UQ for ML Applied to Data Analysis

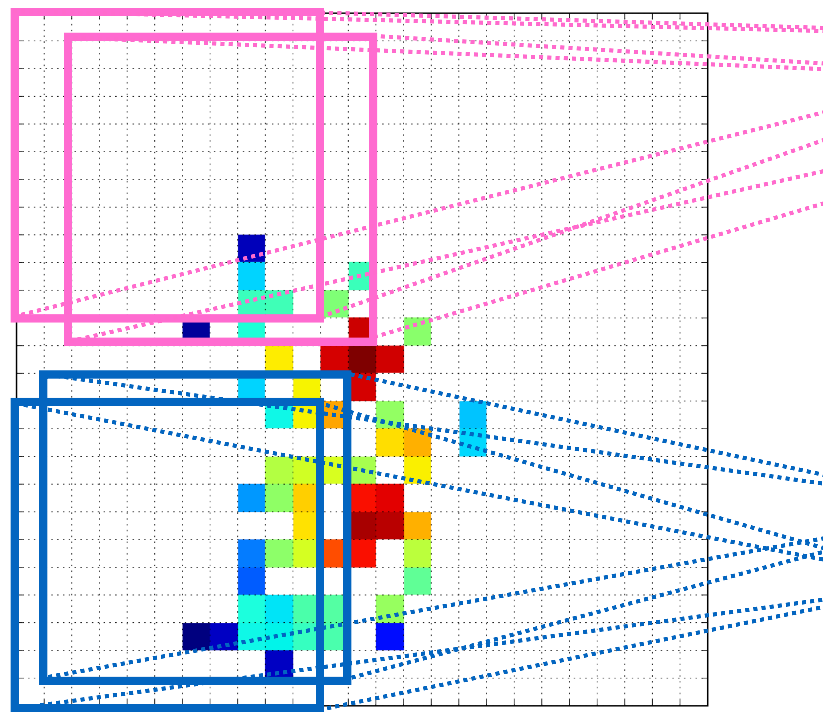
Benjamin Nachman

Lawrence Berkeley National Laboratory

bpnachman.com

bpnachman@lbl.gov

@bpnachman bnachman



BERKELEY LAB

BIDS
BERKELEY INSTITUTE FOR DATA SCIENCE

EICUG Topical Meeting
June 22, 2022

Uncertainties

"But what are the uncertainties on the NN"?

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Goal: let's sharpen this question and explore various cases and related topics.

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Goal: let's sharpen this question and explore various cases and related topics.

Outline:

- Uncertainty Landscape with ML
- Reducing Uncertainties with ML
- Conclusions / Outlook

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Goal: let's sharpen this question and explore various cases and related topics.

Outline:

- **Uncertainty Landscape with ML**
- Reducing Uncertainties with ML
- Conclusions / Outlook

Uncertainties for a NN-based analysis

To keep things simple, let's assume we are targeting some particular type of event S and we want to isolate it from the more common events B.

(a similar story holds for other learning objectives like regression, density estimation, etc., but I'll stick to this for it simplicity and ubiquity)

Uncertainties for a NN-based analysis

Precision / Optimality

Bad use of our data, time, money, etc. but ***not wrong****.*

Uncertainties for a NN-based analysis

Optimal by Neyman-Pearson

Precision / Optimality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Bad use of our data, time, money, etc. but ***not wrong****.*

Note that this is not p(x|S) / p(x|B), however the two are monotonically related to each other.

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN}(\mathrm{x}) \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

The distribution of the (corrected) sim. is not correct.

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Usually, there is no
"uncertainty on the NN" per se.

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

The distribution of the (corrected) sim. is not correct.

U... NN-based analysis

An optimality uncertainty becomes a bias uncertainty when we rely on the NN output to be a likelihood ratio

(see e.g. many examples from likelihood-free inference)

...ality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Usually, there is no "uncertainty on the NN" per se.

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

The distribution of the (corrected) sim. is not correct.

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

limited training statistics

$p_{\mathrm{train}}(x) \neq p_{\mathrm{true}}(x)$

inaccurate training data

$\mathrm{NN(x)}|_{p_{\mathrm{true}}=p_{\mathrm{train}}} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

model/optimization flexibility

Statistical uncertainty | **Systematic uncertainty**

limited prediction statistics

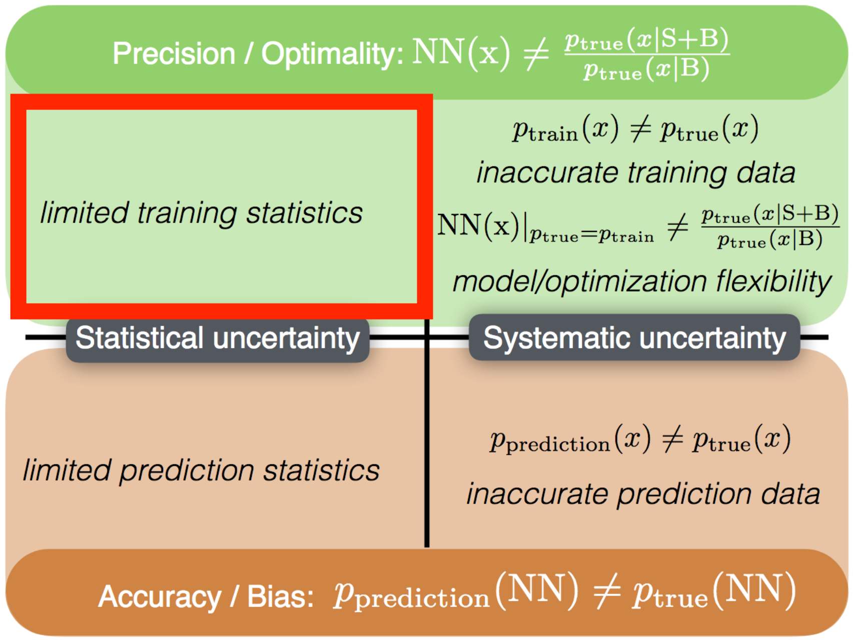
$p_{\mathrm{prediction}}(x) \neq p_{\mathrm{true}}(x)$

inaccurate prediction data

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

1909.03081

How to estimate precision stat. uncerts.



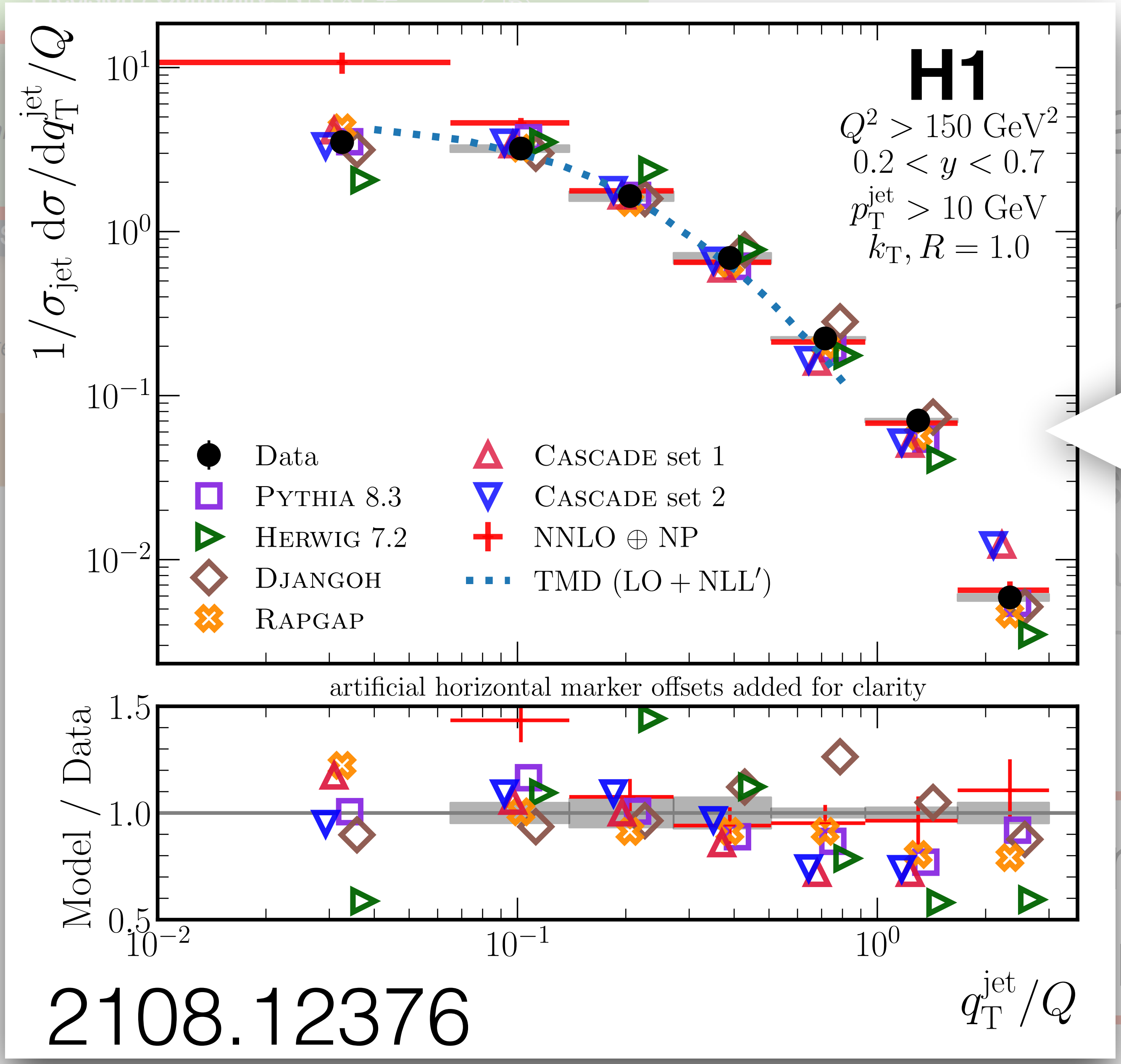
You can always accomplish this by bootstrapping: making pseudo-datasets from resampling and then retraining.

It is important to fix the NN initialization so that you are not also testing your sensitivity to that.

This can be painful because it requires retraining many NNs.

Maybe can accomplish with one Bayesian NN?
See e.g. 1904.10004 for a particle physics example.

How to estimate precision stat. uncerts.



always accomplish this by

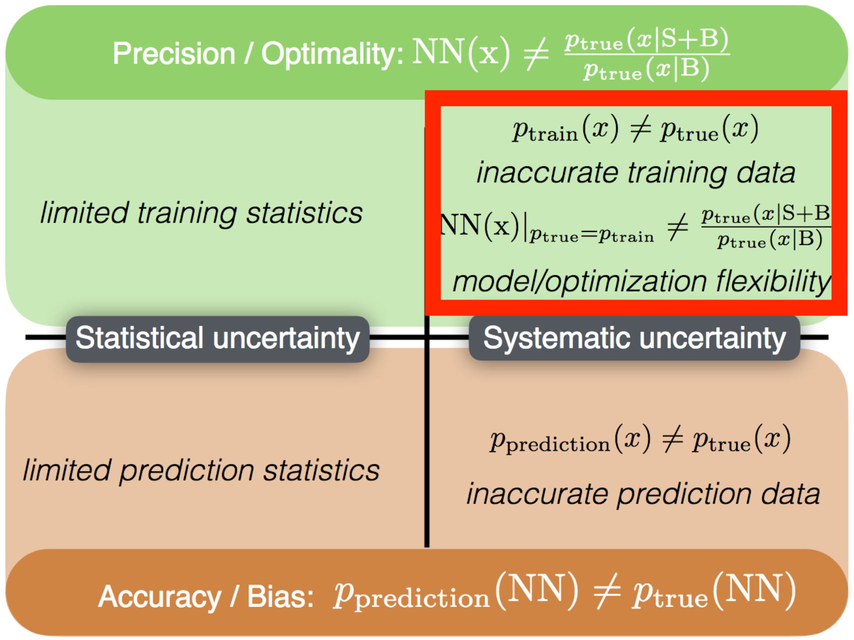
Example of an 8-dimensional ML-based unfolding where we used 100 bootstraps for a 10% uncertainty on the uncertainty

ng your sensitivity to that.

nful because it
ng many NNs.

