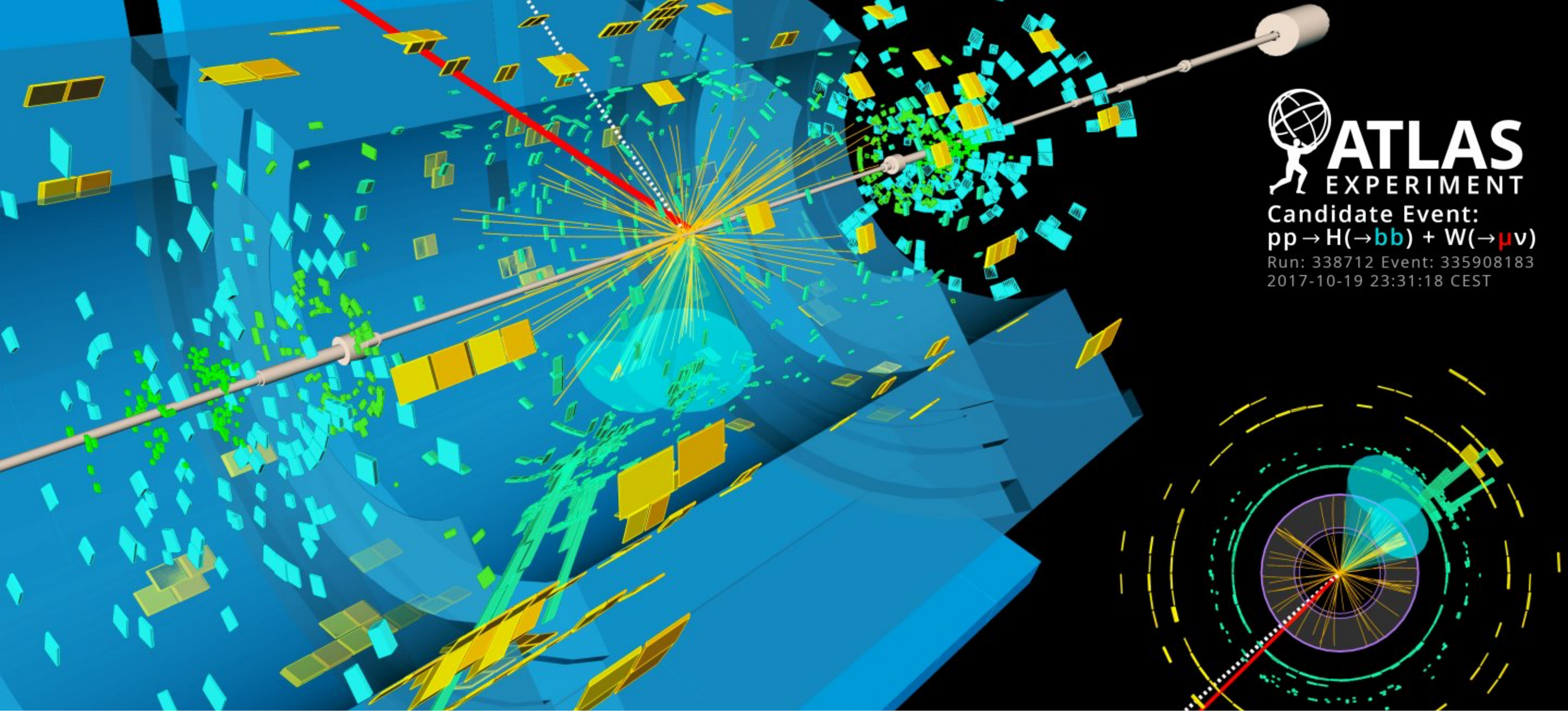


ML in Data Analysis: Systematic Uncertainties with ML

Lecture 4

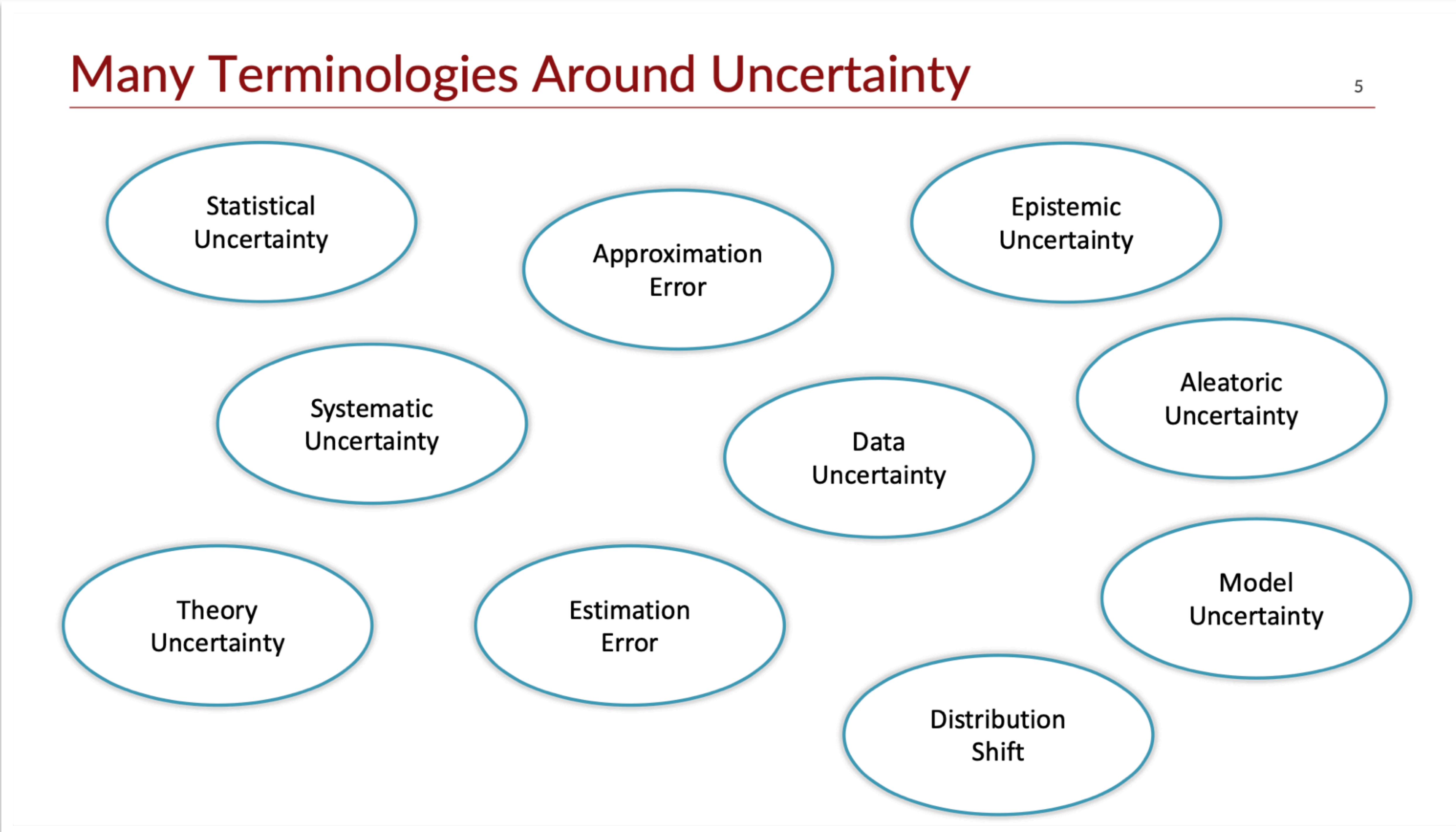
Sofia Vallecorsa | Ilaria Luise

Thematic CERN School of Computing on Machine Learning

18th October 2024

CERN

Introduction



Goal of today's lecture: understand the different concepts and link them together

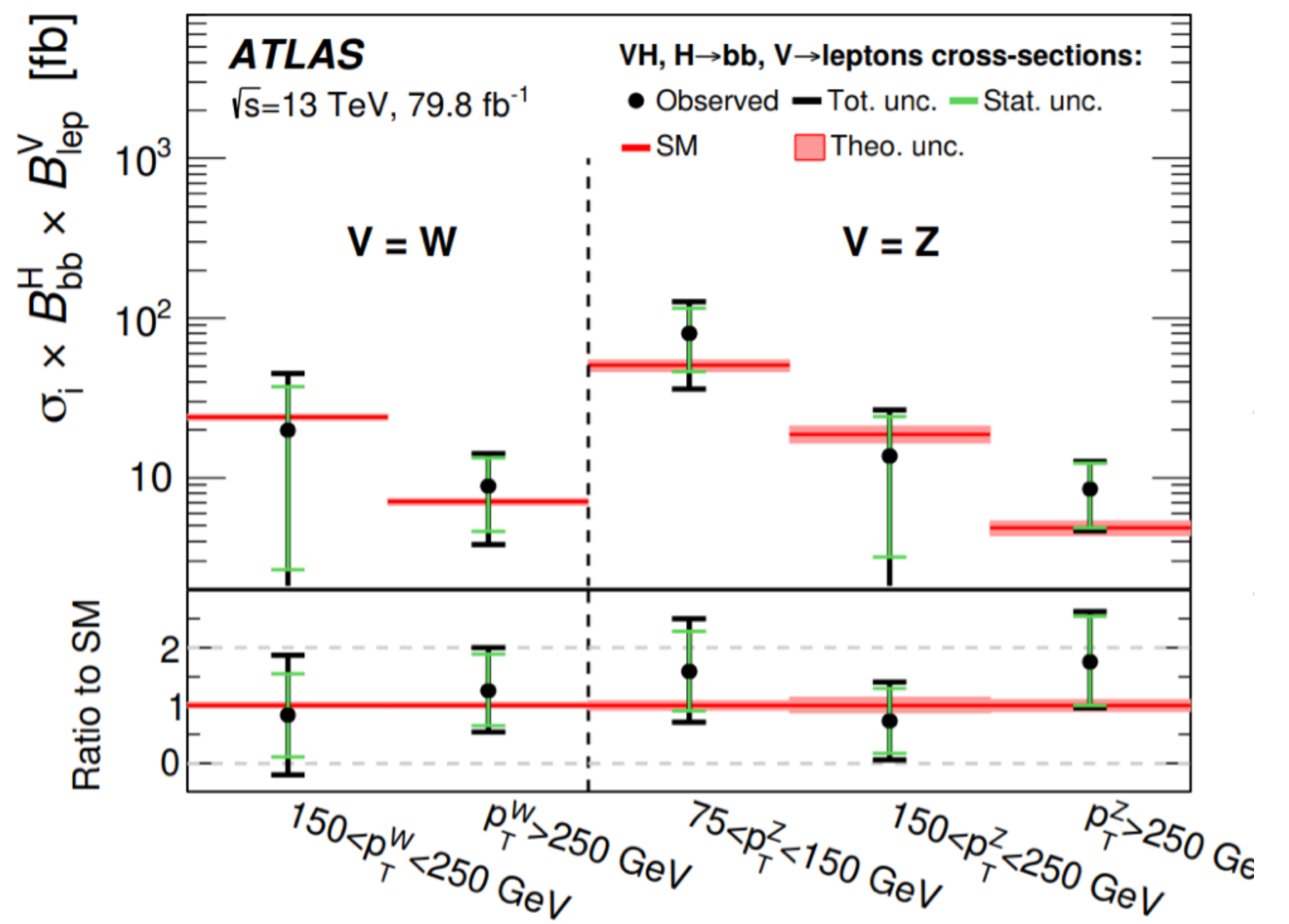
Slide from M. Kagan

Introduction: Why systematics are important?

We are entering in a new era:

After the Higgs boson discovery, the focus shifted toward the **measurement of its properties**:

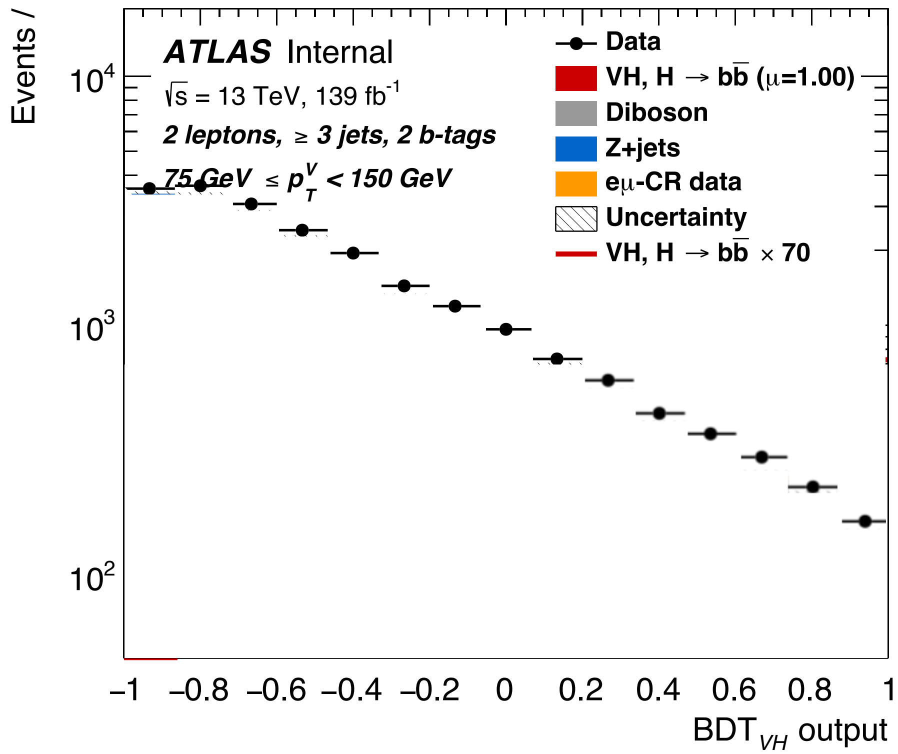
Is this "the Higgs"?

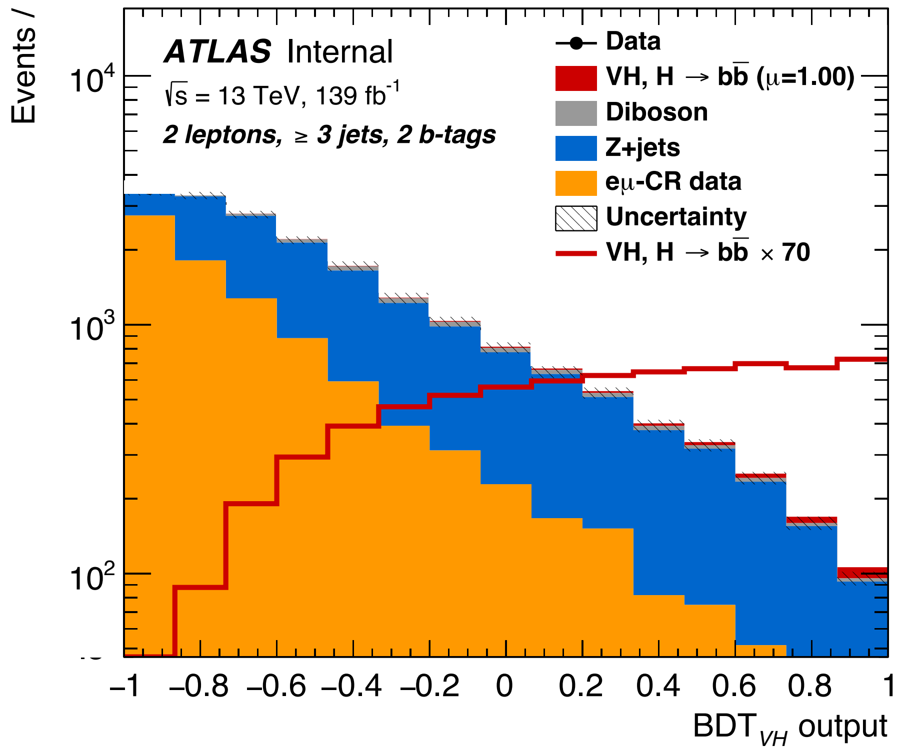


"precise" Higgs measurements → reduce the uncertainties to ***increase the sensitivity*** to tiny BSM induced anomalies.

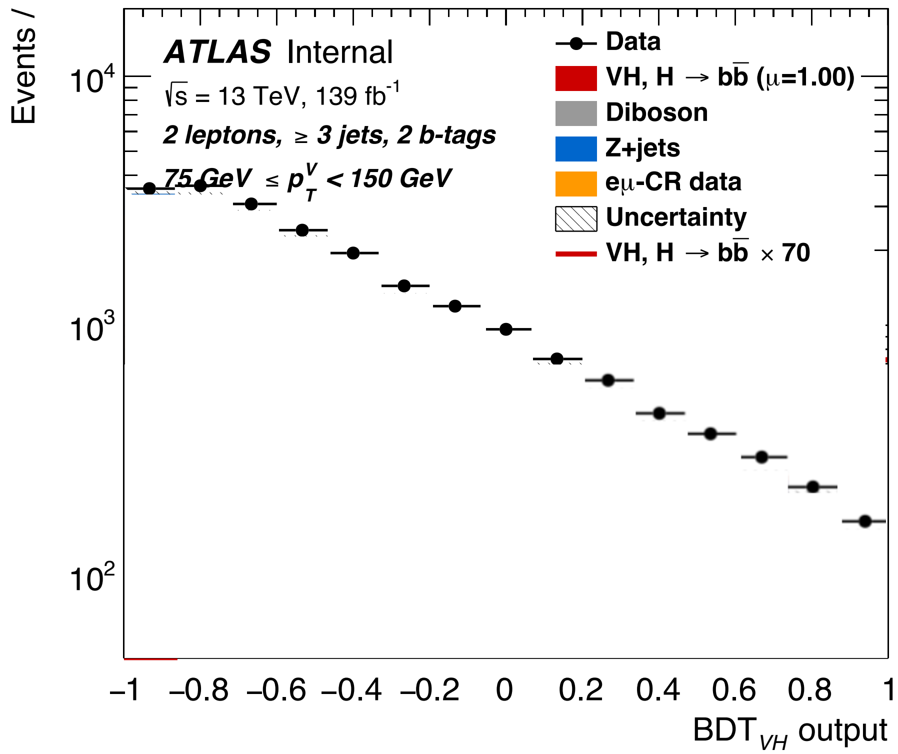
Systematic uncertainties in HEP

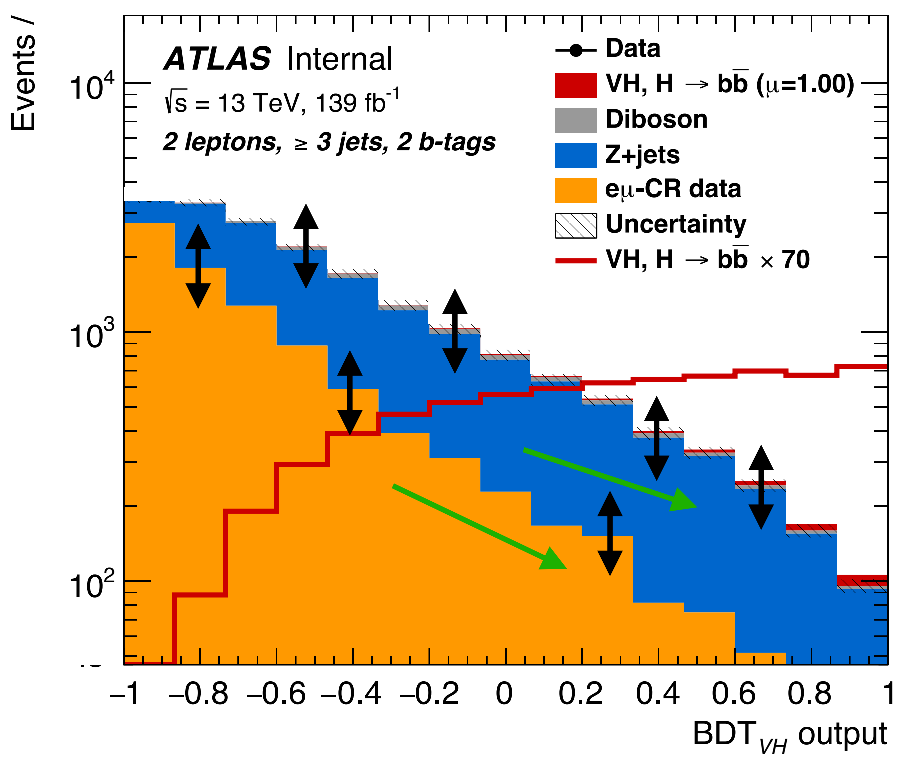
How does a fit (usually) work in HEP?



