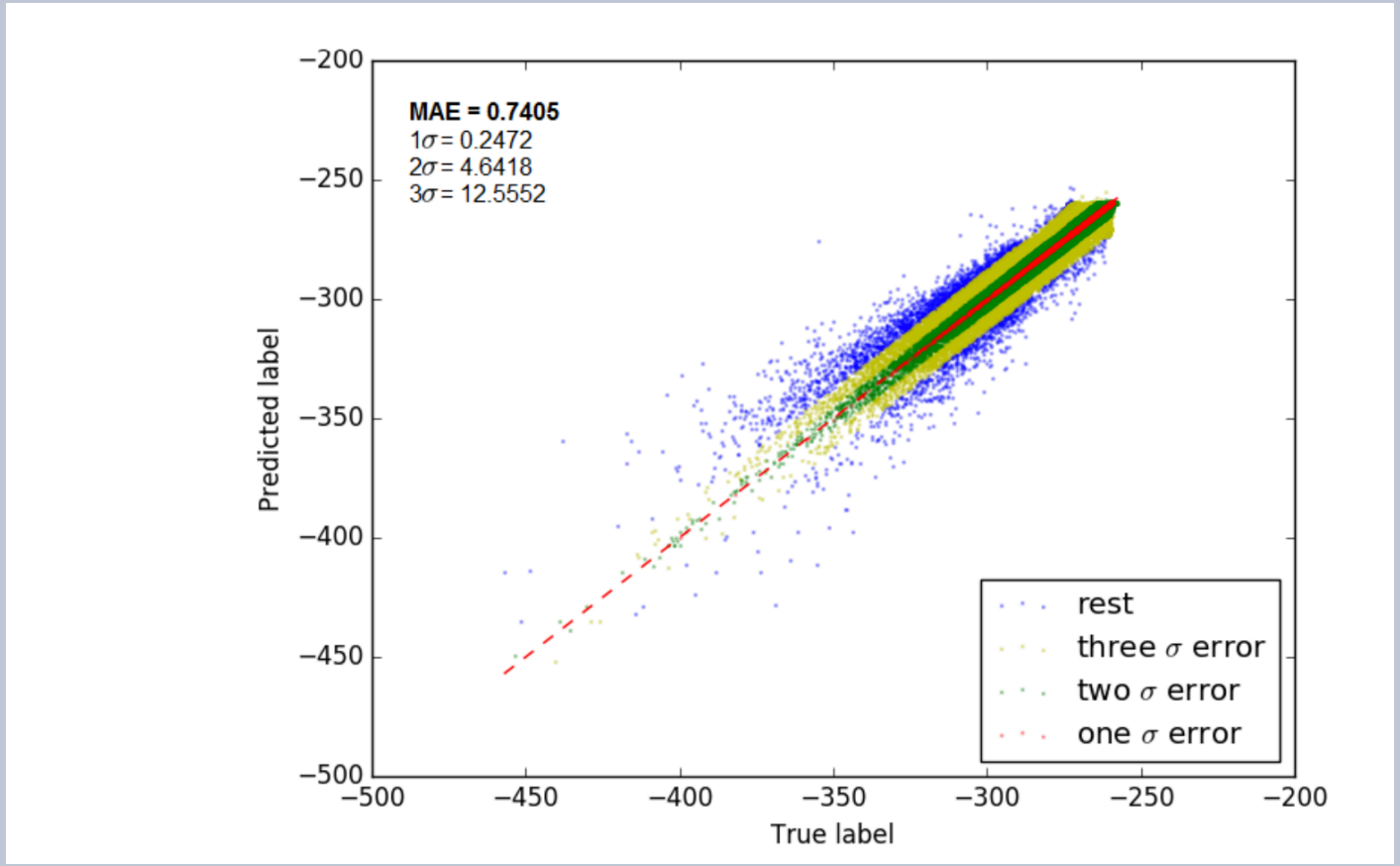
or even better

Likelihood (Model, 20d Model-parameters)

Learning physical models without (2d) simplifications

https://arxiv.org/abs/2002.12220

Example:
Global fits of Gambit
Zenodo data



A seven-dimensional Minimal Supersymmetric Standard Model (MSSM7)

A total of 22.6 million samples were used for the training and evaluation of the models. Data exploration reveals that ≈ 595000 of those samples have a likelihood of 0 whereas all the other samples range from ≈ −450 to −255.

Example 2

Finding solutions in high-d likelihood landscape

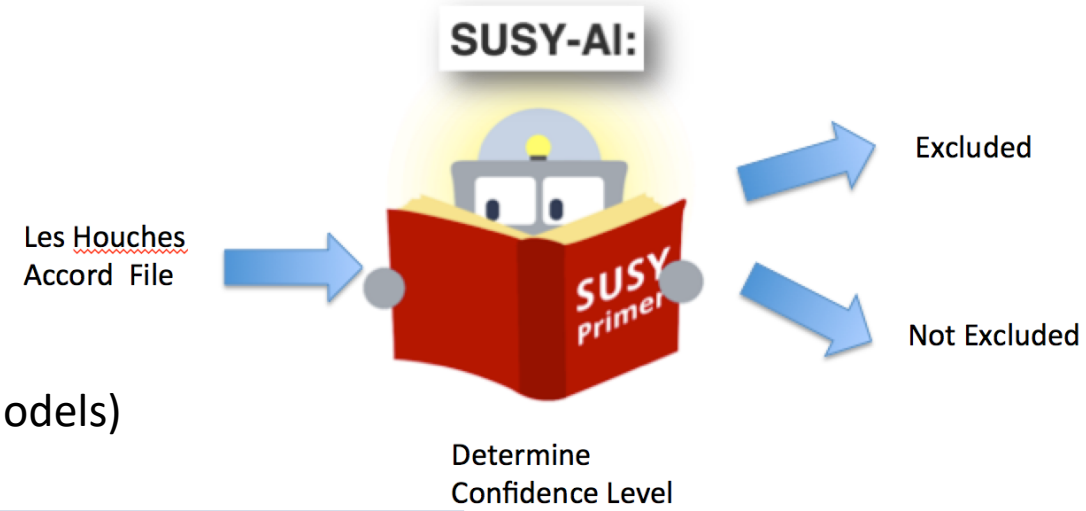
How can we find solutions in realistic
(i.e. high-dimensional) physical models → upcoming darkmachines paper