Maybe can accomplish with one Bayesian NN?
See e.g. 1904.10004 for a particle physics example.

How to estimate precision syst. uncerts.



As with all systematic uncertainties, this is hard to quantify.

One component is due to the modeling of p(x) - more on this later.

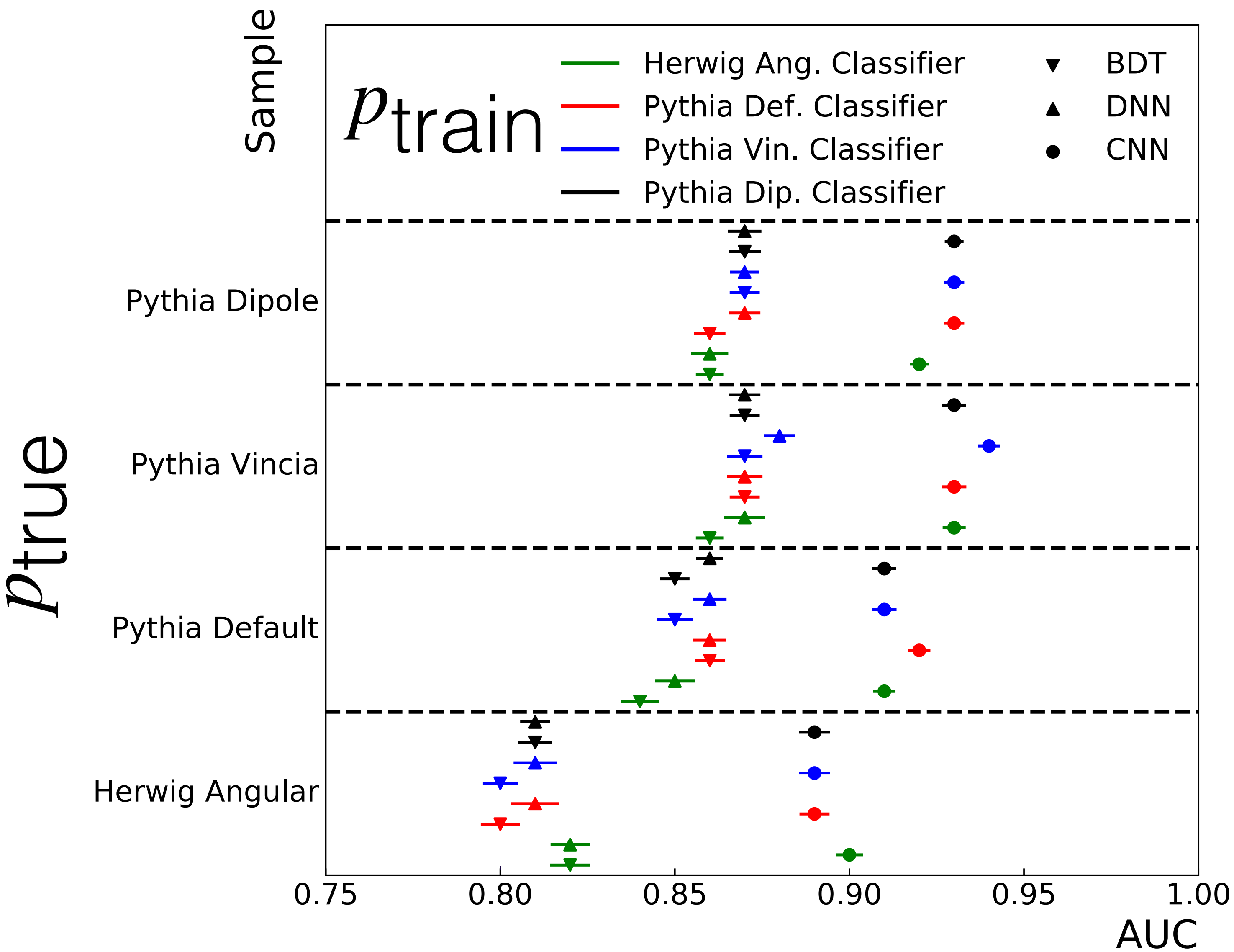
Testing the flexibility of the network requires checking the sensitivity to the architecture (#layers, nodes/layer, etc.), the initialization, the training procedure (#epochs, learning rate, etc.)

How to estimate precision syst. uncerts.

…brief aside #1: just because $p_{\text{train}} \neq p_{\text{true}}$ doesn't mean that there is an "uncertainty".

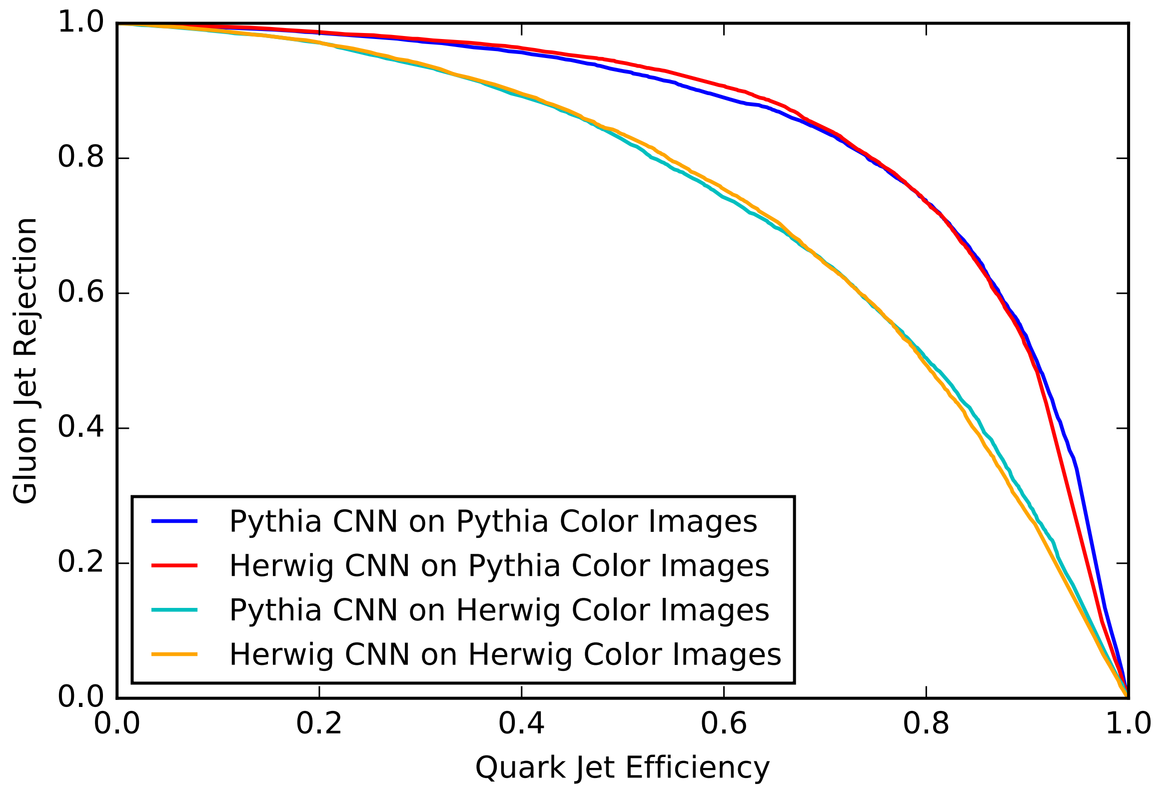
How to estimate precision syst. uncerts.

…brief aside #1: just because $p_{\text{train}} \neq p_{\text{true}}$ doesn't mean that there is an "uncertainty".



