Fast LHC Signal Prediction using Machine Learning

TeVPA 2016 @ CERN

Sebastian Liem
Collaborators: Gianfranco Bertone, Marc Peter Deisenroth, Jong Soo Kim, Roberto Ruiz de Austri, Max Welling

GRAPPA
GRavitation AstroParticle Physics Amsterdam

Calculating signal is slow!

Comparing your favourite model to LHC data means calculating the number of signal events in a specific signal region.

This calculation involves very expensive MC calculations. O(10 min) for a single set of parameter values using fast simulator (e.g. Delphes).

Global scans of high dimensional theories can require O(10^6-10^8) likelihood evaluations.

Model parameters

Event generator

Detector simulation

Cross section

Analysis cuts

$$N_s = L\sigma\epsilon$$

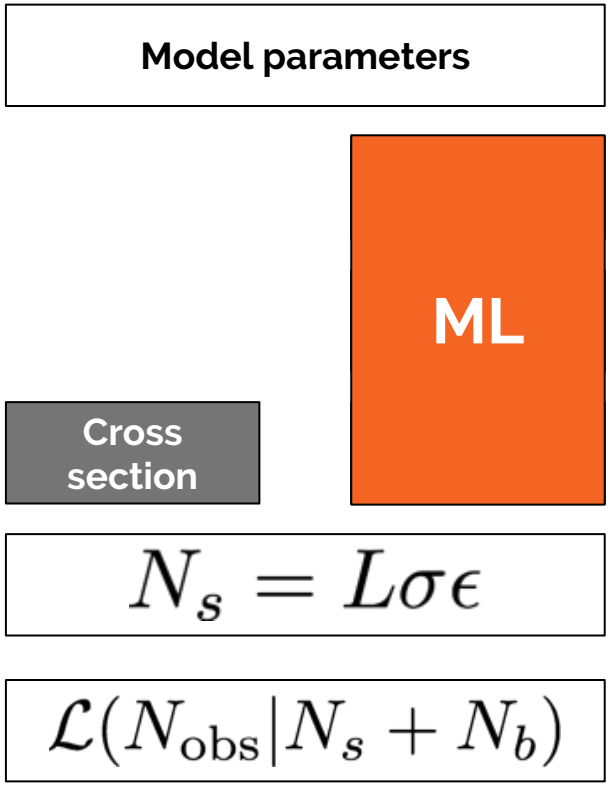
$$\mathcal{L}(N_{\rm obs}|N_s + N_b)$$

Machine learning is fast!

Replace the expensive MC calculation with a **surrogate function**.

$$\epsilon = f(\theta) + \xi$$

In ML language this a supervised regression problem. Many algorithms, we will use Gaussian processes.



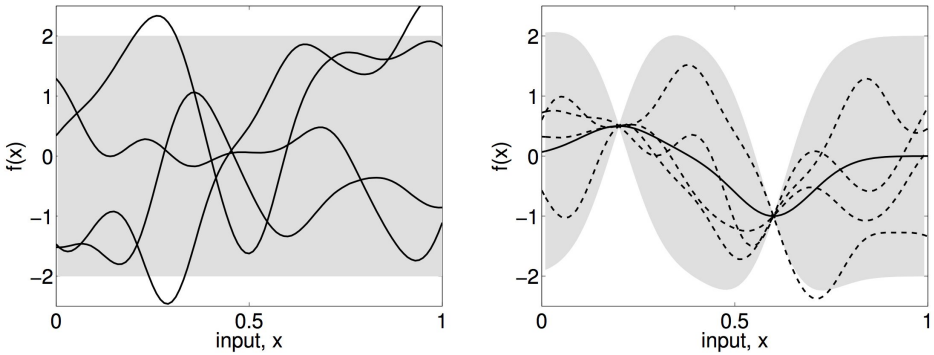
Gaussian Processes

Non-parametric: No need to assume a functional form.
Probabilistic: Produces posteriors, i.e. estimates its own error.
Bayesian: Needs a prior on the type of functions. Kernel/covariance function.

$$k_y(x_p, x_q) \;=\; \sigma_f^2 \exp\left(-\frac{1}{2\ell^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq}.$$

Picking hyperparameters is the *learning* task in Gaussian processes.