Interested in improved classification?
→ See next slides

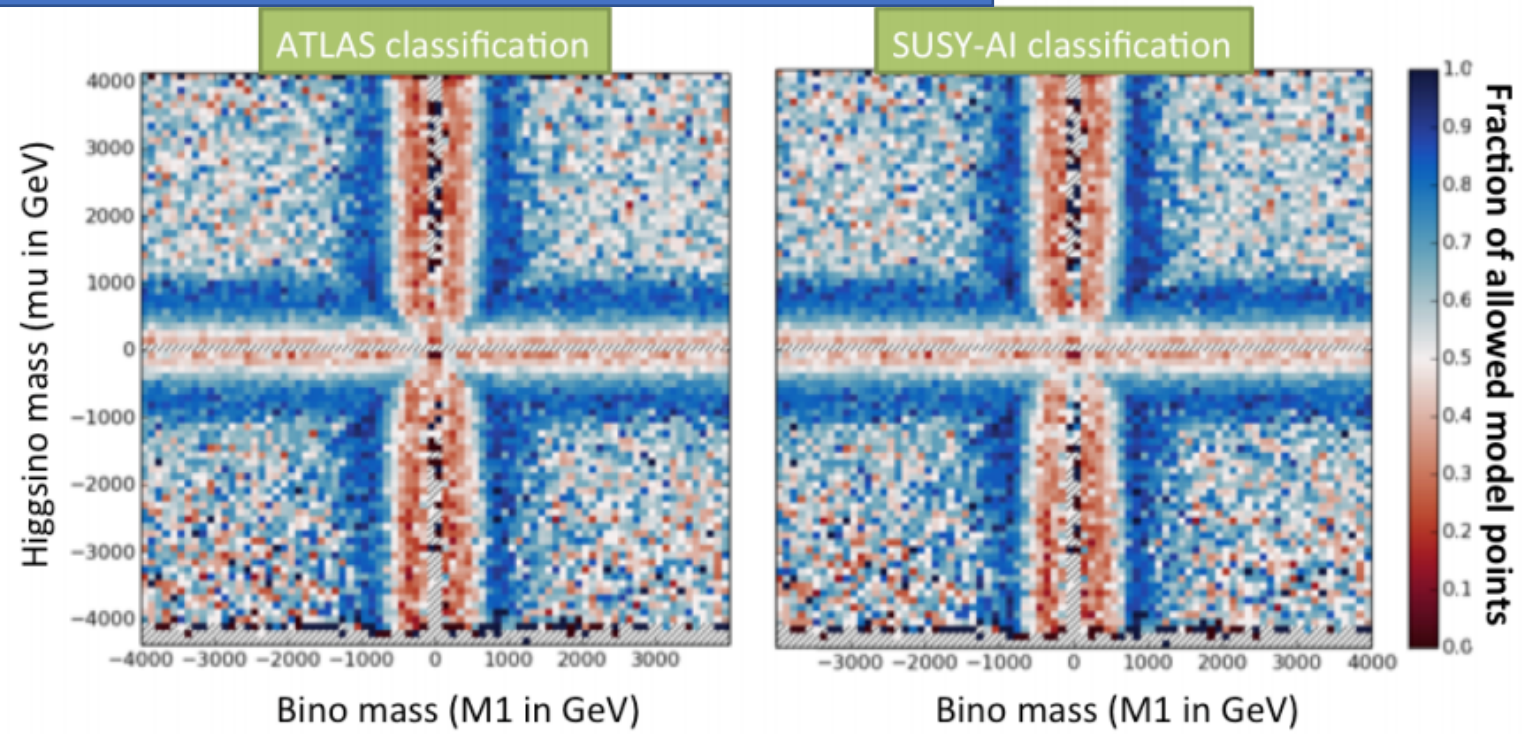
SUSY-AI

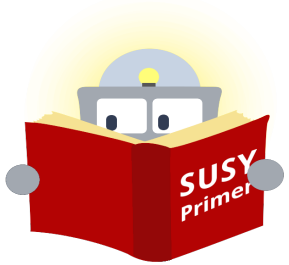
Encoding of model constraints with Machine Learning

Aim: Generic framework (**all** models)



Testing with out-of-bag estimation (remember 0.68!)



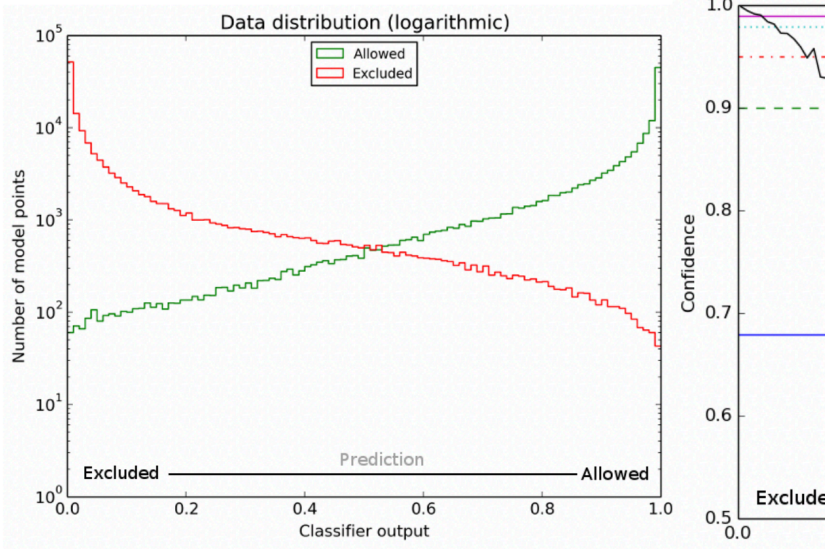


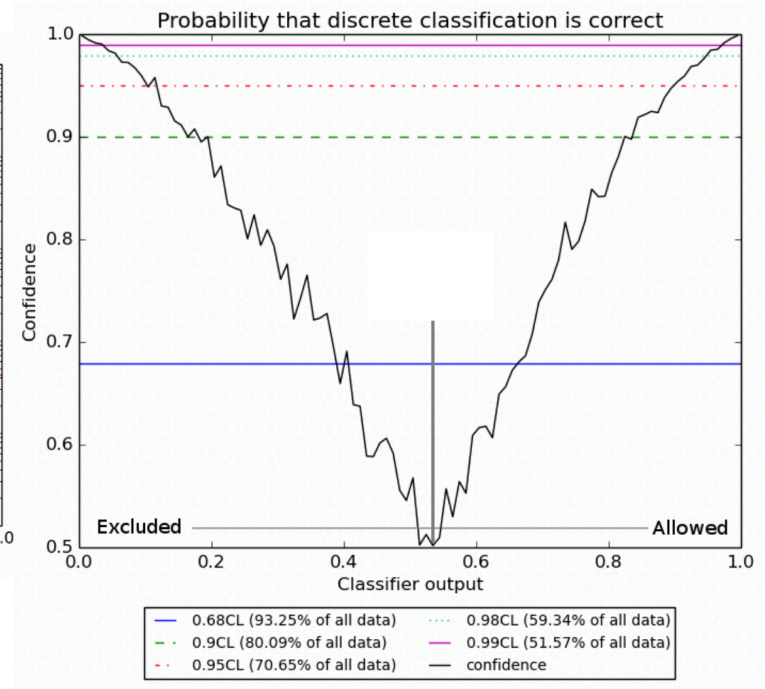
Used training data to learn classification

It determines a **confidence** level of its **classification** using the training data.

→ Need **more points** in regions of low certainty

Ratio of majority class per bin





Active learning

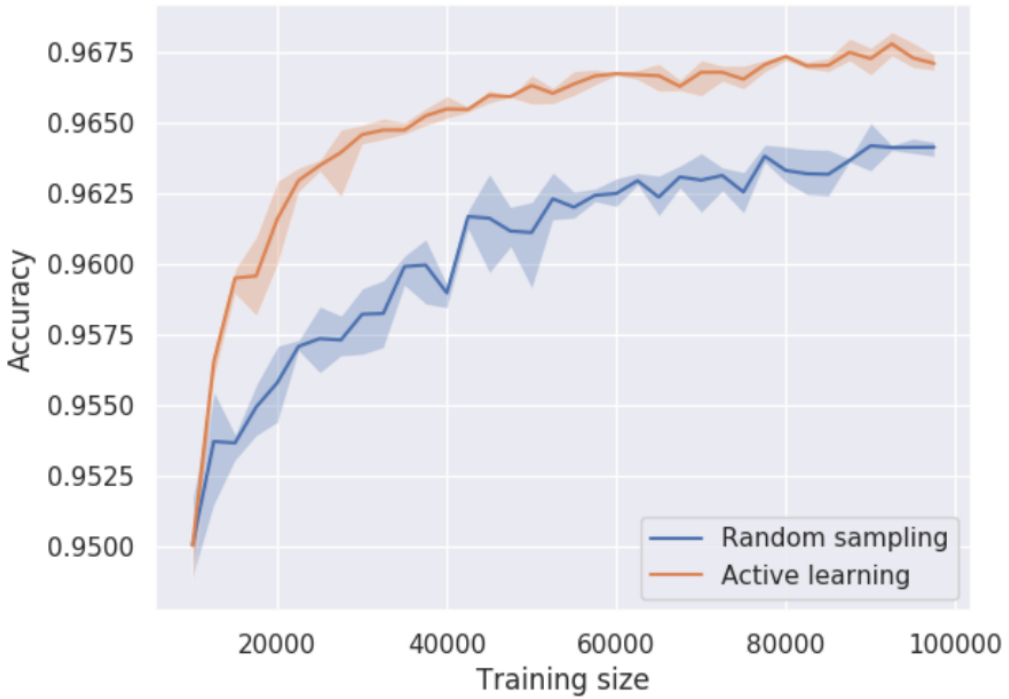


FIG. 5. Accuracy development on model exclusion of the 19-dimensional model for new physics (pMSSM) for random sampling and active learning using a dropout Neural Network with infinite pool. True labeling was provided by a machine learning algorithm trained on model points and labels provided by ATLAS [1]. The gain of active learning with respect to random sampling (as described by Equation 2) is 3 to 4. The bands show the range in which all curves of that colour lay when the experiment was repeated 7 times.