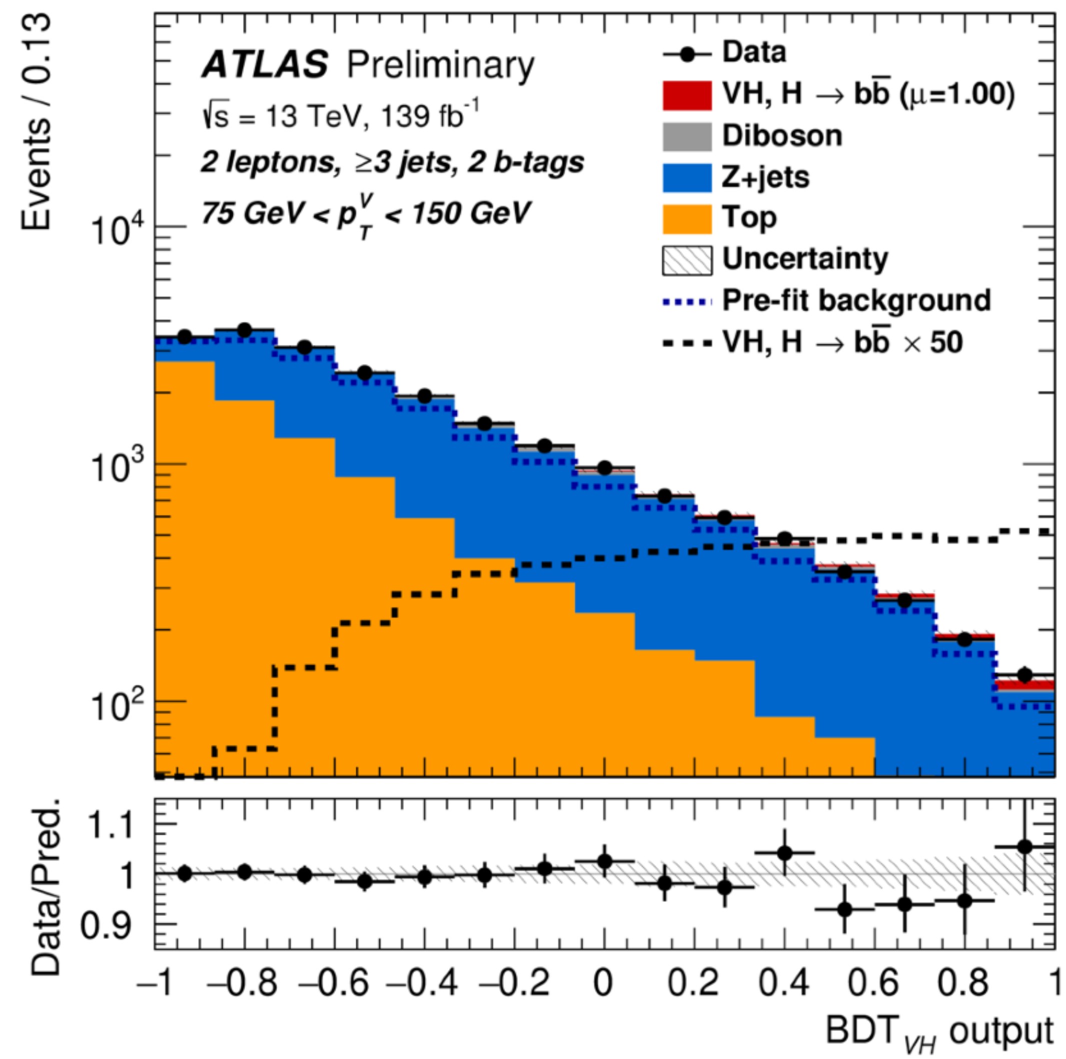


How does a fit (usually) work in HEP?





Note: sometimes the Montecarlo can be replaced by a certain function, like for example for H→γγ

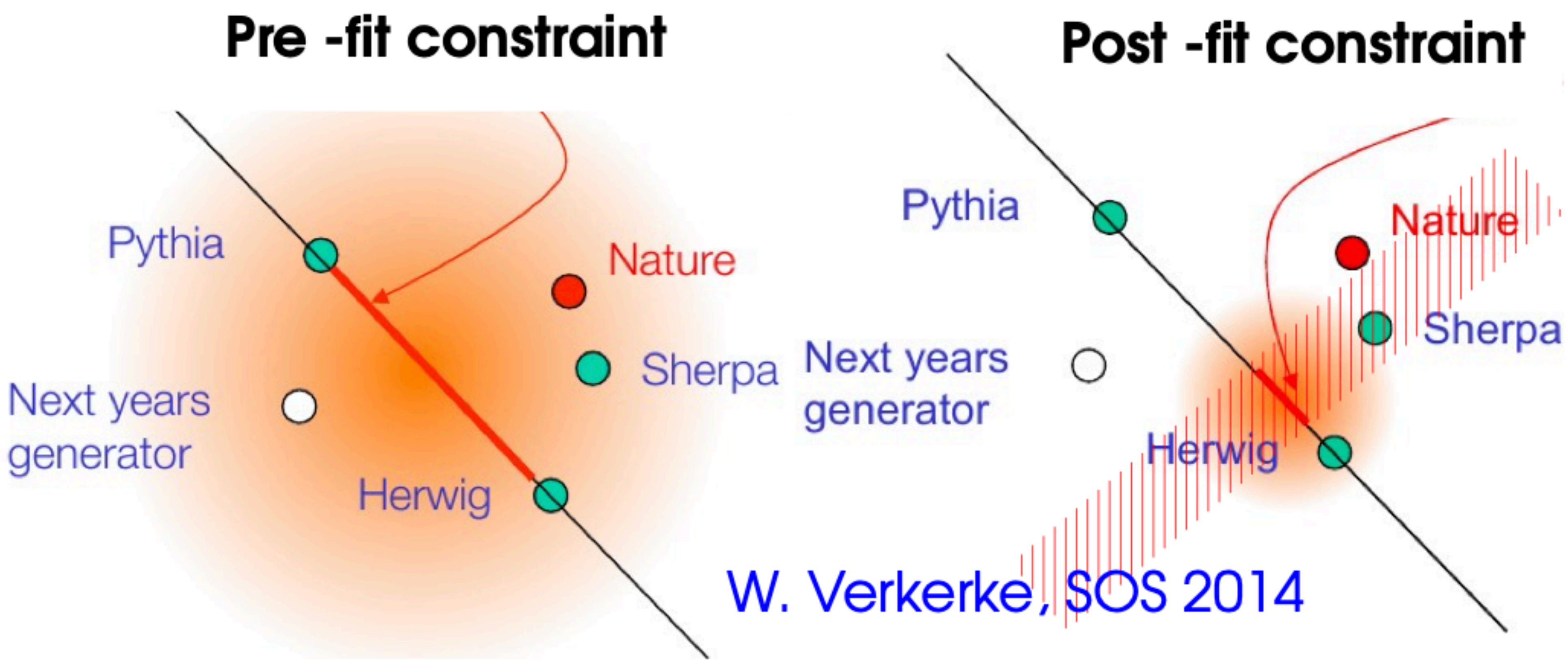


Maximise the likelihood

..within some boundaries

Nuisance parameters

These boundaries are called "nuisance parameters" and define our *__level of uncertainty__* on the montecarlo



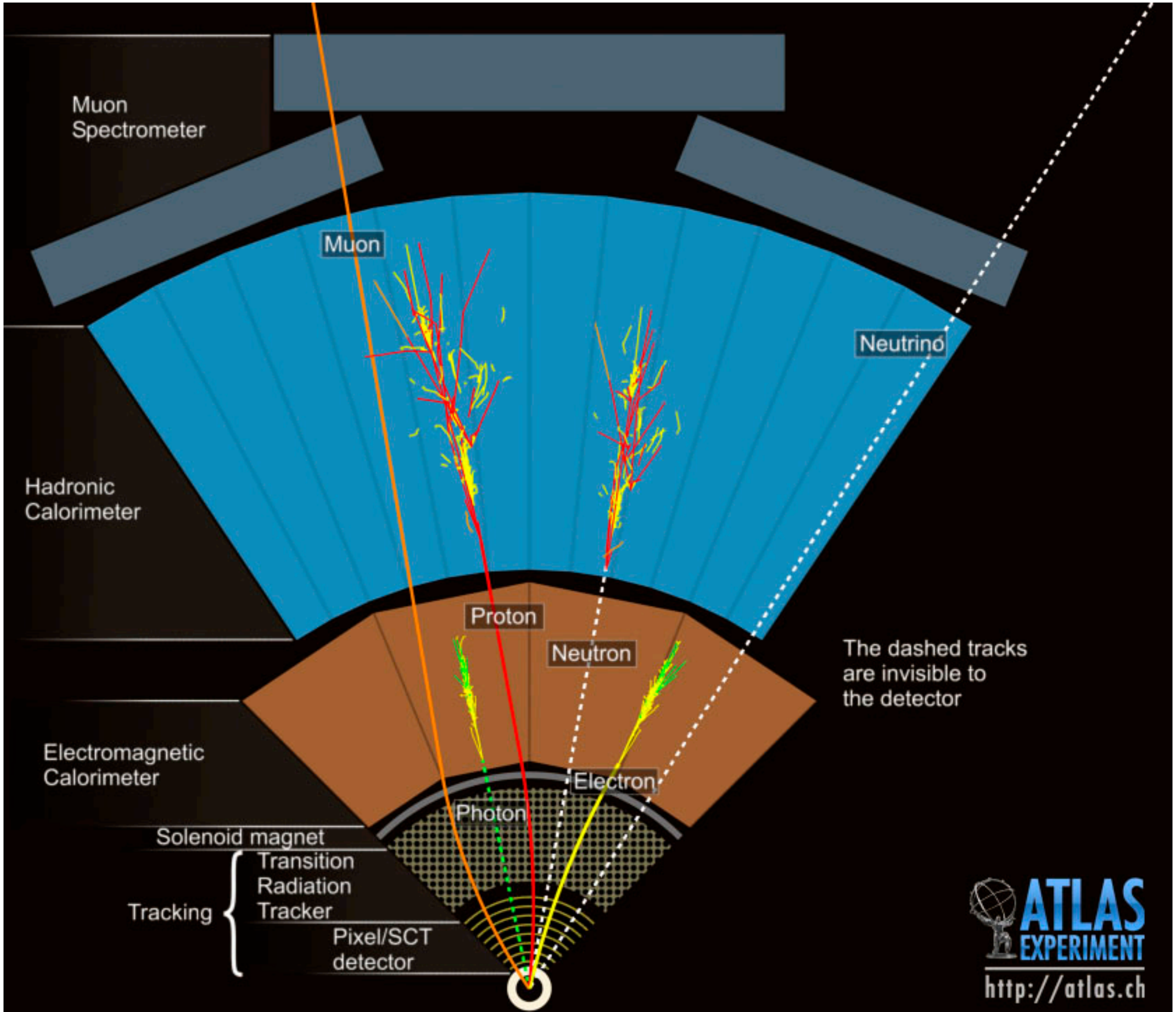
- Account for shape differences
- Account for normalisation effects
- Account for uncertainties in the applied corrections or in the theory
- Account for uncertainties associated with limited data

Types of uncertainty

Total uncertainty

- **Systematic Uncertainties**
  - **Experimental**
  - **Modelling**
- **Statistical Uncertainties**

Experimental uncertainties

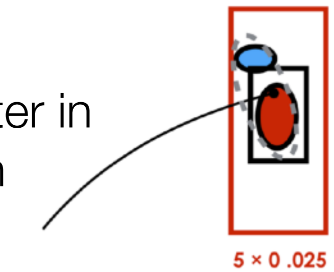


Muons:
Match tracks in the MS and in the ID (combined muons)

$$\chi^2_{match}$$

Electrons:

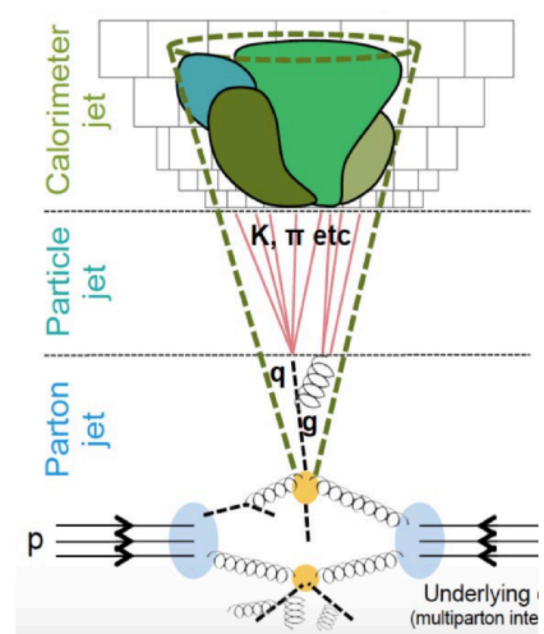
match a cluster in ECAL with an ID track



Jets:

Reconstructed from clusters in the ECAL+HCAL

Anti-k_T algorithm



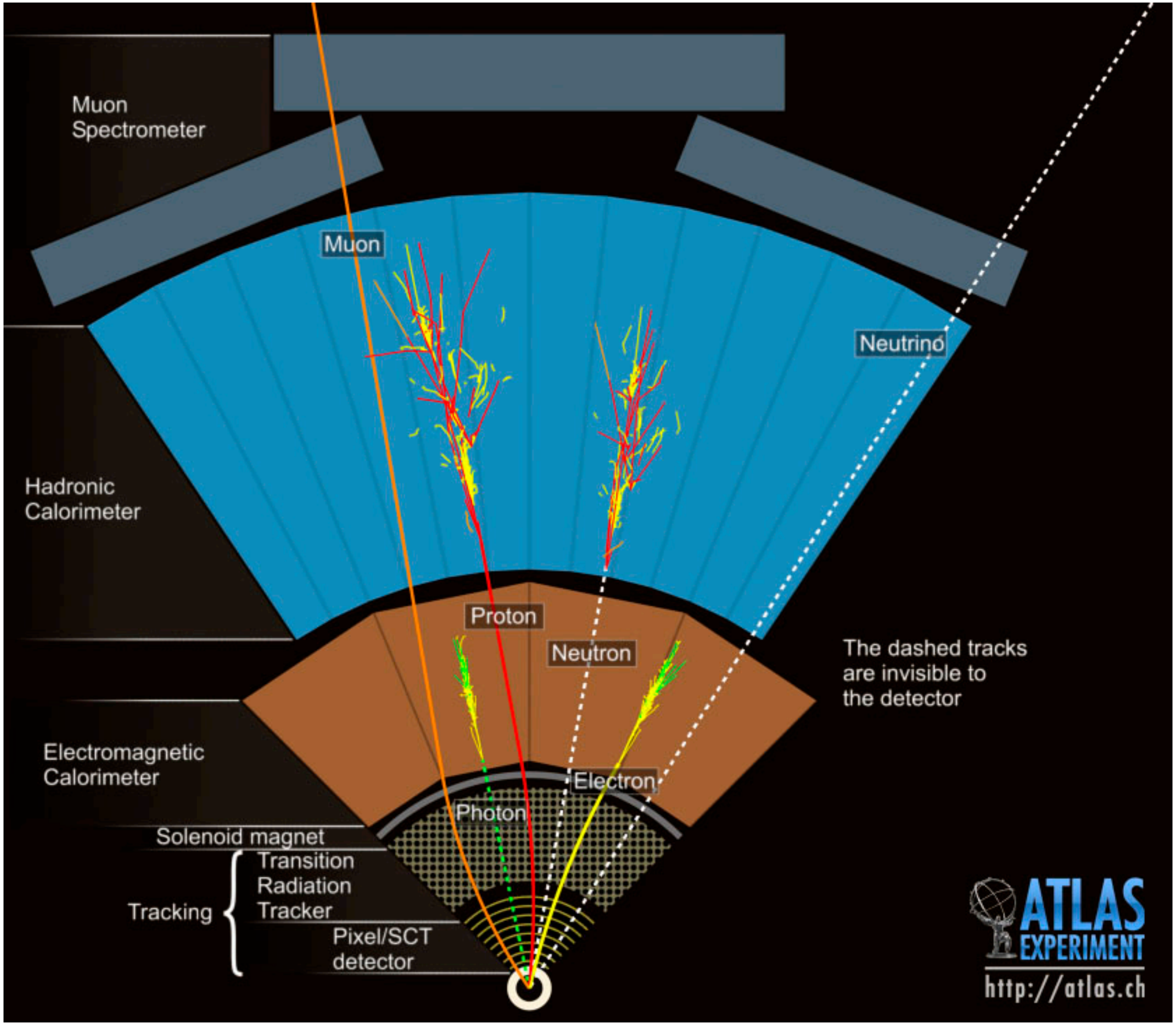
E_T miss:

Missing Transv. Momentum

Momentum imbalance in the transverse plane:

$$\vec{E}_T^{miss} = -\sum_{i \in obj.} \vec{p}_T^{\,i}$$

Experimental uncertainties



Some examples:

- Calibrations
- Identification
- Trigger uncertainties

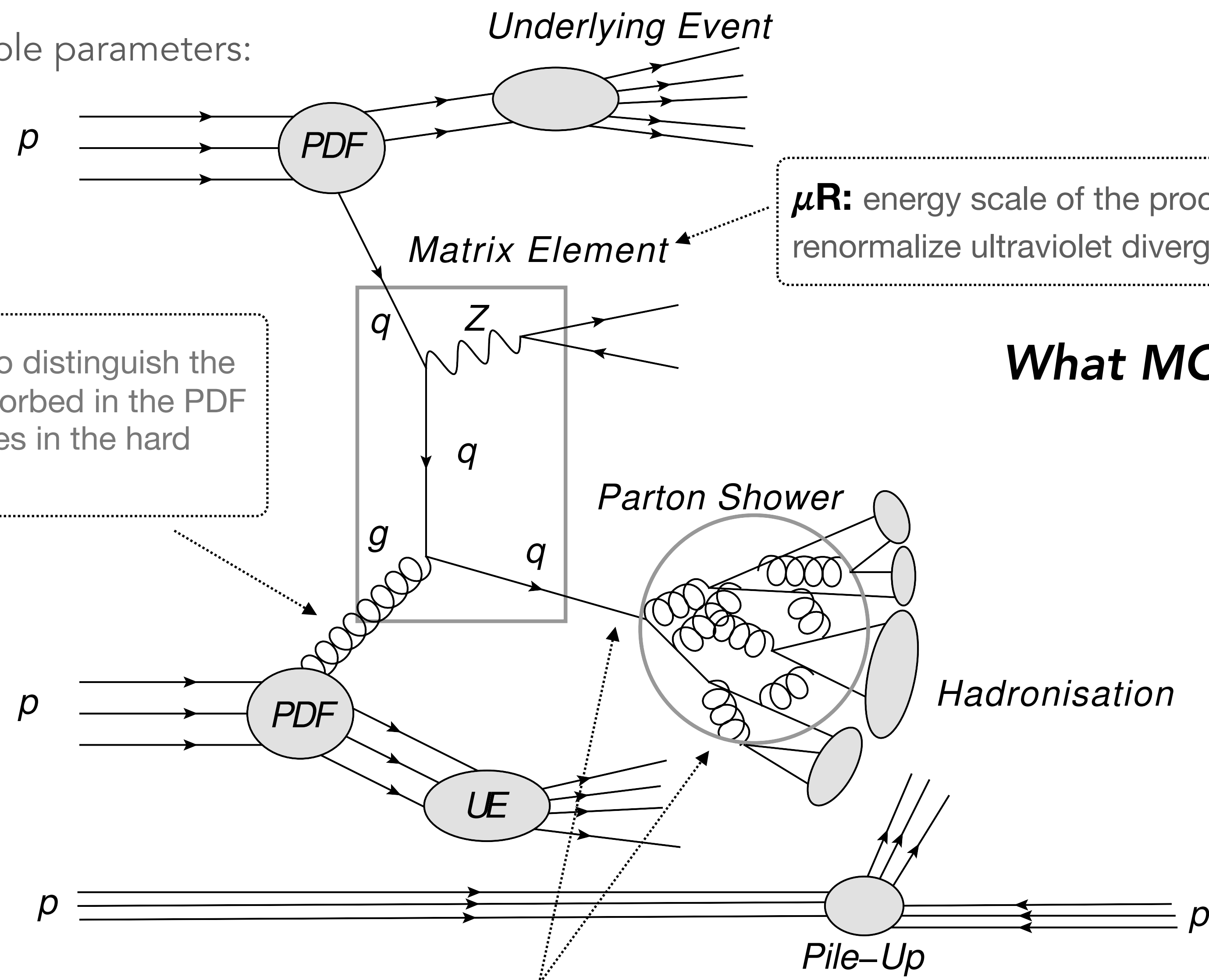
- Jet energy scales
- Flavour tagging

- Energy corrections

Modelling uncertainties

Start building the MC Systematic Model

Configurable parameters:



μR: energy scale of the process (used to renormalize ultraviolet diverg. In loops)

μF: cutoff to distinguish the partons absorbed in the PDF from the ones in the hard scattering.

Matrix element matching scale (CKKW): the scale taken for the calculation of the overlap between jets from the matrix element and the parton shower.

Resummation scale (QSF): the scale used for the resummation of soft gluon emissions.

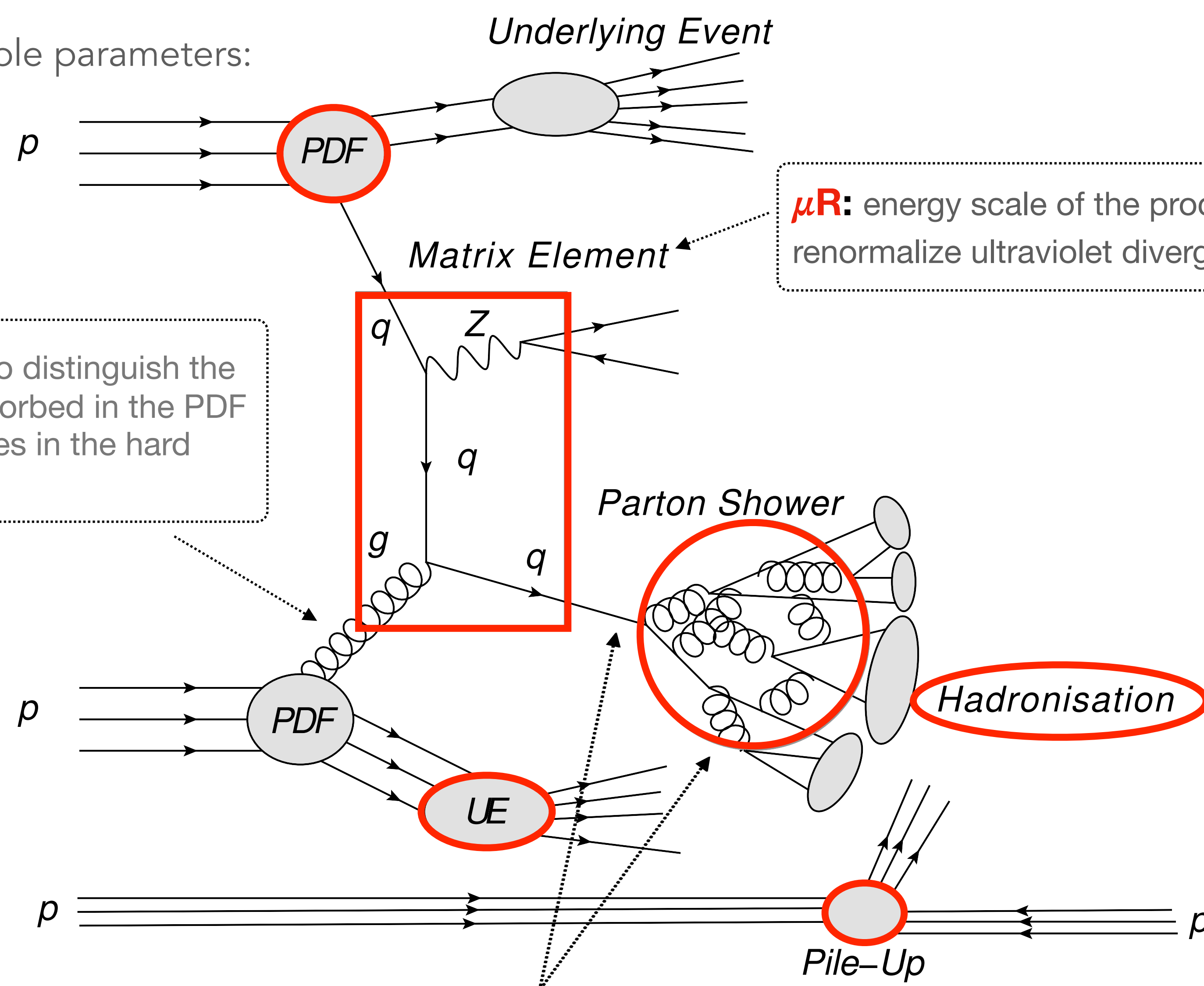
What MC uncertainties should we consider in our Systematic Model?

- Each generator is made up of building blocks tuned using an array of configurable parameters.
- Need to build our systematic model to account for the uncertainties on these configurable parameters.
- give enough freedom to the fit to absorb these potential Data/MC differences.

Modelling uncertainties

Start building the MC Systematic Model

Configurable parameters:



μ**F:** cutoff to distinguish the partons absorbed in the PDF from the ones in the hard scattering.

μ**R:** energy scale of the process (used to renormalize ultraviolet diverg. In loops).

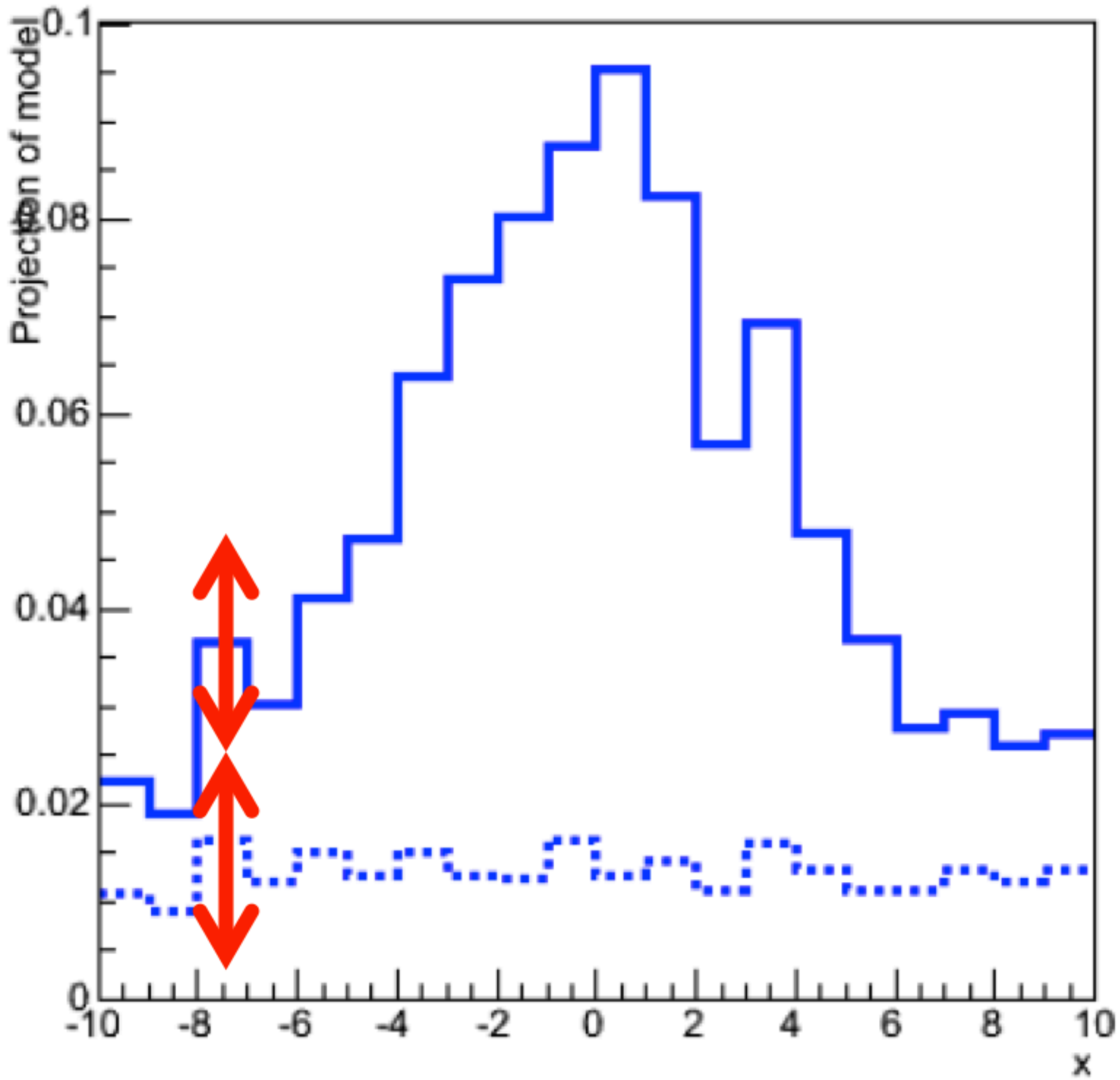
Matrix element matching scale (CKKW): the scale taken for the calculation of the overlap between jets from the matrix element and the parton shower.

Resummation scale (QSF): the scale used for the resummation of soft gluon emissions.

Consider one effect at a time:

- PDF
- Renormalization scale (μR)
- Factorisation scale (μF)
- Matrix Element
- Parton Shower
- Resummation Scale (QSF)
- CKKW
- Underlying Event
- Pile-up (not covered)
- EW corrections (not covered)
- Radiation High/Low

Statistical uncertainties



Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

The fit model:

Simultaneous binned Likelihood fit built across multiple analysis categories:

Nuisance parameters (NP)
Additional parameters
(not of interest).

Hypothesis testing:
$\mu = 0$ → background only hypothesis
$\mu = 1$ → SM signal+background hypothesis

$$\mathscr{L}(\mu, \boldsymbol{\theta}) = \prod_{i=1}^{N_{bins}} \frac{(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta}))} \times \mathscr{L}_{aux}(\boldsymbol{\theta})$$

Parameter of interest:
Signal strength

$$\mu = \frac{\sigma \cdot BR}{\sigma_{SM} \cdot BR_{SM}}$$

Poisson term for each bin

$$\mathscr{L}_{aux}(\boldsymbol{\theta}) = \prod_{j}^{N_{syst}} \frac{1}{\sqrt{2\pi\sigma_j}} \cdot e^{\frac{-(\bar{\theta}_j - \theta_j)^2}{2\sigma_j^2}}$$

Gaussian penalty term for the distribution of the NP priors

Systematic uncertainties are parametrized by **nuisance parameters (NPs)**, constrained with priors:

- *JES, JER, MET*
- *Lepton reco, ID, iso, calibration*
- *b-tagging uncertainties*
- *Lumi, pile-up*
- *Shapes and relative normalizations across regions*
- *Flavor composition uncertainties*
- *Theory uncertainties: PDF, scales, PS/UE*
- *…*

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

A concrete example:

Z + jets	
$Z+ll$ normalisation	18%
$Z+cl$ normalisation	23%
Z + HF normalisation	Floating (2-jet, 3-jet)
$Z+bc$-to-$Z+bb$ ratio	30 – 40%
$Z+cc$-to-$Z+bb$ ratio	13 – 15%
$Z+bl$-to-$Z+bb$ ratio	20 – 25%
0-to-2 lepton ratio	7%
m_{bb}, p_{T}^{V}	S
W + jets	
$W+ll$ normalisation	32%
$W+cl$ normalisation	37%
W + HF normalisation	Floating (2-jet, 3-jet)
$W+bl$-to-$W+bb$ ratio	26% (0-lepton) and 23% (1-lepton)
$W+bc$-to-$W+bb$ ratio	15% (0-lepton) and 30% (1-lepton)
$W+cc$-to-$W+bb$ ratio	10% (0-lepton) and 30% (1-lepton)
0-to-1 lepton ratio	5%
W + HF CR to SR ratio	10% (1-lepton)
m_{bb}, p_{T}^{V}	S
$t\bar{t}$ (all are uncorrelated between the 0+1- and 2-lepton channels)	
$t\bar{t}$ normalisation	Floating (0+1-lepton, 2-lepton 2-jet, 2-lepton 3-jet)
0-to-1 lepton ratio	8%
2-to-3-jet ratio	9% (0+1-lepton only)
W + HF CR to SR ratio	25%
m_{bb}, p_{T}^{V}	S
Single top-quark	
Cross-section	4.6% (s-channel), 4.4% (t-channel), 6.2% (Wt)
Acceptance 2-jet	17% (t-channel), 55% ($Wt(bb)$), 24% (Wt(other))
Acceptance 3-jet	20% (t-channel), 51% ($Wt(bb)$), 21% (Wt(other))
m_{bb}, p_{T}^{V}	S (t-channel, $Wt(bb)$, Wt(other))
Multi-jet (1-lepton)	
Normalisation	60 – 100% (2-jet), 90 – 140% (3-jet)
BDT template	S

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

A concrete example:

Z + jets	
$Z + ll$ normalisation	18%
$Z + cl$ normalisation	23%
Z + HF normalisation	Floating (2-jet, 3-jet)
$Z + bc$-to-$Z + bb$ ratio	30 – 40%
$Z + cc$-to-$Z + bb$ ratio	13 – 15%
$Z + bl$-to-$Z + bb$ ratio	20 – 25%
0-to-2 lepton ratio	7%
m_{bb}, p_T^V	S
W + jets	
$W + ll$ normalisation	32%
$W + cl$ normalisation	37%
W + HF normalisation	Floating (2-jet, 3-jet)
$W + bl$-to-$W + bb$ ratio	26% (0-lepton) and 23% (1-lepton)
$W + bc$-to-$W + bb$ ratio	15% (0-lepton) and 30% (1-lepton)
$W + cc$-to-$W + bb$ ratio	10% (0-lepton) and 30% (1-lepton)
0-to-1 lepton ratio	5%
W + HF CR to SR ratio	
m_{bb}, p_T^V	
$t\bar{t}$ (all are uncorrelated betw	
$t\bar{t}$ normalisation	Floating
0-to-1 lepton ratio	
2-to-3-jet ratio	
W + HF CR to SR ratio	
m_{bb}, p_T^V	
Singl	
Cross-section	4.6%
Acceptance 2-jet	17% (t-
Acceptance 3-jet	20% (t-
m_{bb}, p_T^V	
Multi-jet (1-lepton)	
Normalisation	60 – 100% (2-jet), 90 – 140% (3-jet)
BDT template	S

Signal	
Cross-section (scale)	0.7% (qq), 27% (gg)
Cross-section (PDF)	1.9% ($qq \to WH$), 1.6% ($qq \to ZH$), 5% (gg)
$H \to b\bar{b}$ branching fraction	1.7%
Acceptance from scale variations	2.5 – 8.8%
Acceptance from PS/UE variations for 2 or more jets	2.9 – 6.2% (depending on lepton channel)
Acceptance from PS/UE variations for 3 jets	1.8 – 11%
Acceptance from PDF+α_S variations	0.5 – 1.3%
m_{bb}, p_T^V, from scale variations	S
m_{bb}, p_T^V, from PS/UE variations	S
m_{bb}, p_T^V, from PDF+α_S variations	S
p_T^V from NLO EW correction	S

A concrete example:

Z + jets	
$Z+ll$ normalisation	18%
$Z+cl$ normalisation	23%
Z + HF normalisation	Floating (2-jet, 3-jet)
$Z+bc$-to-$Z+bb$ ratio	30 – 40%
$Z+cc$-to-$Z+bb$ ratio	13 – 15%
$Z+bl$-to-$Z+bb$ ratio	20 – 25%
0-to-2 lepton ratio	7%
m_{bb}, p_T^V	S
W + jets	
$W+ll$ normalisation	32%
$W+cl$ normalisation	37%
W + HF normalisation	Floating (2-jet, 3-jet)
$W+bl$-to-$W+bb$ ratio	26% (0-lepton) and 23% (1-lepton)
$W+bc$-to-$W+bb$ ratio	15% (0-lepton) and 30% (1-lepton)
$W+cc$-to-$W+bb$ ratio	10% (0-lepton) and 30% (1-lepton)
0-to-1 lepton ratio	5%
W + HF CR to SR ratio	
m_{bb}, p_T^V	
$t\bar{t}$ (all are uncorrelated betw	
$t\bar{t}$ normalisation	Floating
0-to-1 lepton ratio	
2-to-3-jet ratio	
W + HF CR to SR ratio	
m_{bb}, p_T^V	
Singl	
Cross-section	4.6%
Acceptance 2-jet	17% (t-
Acceptance 3-jet	20% (t-
m_{bb}, p_T^V	
Multi-jet (1-lepton)	
Normalisation	60 – 100% (2-jet), 90 – 140% (3-jet)
BDT template	S