To deal with large training dataset, we use Distributed Gaussian Processes
Deisenroth & Ng, arXiv:1502.02843



(a), prior (b), posterior

Figure 1.1: Panel (a) shows four samples drawn from the prior distribution. Panel (b) shows the situation after two datapoints have been observed. The mean prediction is shown as the solid line and four samples from the posterior are shown as dashed lines. In both plots the shaded region denotes twice the standard deviation at each input value x.

"Gaussian Processes for Machine Learning"
Rasmussen & Williams, 2006
www.gaussianprocess.org

Natural SUSY

Natural because it stabilizes the electroweak scale without fine-tuning.

Only few SUSY states needs to be light.

$$\theta = \{\tan\beta,\ \mu,\ M_3,\ m_{\tilde{Q}_t},\ m_{\tilde{t}_R},\ A_t\}$$

Low-dimensional yet realistic theory.

We already had the training data from previous paper.

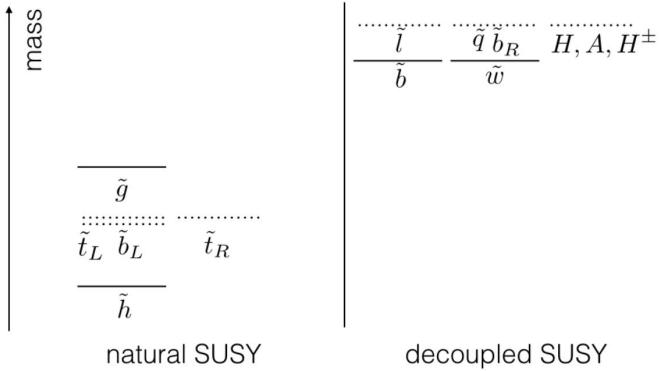
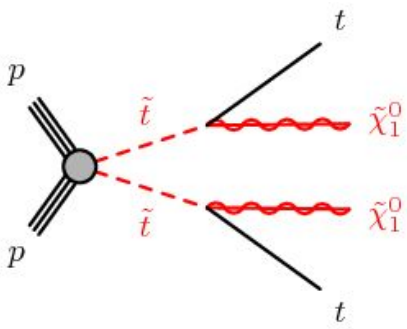


FIG. 1. The minimal natural SUSY mass spectrum on the left while the remaining supersymmetric particles are decoupled on the right.

"Natural SUSY: Now or Never?" Kim et al. arXiv:1606.06738

Two Signal Regions



Defined in ATLAS-PHYS-PUB-2013-011

Looking direct stop production with HL-LHC, 14 TeV with 3000 fb^{-1}

Stops decay typically to top or b quarks, W/Z or Higgs bosons, and a LSP. Multiple jets, b-jets, large MET, possibly leptons.

1-lepton

MET > 750 GeV
m_T(lepton, MET) > 550 GeV

Total bkg: 21.1 ± 5.9

0-lepton

MET > 800 GeV
m_T(b-jet, MET) > 400 GeV

Total bkg: 12.2 ± 3.9

Training the Gaussian Processes

18647 models split into 16647 models for training and 2000 models for testing.

O(10 min) to train per signal region.

The lunch is not free, just cheaper!

SPheno, Pythia, NLLFAST, CheckMATE, Delphes etc. still needed to generate training data.