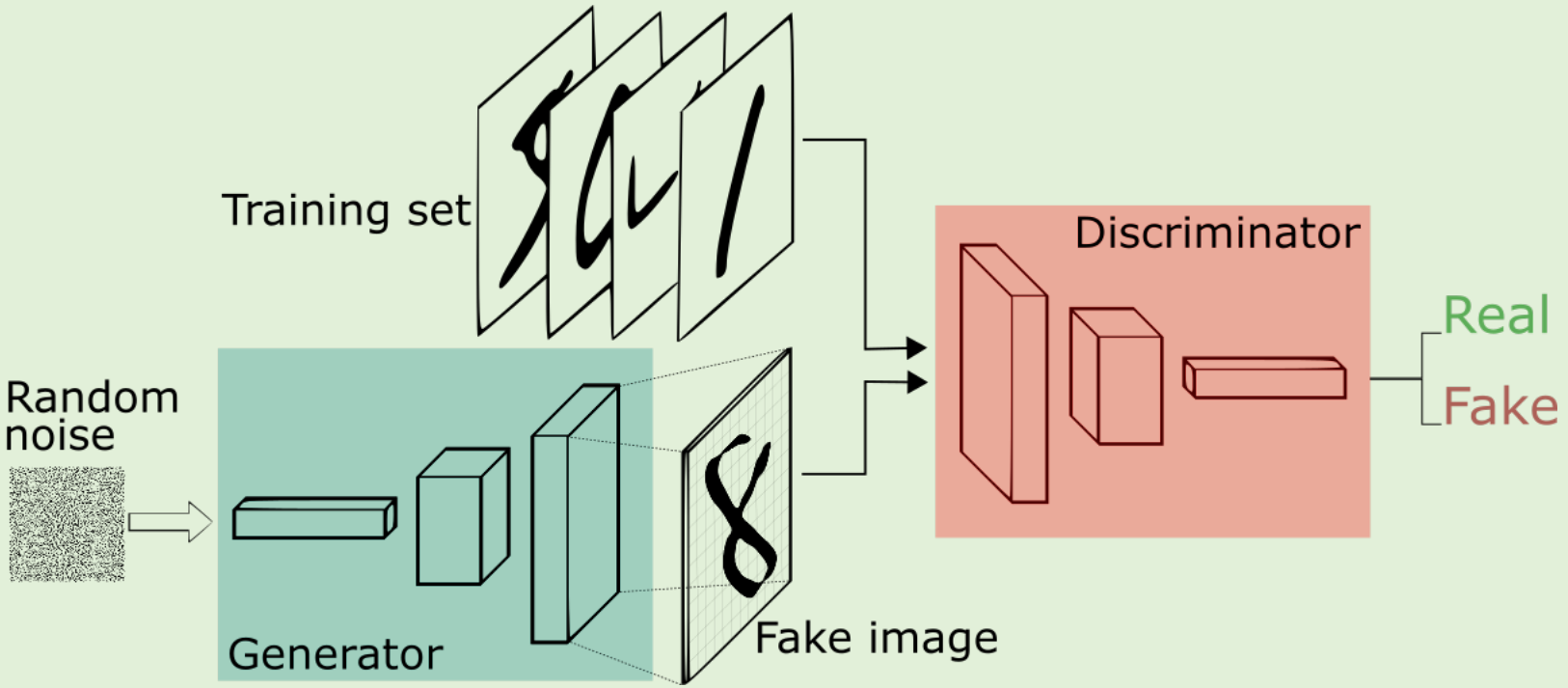
arXiv:1905.08628 , mainly
Bob Stienen

Query-by-Dropout-Committee

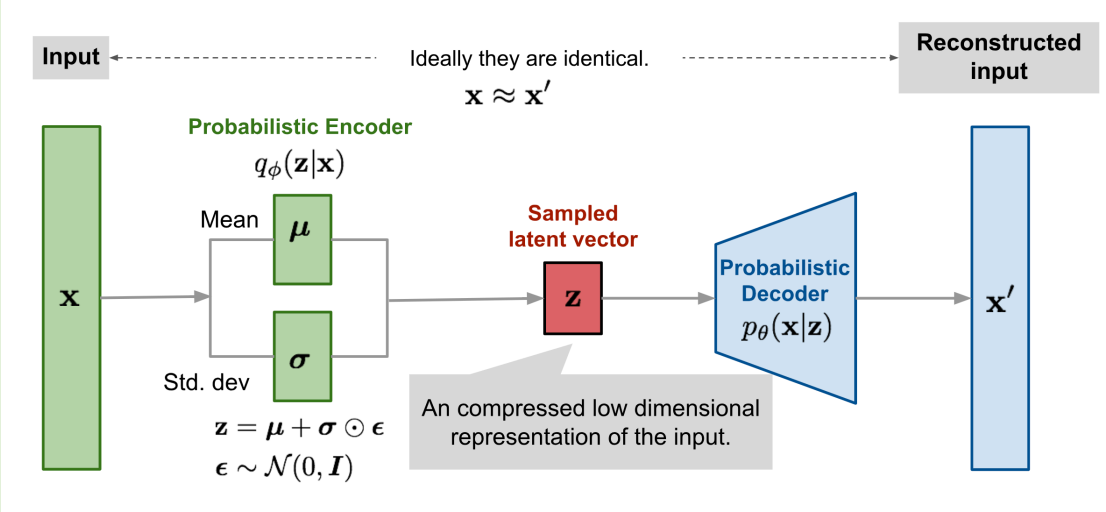
Example 3

Network based generators for HEP

Generative Adversarial Networks (Gans), autoencoders, flow models



Details: B-VAE



1901.00875

Together with an “information buffering” of the latent space we could find optimal event sampling properties for a B-VAE.

The “buffer” collects observations $Z = \{z_1, \ldots, z_m\}$ by sampling $q_\phi(z|X_L)$ where $X_L \subset X$ is a subset of real Monte Carlo events. The distribution over the latent vector z is then given by:

$$p_{\phi,X_L}(z) = \sum_{i=1}^{m} q_\phi(z|x^i)p(x^i) \text{ with } \mathrm{p}(\mathrm{x}^\mathrm{i}) = \frac{1}{\mathrm{m}}.$$

To avoid overtraining the “information buffering” needs to introduce new hyperparameters alpha and gamma, allowing more variance in sampling the latent vector z via:

$$z \sim q_\phi(z|X_L)$$

$$z^i \sim \begin{cases} \mathcal{N}(\mu^i, \alpha\sigma^{2,i}) & \text{if } \sigma < \sigma_T \\ \mathcal{N}(\mu^i, \sigma^{2,i}) & \text{else} \end{cases},$$