Cross-section (scale)	
Cross-section (PDF)	gg)
$H \to b\bar{b}$ branching fraction	
Acceptance from scale variat	
Acceptance from PS/UE variations for 2 or more jets	2.9 – 6.2% (depending on lepton channel)
Acceptance from PS/UE variations for 3 jets	1.8 – 11%
Acceptance from PDF+α_S variations	0.5 – 1.3%
m_{bb}, p_T^V, from scale variations	S
m_{bb}, p_T^V, from PS/UE variations	S
m_{bb}, p_T^V, from PDF+α_S variations	S
p_T^V from NLO EW correction	S

ZZ	
Normalisation	20%
0-to-2 lepton ratio	6%
Acceptance from scale variations	10 – 18%
Acceptance from PS/UE variations for 2 or more jets	6%
Acceptance from PS/UE variations for 3 jets	7% (0-lepton), 3% (2-lepton)
m_{bb}, p_T^V, from scale variations	S (correlated with WZ uncertainties)
m_{bb}, p_T^V, from PS/UE variations	S (correlated with WZ uncertainties)
m_{bb}, from matrix-element variations	S (correlated with WZ uncertainties)
WZ	
Normalisation	26%
0-to-1 lepton ratio	11%
Acceptance from scale variations	13 – 21%
Acceptance from PS/UE variations for 2 or more jets	4%
Acceptance from PS/UE variations for 3 jets	11%
m_{bb}, p_T^V, from scale variations	S (correlated with ZZ uncertainties)
m_{bb}, p_T^V, from PS/UE variations	S (correlated with ZZ uncertainties)
m_{bb}, from matrix-element variations	S (correlated with ZZ uncertainties)
WW	
Normalisation	25%

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

A concrete example:

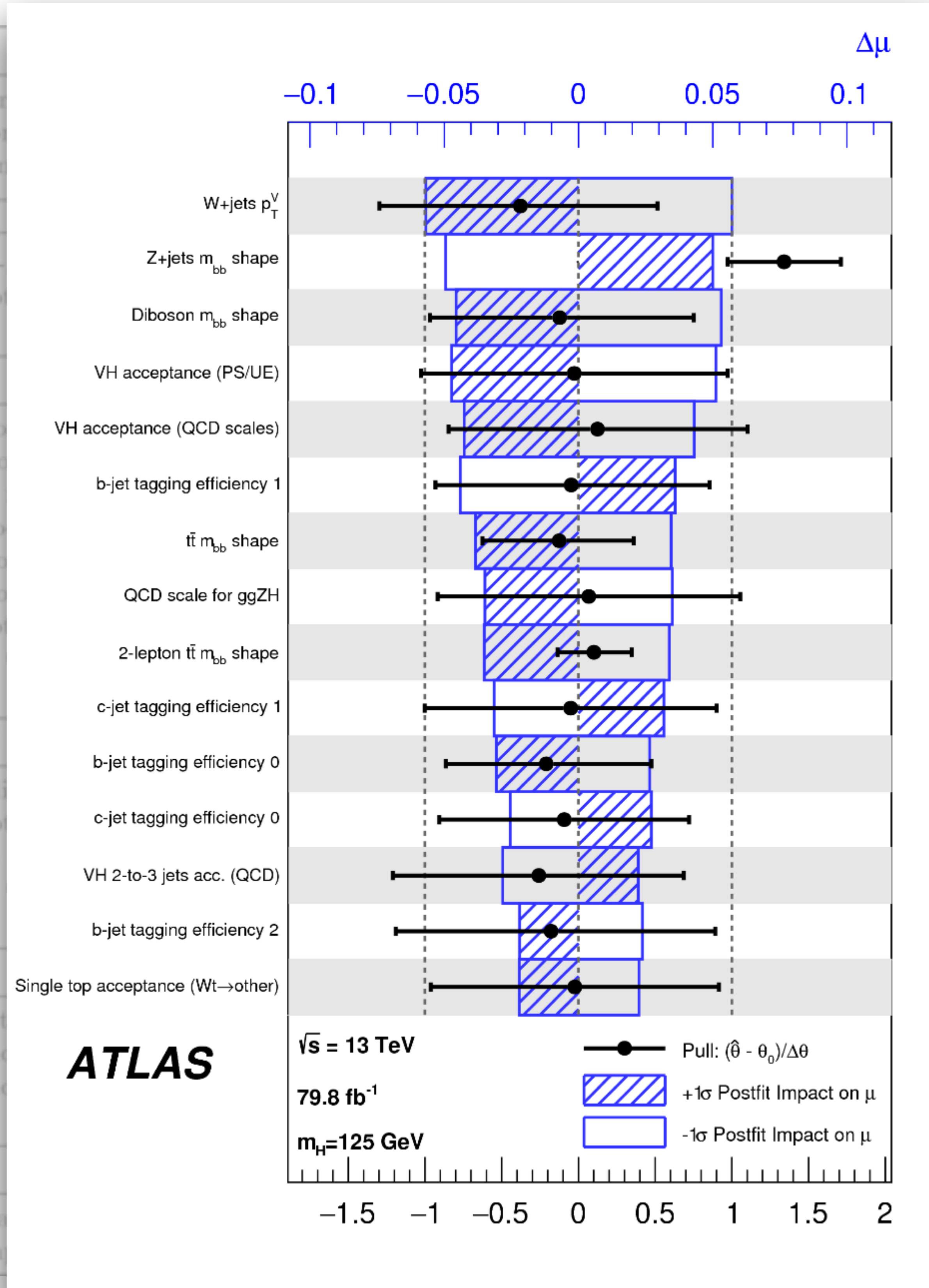
Z + jets	
$Z + ll$ normalisation	18%
$Z + cl$ normalisation	23%
Z + HF normalisation	Floating (2-jet, 3-jet)
$Z + bc$-to-$Z + bb$ ratio	30 – 40%
$Z + cc$-to-$Z + bb$ ratio	13 – 15%
$Z + bl$-to-$Z + bb$ ratio	20 – 25%
0-to-2 lepton ratio	7%
m_{bb}, p_{T}^{V}	S
W + jets	
$W + ll$ normalisation	32%
$W + cl$ normalisation	37%
W + HF normalisation	Floating (2-jet, 3-jet)
$W + bl$-to-$W + bb$ ratio	26% (0-lepton) and 23% (1-lepton)
$W + bc$-to-$W + bb$ ratio	15% (0-lepton) and 30% (1-lepton)
$W + cc$-to-$W + bb$ ratio	10% (0-lepton) and 30% (1-lepton)
0-to-1 lepton ratio	5%
W + HF CR to SR ratio	
m_{bb}, p_{T}^{V}	
$t\bar{t}$ (all are uncorrelated betw	
$t\bar{t}$ normalisation	Floating
0-to-1 lepton ratio	
2-to-3-jet ratio	
W + HF CR to SR ratio	
m_{bb}, p_{T}^{V}	
Singl	
Cross-section	4.6%
Acceptance 2-jet	17% (t-
Acceptance 3-jet	20% (t-
m_{bb}, p_{T}^{V}	
Multi-jet (1-lepton)	
Normalisation	60 – 100% (2-jet), 90 – 140% (3-jet)
BDT template	S

Normalisation
0-to-2 lepton ratio
Acceptance from scale variatio
Acceptance from PS/UE varia
Acceptance from PS/UE varia
m_{bb}, p_{T}^{V}, from scale variations
m_{bb}, p_{T}^{V}, from PS/UE variatio
m_{bb}, from matrix-element vari
Normalisation
0-to-1 lepton ratio
Acceptance from scale variatio
Acceptance from PS/UE varia
Acceptance from PS/UE varia
m_{bb}, p_{T}^{V}, from scale variations
m_{bb}, p_{T}^{V}, from PS/UE variatio
m_{bb}, from matrix-element vari
Normalisation

Cross-section (scale)
Cross-section (PDF)
$H \to b\bar{b}$ branching fraction
Acceptance from scale variat
Acceptance from PS/UE variations for 2 or more je
Acceptance from PS/UE variations for 3 jets
Acceptance from PDF+α_{S} variations
m_{bb}, p_{T}^{V}, from scale variations
m_{bb}, p_{T}^{V}, from PS/UE variations
m_{bb}, p_{T}^{V}, from PDF+α_{S} variations
p_{T}^{V} from NLO EW correction

Source of uncertainty		σ_μ
Total		0.259
Statistical		0.161
Systematic		0.203
Experimental uncertainties		
Jets		0.035
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.014
Leptons		0.009
b-tagging	b-jets	0.061
	c-jets	0.042
	light-flavour jets	0.009
	extrapolation	0.008
Pile-up		0.007
Luminosity		0.023
Theoretical and modelling uncertainties		
Signal		0.094
Floating normalisations		0.035
Z + jets		0.055
W + jets		0.060
$t\bar{t}$		0.050
Single top quark		0.028
Diboson		0.054
Multi-jet		0.005
MC statistical		0.070

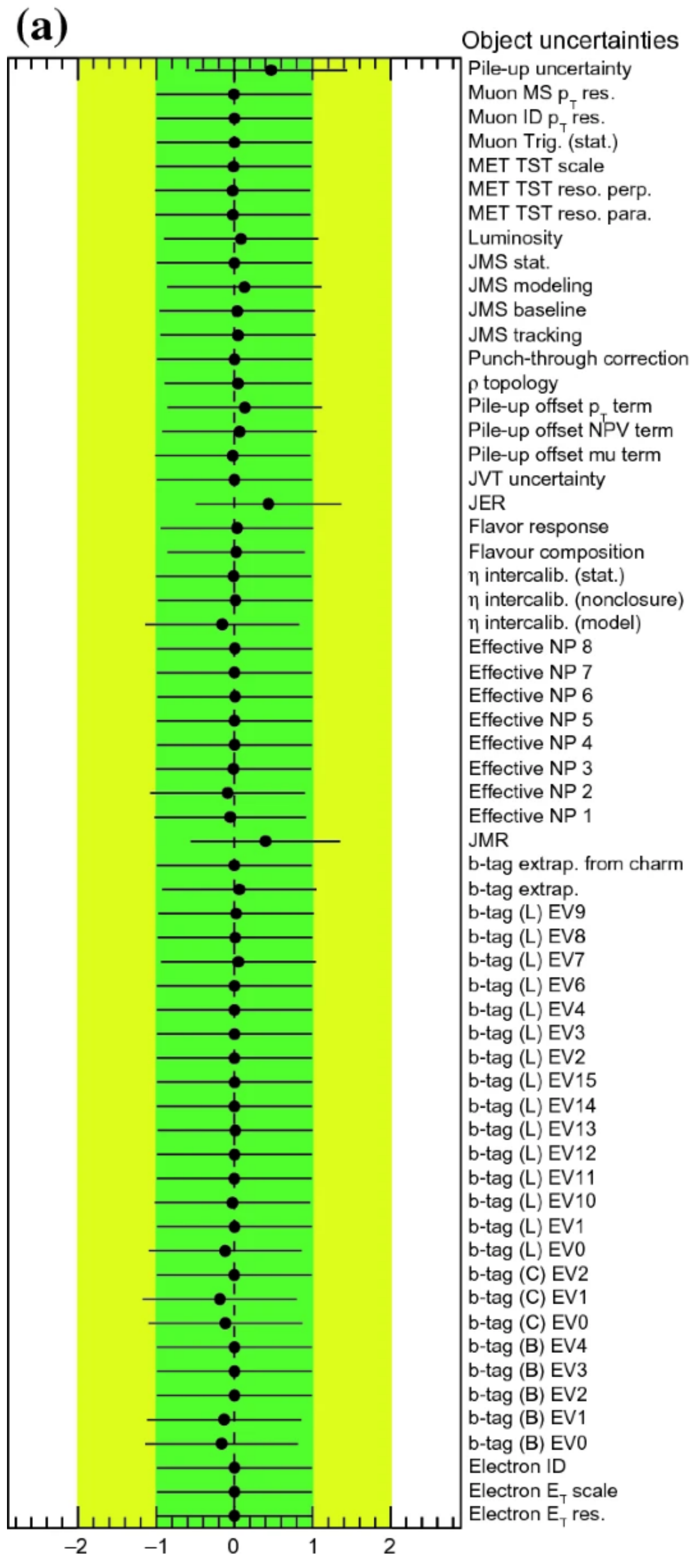
A concrete example:

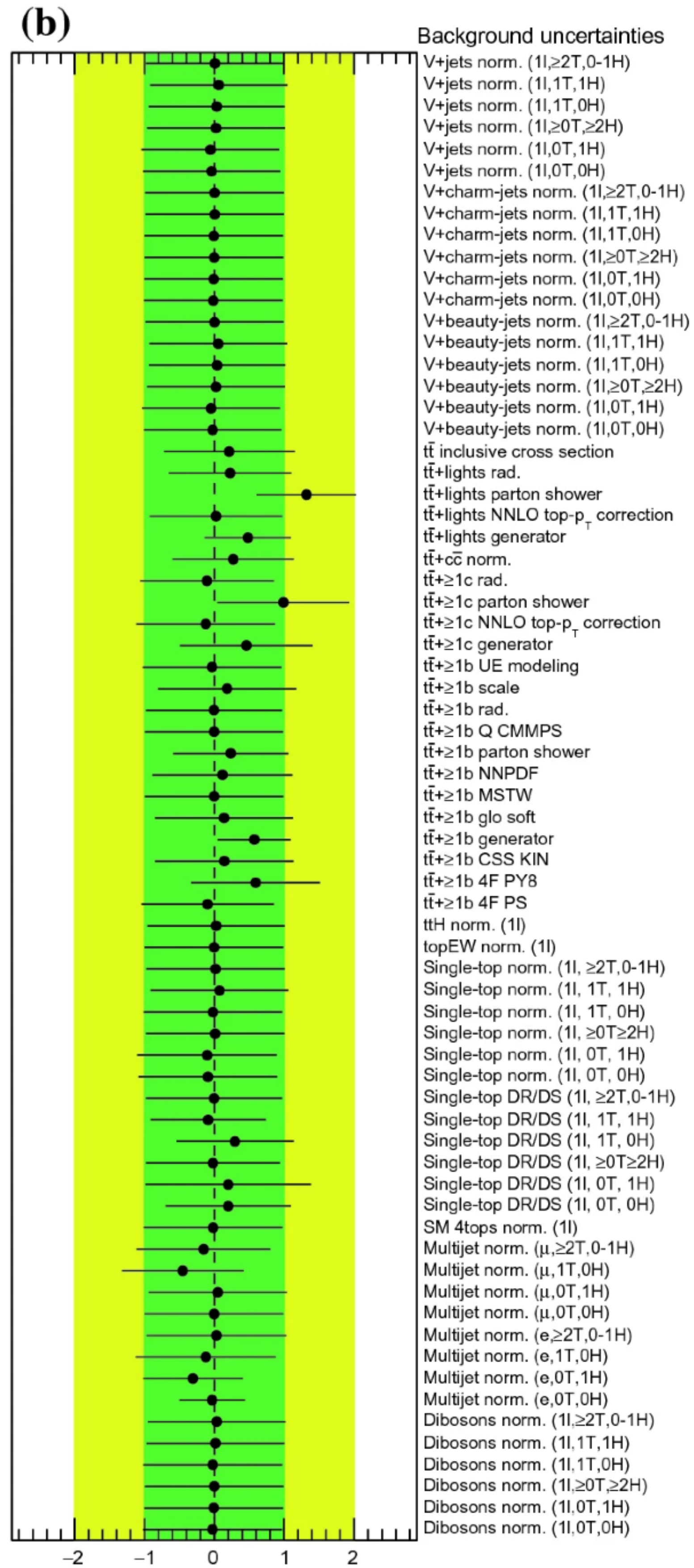


Source of uncertainty		σ_μ
Total		0.259
Statistical		0.161
Systematic		0.203
Experimental uncertainties		
Jets		0.035
E_T^{miss}		0.014
Leptons		0.009
b-tagging	b-jets	0.061
	c-jets	0.042
	light-flavour jets	0.009
	extrapolation	0.008
Pile-up		0.007
Luminosity		0.023
Theoretical and modelling uncertainties		
Signal		0.094
Floating normalisations		0.035
Z + jets		0.055
W + jets		0.060
$t\bar{t}$		0.050
Single top quark		0.028
Diboson		0.054
Multi-jet		0.005
MC statistical		0.070

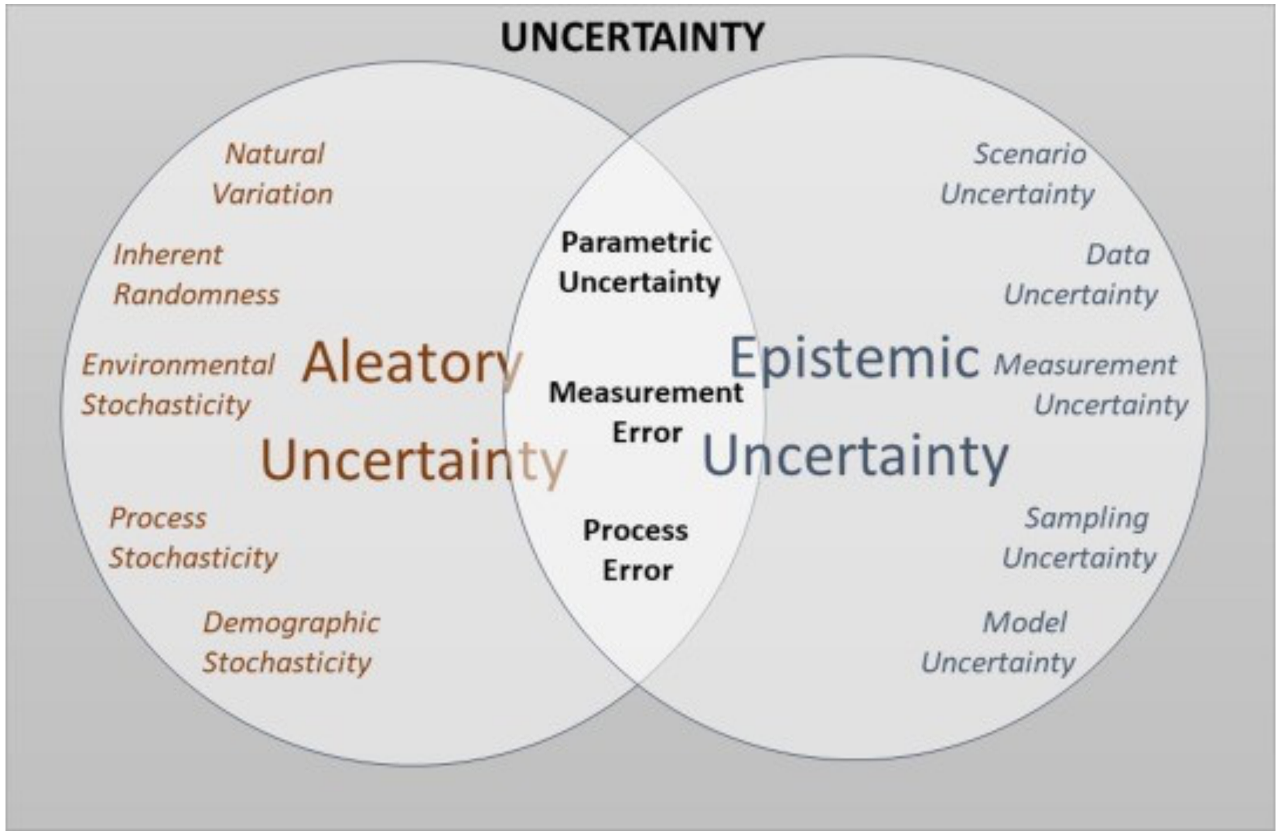
A concrete example:

The “pull plot”





Uncertainties in Machine Learning



Introduction

A review of uncertainty quantification in deep learning: Techniques, applications and challenges, M. Abdar et al., https://doi.org/10.1016/j.inffus.2021.05.008

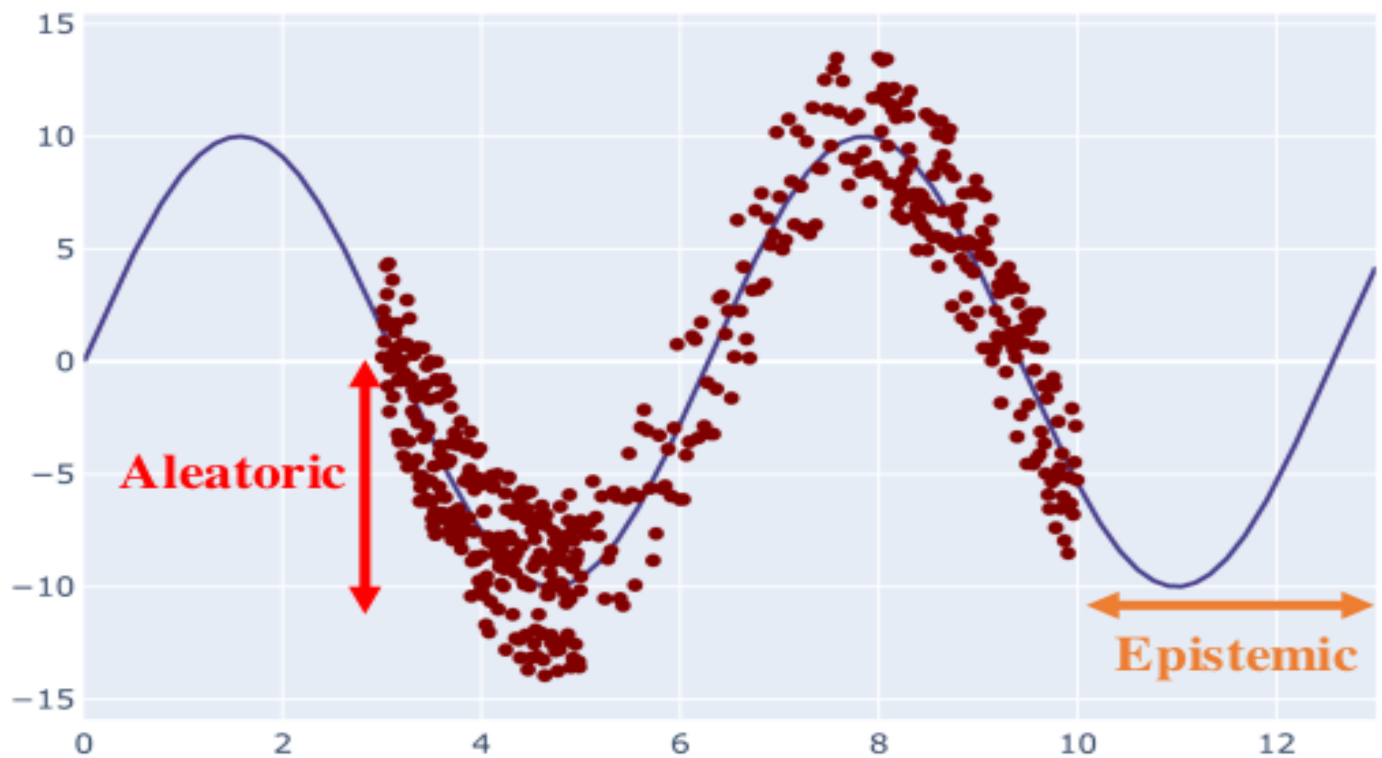
Let x an input point, f_ω a predictive model with parameters ω

Objective: Quantifying the uncertainty on the prediction $f_\omega(x)$

Predictive uncertainty

Aleatoric uncertainty
Uncertainty related to the data

Epistemic uncertainty
Uncertainty related to the model



Representation of the total **predictive uncertainty** by a probability distribution

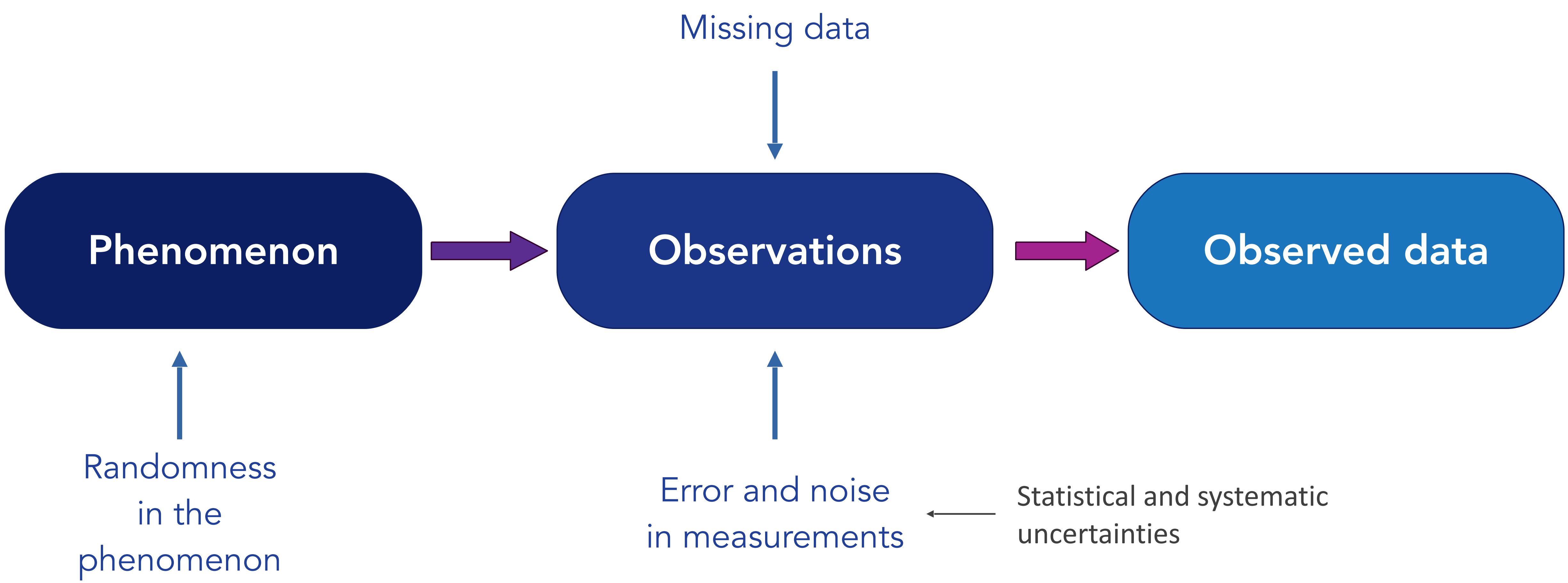
$$p(y^* \mid x, D) = \int_\omega \underbrace{p(y^* \mid x, \omega)}_{\text{Aleatoric}} \underbrace{p(\omega \mid D)}_{\text{Epistemic}} d\omega$$

x : input data point
ω : model parameters
y^* : possible output
D : Training dataset

Slide from G. Daniel

Aleatoric uncertainties

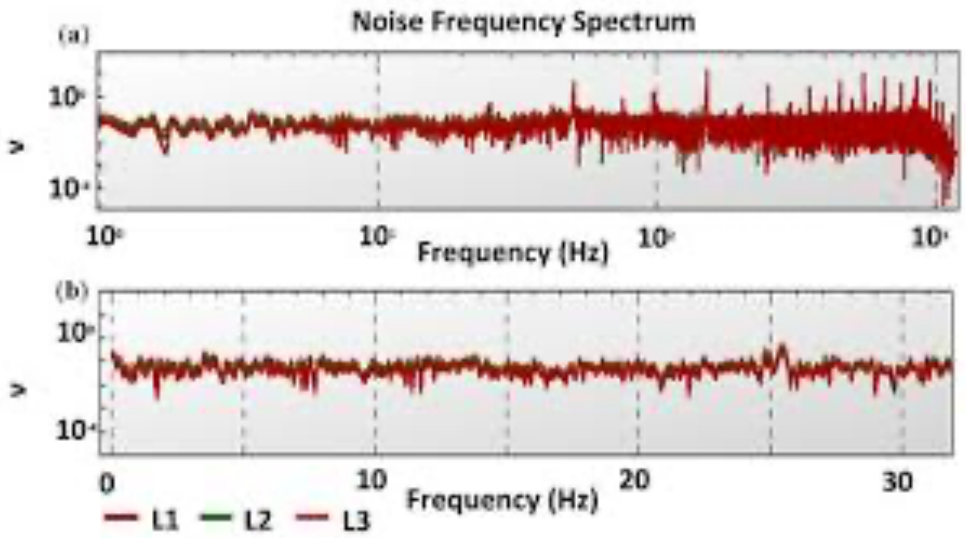
Uncertainty intrinsic within the data, irreducible by improving the model or increasing the dataset
A larger dataset does not reduce aleatoric uncertainty,
but it helps to give a better estimation!



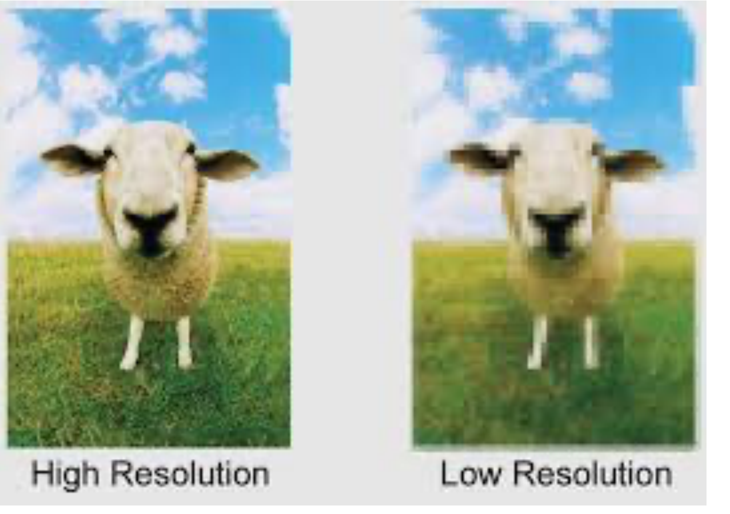
We can reduce the aleatoric uncertainty **by improving the measurement (reducing the error or noise)** for instance.

Slide from G. Daniel

Aleatoric uncertainties: examples

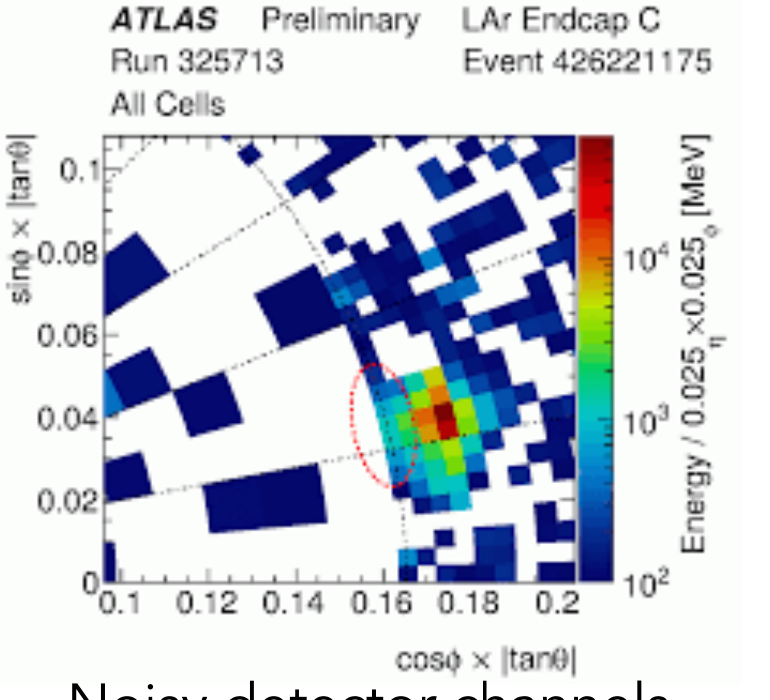


Noisy spectra



Noisy images

Text from social media



Noisy detector channels
(e.g. for reconstruction)

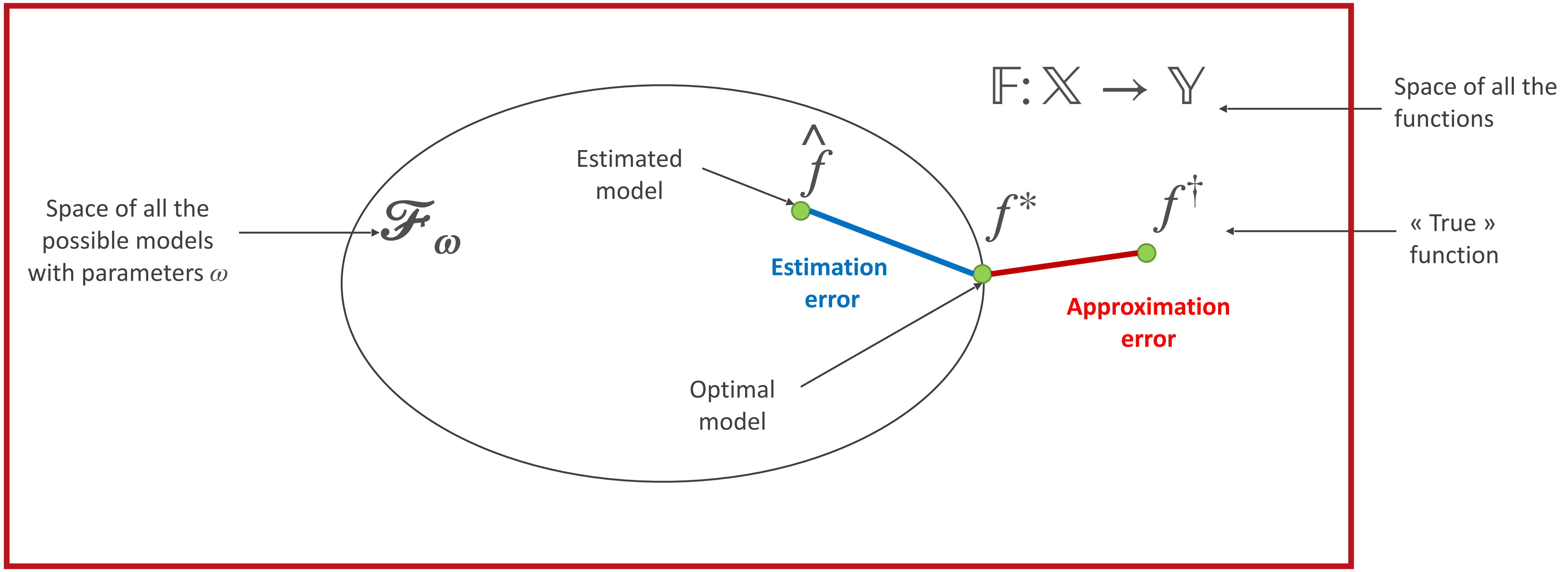
Slide from G. Daniel

Epistemic uncertainties

Represents the lack of « knowledge » or « understanding » of a model on a specific input data point

Two main origins of epistemic uncertainty for machine learning models:

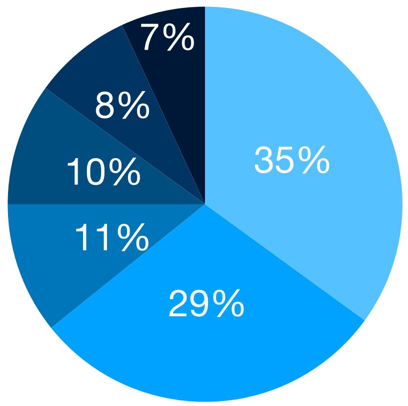
- **Estimation error**: the training dataset is just a sample of all the possible observable data
- **Approximation error**: no model can approximate perfectly the unknown « true » function



It can be possible to reduce epistemic uncertainty by using more data and increasing the model complexity

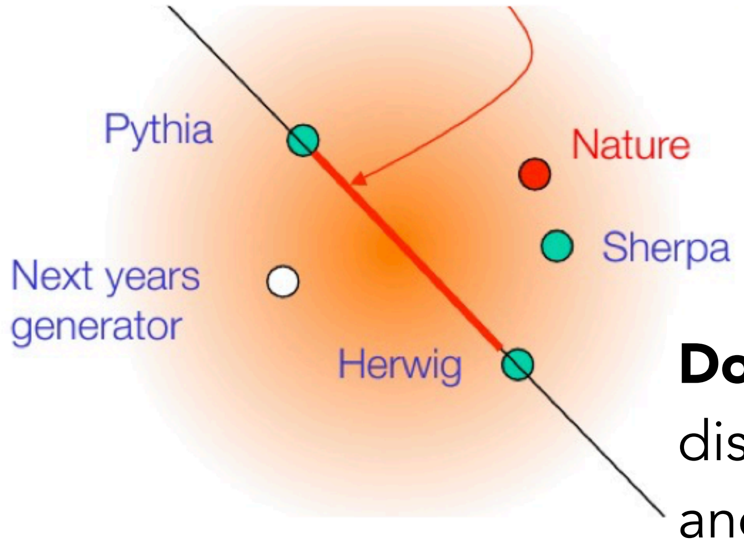
Epistemic uncertainties: examples

Epistemic uncertainty refers to the uncertainty of the model (epistemology is the study of knowledge) and is **often due to a lack of training data**.