2204.03812

1612.01551

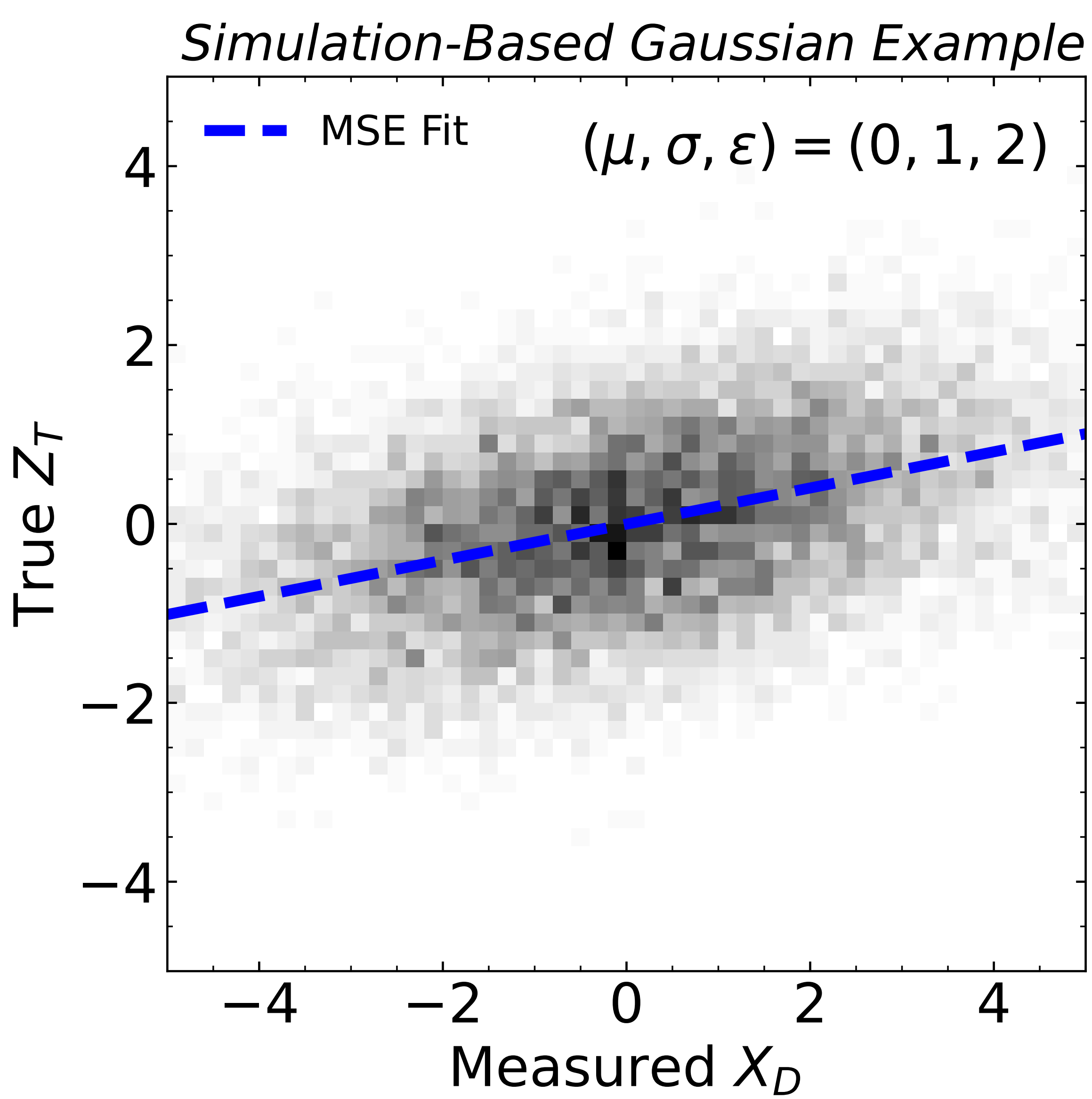


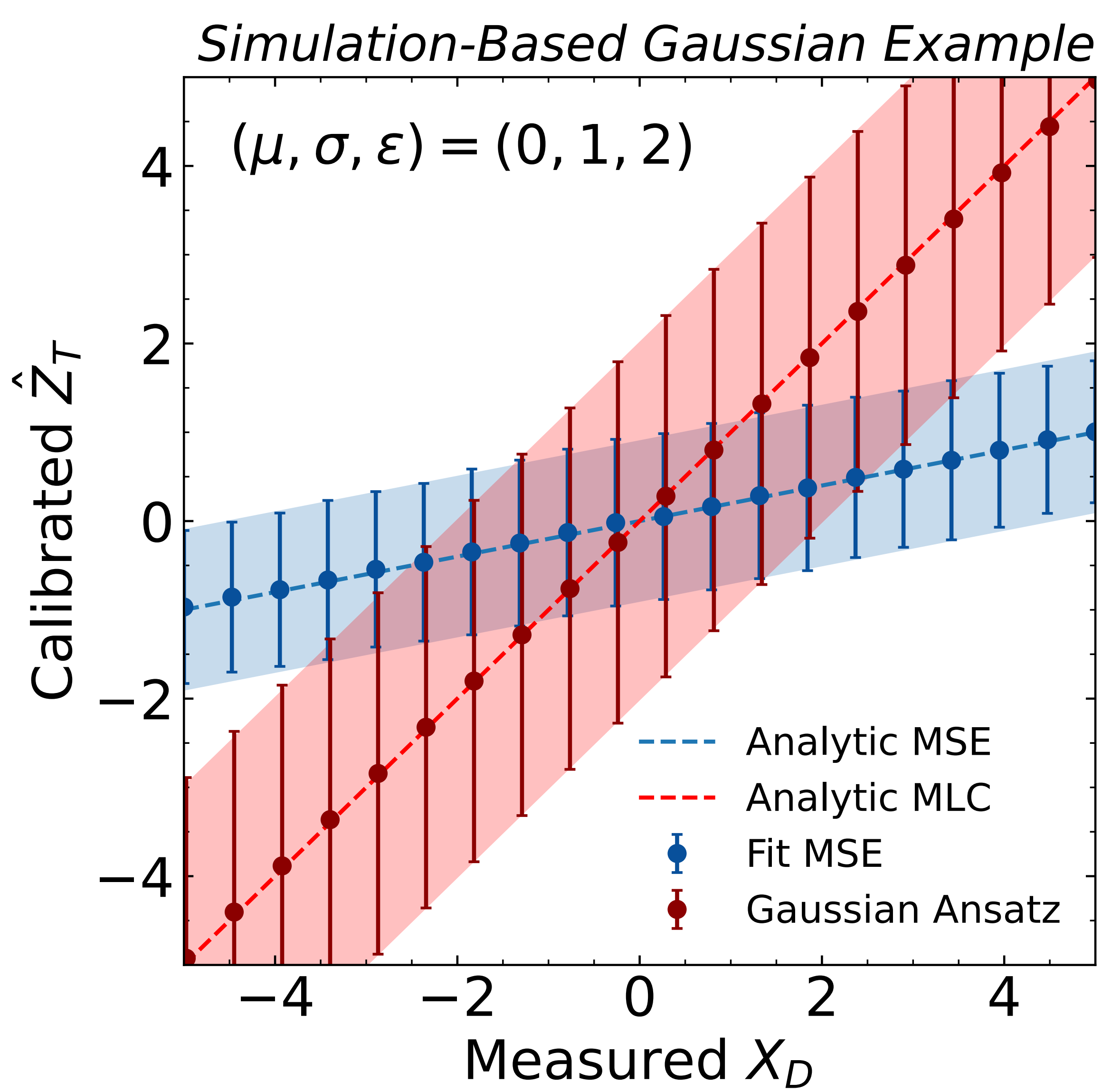
How to estimate precision syst. uncerts.

…brief aside #2: just because $p_{\text{train}} = p_{\text{true}}$ doesn't mean that the uncertainty is zero!

How to estimate precision syst. uncerts.

…brief aside #2: just because $p_{\text{train}} = p_{\text{true}}$ doesn't mean that the uncertainty is zero!

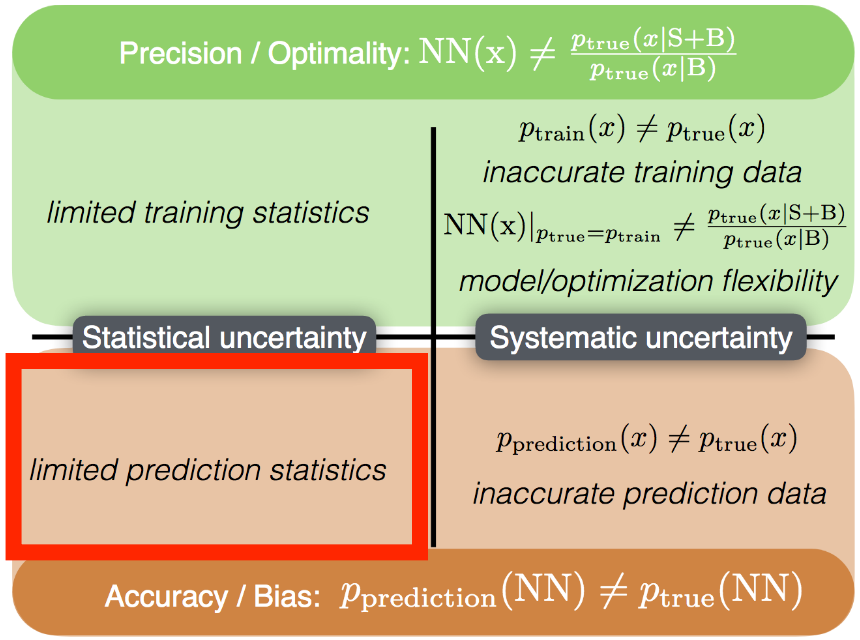




2205.05084

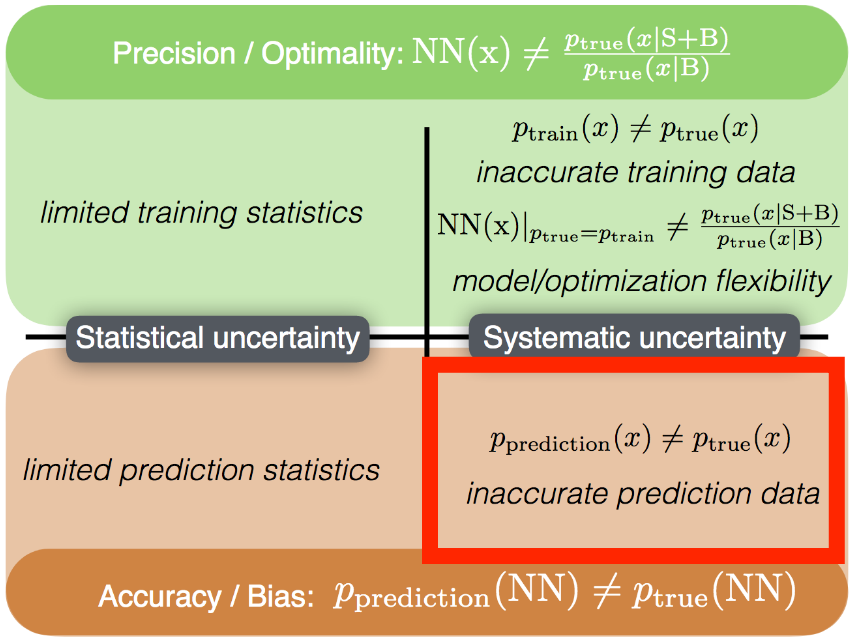
MLC, Gaussian Ansatz ~ Maximum Likelihood

How to estimate bias stat. uncerts.



Can be estimated via bootstrapping. Less painful here because the NN's are fixed.

How to estimate bias syst. uncerts.



This is the trickiest one…

…because we need the uncertainty on the modeling of x and x can be high-dimensional!

In many cases, the uncertainties factorize, e.g. the uncertainty on two photon energies can be decomposed into the uncertainty on each photon.

However, in many cases, we simply do not know the full uncertainty model (= nuisance parameters and their distribution)

High-dimensional Bias Uncertainties

One word of caution: current paradigm for uncertainties may be too naive for high-dimensional analysis!

(truly end-to-end)

e.g. for some uncertainties, we often compare two different models - one nuisance parameter.

High-dimensional Bias Uncertainties

One word of caution: current paradigm for uncertainties may be too naive for high-dimensional analysis!

(truly end-to-end)

e.g. for some uncertainties, we often compare two different models - one nuisance parameter.

How can we even see how sensitive we are to high-dimensional effects?

High-dimensional Bias Uncertainties

One word of caution: current paradigm for uncertainties may be too naive for high-dimensional analysis!

(truly end-to-end)

e.g. for some uncertainties, we often compare two different models - one nuisance parameter.

How can we even see how sensitive we are to high-dimensional effects?

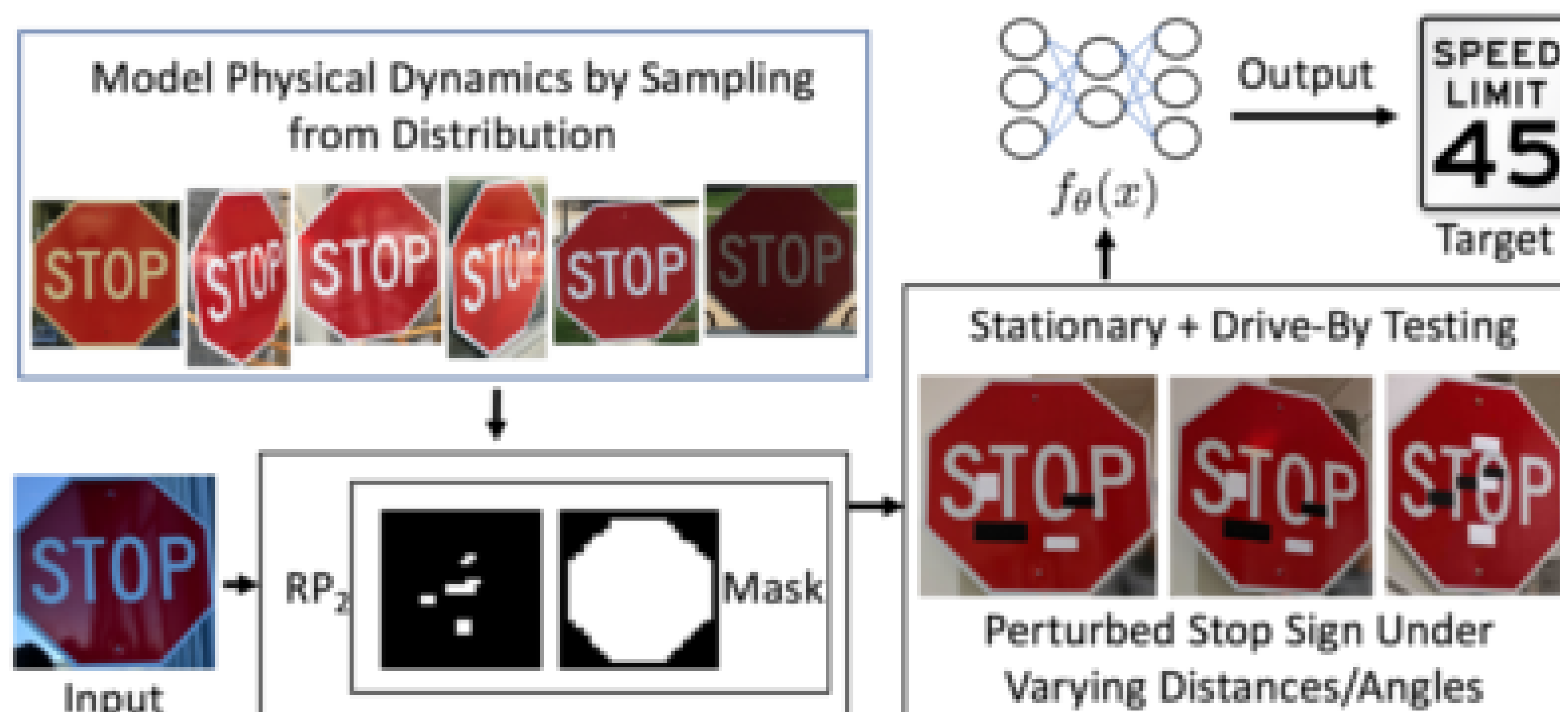
One (worse case?) Answer: borrow tools from AI Safety

AI Safety

There is a vast literature on how easy it is to “attack” a NN.