Models uniformly sampled from these ranges:

$$
\begin{aligned}
0.1\,\mathrm{TeV} \leq\ & |\mu| \leq 1.0\,\mathrm{TeV}, \\
0.1\,\mathrm{TeV} \leq\ & m_{\tilde{Q}_t} \leq 2.0\,\mathrm{TeV}, \\
0.1\,\mathrm{TeV} \leq\ & m_{\tilde{t}_R} \leq 2.0\,\mathrm{TeV}, \\
0.1\,\mathrm{TeV} \leq\ & |M_3| \leq 3.0\,\mathrm{TeV}, \\
& |A_t| \leq 3.0\,\mathrm{TeV}, \\
1 \leq\ & \tan\beta \leq 20.
\end{aligned}
$$

All models avoid LEP-II chargino limit, and have reasonable Higgs boson mass.

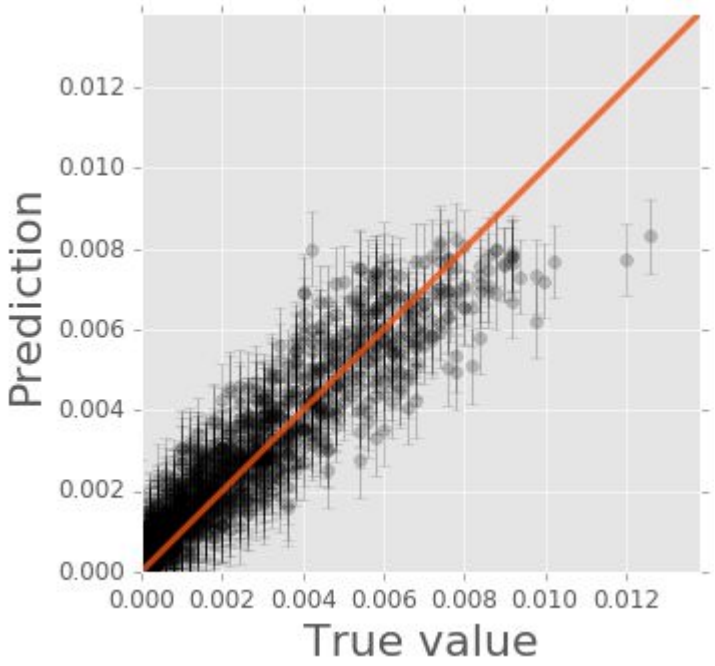
Is it fast?

