

# Contents

<b>1</b>	<b>Results</b>	<b>1</b>
1.1	Learning Curves . . . . .	1
1.2	Comparison with Prospino and NLL-fast . . . . .	3
1.2.1	Relative Deviance . . . . .	3
1.2.2	Cross Sections . . . . .	5
1.3	The Optimal Model . . . . .	7
1.3.1	Times and Kernels . . . . .	10
1.3.2	Model Size . . . . .	10



# Chapter 1

## Results

In this chapter the results of training Gaussian processes on the MSSM-24 dataset are shown. The settings are the cumulative settings found in the previous chapter, and distributed Gaussian processes are used to scale the problem to large datasets. Learning curves as a function of number of experts are plotted for the cumulative settings. Plots of the relative deviances from the **Prospino** data are shown, and the resulting predictions are compared to data from **Prospino** and **NLL-fast**. Finally, the optimal model with respect to predictive capabilities, model size and computation times is discussed.

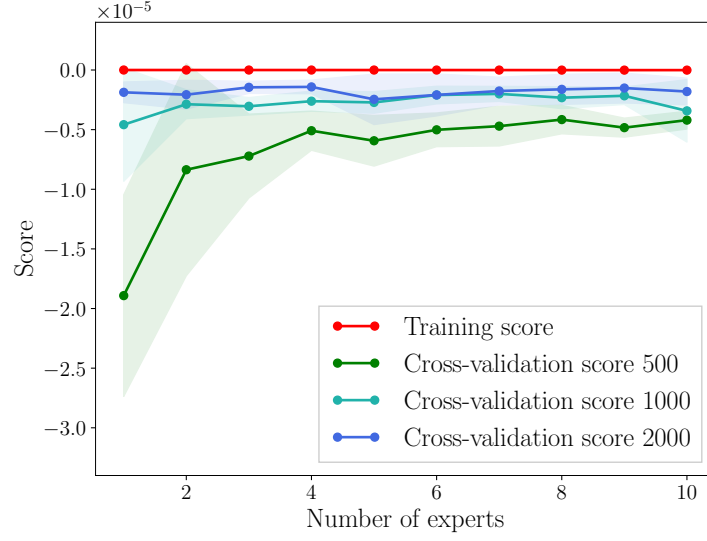
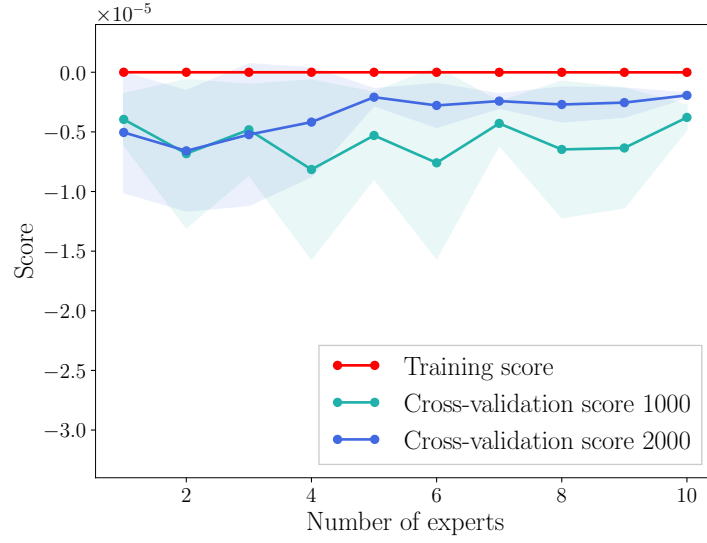
### 1.1 Learning Curves

Learning curves for the cumulative settings from Sec. ?? are shown in Fig. 1.1 for  $\tilde{d}_L \tilde{u}_L$  and  $\tilde{d}_L \tilde{d}_L$ . The experts are trained with 500, 1000 and 2000 points per expert, and learning curves are calculated according to the method described in Sec. ??.