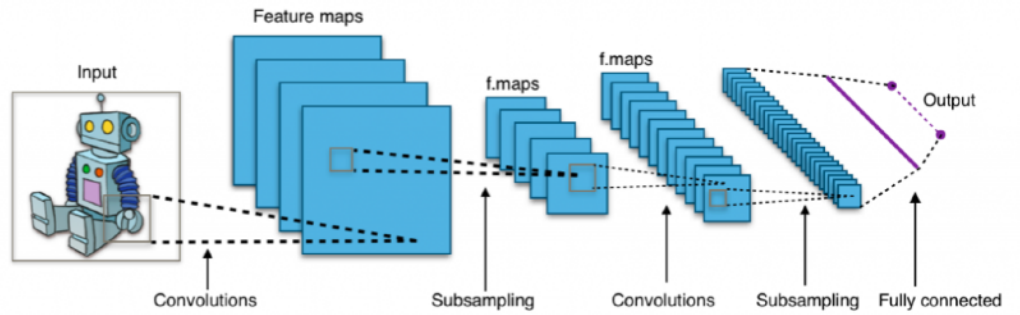


Rare or underrepresented occurrences in a dataset

Rare words in a text dataset



Domain shift: differences in distribution between data and Montecarlo or between test and training datasets

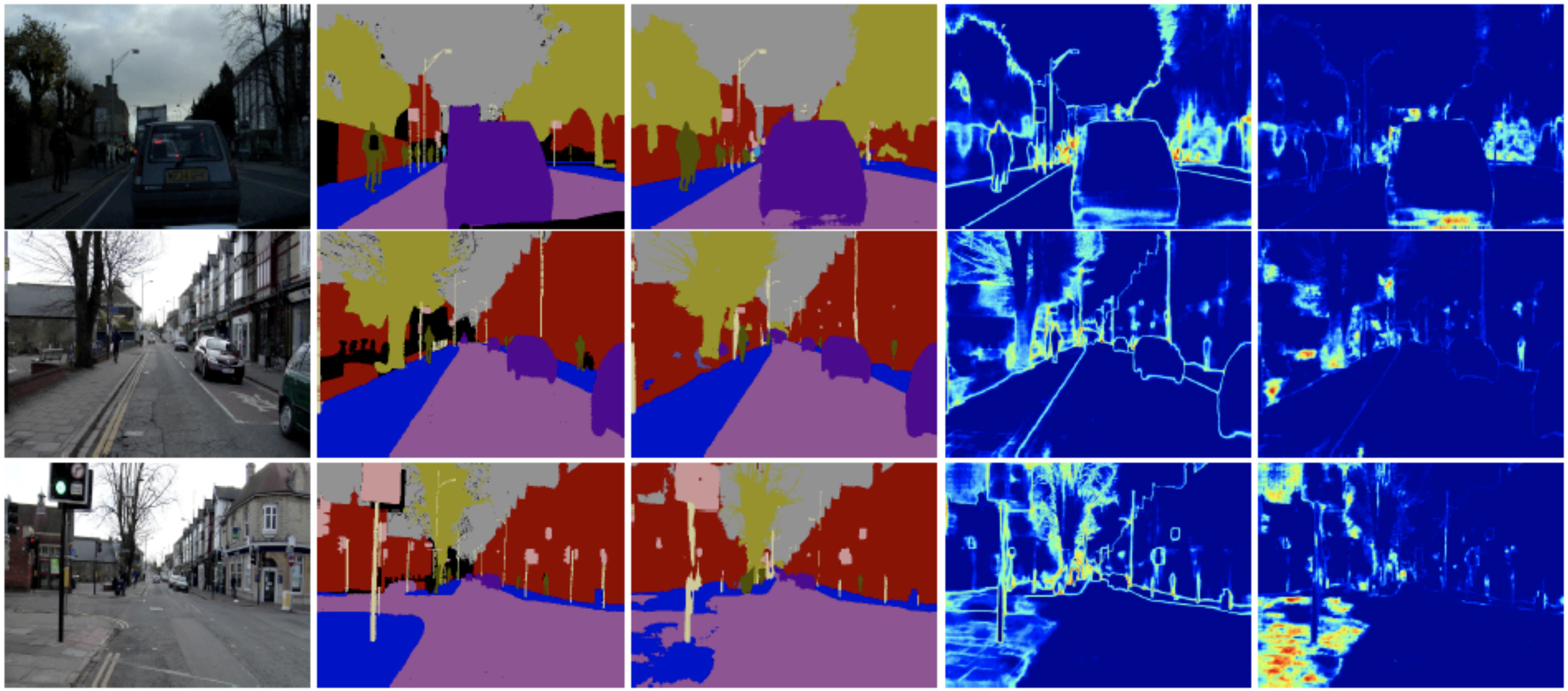


Choice of the ML architecture

Slide from G. Daniel

Uncertainties example

the model fails to segment the footpath due to increased epistemic uncertainty, but not aleatoric uncertainty



(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

Kendal, Gal, NIPS 2017, https://papers.nips.cc/paper/2017/file/2650d6089a6d640c5e85b2b88265dc2b-Paper.pdf

Can we match these uncertainties with what we have seen in HEP analyses?

- Aleatoric uncertainties
- Epistemic uncertainties

- Experimental uncertainties
- Modelling uncertainties
  - Shape uncertainties (change in distribution)
  - From limited knowledge of the distribution
- Statistical uncertainties

Final answer (debatable, but still..):

Machine Learning	HEP
Aleatoric uncertainty • "Statistical" / "Data" Uncertainty • Uncertainty Inherent to data • Not reduced w/ more data	Detector Noise Resolutions
	Stat. errors in HEP ?
Epistemic uncertainty • "Model" Uncertainty • Uncertainty from Imperfect knowledge • Reduces with more data	Systematic errors induced by ML model training on finite stats.
Domain Shift • Imperfect model of data generation process	Systematic Uncertainties from data / simulation differences

***Even within the ML community, these terms can be ambiguous**

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

How to reduce uncertainties:

How might we **reduce** uncertainty? (ML perspective)

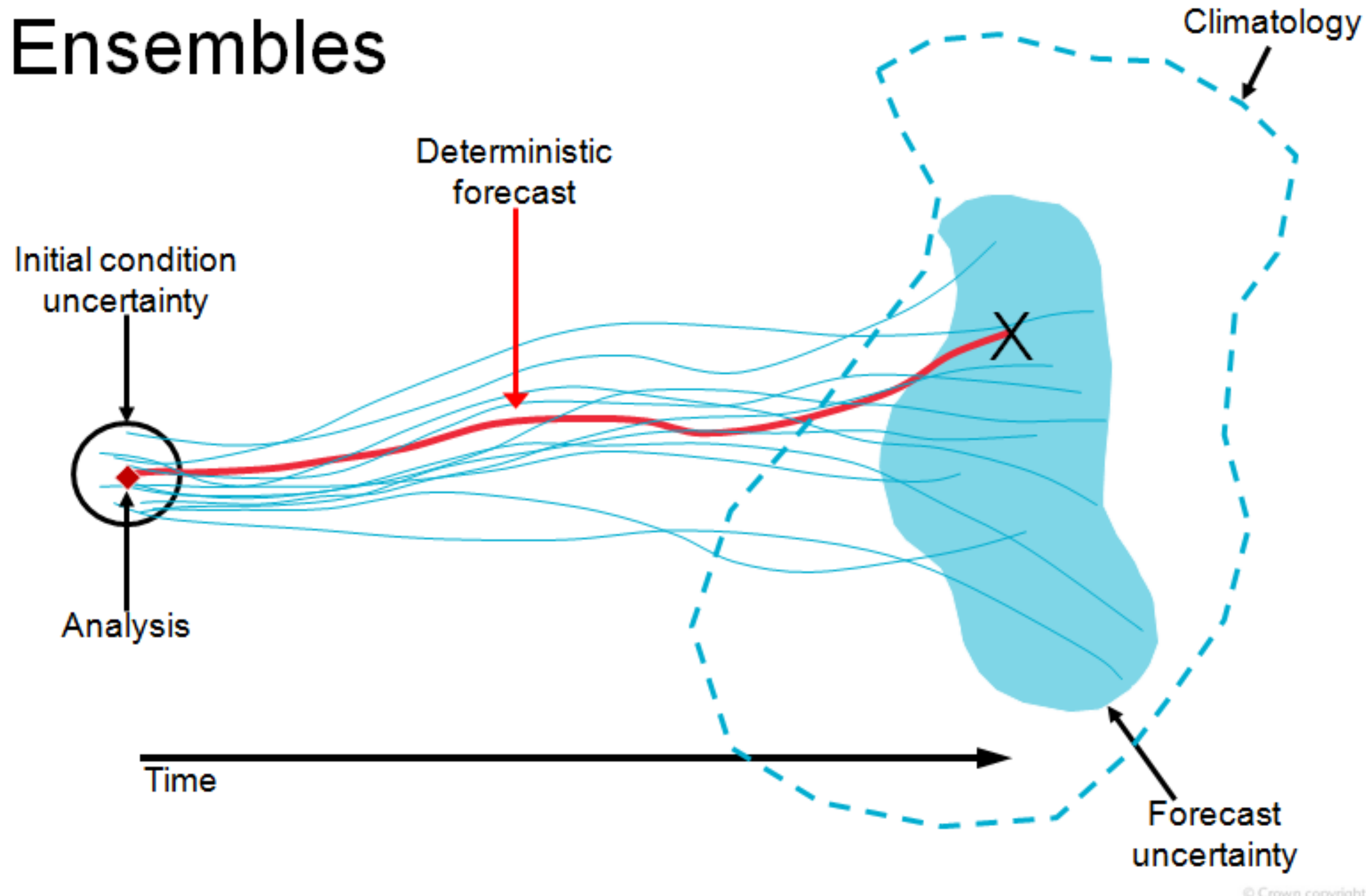
Uncertainty *about* the model (its structure and parameters)	**Initial condition uncertainty**	**Uncertainty due to limitations of the model** (modelled as stochastic dynamics)
Use **more historical data** and **compute** for model selection and parameter learning. More data-efficient and compute-efficient model architectures and learning methods	Assimilate **more observations** (and more precise obs) Better assimilation methods (could be ML-based) Better models used for assimilation (see <-- and -->)	*Subject to enough data*: allow the model more: • **Learning capacity** (parameter count, ...) • **Computational capacity** (resolution, latent size, message-passing steps, ...) • **State representation capacity** (resolution, latent size, ...)

Limits of predictability: we expect some uncertainty is irreducible, for anything short of a perfect model and perfect initial conditions

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

Deterministic vs stochastic models

- For ML models, stochasticity is bound up with physical realism.
- Much easier to produce realistic outputs from a stochastic ML model (‘generative model’) than a deterministic ML model.
- Technical tip: Deterministic ML loss functions without physical constraints will tend to blurry the hedge of uncertainty



Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

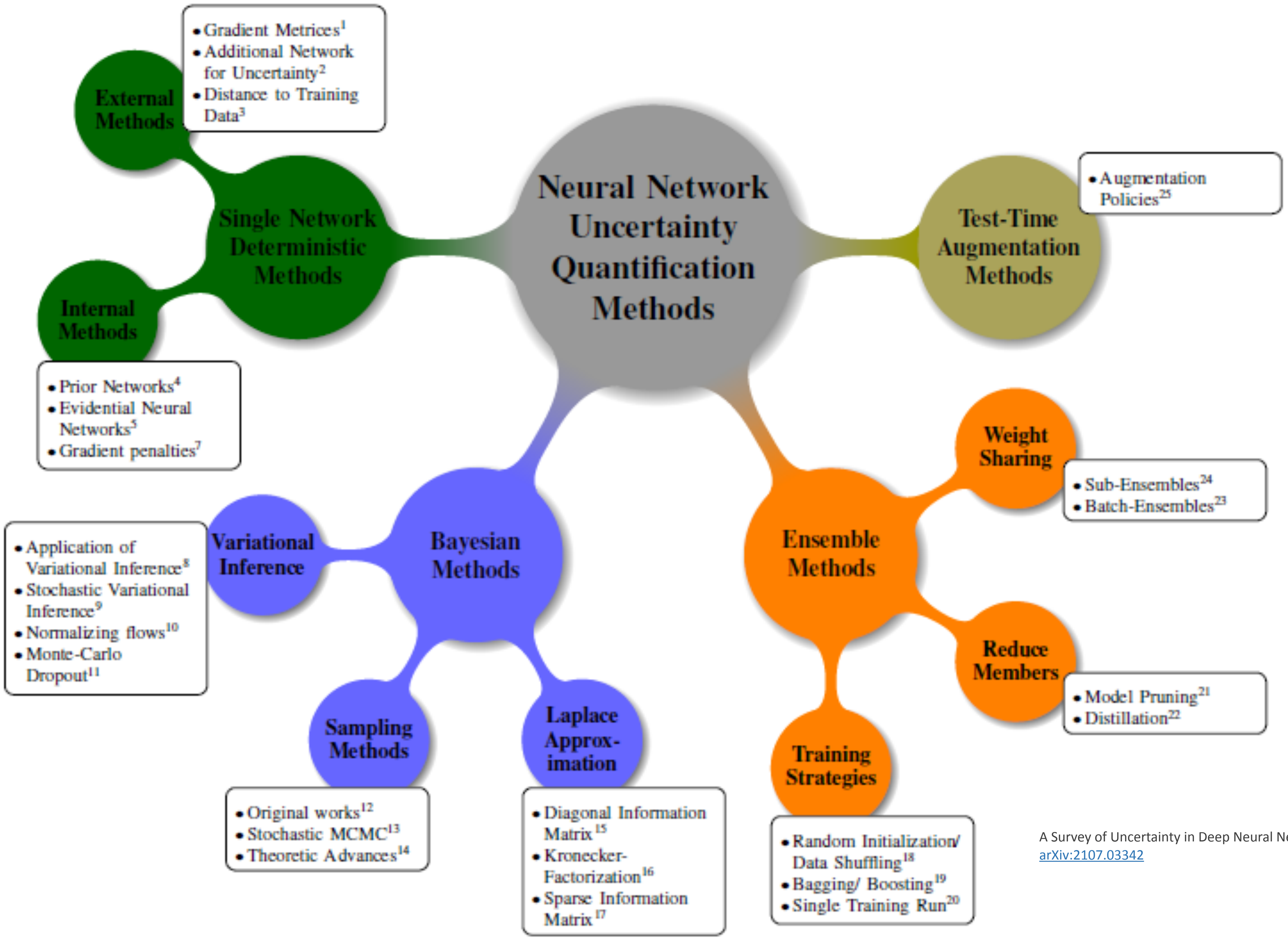
How to represent uncertainties

How might we **represent** uncertainty? (ML perspective)

Uncertainty *about* the model (its structure and parameters)	**Initial condition uncertainty**	**Uncertainty *due to* limitations of the model** (modelled as stochastic dynamics)
Bayesian ML methods: • to obtain approximate posterior over parameters • or over model structures Ad-hoc multi-model ensembles: • **trained from multiple random initializations** • trained on different resampled datasets ...	**Ensemble data assimilation** **Ad-hoc initial perturbations** End-to-end ML model conditioning directly on obs ...	**Probabilistic generative models** (**Diffusion**, GANs, VAEs, flows, scoring-rule minimization, ...) Ad-hoc perturbations at each timestep ...

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

A list

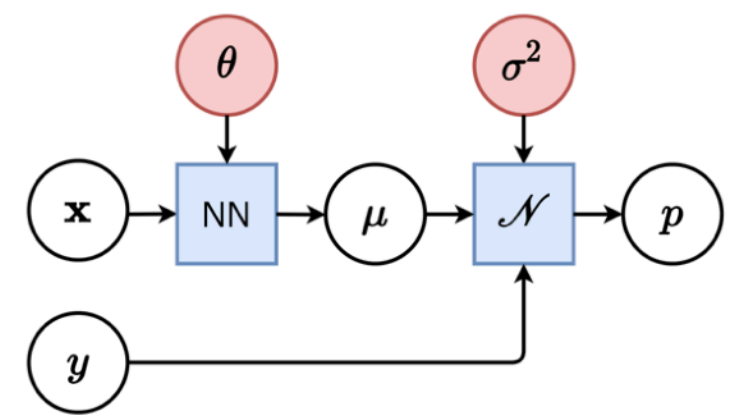


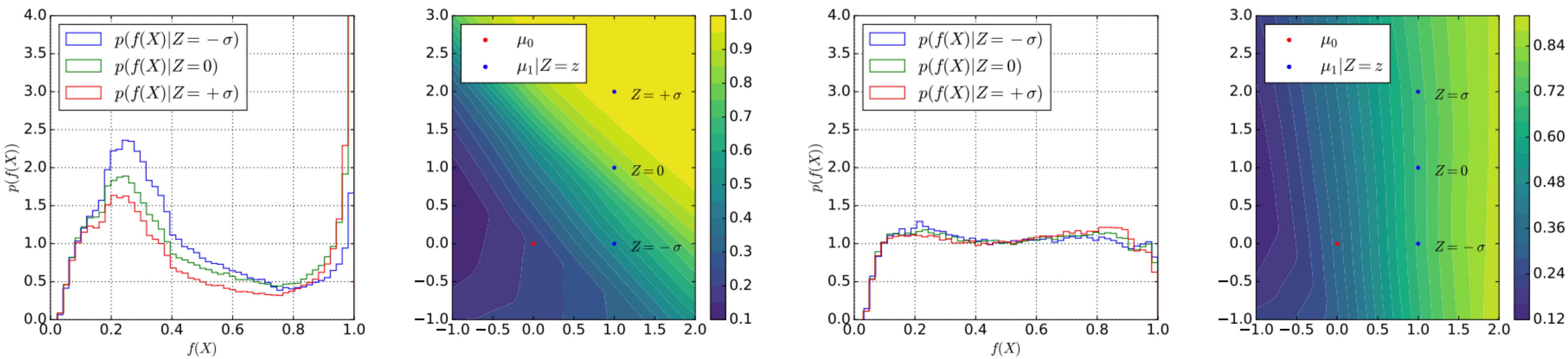
A Survey of Uncertainty in Deep Neural Networks, J. Gawlikowski et al,, arXiv:2107.03342

Example: Learning Systematics

- **Ex. Regression**: model aleatoric uncertainty in the output by modelling the conditional distribution as a Normal distribution