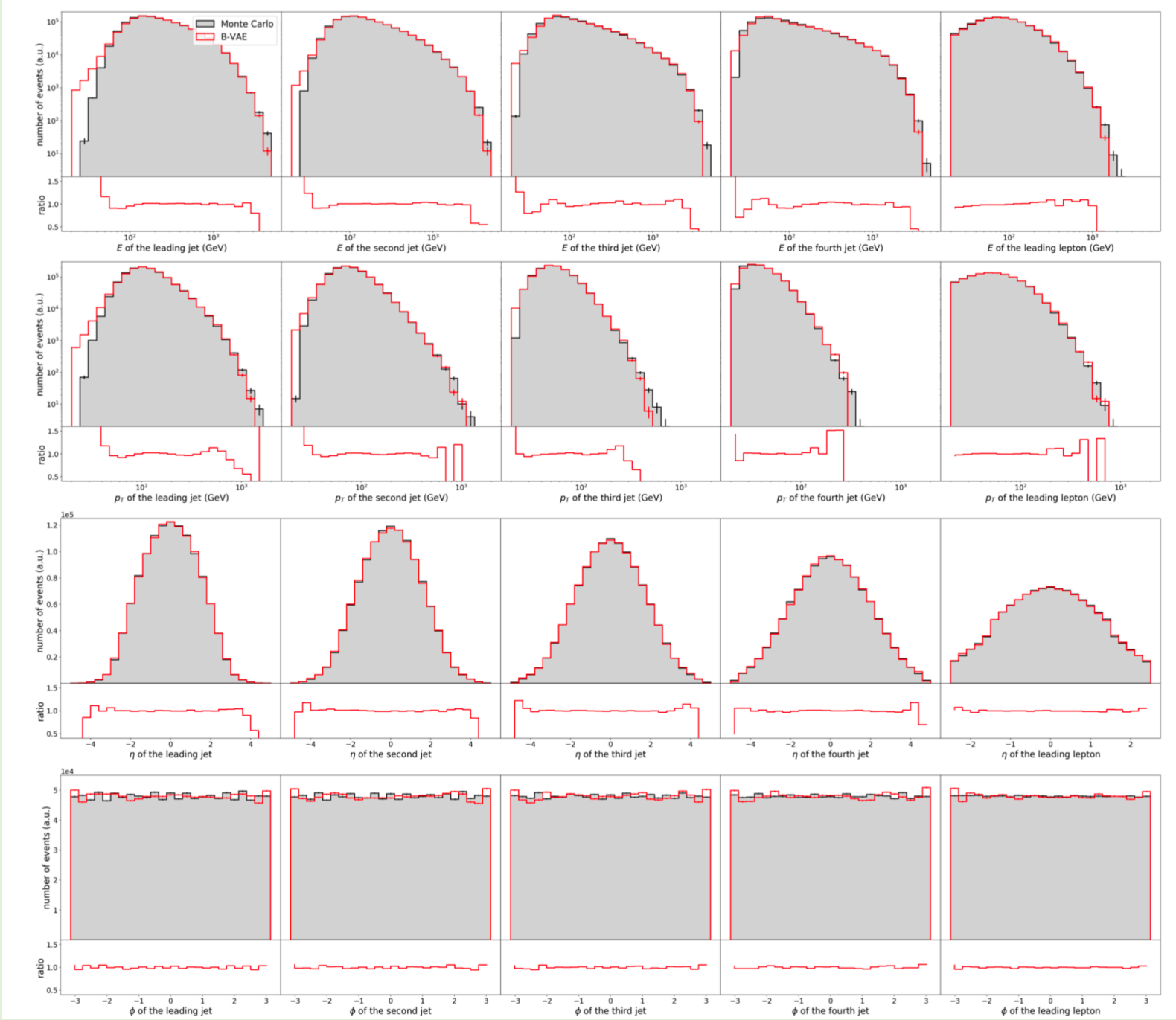
$$\left(z^i\right)_{j=1}^{\dim z} \sim \left(\mathcal{N}\left(\mu_j^i, \alpha_j\sigma_j^{2,i} + \gamma_j\right)\right)_{j=1}^{\dim z},$$

B-VAE sampling top-top events

→ Best way in paper to ”generate” toptop events

Tested various GAN architectures:
We could not get GANs to do this …

(beta = 10^-6, gamma=0.1, dim=20)



Example 4

Anomaly detection

Find a signal of new physics without knowing the signal

What is the objective ?
What needs to be optimized ?
Which approaches are possible ?

Various possible appraches
→In our example we assume that we “know the standard model”

B-VAE sampling top-top events

→ Can reconstruct toptop events

Way to ask for anomaly is:
Is the event in the simulation ?
(here e.g. can we reconstruct it ?)

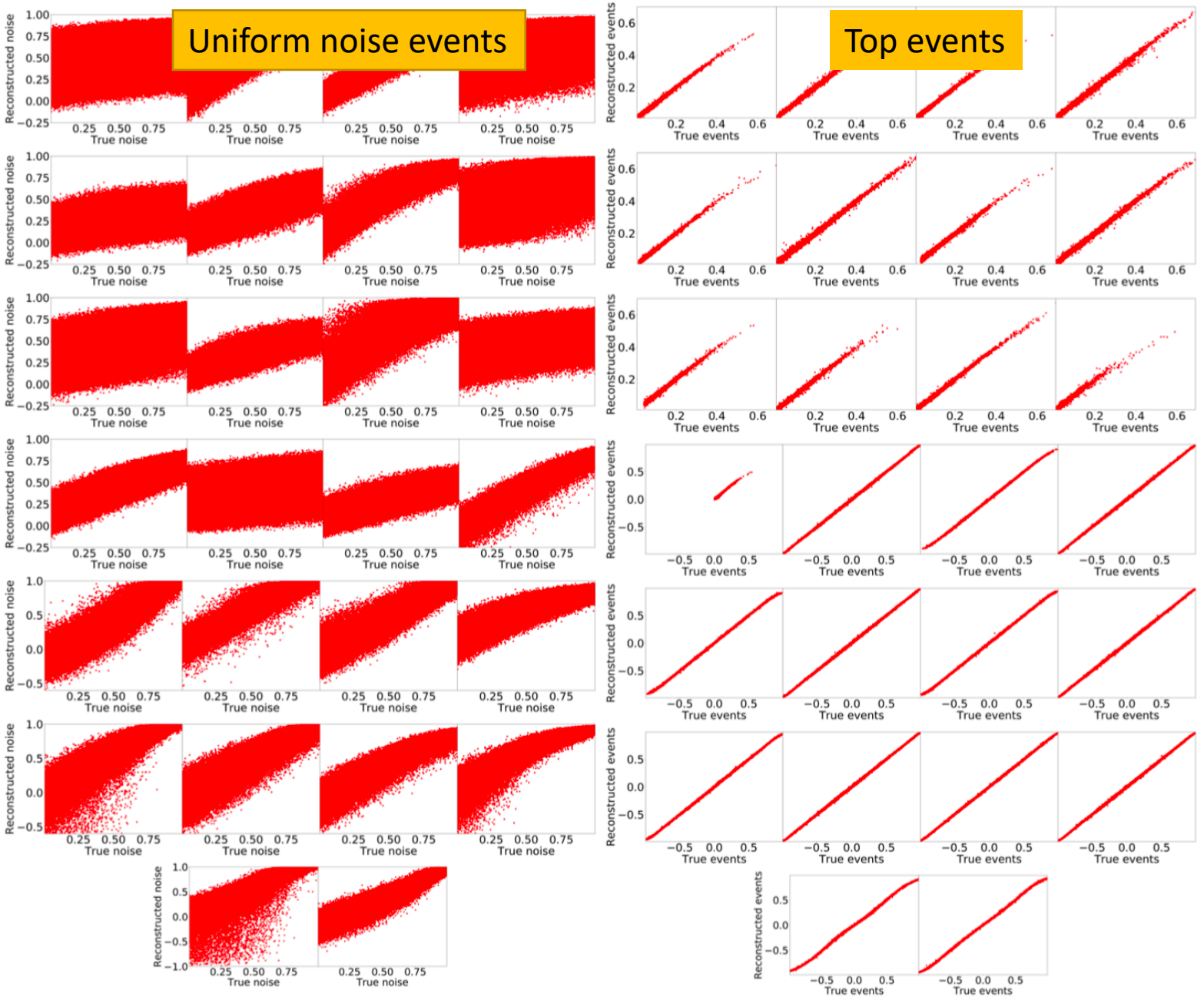


FIG. 1: Input vs. Reconstruction of uniform noise $x \sim U(0,1)$ (first four columns) and real events (last four columns) for a VAE with $\dim(z) = 20$ and $B = 10^{-6}$.