The training scores for the estimators of  $\tilde{d}_L \tilde{d}_L$  and  $\tilde{d}_L \tilde{u}_L$  are 1, indicating that neither model is underfitting. The validation curves for both processes converge towards 1, albeit faster and for less training points per expert for  $\tilde{d}_L \tilde{d}_L$  than for  $\tilde{d}_L \tilde{u}_L$ . In both cases the training and validation scores are very high, even for few experts. Adding more data, both in the form of more experts and more points per expert, give higher validation scores. Predictions in this chapter are therefore done with the largest reasonable<sup>1</sup> models, of 10 experts with 5000 and 8000 training points each, depending on whether or not the models need to be stored.

---

<sup>1</sup>Taking into account computation times, matrix sizes and model sizes

(a) The process  $\tilde{d}_L \tilde{d}_L$ (b) The process  $\tilde{d}_L \tilde{u}_L$ 

**Figure 1.1:** Learning curves as a function of number of experts, with 500, 1000 and 2000 training points per expert for the processes (a)  $\tilde{d}_L \tilde{d}_L$  and (b)  $\tilde{d}_L \tilde{u}_L$ . For  $\tilde{d}_L \tilde{d}_L$  the validation score for 500 training points is not shown, as the uncertainty is very large. The  $k$ -fold cross validation uses  $R^2$ -score as described in Sec. ??, and here  $R^2 - 1$  is plotted.

## 1.2 Comparison with Prospino and NLL-fast

In this section plots of the relative deviance distributions as defined in Sec. ?? are shown for the squark pair-production cross sections from the MSSM-24 and CMSSM datasets. Cross sections predicted by the DGP are compared to cross sections calculated using **Prospino**, and **NLL-fast** where this is possible.

The settings used in this chapter are

- 10 GP experts with 8000 or 5000 training points from the MSSM-24 dataset each
- 20 000 test points
- Features  $m_{\tilde{g}}, m_{\tilde{q}_i}, \bar{m}$  ( $m_{\tilde{g}}, m_{\tilde{q}_i}, m_{\tilde{q}_j}, \bar{m}$ ) for the process  $\tilde{q}_i \tilde{q}_j$  where  $i = j$  ( $i \neq j$ )<sup>2</sup>
- The Matérn kernel with  $\nu = 1.5$  and a white noise term

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp \left( 1 + \sqrt{3} [(\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j)]^{1/2} \right) \quad (1.1)$$

$$\times \exp \left( \sqrt{3} [(\mathbf{x}_i - \mathbf{x}_j)^T M (\mathbf{x}_i - \mathbf{x}_j)]^{1/2} \right) + \sigma_n^2 \delta_{ij}, \quad (1.2)$$

where  $M = \text{diag}(\ell)^{-2}$ .

- A cut on the cross sections  $\sigma > \sigma_{cut} = 10^{-16}$  fb

All first- and second generation squarks are considered;  $m_{\tilde{u}_L}, m_{\tilde{d}_L}, m_{\tilde{s}_L}, m_{\tilde{c}_L}, m_{\tilde{u}_R}, m_{\tilde{d}_R}, m_{\tilde{s}_R}$  and  $m_{\tilde{c}_R}$ . These make up 36 different processes for squark pair production. Gaussian processes were used on all processes.