- Generative models –based uncertainty learning

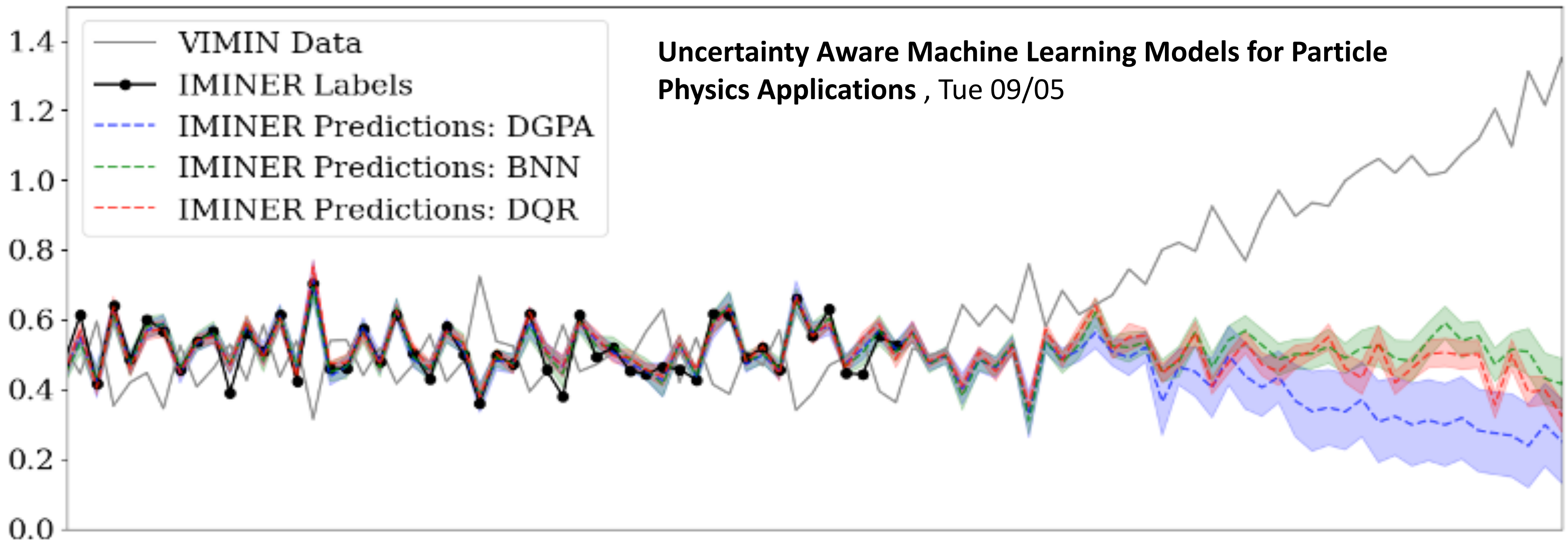




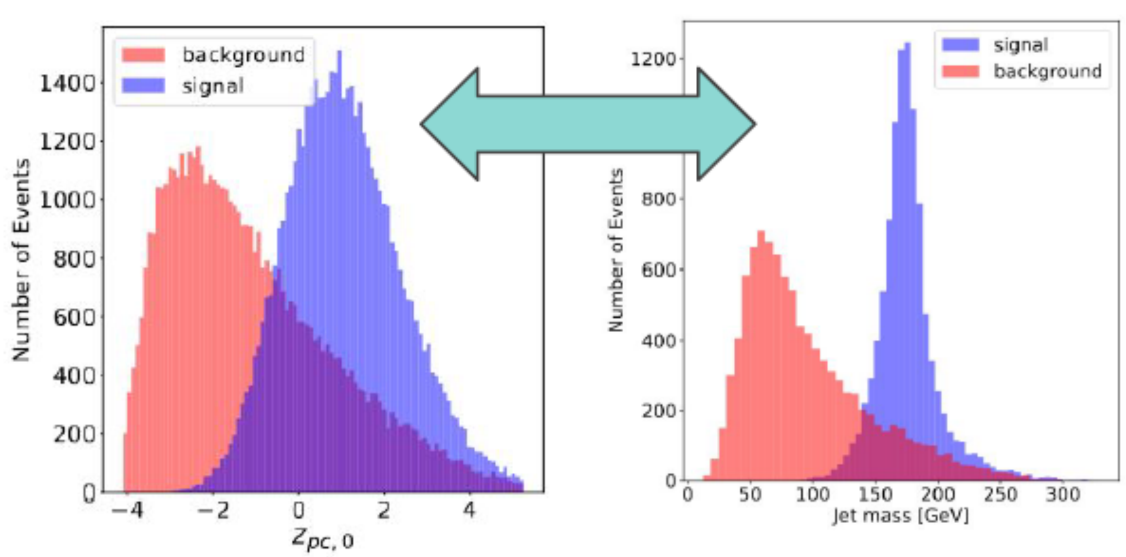
Louppe, Gilles, Michael Kagan, and Kyle Cranmer. "Learning to pivot with adversarial networks." arXiv:1611.01046 (2016).

Interpretability

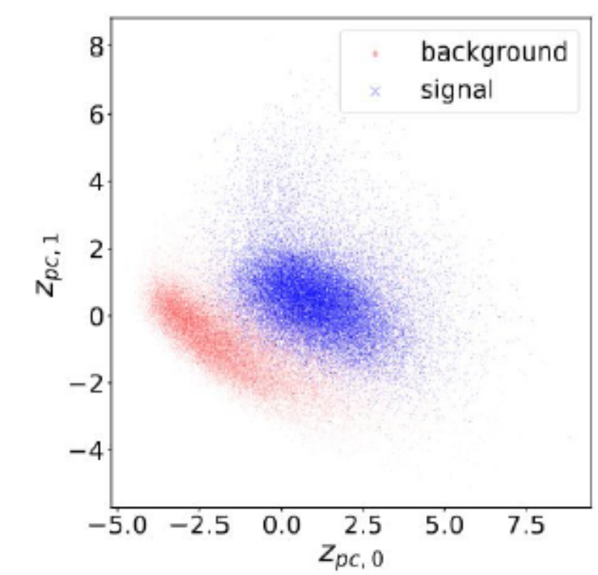
Uncertainty Aware Machine Learning Models for Particle Physics Applications , Tue 09/05



Interpretability Inspires: Explainable AI for DNN Top Taggers, CHEP2023



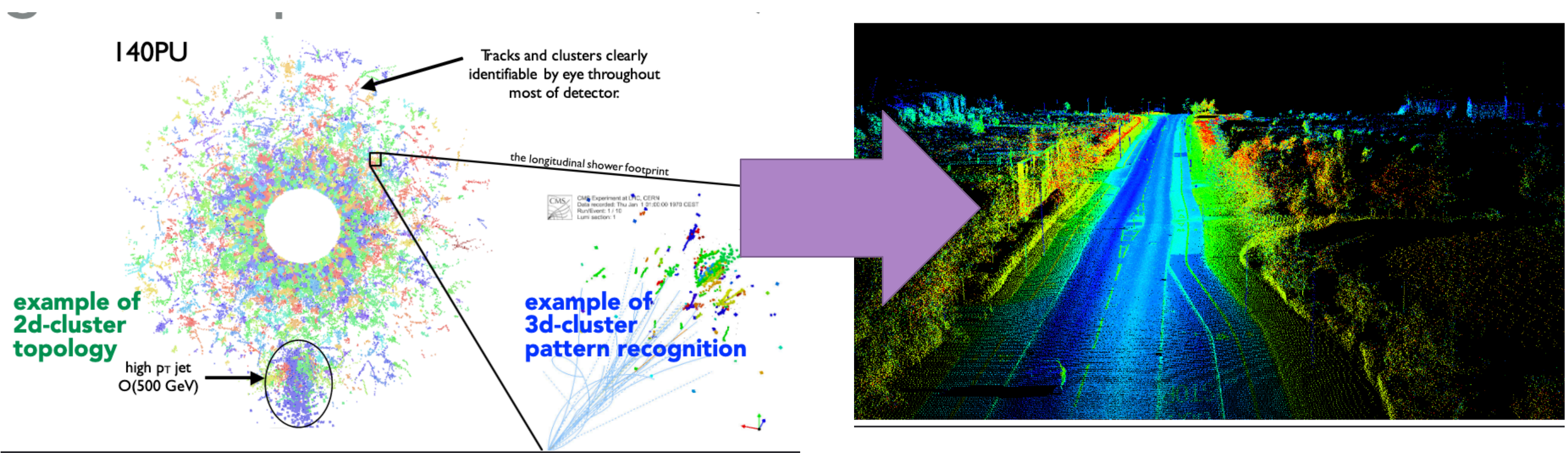
Jet class information is encoded in the correlation structure of the latent spaces



FAIR principles

FAIR:

Findability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets



Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

Some resources:

- PHYSTAT seminar: On relating Uncertainties in Machine Learning and HEP [link]

- Uncertainties workshop at Learning to Discover

- Great new ML review in PDG: [Cranmer, Seljak, Terao, 2021]

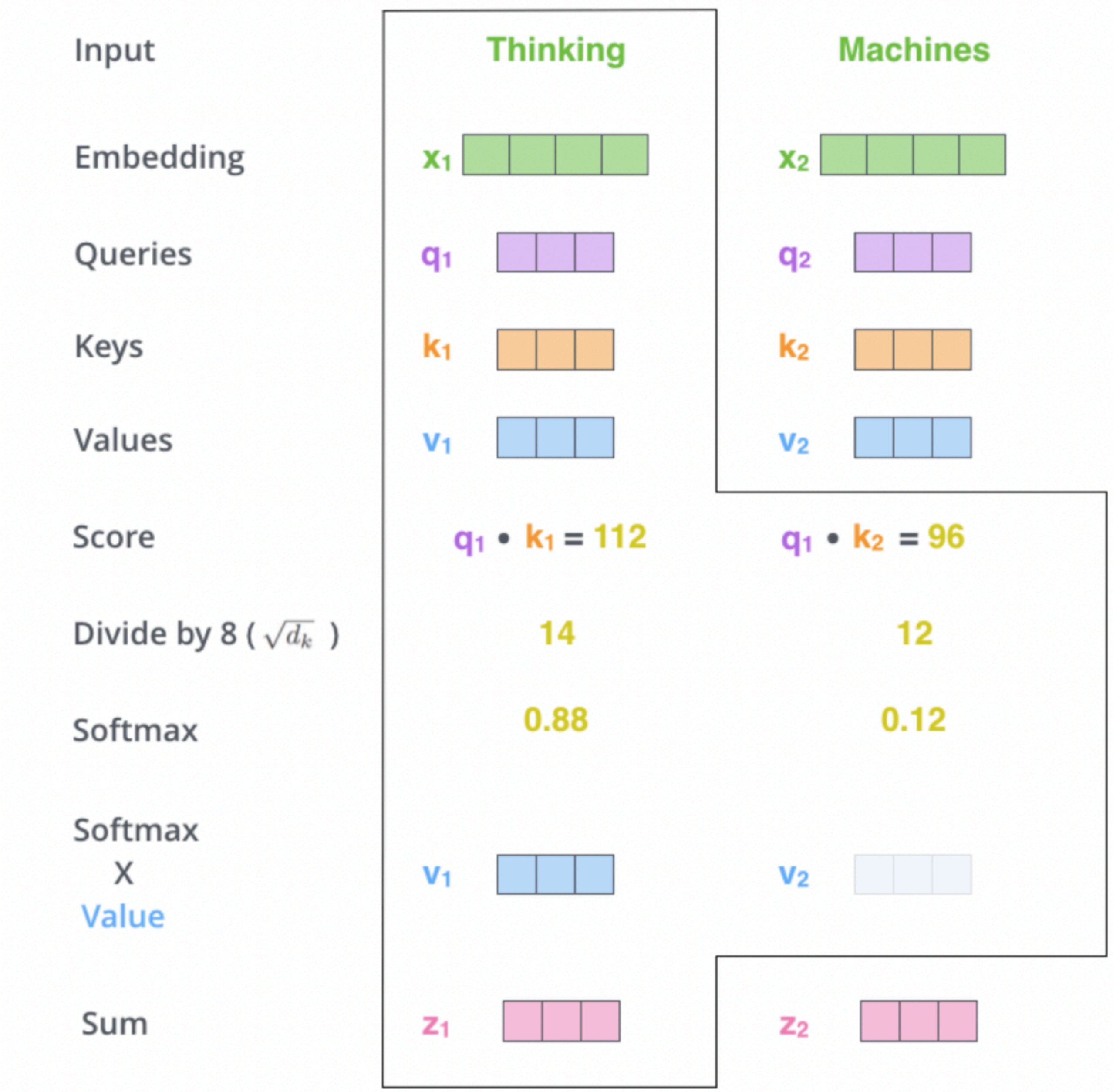
- Snowmass paper on uncertainty for ML in HEP: [2208:03284]

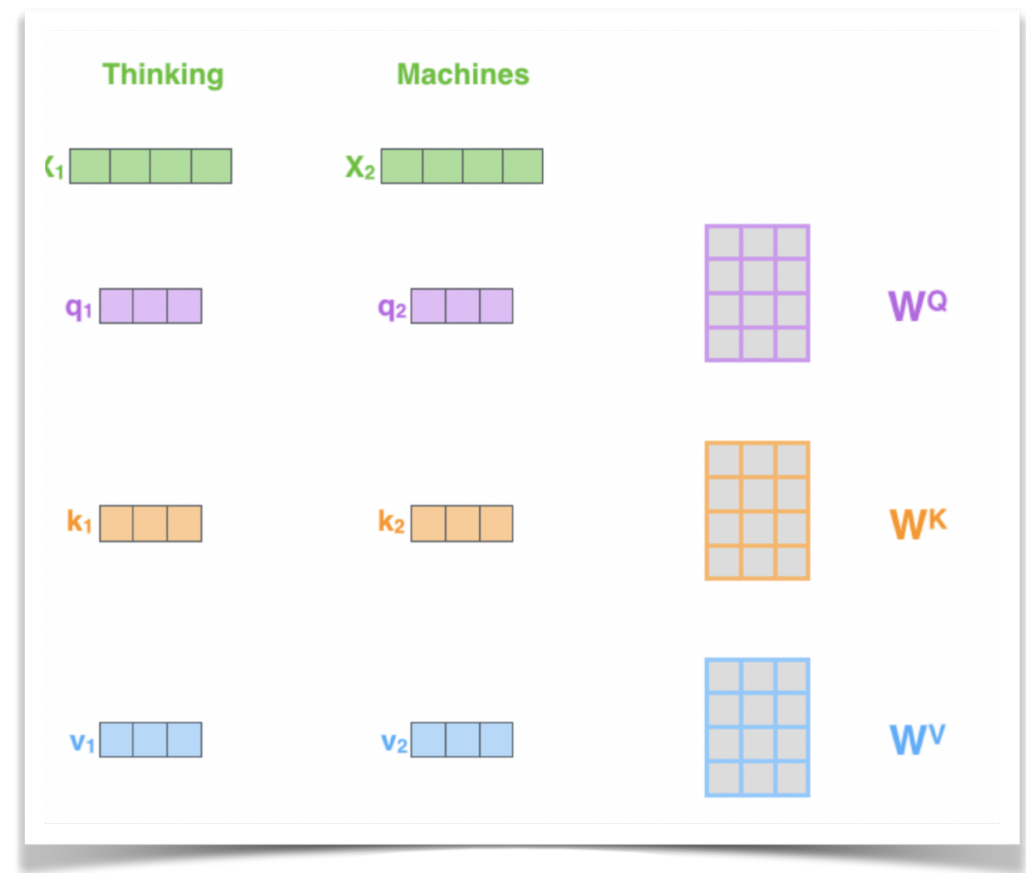
- Book Chapter: [Dorigo, de Castro Manzano]

Backup

Keys, queries, values

http://jalammar.github.io/illustrated-transformer/

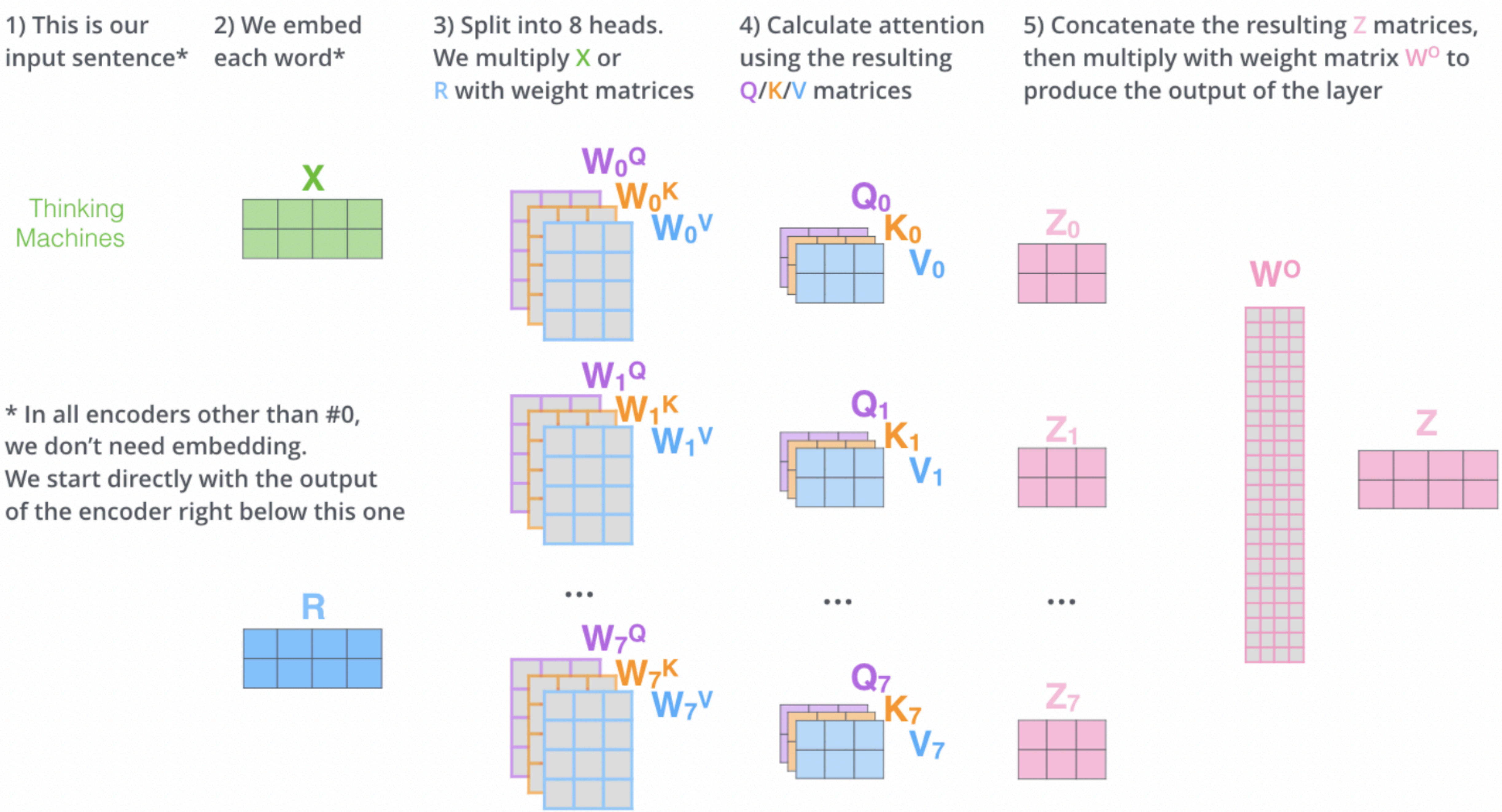




Multiplying x1 by the WQ weight matrix produces q1, the "query" vector associated with that word. We end up creating a "query", a "key", and a "value" projection of each word in the input sentence.