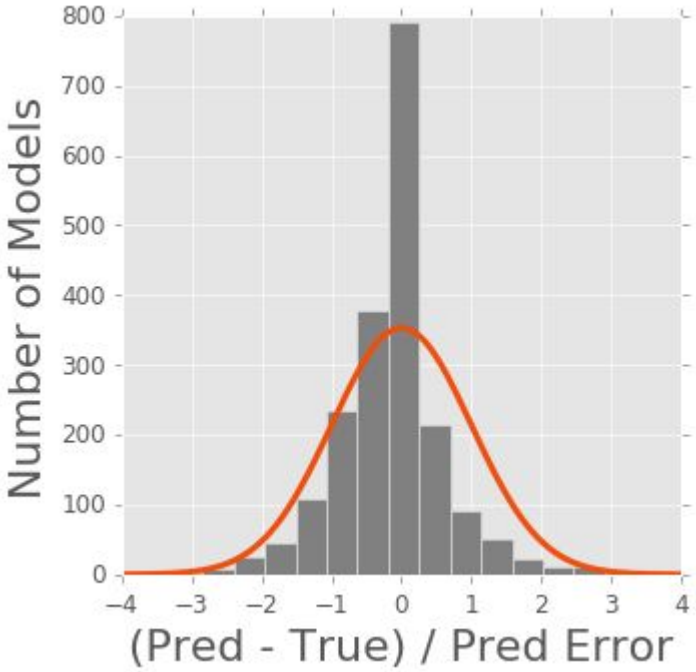
0.06 sec/eval

4 GHz single core, unoptimized Python

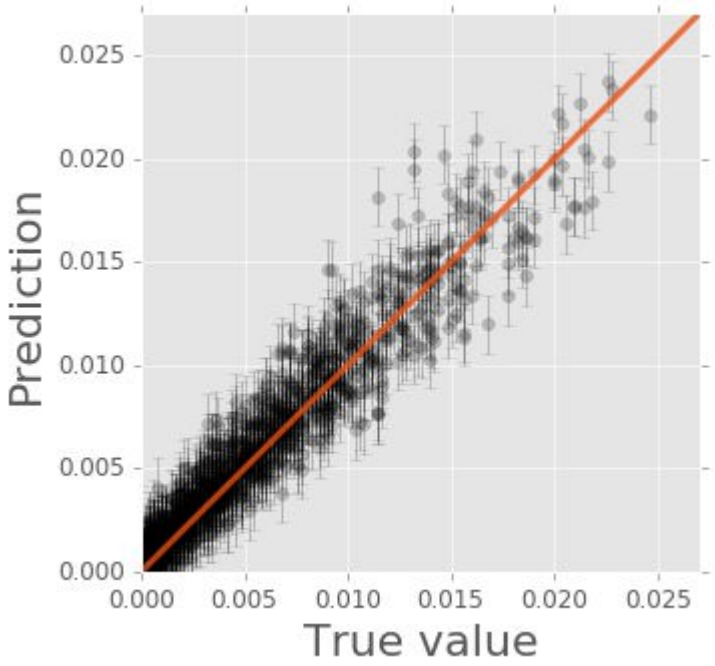
Compare with O(10 min) for MC approach.

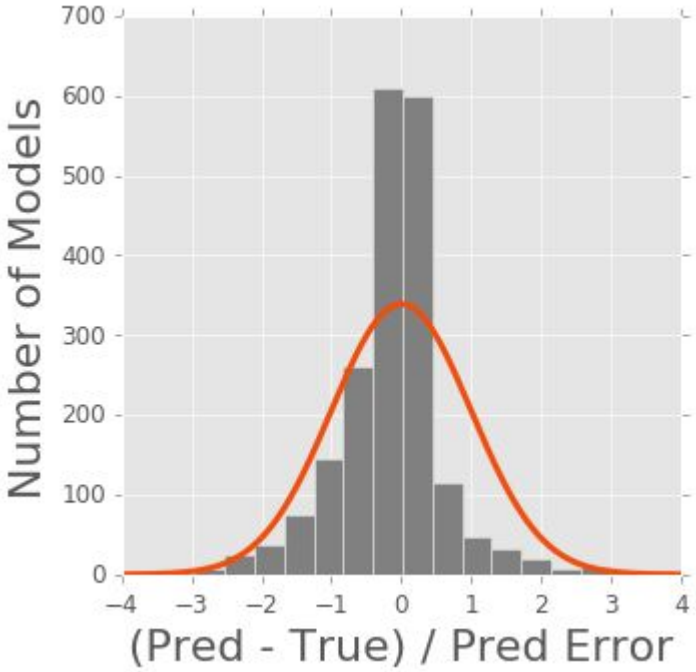
Is it right?





0-lepton signal region





1-lepton signal region

Proof-of-concept Reconstruction

Assume one particular model is true, and measured at the LHC.

Try to reconstruct the parameters using our two signal regions.

Reasonable Higgs mass, LEP chargino limit.