We have 4 new methods in pipeline, about 20 methods on arxiv since 2018

It is not clear which one works best

-> Need comparison with a large set of different data sets

(which will include different signal models)

Example 5

Les Houches 2019 challenges

Model-Independent Signal Detection: A Challenge using Benchmark Monte Carlo Data and Machine Learning

M. van Beekveld, S. Caron, A. De Simone, A. Farbin, L. Hendriks, A. Jueid, A. Leinweber, J. Mamuzic, E. Merényi, A. Morandini, C. Nellist, S. Otten, M. Pierini, R. Ruiz de Austri, S. Sekmen, J. Schouwenberg, R. Vilalta, M. White

Abstract

We discuss model-independent signal detection algorithms, with a particular focus on approaches that are based on unsupervised machine learning. We also offer a set of simulated LHC events, corresponding to $10\,\mathrm{fb}^{-1}$ of data. These events can be used as a benchmark dataset, for example for the comparison of signal detection algorithms. We explain the main features, the data format and describe the use of this data for an upcoming data challenge. The data is available at the webpage https://www.phenoMLdata.org.

https://arxiv.org/pdf/2002.12220.pdf

SM processes			
Physics process	Process ID	σ (pb)	N_{tot} ($N_{10\,\text{fb}^{-1}}$)
$pp \to jj$	njets	$19718_{H_T>600\,\text{GeV}}$	415331302 (197179140)
$pp \to W^{\pm}(+2j)$	w_jets	$10537_{H_T>100\,\text{GeV}}$	135692164 (105366237)
$pp \to \gamma(+2j)$	gam_jets	$7927_{H_T>100\,\text{GeV}}$	123709226 (79268824)
$pp \to Z(+2j)$	z_jets	$3753_{H_T>100\,\text{GeV}}$	60076409 (37529592)
$pp \to t\bar{t}(+2j)$	ttbar	541	13590811 (5412187)
$pp \to W^{\pm}t(+2j)$	wtop	318	5252172 (3176886)
$pp \to W^{\pm}\bar{t}(+2j)$	wtopbar	318	4723206 (3173834)
$pp \to W^{+}W^{-}(+2j)$	ww	244	17740278 (2441354)
$pp \to t$+jets$(+2j)$	single_top	130	7223883 (1297142)
$pp \to \bar{t}$+jets$(+2j)$	single_topbar	112	7179922 (1116396)
$pp \to \gamma\gamma(+2j)$	2gam	47.1	17464818 (470656)
$pp \to W^{\pm}\gamma(+2j)$	Wgam	45.1	18633683 (450672)
$pp \to ZW^{\pm}(+2j)$	zw	31.6	13847321 (315781)
$pp \to Z\gamma(+2j)$	Zgam	29.9	15909980 (299439)
$pp \to ZZ(+2j)$	zz	9.91	7118820 (99092)
$pp \to h(+2j)$	single_higgs	1.94	2596158 (19383)
$pp \to t\bar{t}\gamma(+1j)$	ttbarGam	1.55	95217 (15471)
$pp \to t\bar{t}Z$	ttbarZ	0.59	300000 (5874)
$pp \to t\bar{t}h(+1j)$	ttbarHiggs	0.46	200476 (4568)
$pp \to \gamma t(+2j)$	atop	0.39	2776166 (3947)
$pp \to t\bar{t}W^{\pm}$	ttbarW	0.35	279365 (3495)
$pp \to \gamma\bar{t}(+2j)$	atopbar	0.27	4770857 (2707)
$pp \to Zt(+2j)$	ztop	0.26	3213475 (2554)
$pp \to Z\bar{t}(+2j)$	ztopbar	0.15	2741276 (1524)
$pp \to t\bar{t}t\bar{t}$	4top	0.0097	399999 (96)
$pp \to t\bar{t}W^{+}W^{-}$	ttbarWW	0.0085	150000 (85)

Table 2: Generated background processes (first column) with the corresponding identification (second column), the LO cross section σ in pb (third column) and the total number of generated events N_{tot} (fourth column). In the last column, we also indicate the number of events corresponding to 10 fb^{-1} of data ($N_{10\,\text{fb}^{-1}}$).

Simulation settings

- MG5_aMC@NLO v6.3.2 (Madgraph) interfaced with Pythia 8.2
- quick detector simulation is performed with Delphes 3 using a modified version of the ATLAS detector card.
- Pileup is not included in this dataset.

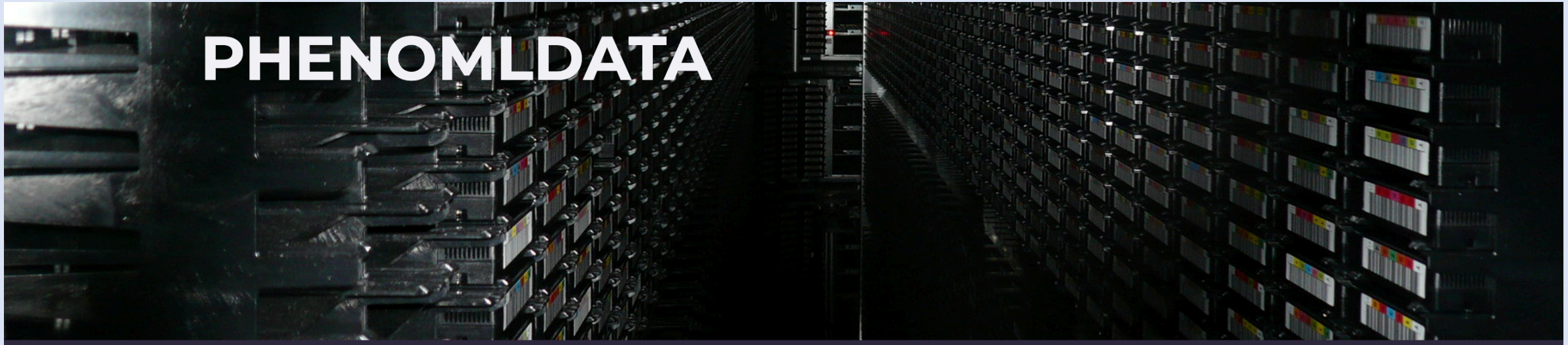
- Event is stored if one of the following objects are found:
  - At least one (b)-jet with transverse momentum $p_T > 60$ GeV and pseudorapidity $|\eta| < 2.8$, or
  - at least one electron with $p_T > 25$ GeV and $|\eta| < 2.47$, except for $1.37 < |\eta| < 1.52$, or
  - at least one muon with $p_T > 25$ GeV and $|\eta| < 2.7$, or
  - at least one photon with $p_T > 25$ GeV and $|\eta| < 2.37$.

Data format

- Available as .csv file, i.e. text on zenodo:

```
event ID; process ID; event weight; MET; METphi; obj1, E1, pt1, eta1,
                phi1; obj2, E2, pt2, eta2, phi2;  ...
```