Tranformers

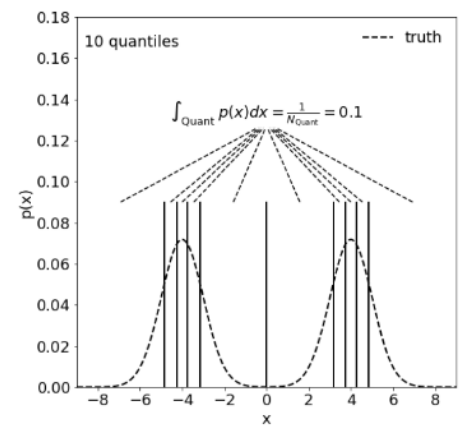
http://jalammar.github.io/illustrated-transformer/

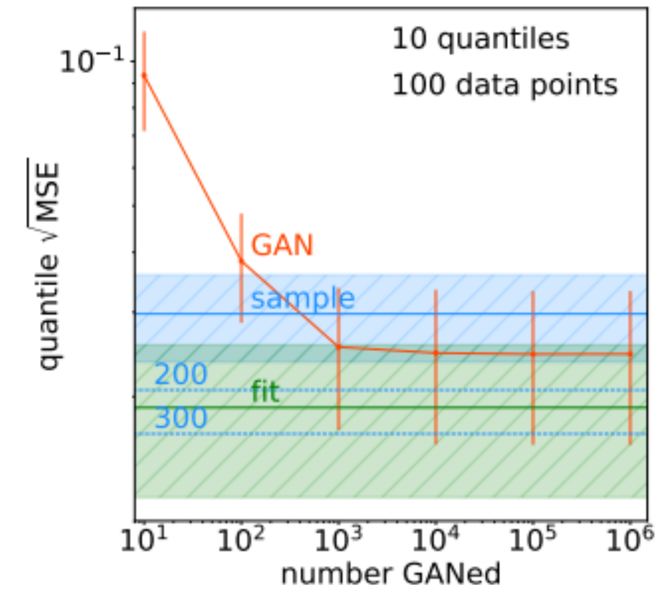


Systematics: training dataset size

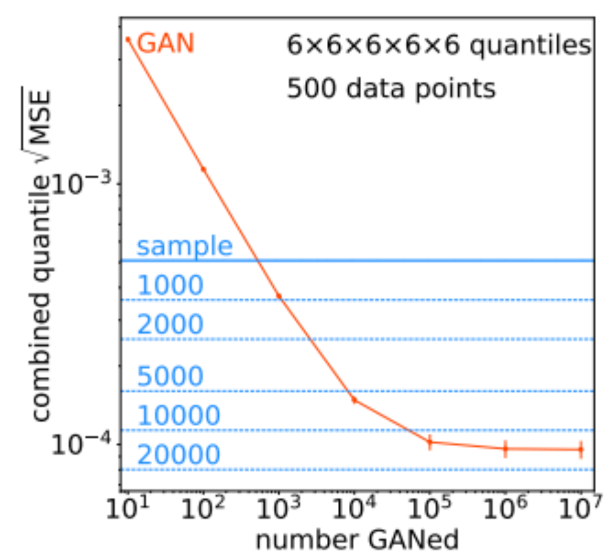
arxiv:2008.06545
NEW: arxiv:2202.073

- If a GAN is trained on **N** data points, how many **new** points can be drawn?
- GAN can describe distribution better than training data
- Needs 10,000 GAN points to match 150 true points
- In terms of **information**:
  - **sample**: only data points
  - **fit**: data + true function
  - **GAN**: data + smooth, continuous function





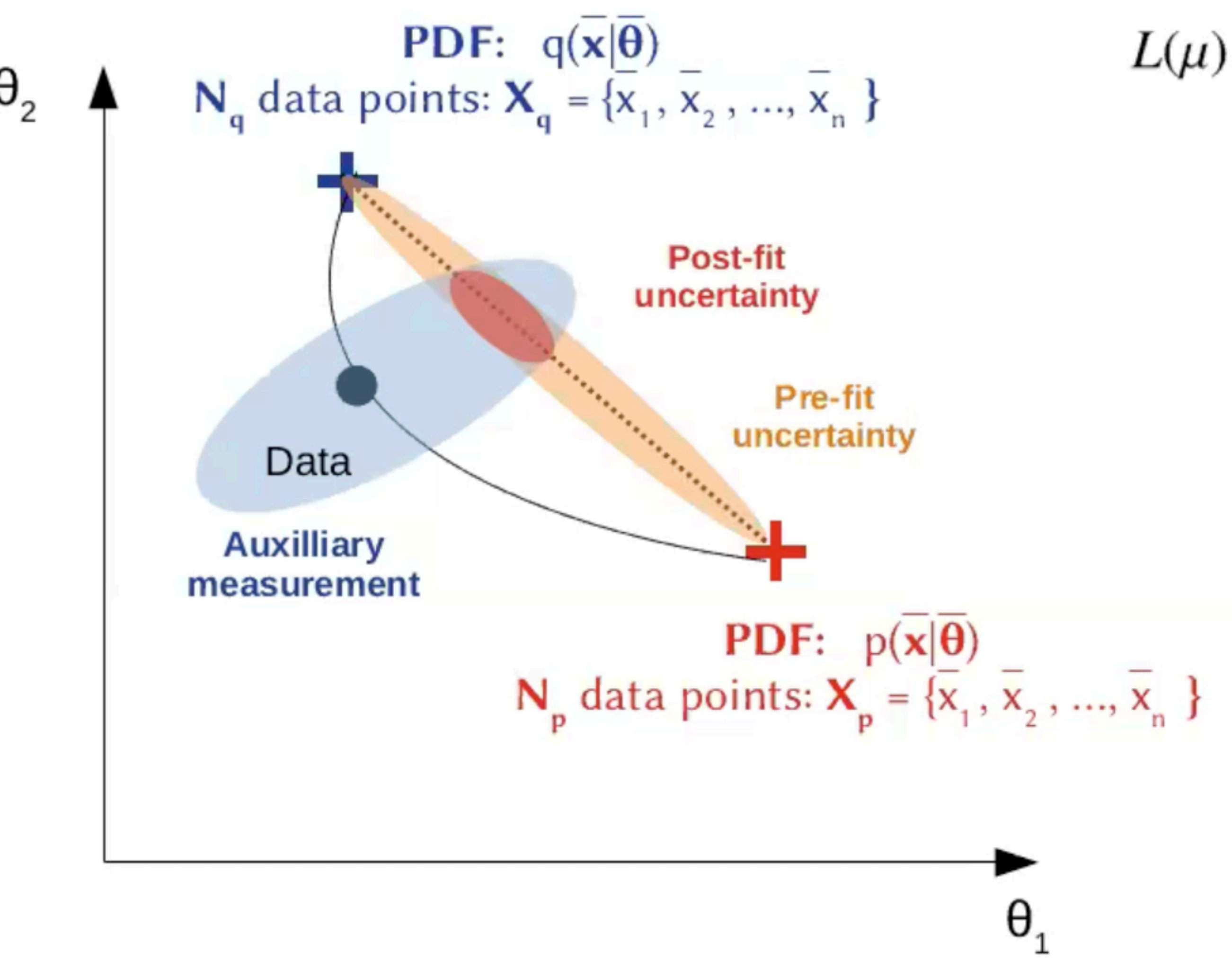
Generalisation to multi-dimensional problem



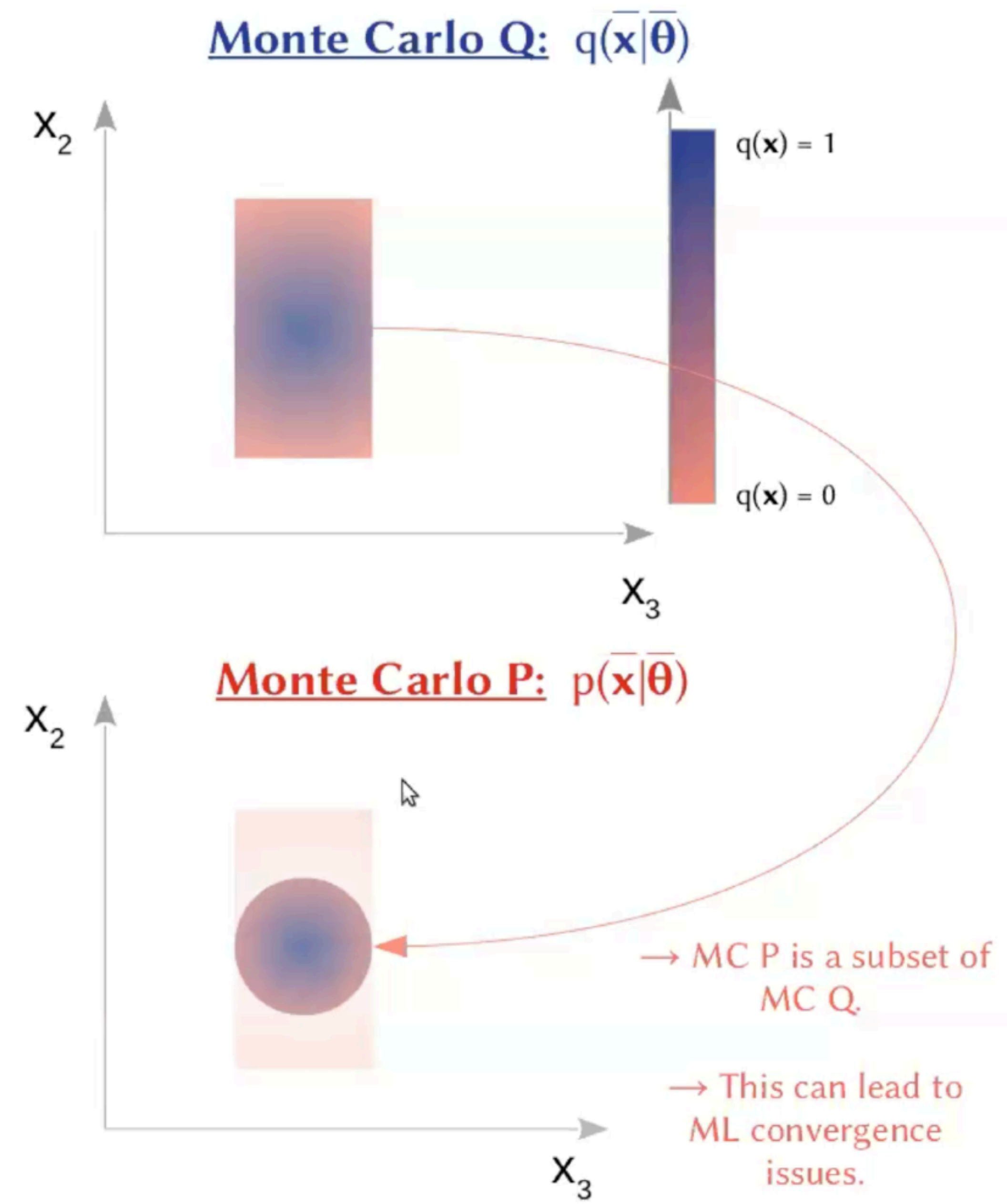
Most physics data sets described by continuous function → GAN can interpolate

Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

Bonus: Montecarlo reweighting with NN



$L(\mu)$:

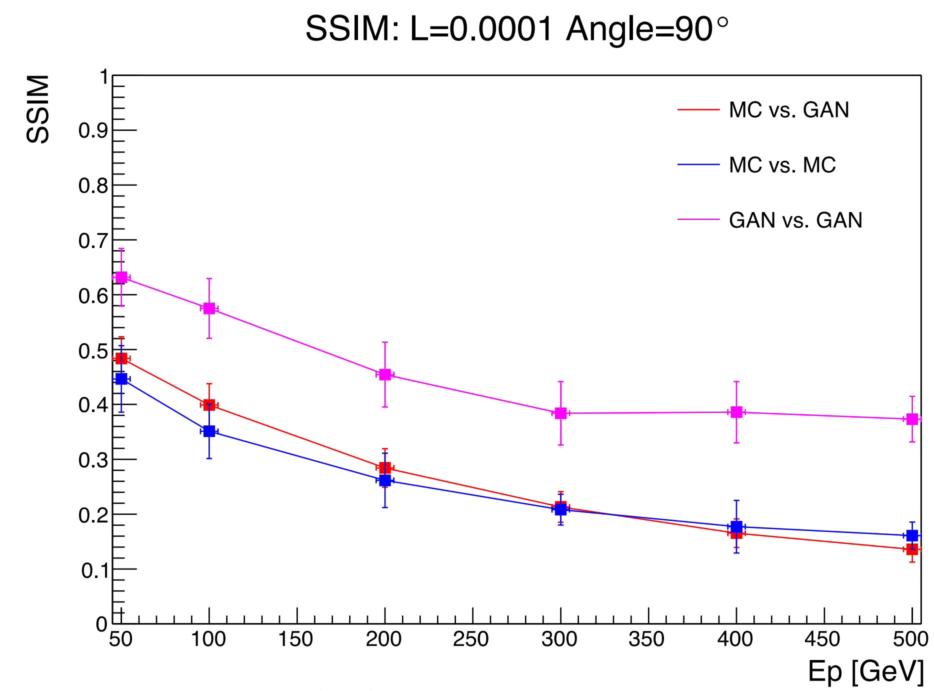


Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch

Systematic uncertainties: image similarity

GAN can exhibit **mode-collapse** or **mode-drop**

How much **diversity** in the generated sample?



$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) = 1 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$

- Use the **Structural Similarity Index**

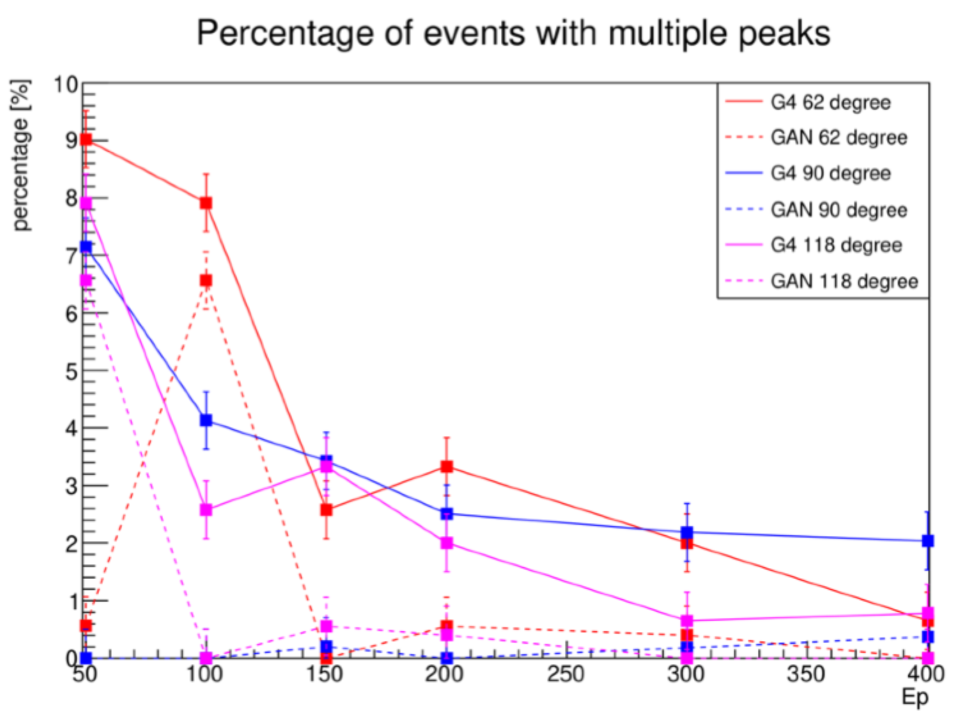
$$\mathrm{SSIM}(\boldsymbol{x}, \boldsymbol{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

where $\boldsymbol{x}, \boldsymbol{y}$ are two samples to be compared

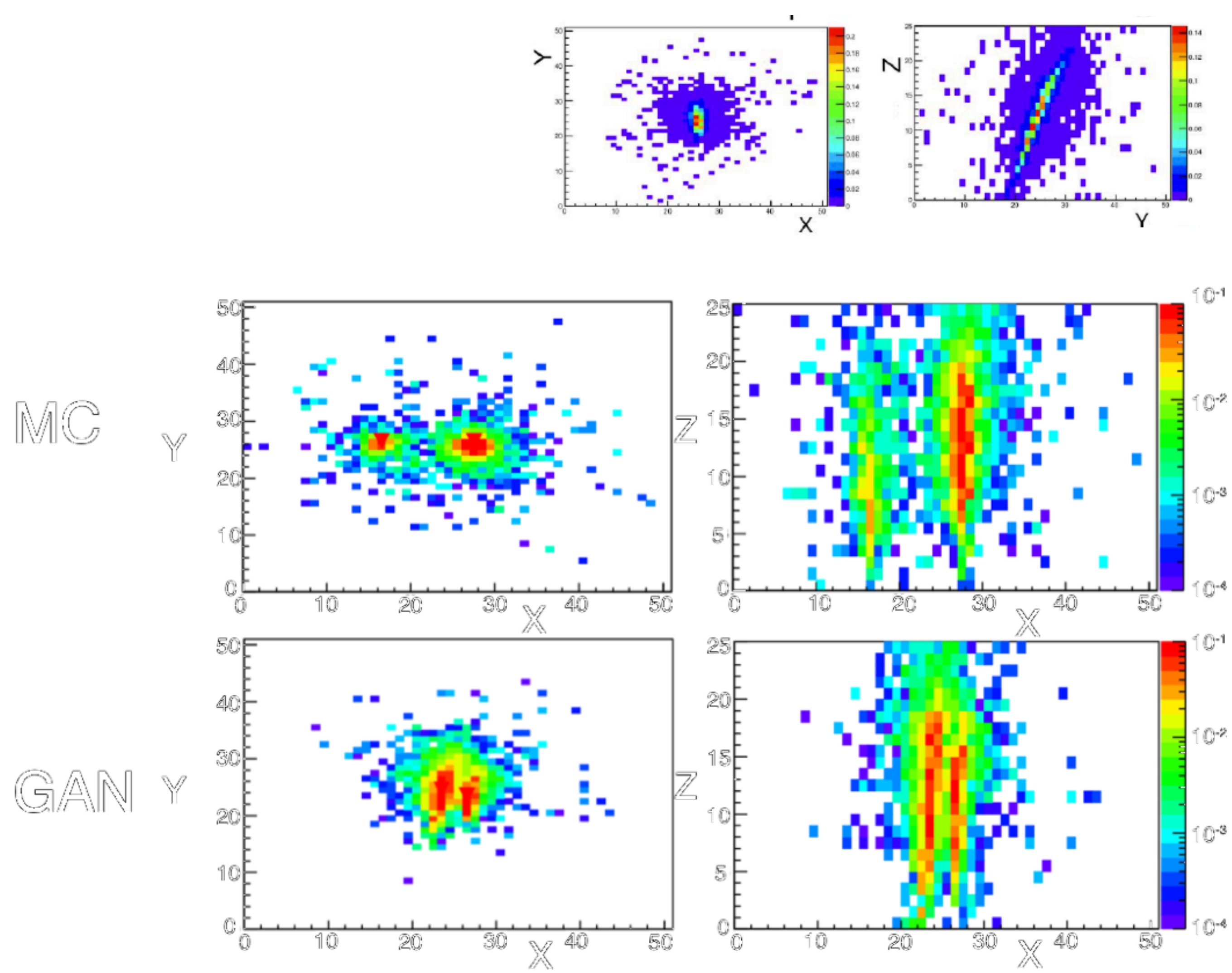
- Calculated on sliding windows, then averaged.
- Ours is a 3D problem: SSIM computed in ***xy* plane**, 3rd dimension is **channel**
- Adjust C1-C2 to the pixel dynamic range

Systematics: rare events

It is important to reproduce correctly the topology and occurrence of rare events



"Standard"



Sofia Vallecorsa, Ilaria Luise CERN - sofia.vallecorsa@cern.ch | ilaria.luise@cern.ch