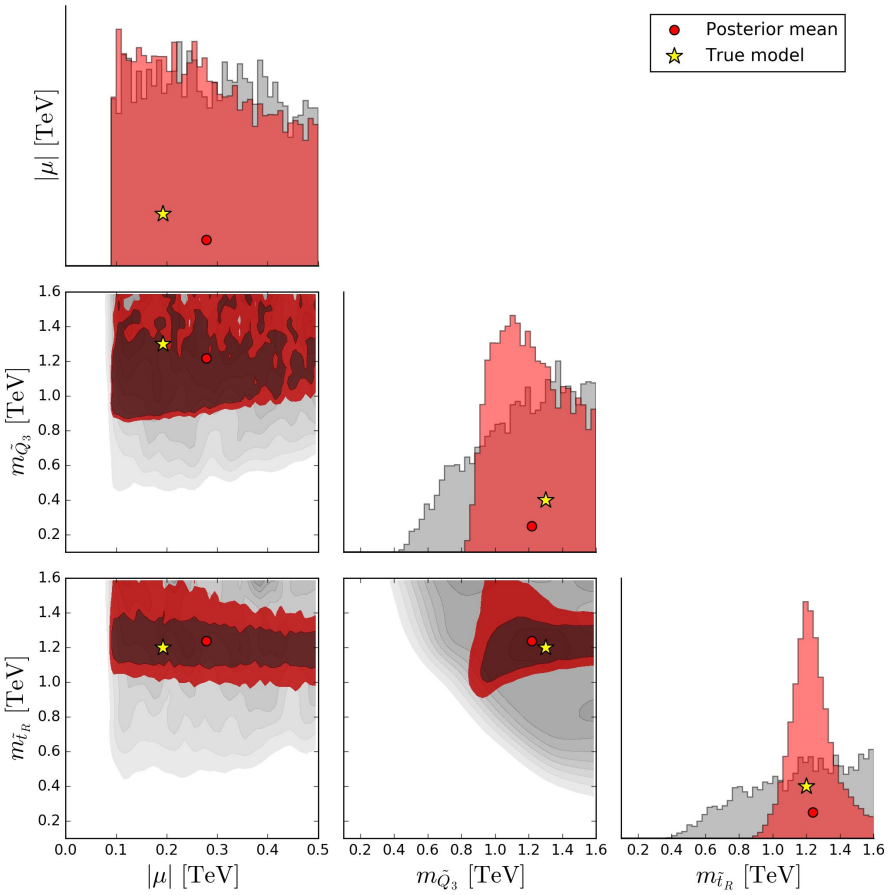
Benchmark model has

63.2 events in 0-lepton SR

161.8 events in 1-lepton SR

(Bkg: 12.2 ± 3.9 and 21.1 ± 5.9)

No MC performed!
No event generation.
No detector simulation.
~100k likelihood eval.



Reconstruction, 3D scan

Outlook

Investigate the noise model of our MC calculators. Other kernels or Student-t processes are possibilities.

Extend to MSSM, other theories, more SRs. Online training.

Also collaborating with the iDark group of Sascha Caron (authors of SUSY-AI). iDark = ML + HEP + astrophysics expertise.

To conclude...