- Full event information: Several 100 Terabyte → hope to store at CERN open data



PHENOMLDATA

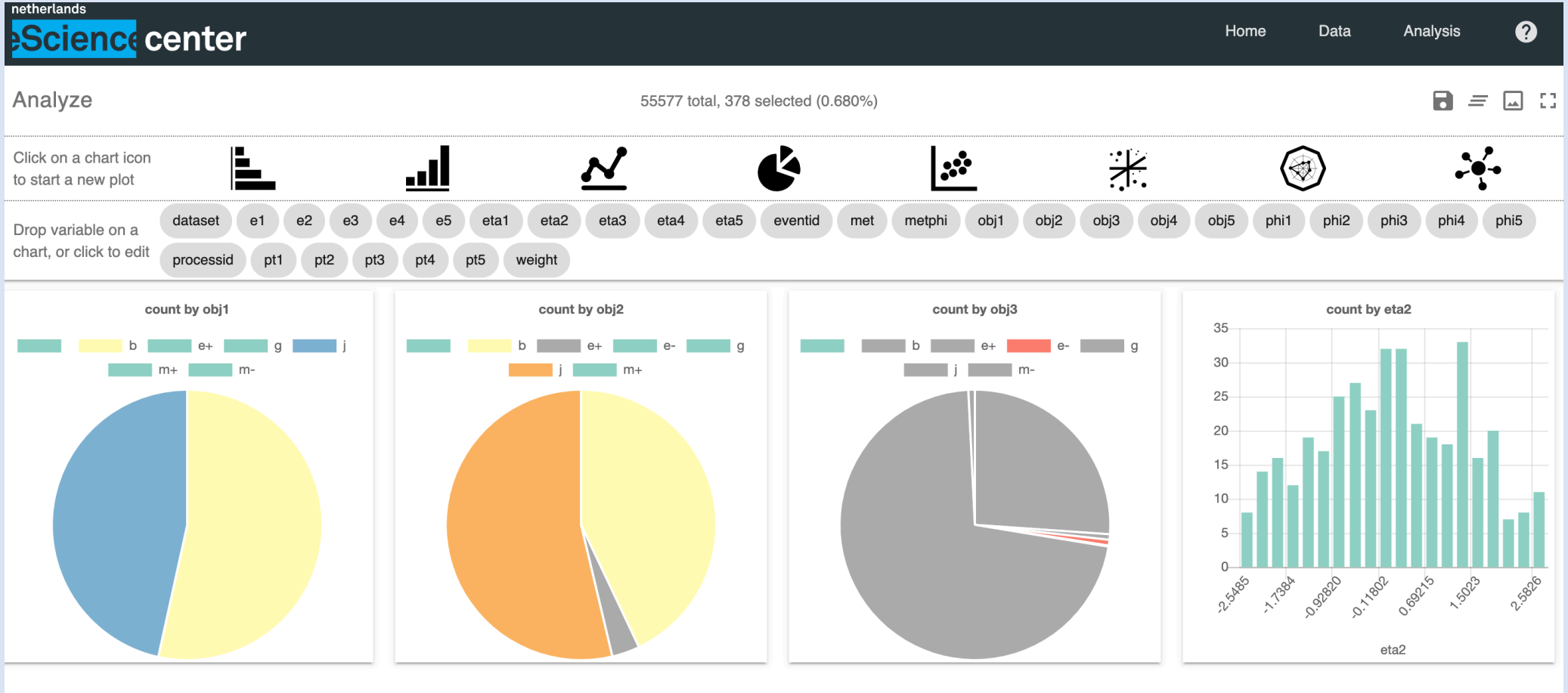
Datasets / repositories

Regression / classification dataset

LHC simulation challenge

Dataset from 2002.12220. This dataset can be visualised online in the SPOT visualisation tool

Online Data visualisation via SPOT



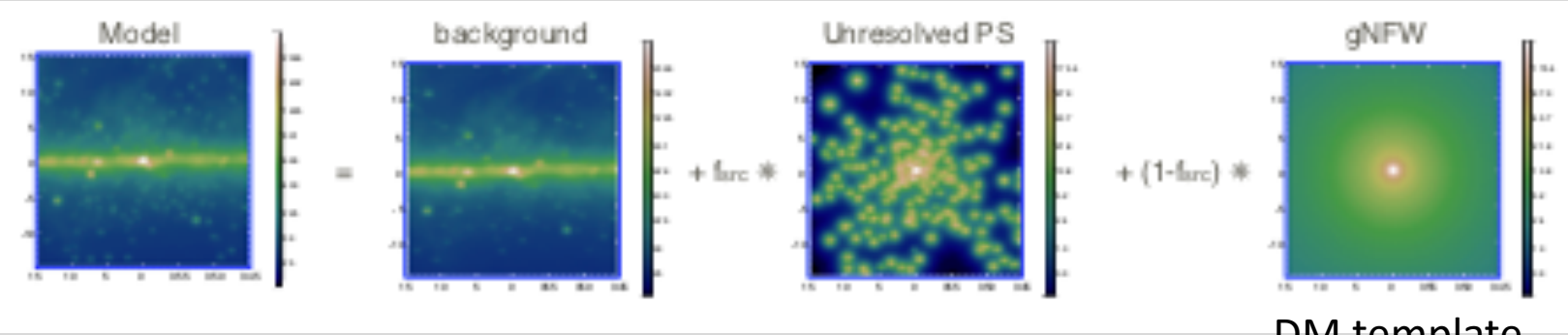
Example 6

Astrophysics example: Gamma-ray from Galactic Center

Still not completely resolved question: Is the “Galactic Center” excess due to additional point sources or diffuse emission.

2017: First try with Deep Convolutional Network https://arxiv.org/abs/1708.06706

One energy bin , i.e. one picture → Output fraction of point sources



DM template

2017: f_src fitted to ”half” the GC template picture

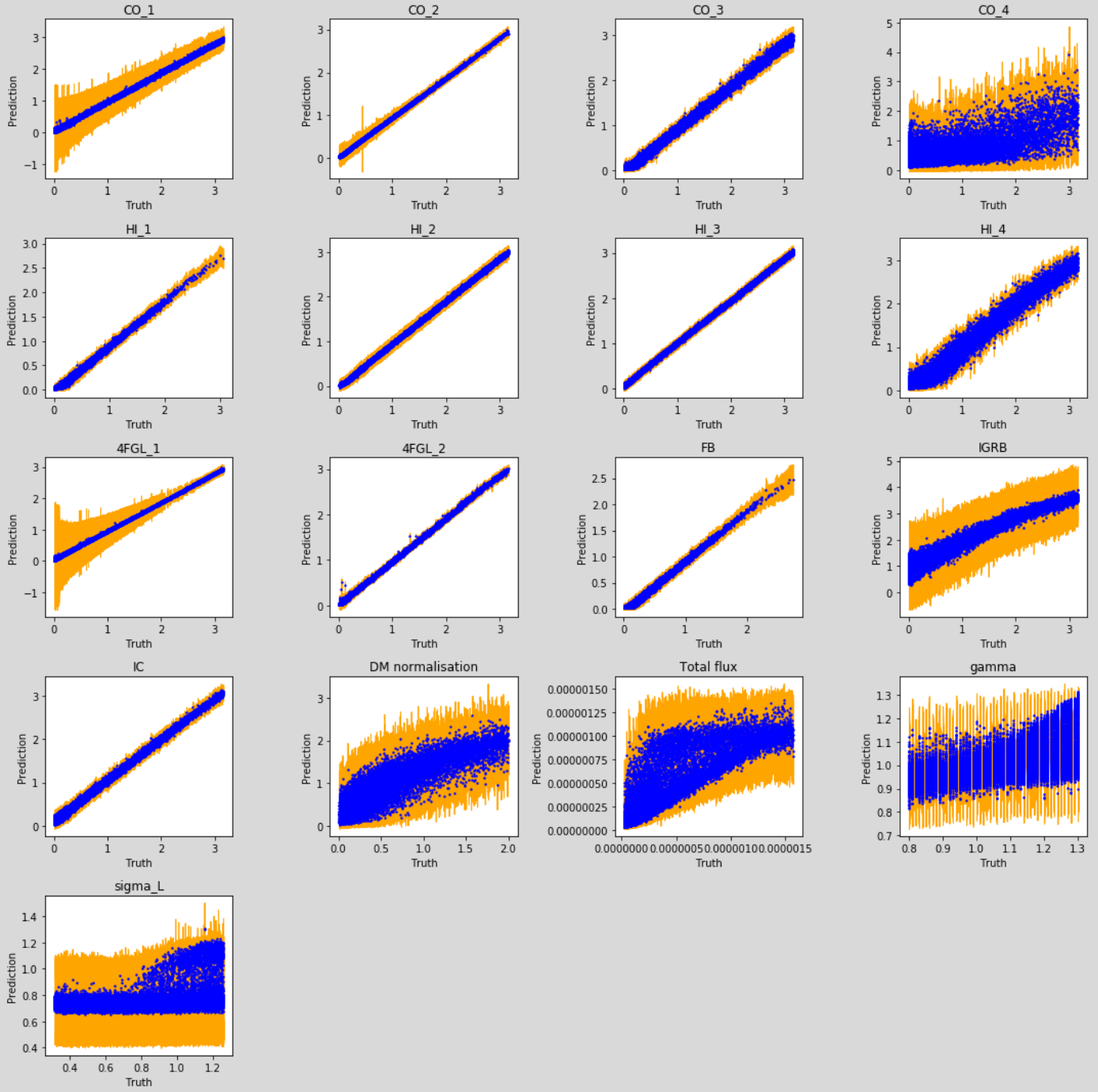
2020: Bayesian Conv Network including uncertainty to estimate for 16-20 parameters simultaneously

Input: 2d images with 5 energy bins (colours)