They want to know: how subtle can an attack be and still significantly impact the output.

We know (hope?!) that nature is not evil, but these tools can help us probe the high-dimensional sensitivity of our NNs.



K. Eykholt et. al, 1707.08945

Bounding high-dim. uncerts: strategy

J = collision event (in all of its high-dimensional glory)

f = fixed classifier for S vs. B

Loss

$$\mathcal{L}_{\text{sig}} = \log(1 - f(g(J))),$$

$$\mathcal{L}_{\text{bg}} = \lambda_{\text{cls}}(f(J) - f(g(J)))^2 + \sum_i \lambda_{\text{obs}}^{(i)}(\mathcal{O}^{(i)}(J) - \mathcal{O}^{(i)}(g(J))^2$$

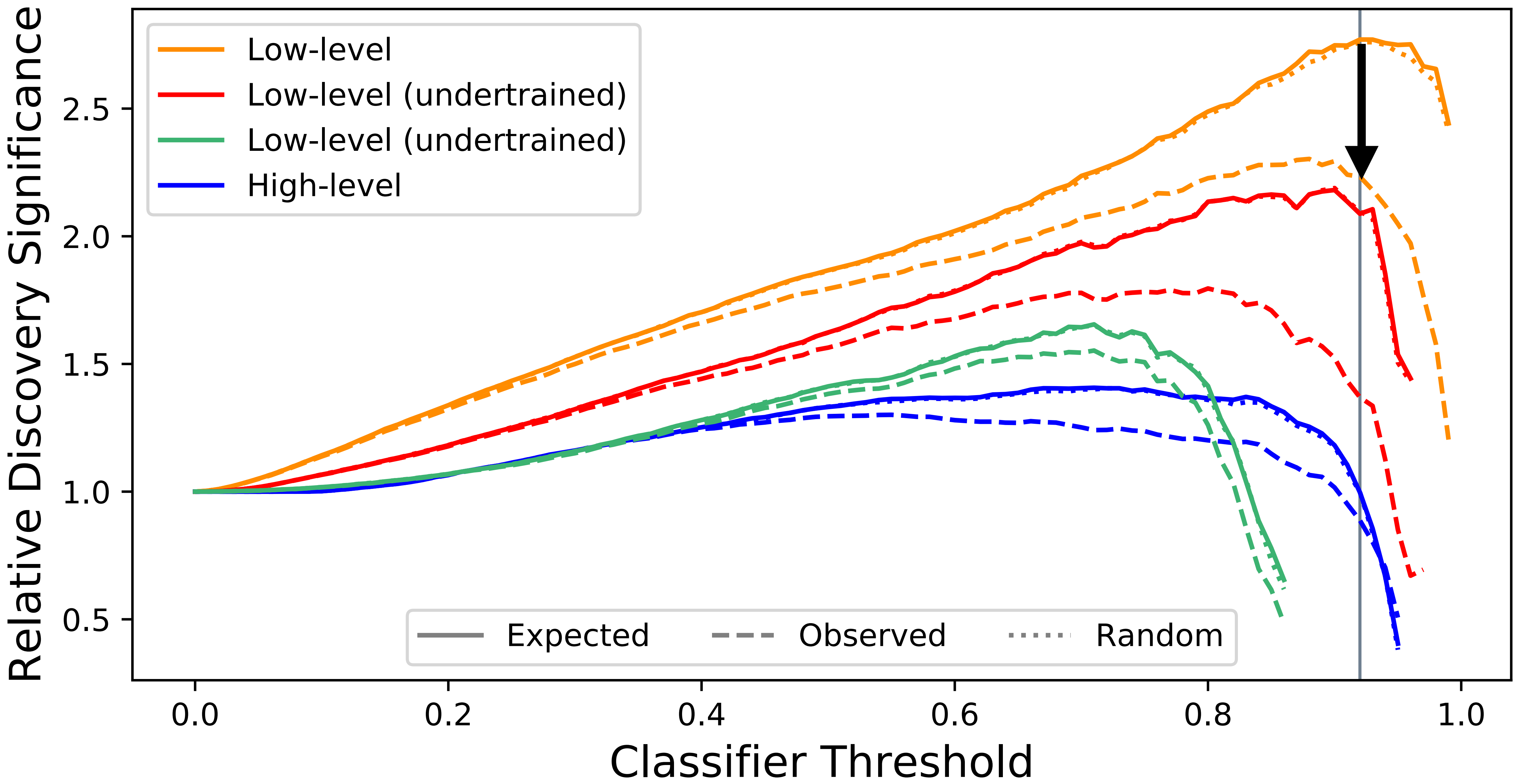
1910.08606 (see also 2203.13890)

g is a learned NN that maps J to $J + \delta J$.

O(J) are observables that are validated in a control region.

High-dimensional Uncertainty

1910.08606



“worst-case uncertainty”

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Goal: let's sharpen this question and explore various cases and related topics.

Outline:

- Uncertainty Landscape with ML
- Reducing Uncertainties with ML
- Conclusions / Outlook

Uncertainties

"But what are the uncertainties on the NN"?

- question asked by every reviewer

Goal: let's sharpen this question and explore various cases and related topics.

Outline:

- Uncertainty Landscape with ML
- **Reducing Uncertainties with ML**
- Conclusions / Outlook

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN}(\mathrm{x}) \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

limited training statistics

$p_{\mathrm{train}}(x) \neq p_{\mathrm{true}}(x)$

inaccurate training data

$\mathrm{NN}(\mathrm{x})|_{p_{\mathrm{true}}=p_{\mathrm{train}}} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

model/optimization flexibility

Statistical uncertainty | **Systematic uncertainty**

limited prediction statistics

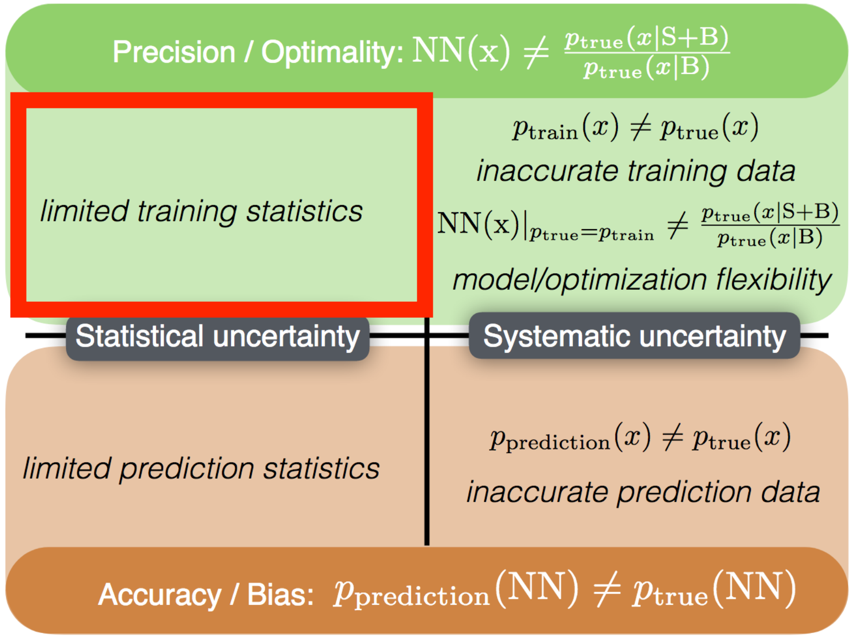
$p_{\mathrm{prediction}}(x) \neq p_{\mathrm{true}}(x)$

inaccurate prediction data

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

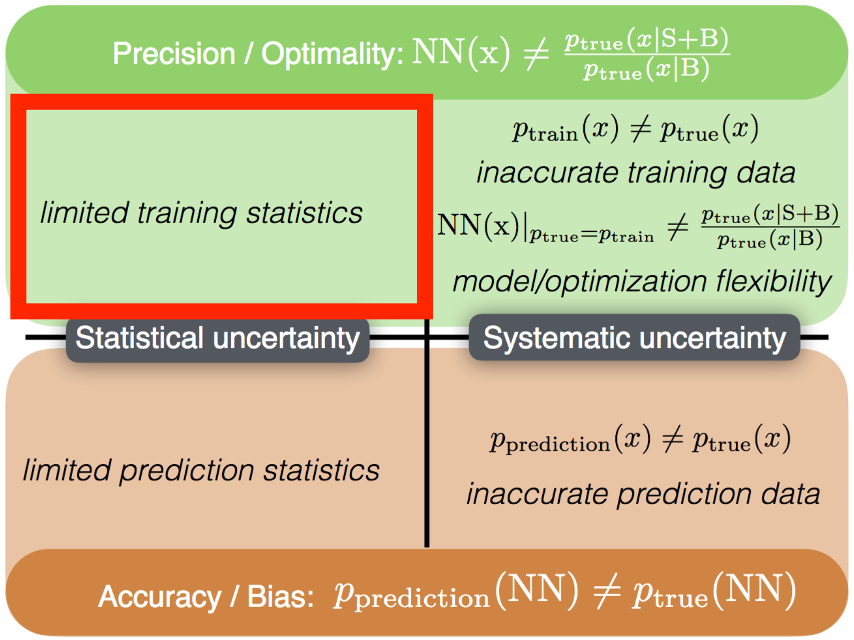
1909.03081

How to reduce precision stat. uncerts.



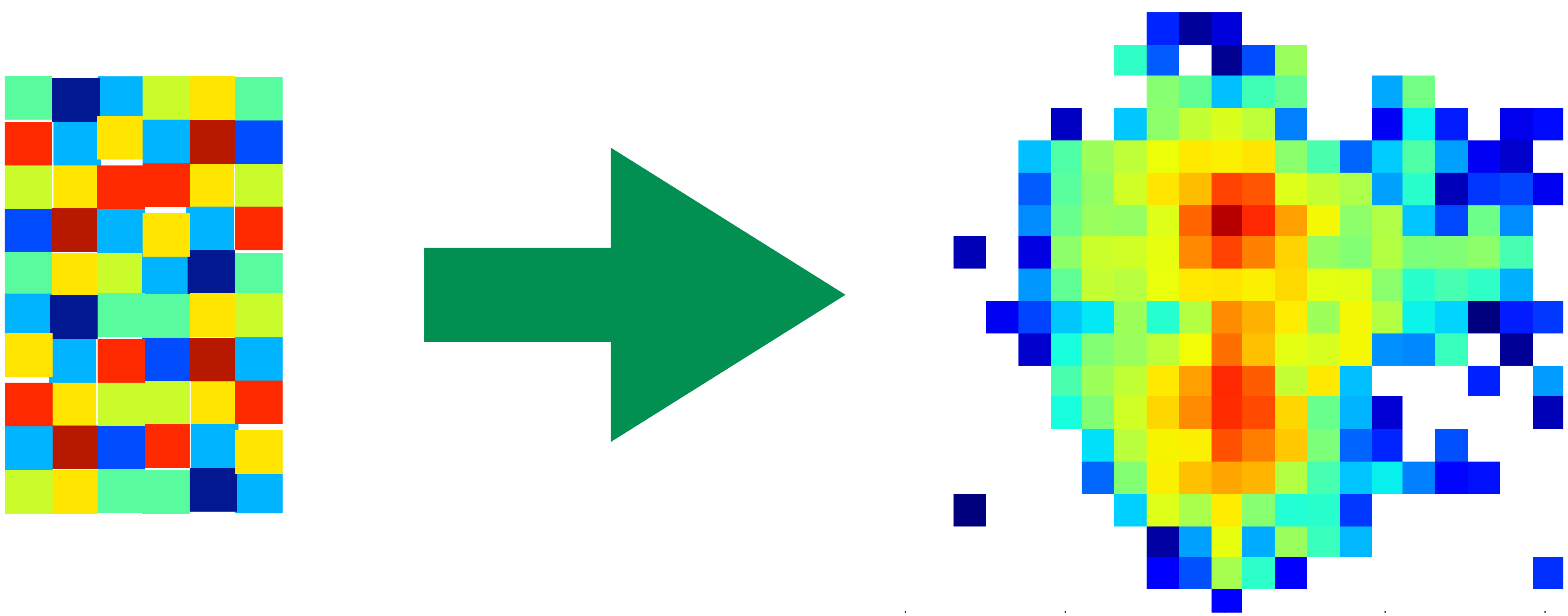
Train with more events!