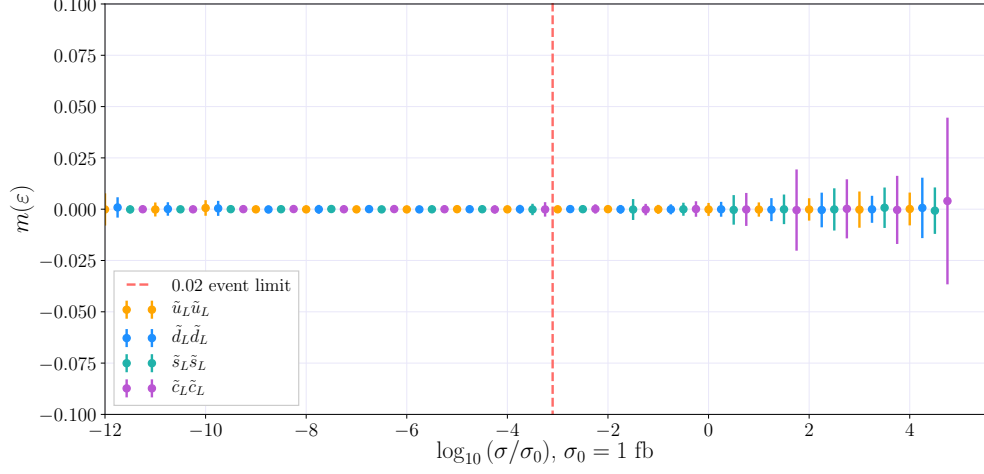
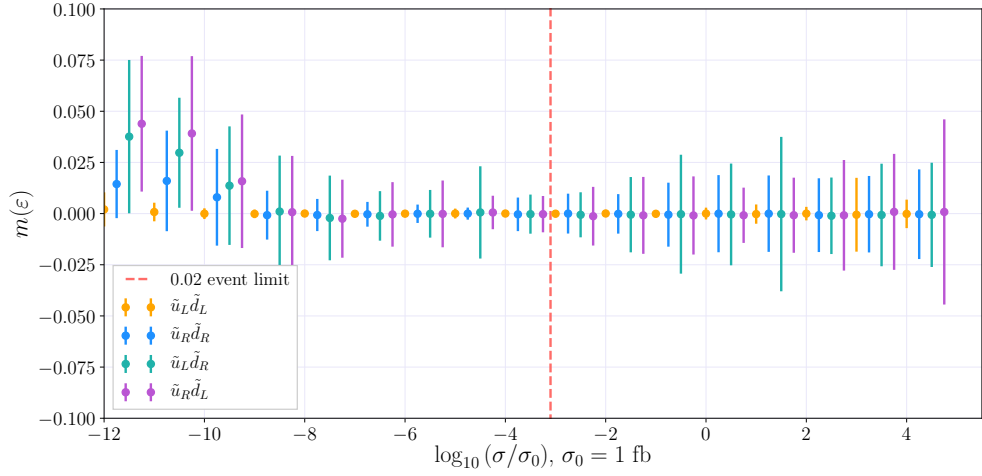
### 1.2.1 Relative Deviance

#### MSSM-24

A selection of the relative deviance plots are shown in Fig. 1.3. In Fig. 1.1a the relative deviance distributions are shown for all processes with equal-flavour left-handed squarks. These processes only have 3 features, as opposed to 4, which appear to make them easier to learn for the GP. The GP estimators for the corresponding right-handed processes are almost identical as functions of  $m_{\tilde{q}_R}$ . This is because **Prospino** calculates cross sections for strong interactions, and the NLO terms only contain QCD corrections, as discussed in Sec. ?. Since there is no electroweak correction, there is no distinction between left- and righthanded squarks in the calculation. They also share the same underlying pdf from the

---

<sup>2</sup>Same flavour quarks with different chiralities, *e.g.*  $\tilde{d}_L \tilde{d}_R$ , are regarded as different  $i \neq j$ , because  $m_{\tilde{d}_R} \neq m_{\tilde{d}_L}$ .

(a) The processes  $\tilde{u}_L \tilde{u}_L$ ,  $\tilde{d}_L \tilde{d}_L$ ,  $\tilde{s}_L \tilde{s}_L$  and  $\tilde{c}_L \tilde{c}_L$ .(b) The processes  $\tilde{u}_L \tilde{d}_L$ ,  $\tilde{u}_R \tilde{d}_R$ ,  $\tilde{u}_L \tilde{d}_R$  and  $\tilde{u}_R \tilde{d}_L$ .