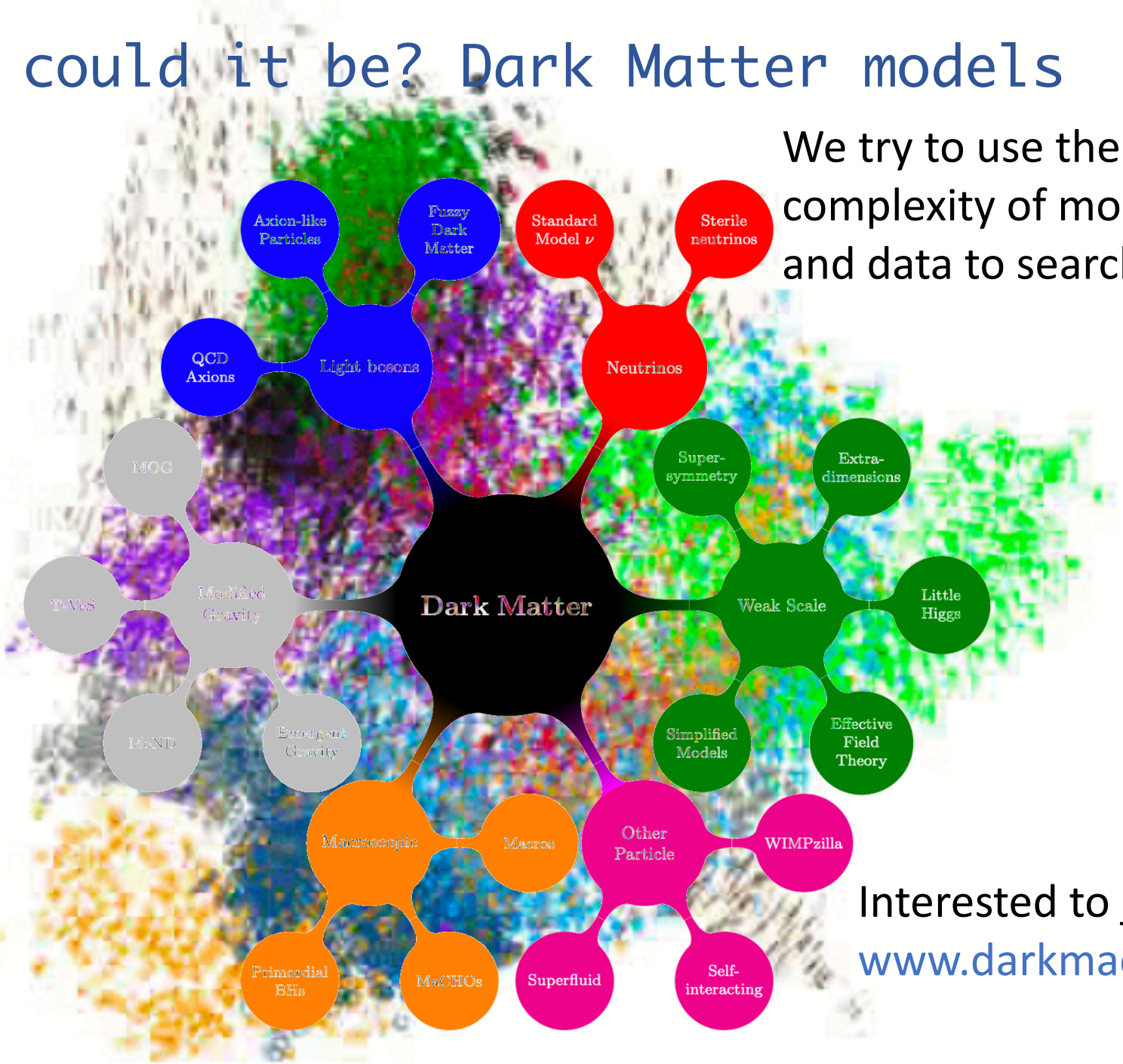
Build 17-20 dimensional model

Example of 17d Parameter estimation of simulation

Summer **2020**: Predict 19d values including uncertainties for “real image”



What could it be? Dark Matter models



We try to use the full complexity of models and data to search for new physics

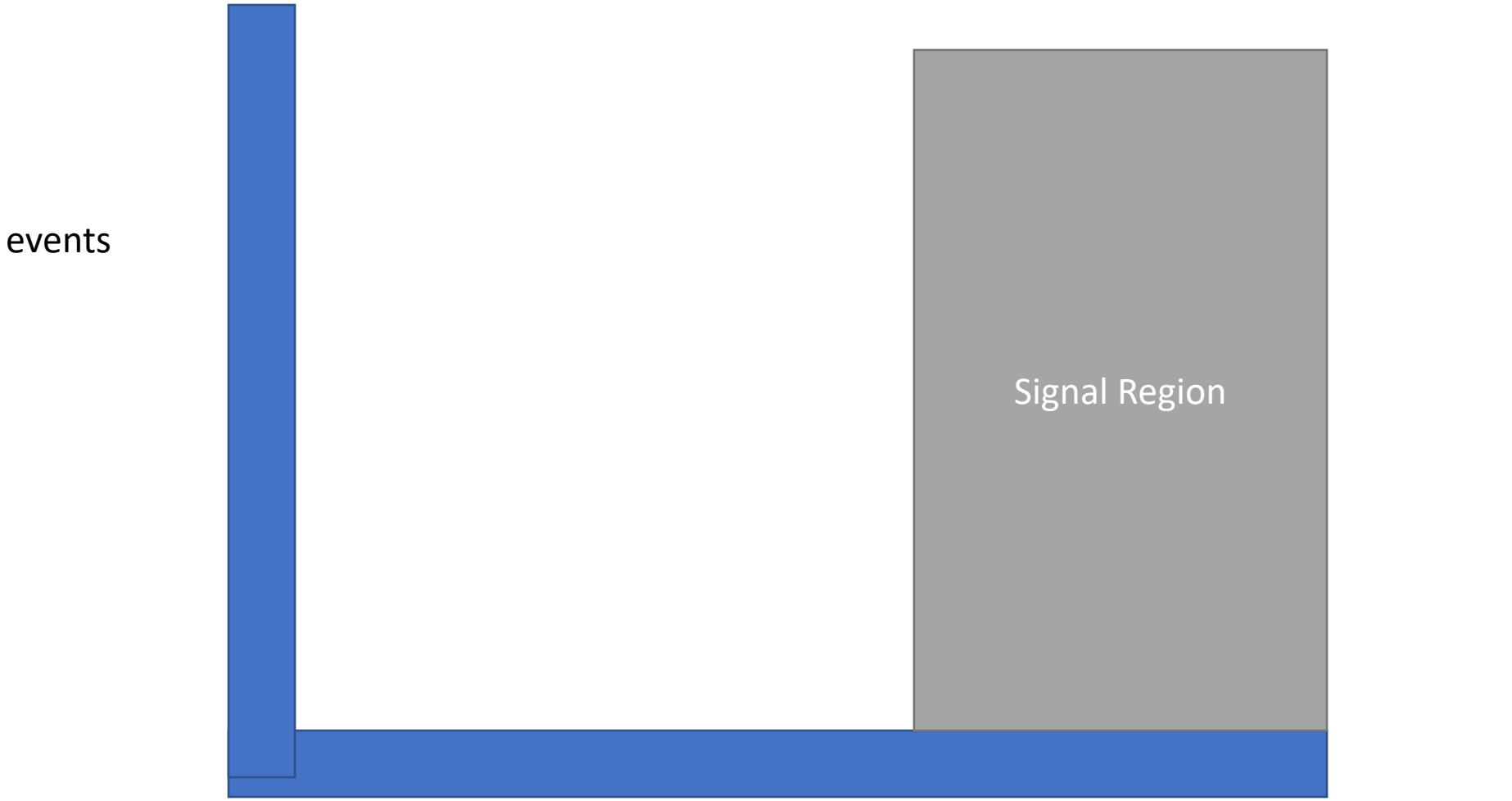
Interested to join & help:
www.darkmachines.org

Ideas and review ?