How to reduce precision stat. uncerts.



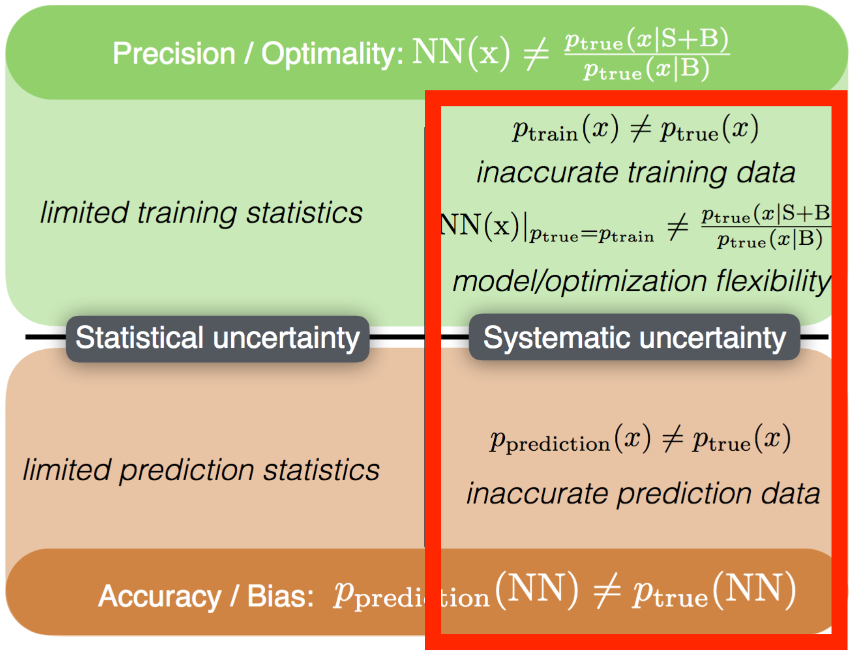
Train with more events!

…maybe use NN's to help with that

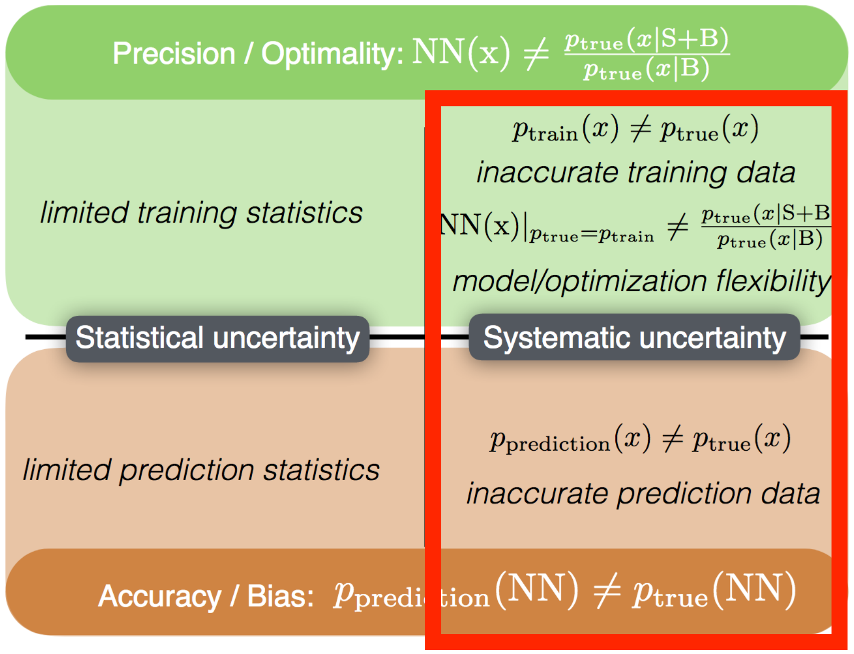


See 1705.02355 and the papers that cite it; also see 2008.06545 and 2202.07352 to study the amplification factor.

Case where we don't have p_{true}



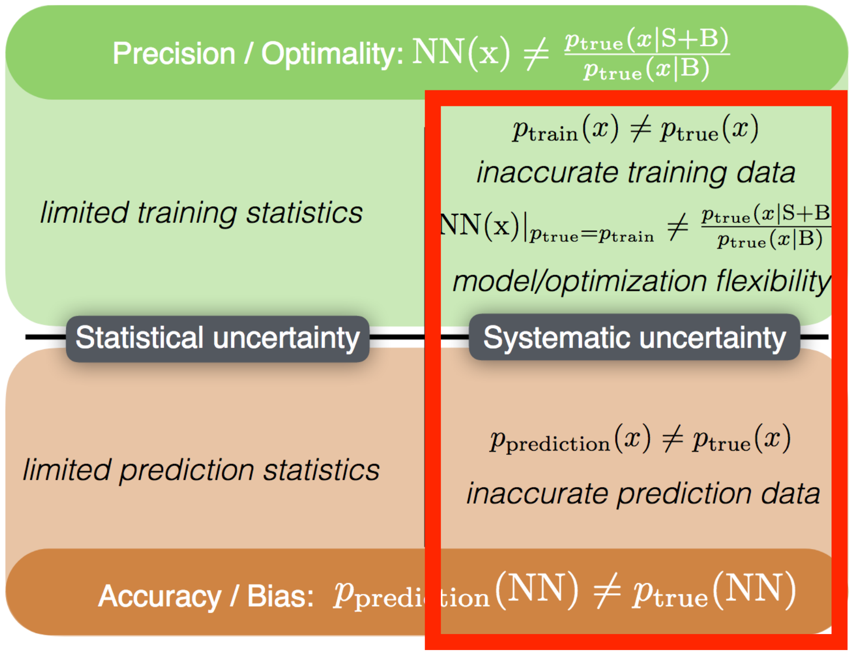
Case where we don't have p_{true}



Option 1 ("Decorrelation"):

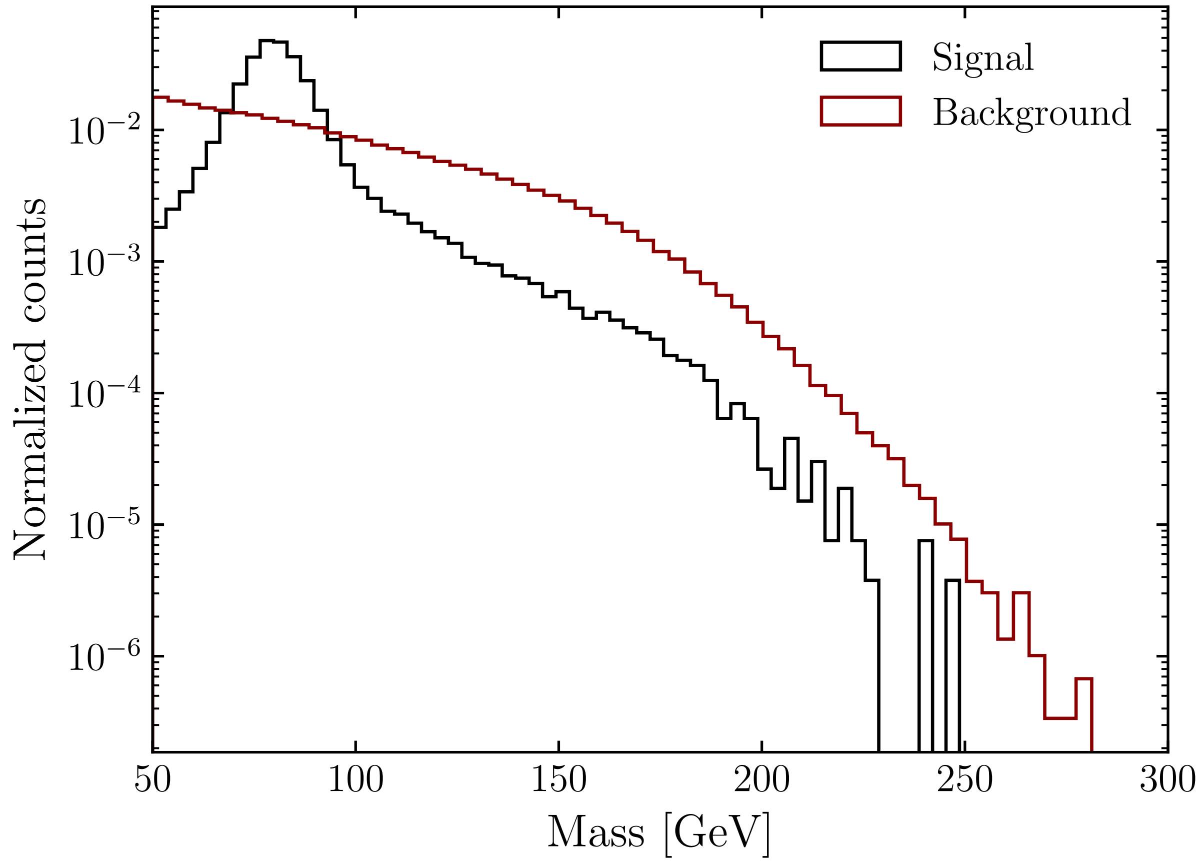
Might be possible to reduce uncertainties or at least alleviate analysis complexity by making your NN independent of known nuisance parameters.

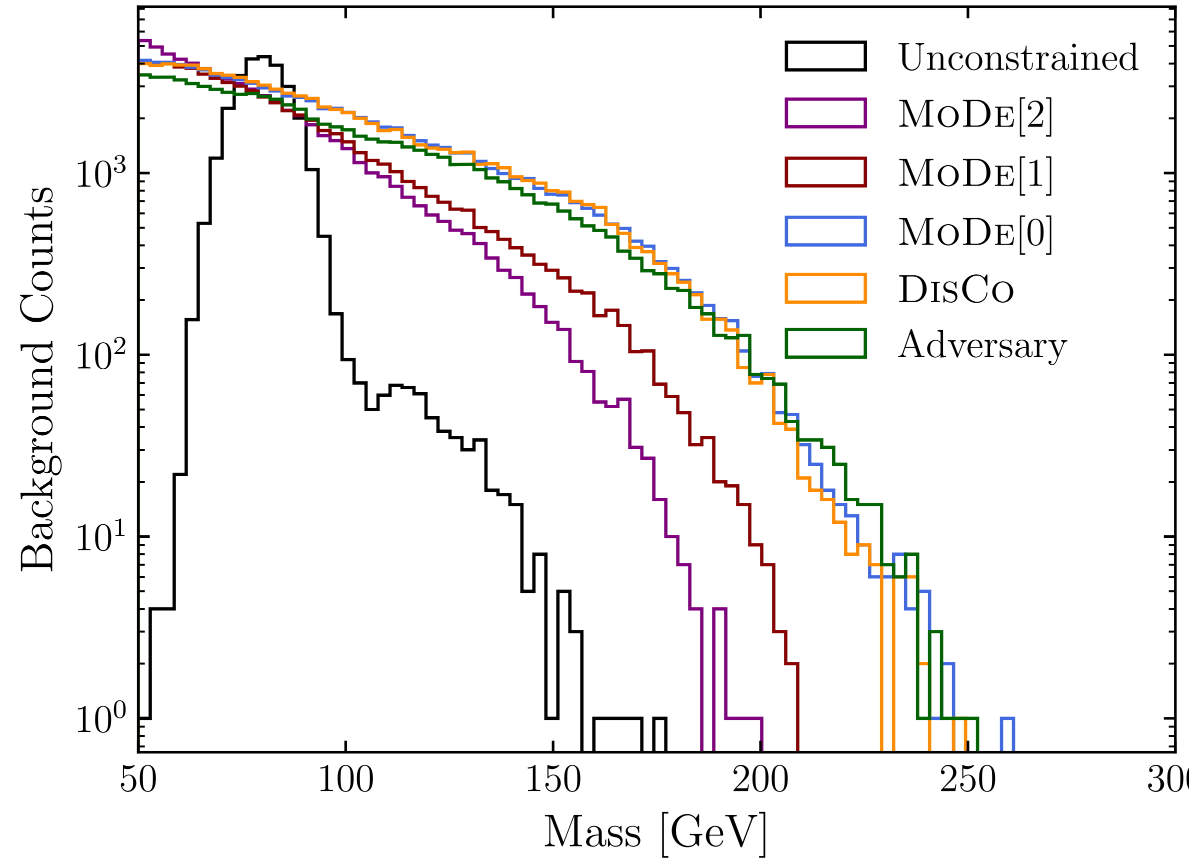
Case where we don't have p_{true}



Option 1 ("Decorrelation"):

Might be possible to reduce uncertainties or at least alleviate analysis complexity by making your NN independent of known nuisance parameters.