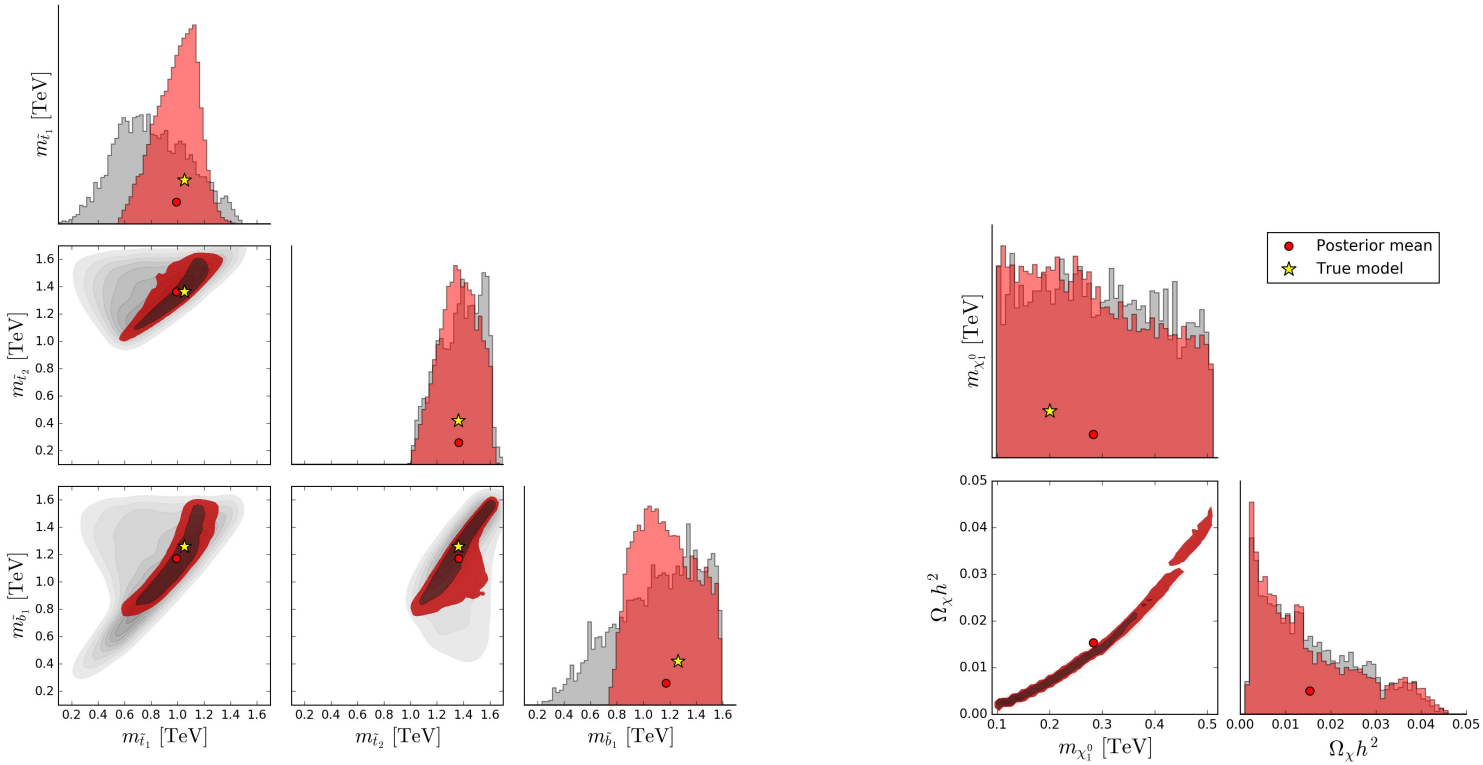
Gaussian processes makes LHC inference really really fast.

Bypasses event generation, detector simulation.

The speed enables a fast reconstruction of theory parameters, and eventually dark matter properties.

The method is completely generic, with many possible applications.

Backup slides

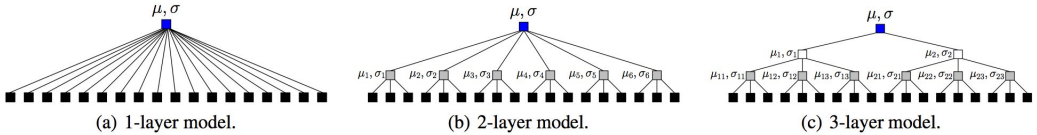


Reconstruction, 3D scan

Distributed Gaussian Processes

The standard Gaussian process scales badly with ***N*** the size of the training dataset. It involves inverting ***N***x***N*** matrices.

We use **distributed** Gaussian processes to avoid this. The data is randomly partitioned and on each partition a Gaussian process is defined. Predictions from each process is then combined.



(a) 1-layer model. (b) 2-layer model. (c) 3-layer model.

Figure 1. Computational graphs of hierarchical PoE models. Main computations are at the leaf nodes (GP experts, black). All other nodes recombine computations from their direct children. The top node (blue) computes the overall prediction.

"Distributed Gaussian Processes"
Deisenroth & Ng, arXiv:1502.02843

Benchmark

$$\mu = -192.5$$
$$M_3 = 2100$$
$$m_{\tilde{Q}_3} = 1300$$
$$m_{\tilde{t}_R} = 1200$$
$$\tan\beta = 8.43$$
$$A_t = 2000$$