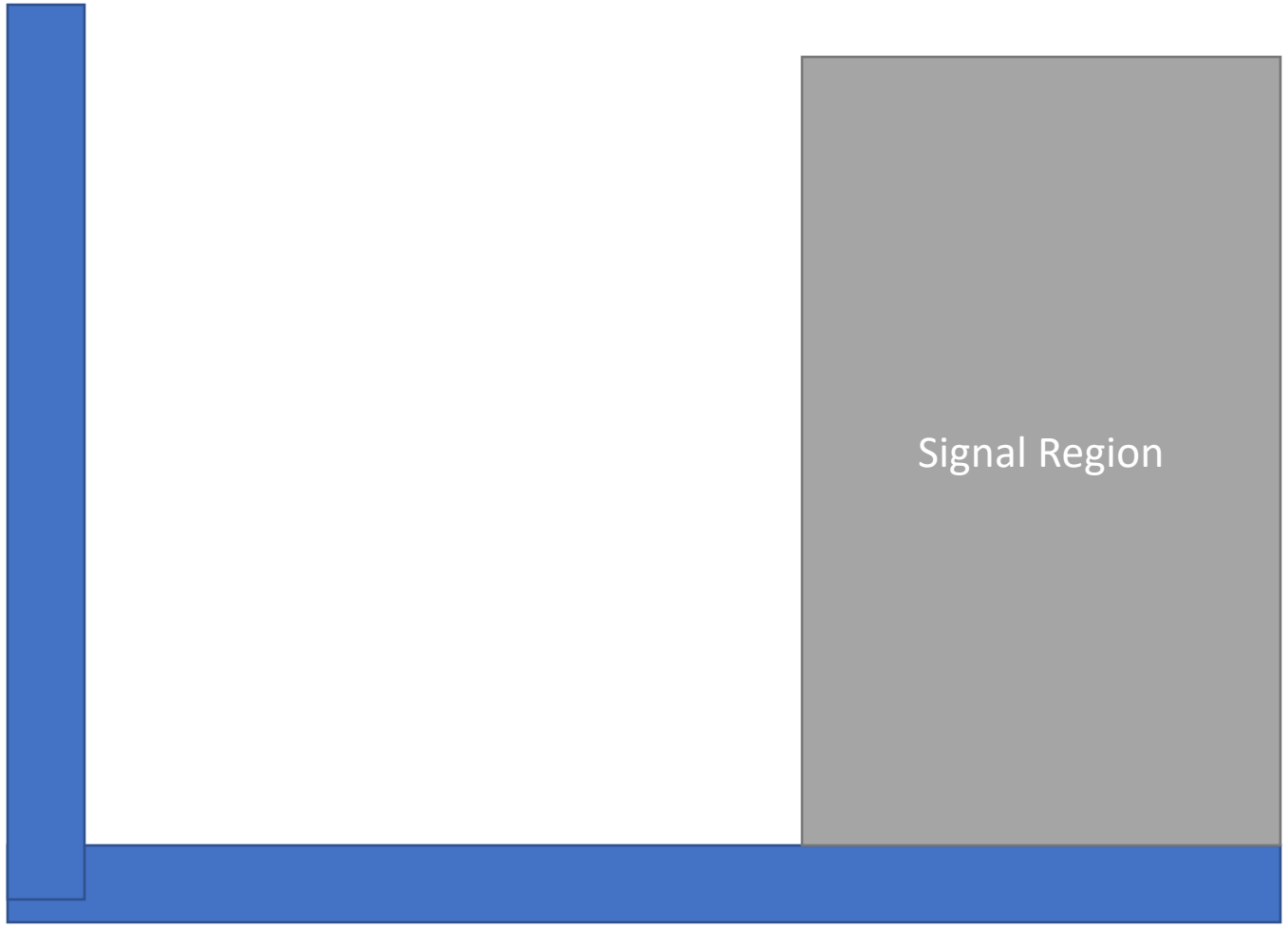
Les Houches 2019 Physics at TeV Colliders: New Physics Working Group Report,
“Model-Independent Signal Detection: A Challenge using Benchmark Monte Carlo Data and Machine Learning”
https://arxiv.org/pdf/2002.12220.pdf

More general title ”Model-independent Signal Detection”
→Search for Signal of new physics without assuming a signal model
→First paper at LHC “General Search” (Jeroen Schouwenberg et al)

How to define a signal region ?



How to define a signal region ?



No difference to other analyses

Medium term (B-VAE)

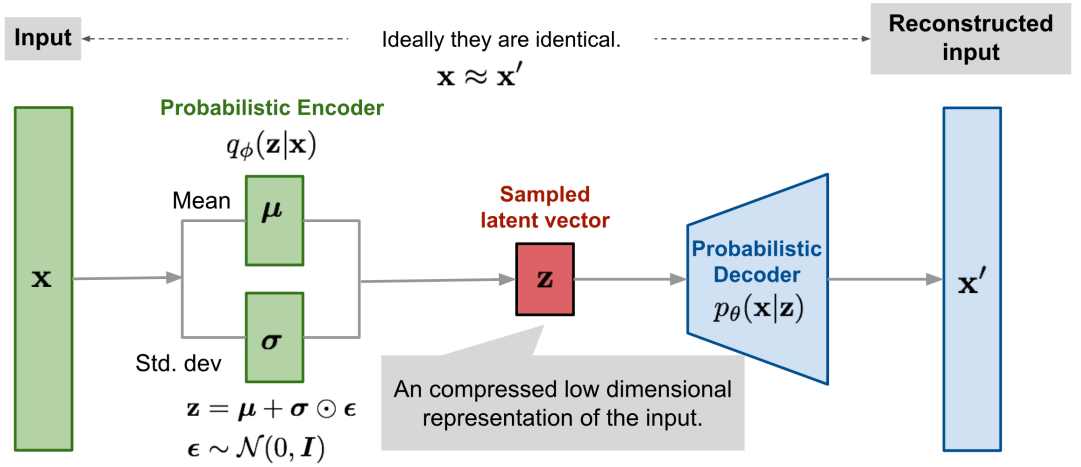
- We proposed to a Variational Autoencoder which we call B-VAE in https://arxiv.org/abs/1901.00875
- **What is this ?** The maximization of the variational lower bound is turned into the minimization of the positive D_KL and MSE such that the loss function of the VAE ∝D_KL+MSE. One can now introduce a parameter to control the importance of D_KL and MSE.

The so called “beta VAE” proposed to apply B>1 to disentangling the latent space, arXiv e-prints , arXiv:1804.03599 (2018)1,264 arXiv:1804.03599 [stat.ML].

We proposed in 1901.00875 to use B<<1 to enhance the ability to reconstruct:

$$L = \frac{1}{M}\sum_{i=1}^{M}(1-B)\cdot MSE + B\cdot D_{KL}$$

B-VAE



Together with an “information buffering” of the latent space we could find optimal event sampling properties for a B-VAE.

The “buffer” collects observations $Z = \{z_1, \ldots, z_m\}$ by sampling $q_\phi(z|X_L)$ where $X_L \subset X$ is a subset of real Monte Carlo events. The distribution over the latent vector z is then given by:

$$p_{\phi,X_L}(z) = \sum_{i=1}^{m} q_\phi(z|x^i)p(x^i) \text{ with } \mathrm{p(x^i)} = \frac{1}{\mathrm{m}}.$$

To avoid overtraining the “information buffering” needs to introduce new hyperparameters alpha and gamma, allowing more variance in sampling the latent vector z via:

$$z \sim q_\phi(z|X_L)$$

$$z^i \sim \begin{cases} \mathcal{N}(\mu^i, \alpha\sigma^{2,i}) & \text{if } \sigma < \sigma_T \\ \mathcal{N}(\mu^i, \sigma^{2,i}) & \text{else} \end{cases},$$

$$\left(z^i\right)_{j=1}^{\dim z} \sim \left(\mathcal{N}\left(\mu_j^i, \alpha_j\sigma_j^{2,i} + \gamma_j\right)\right)_{j=1}^{\dim z},$$

Medium term (darkmachines)

- Compare anomaly detection strategies on event level
- First step: Made dataset (> 1 Billion events), see extra slides for additional info
- We might start from a small subset (4 top vs 2 top) with a group effort
- Currently planning how to organise comparison