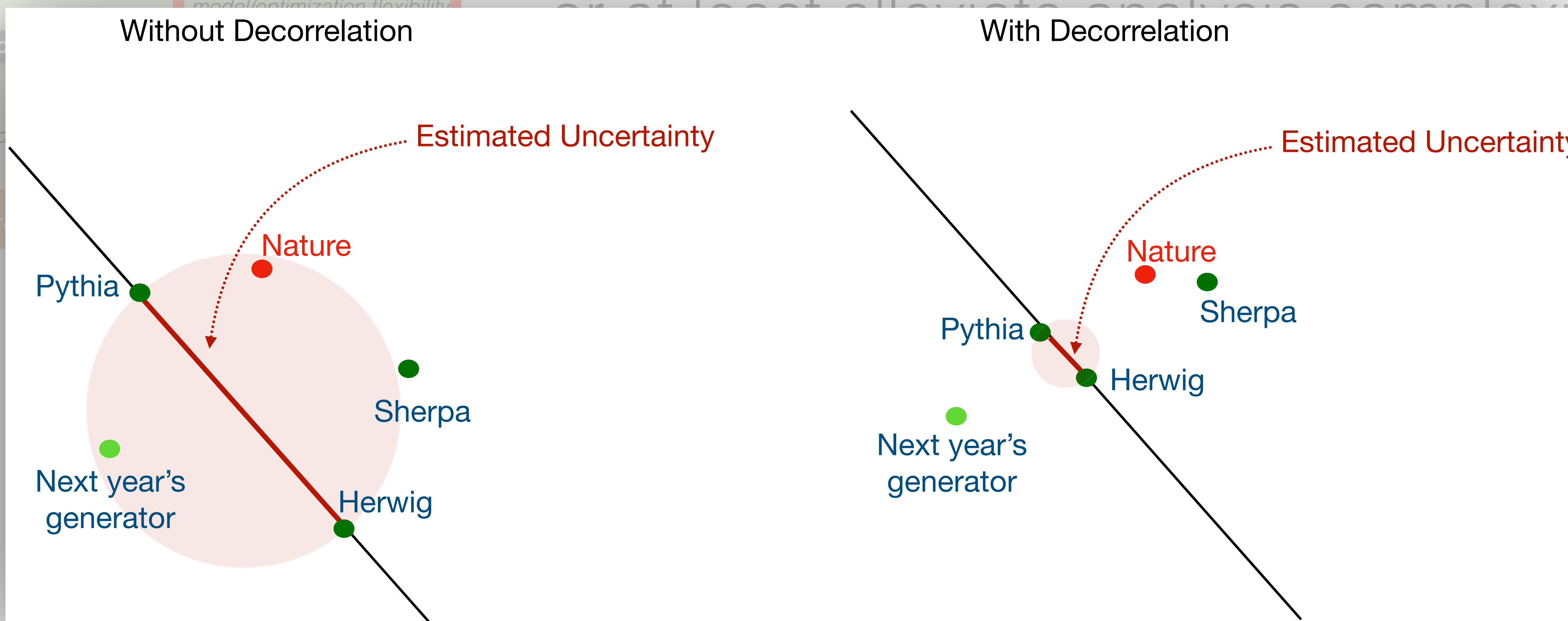


See 2010.09745, 2001.05310, 1611.01046

Case where we don't have p_{true}

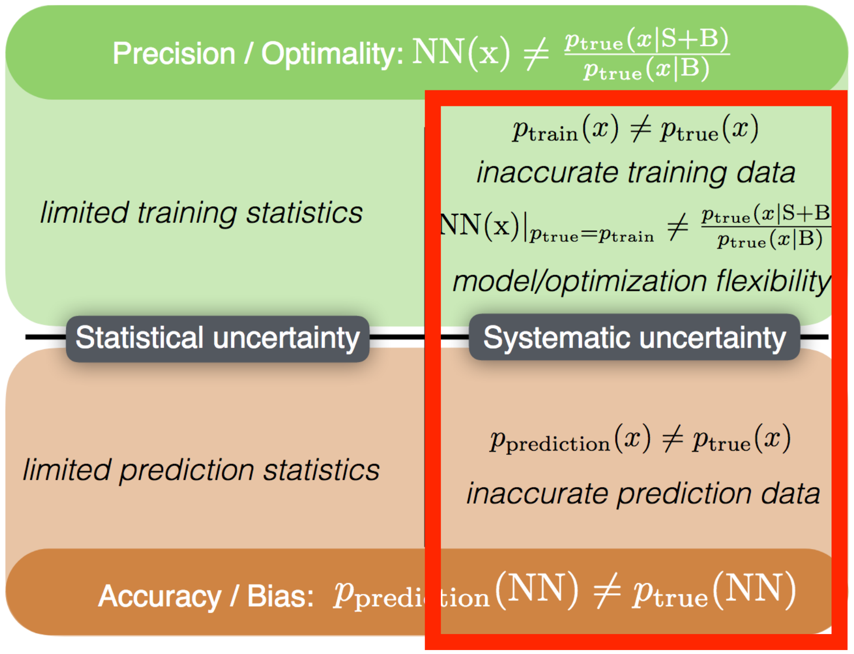
Be careful about theory uncertainties!

Might be possible to reduce uncertainties



2109.08159

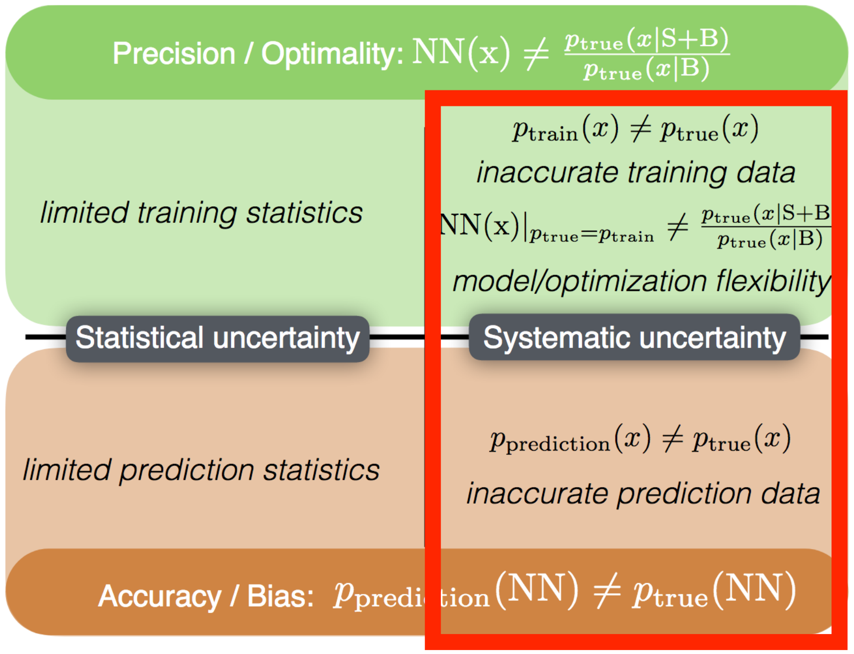
Case where we don't have p_{true}



Option 2 ("uncertainty-awareness"):

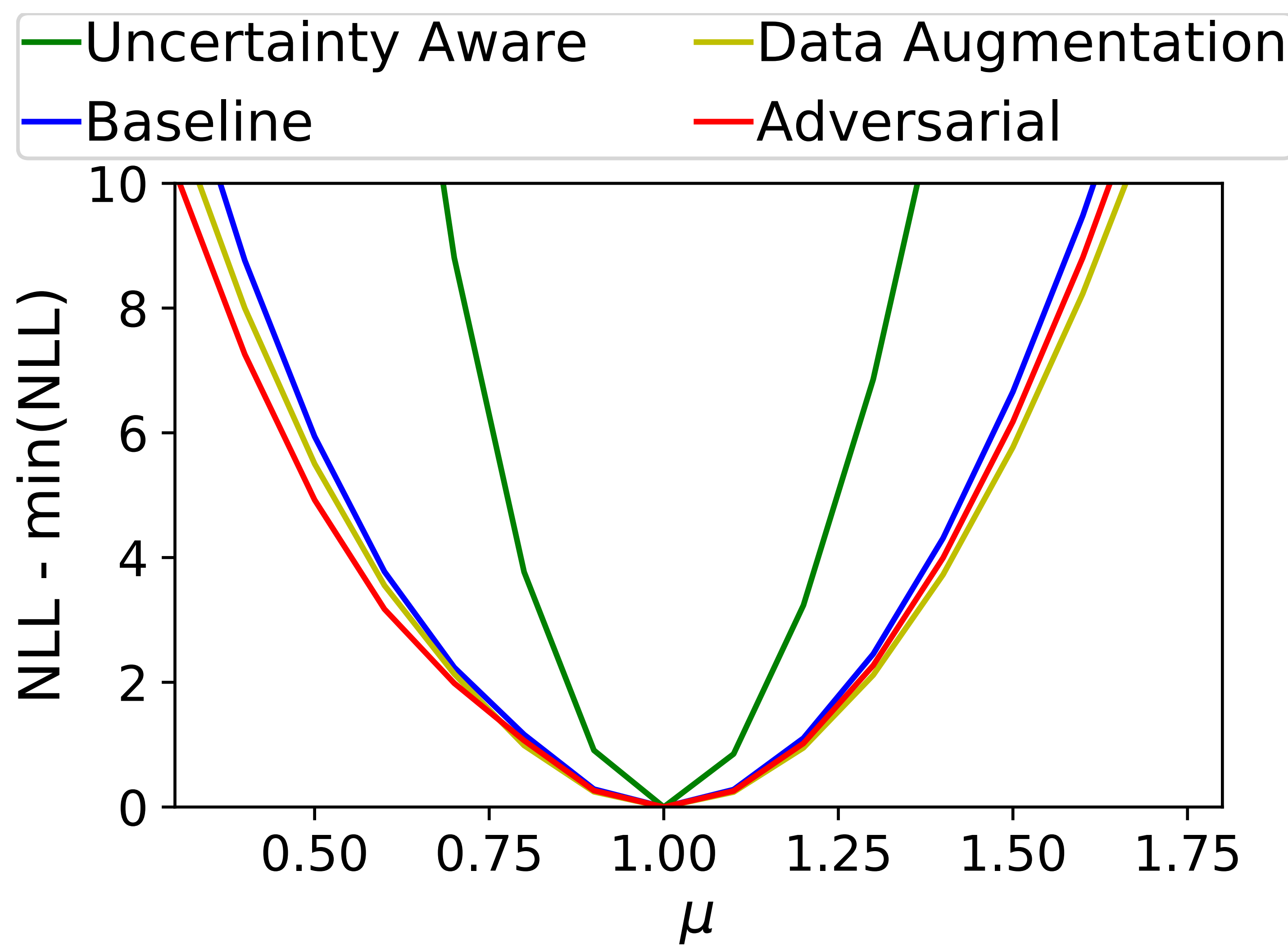
Let the analysis depend explicitly on the nuisance parameters and profile.

Case where we don't have p_{true}



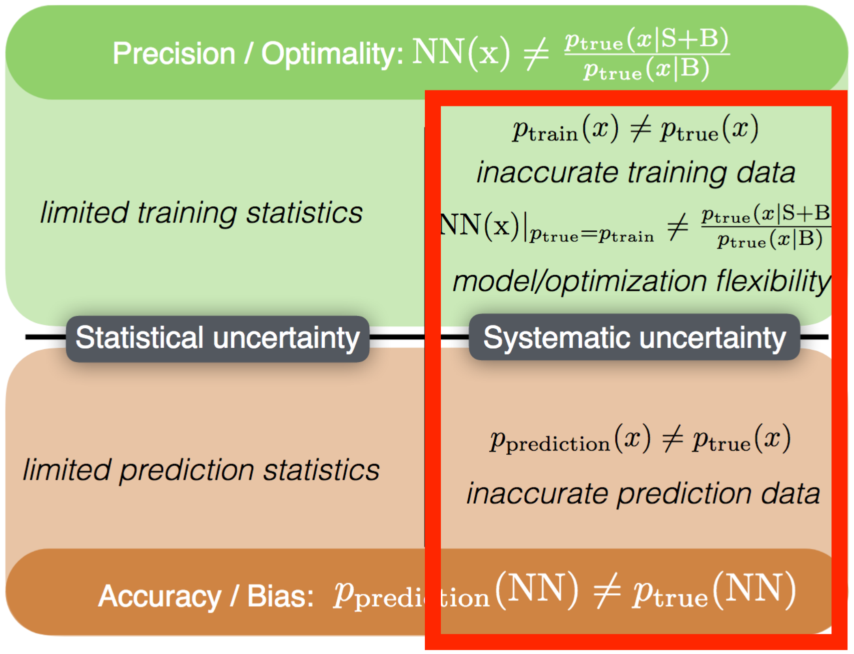
Option 2 ("uncertainty-awareness"):

Let the analysis depend explicitly on the nuisance parameters and profile.



2105.08742 (and refs within)

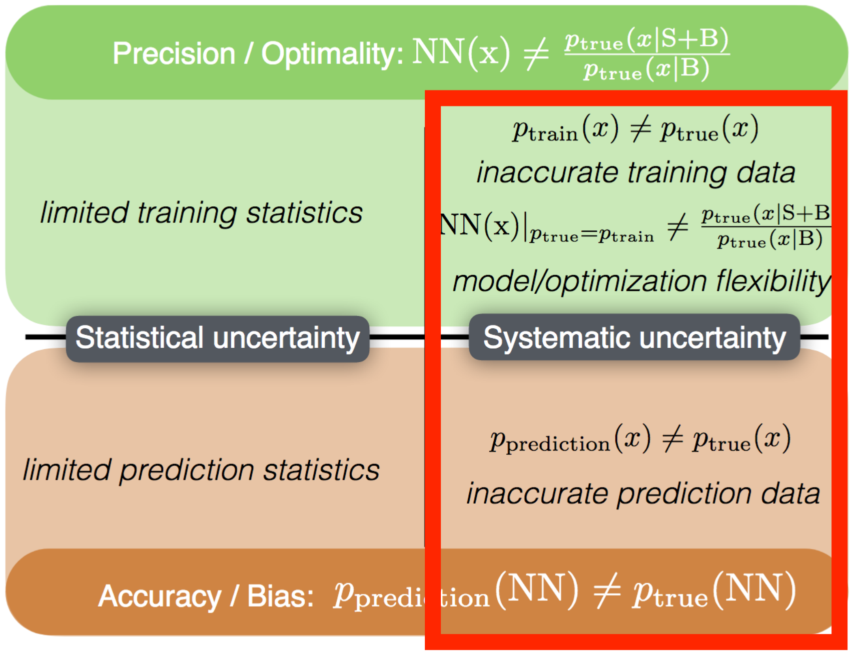
Case where we don't have p_{true}



Option 3 ("Inference-awareness"):

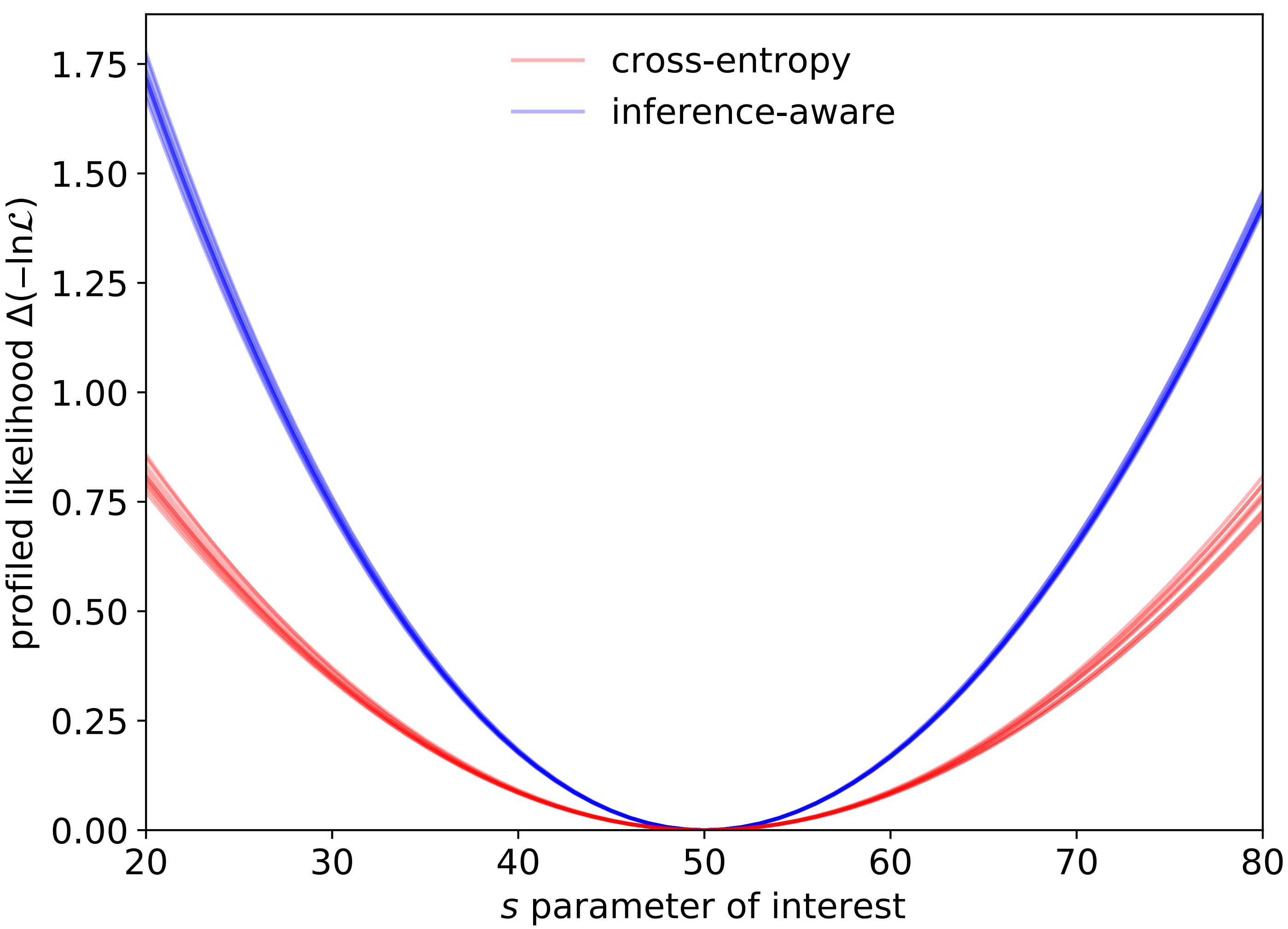
Optimize the final statistic, including all systematic uncertainties.

Case where we don't have p_{true}



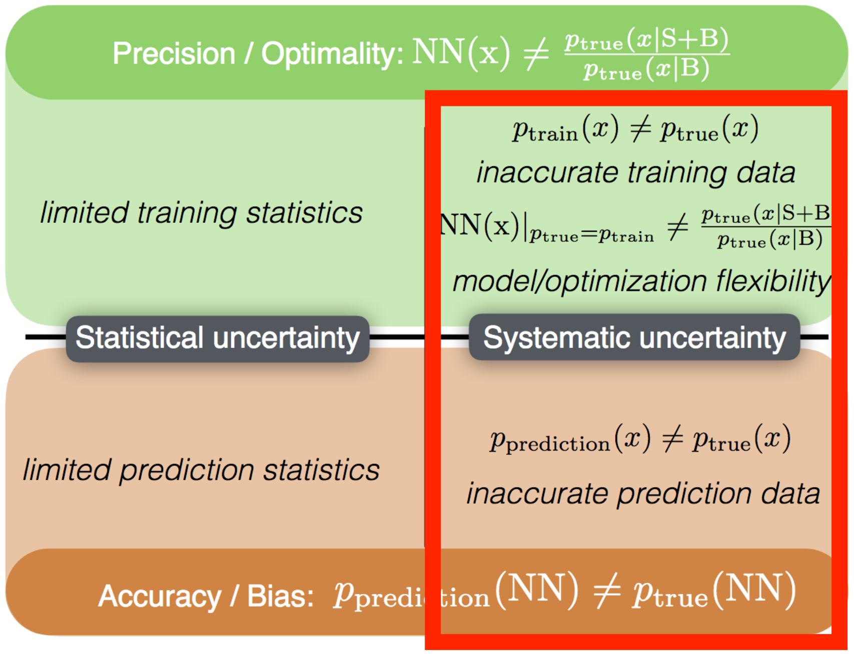
Option 3 ("Inference-awareness"):

Optimize the final statistic, including all systematic uncertainties.



1806.04743 and papers that cite it

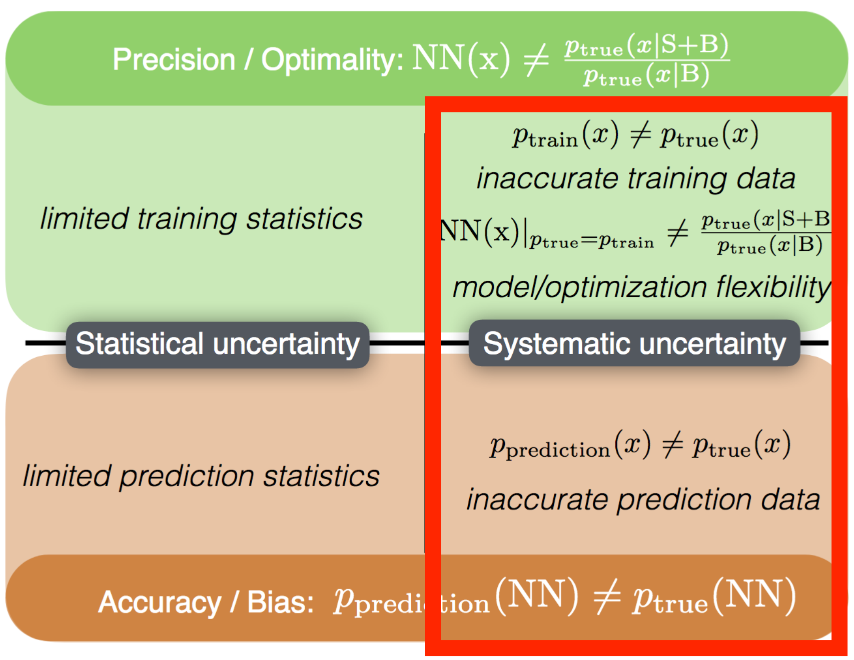
Case where we don't have p_{true}



Option 4:

Learn directly from data and avoid simulations altogether !

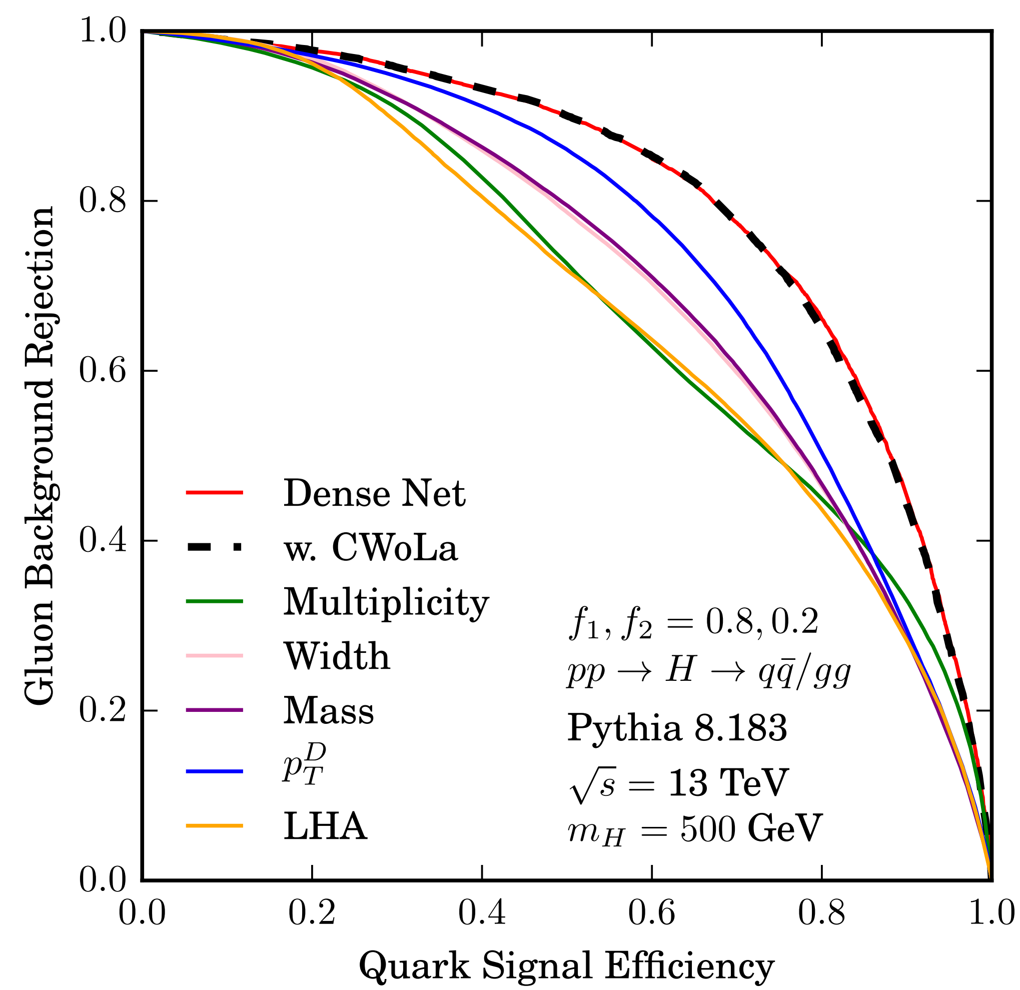
Case where we don't have p_{true}



Option 4:

Learn directly from data and avoid simulations altogether !

Dashed line = ~no labels!



e.g. 1708.02949. Of course, this is not always possible and there may be other uncertainties related to the learning assumptions.

Uncertainties for a NN-based analysis

Precision / Optimality: $\mathrm{NN(x)} \neq \frac{p_{\mathrm{true}}(x|\mathrm{S+B})}{p_{\mathrm{true}}(x|\mathrm{B})}$

Statistical uncertainty	Systematic uncertainty
limited training statistics	$p_{\mathrm{train}}(x) \neq p_{\mathrm{true}}(x)$ *inaccurate training data* $\mathrm{NN(x)}\|_{p_{\mathrm{true}}=p_{\mathrm{train}}} \neq \frac{p_{\mathrm{true}}(x\|\mathrm{S+B})}{p_{\mathrm{true}}(x\|\mathrm{B})}$ *model/optimization flexibility*
limited prediction statistics	$p_{\mathrm{prediction}}(x) \neq p_{\mathrm{true}}(x)$ *inaccurate prediction data*

Accuracy / Bias: $p_{\mathrm{prediction}}(\mathrm{NN}) \neq p_{\mathrm{true}}(\mathrm{NN})$

1909.03081

Uncertainties

“But what are the uncertainties on the NN”?

- question asked by every reviewer

Goal: let’s sharpen this question and explore various cases and related topics.

Outline:

- Uncertainty Landscape with ML
- Reducing Uncertainties with ML
- Conclusions / Outlook

Uncertainties

“But what are the uncertainties on the NN”?

- question asked by every reviewer

Goal: let’s sharpen this question and explore various cases and related topics.

Outline:

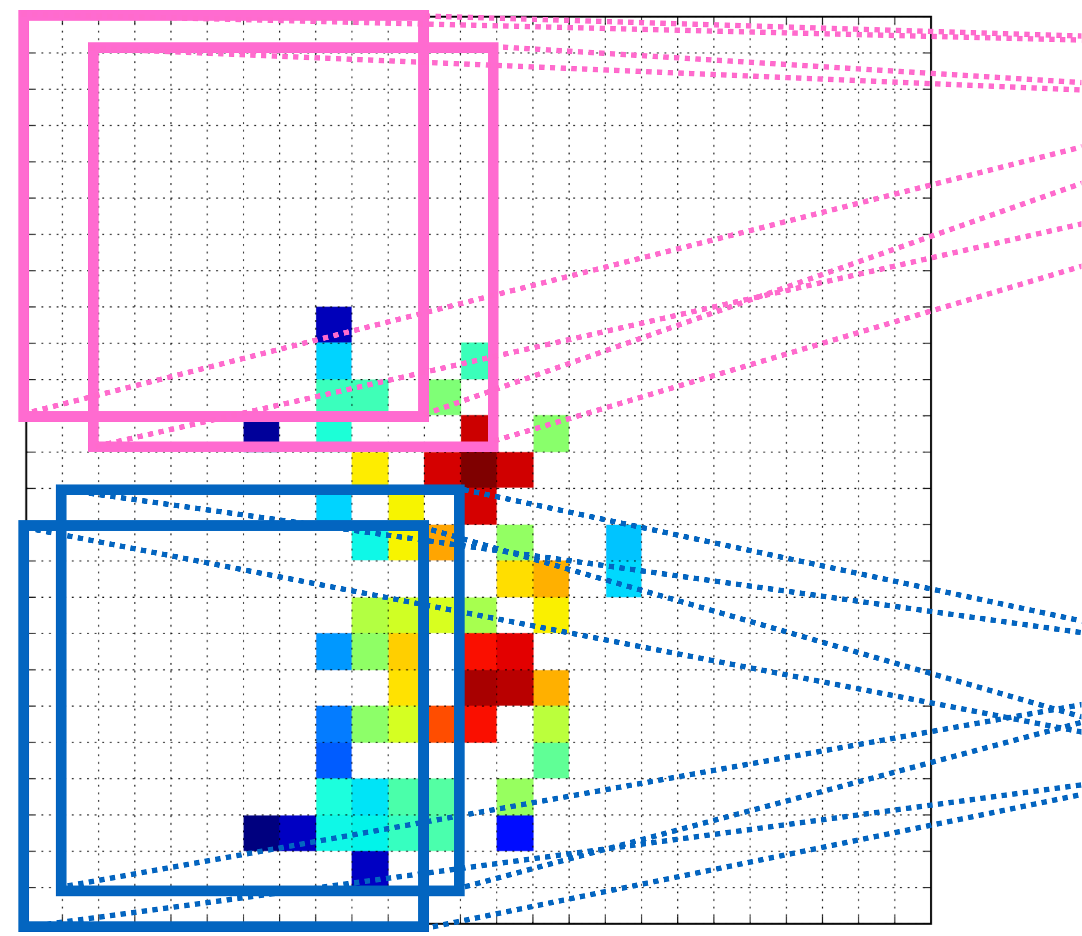
- Uncertainty Landscape with ML
- Reducing Uncertainties with ML
- **Conclusions / Outlook**

Conclusions and Outlook

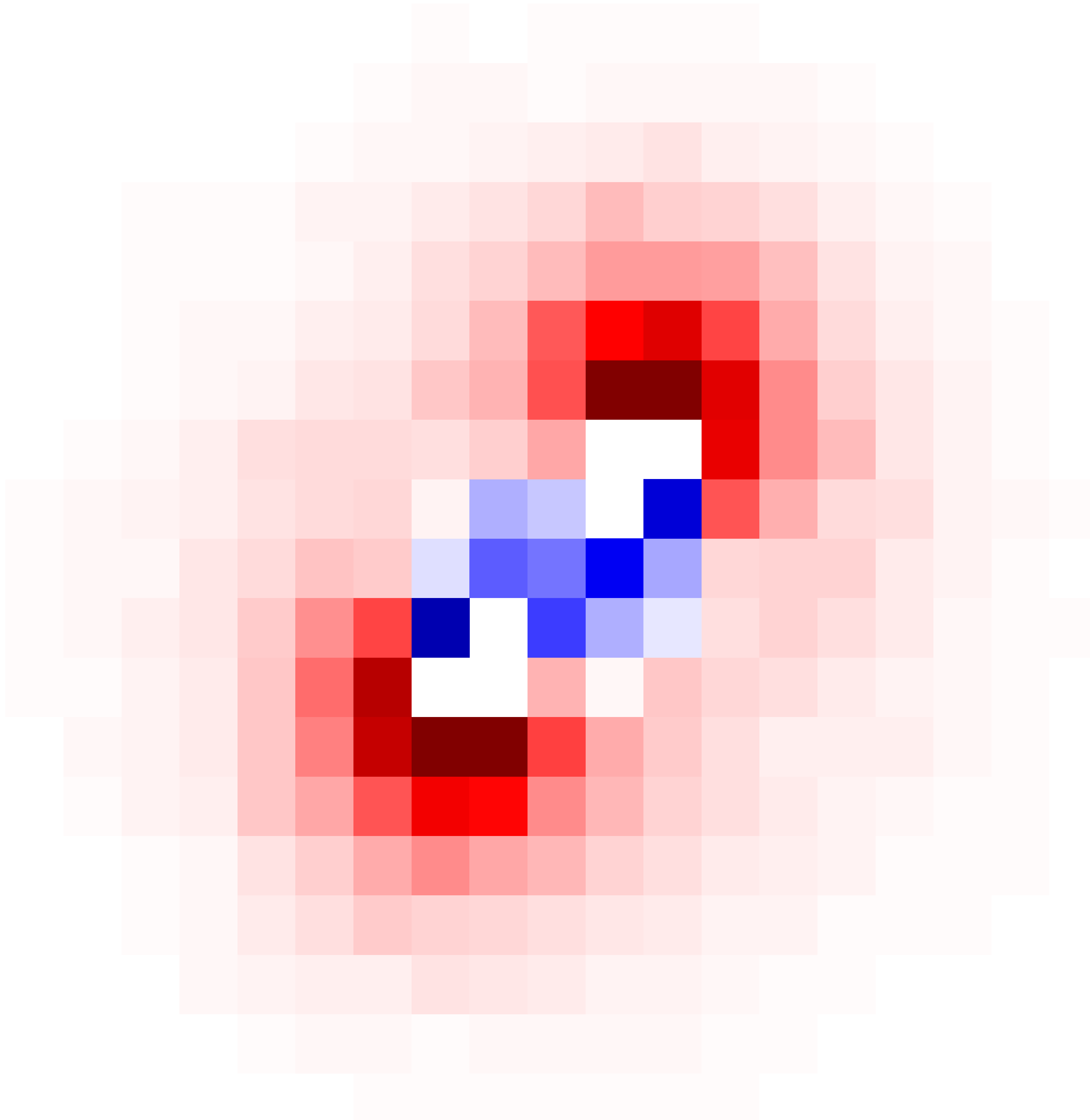
AI/ML has a great potential to **enhance**, **accelerate**, and **empower** HEP/NP analyses

In order to make the best use of these tools, we need to ensure that they are **robust**

A tool is only as good as its calibration !



Hopefully I have helped clarify a little when we are uncertain with ML-based analyses and how to improve for the future!



Fin.