**Figure 1.2:** Relative deviance distributions as a function of the logarithm of the normalized cross sections  $\log_{10} \sigma/\sigma_0$ . Ten experts with 8000 training points from the MSSM-24 dataset each were combined on 20 000 test points from the MSSM-24 dataset for processes (a)  $\tilde{u}_L \tilde{u}_L$ ,  $\tilde{d}_L \tilde{d}_L$ ,  $\tilde{s}_L \tilde{s}_L$  and  $\tilde{c}_L \tilde{c}_L$ ; and (b)  $\tilde{u}_L \tilde{d}_L$ ,  $\tilde{u}_R \tilde{d}_R$ ,  $\tilde{u}_L \tilde{d}_R$  and  $\tilde{u}_R \tilde{d}_L$ .

quark, *e.g.* the pdf's used for calculating  $\tilde{d}_R\tilde{d}_R$  and  $\tilde{d}_L\tilde{d}_L$  is the one for the  $d$ -flavour squark.

The prediction for equal squarks in Fig. 1.2a is very stable and very good. All relative deviance distributions have a mean of approximately zero, and a standard deviation well within the desired value of 10%. The highest cross sections have been excluded from the plots, as there are few training and test points there, and so the prediction has a very large uncertainty.

In Fig. 1.2b the relative deviance distributions are shown for the process  $\tilde{u}\tilde{d}$  for different chirality combinations. The expression for the cross section depends on the chirality combinations, as discussed in Sec. ???. The prediction for  $\tilde{d}_L\tilde{u}_L$  is quite superior to the other chirality combinations, possibly because of the previous argument of the masses  $m_{\tilde{d}_L}$  and  $m_{\tilde{u}_L}$  being very strongly correlated, as discussed in Sec. ???. The prediction for  $\tilde{d}_R\tilde{u}_R$  is also better than the processes with different chiralities, but not as good as  $\tilde{u}_L\tilde{d}_L$ . All processes in Fig. 1.2b have 4 features, which seems to make the function more difficult to predict. The mean values of the relative deviance distributions are still close to zero, particularly for cross sections above the 0.02 event limit. Standard deviations are larger than for equal squark processes, but well within 5% for all cross sections larger than  $> 10^{-8}$  fb.

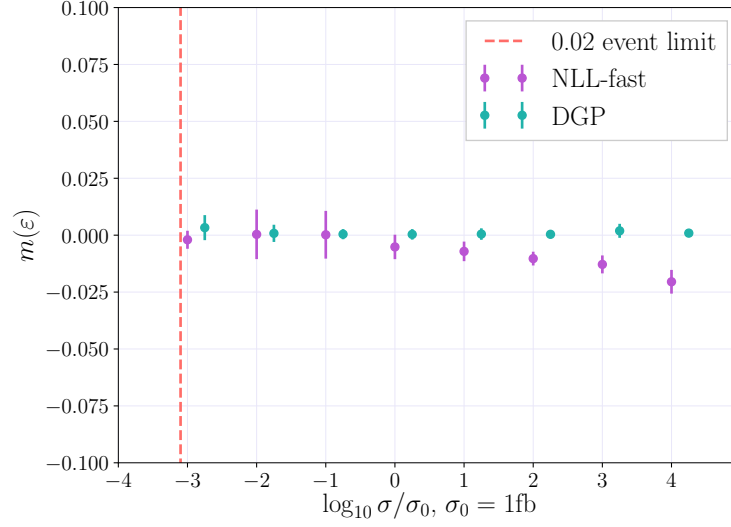
## CMSSM

The DGPs trained on MSSM-24 data are also tested on an CMSSM data set and compared to cross sections calculated by **NLL-fast** for the same parameter points. The true values are the sums of cross sections for all 36 processes for each parameter point, calculated by **Prospino**. The DGPs estimate the cross sections for each of the 36 processes as well, and these are summed for each parameter point.

The resulting relative deviance distributions are shown in Fig. 1.3. The cross sections calculated by **NLL-fast** are quite close to the true values for the CMSSM data. This is because the squark masses in CMSSM have much smaller splittings than in MSSM-24, as discussed in Sec. ??, and **NLL-fast** assumes degenerate squark masses. For large cross sections, however, **NLL-fast** predicts values that are too large, while the DGPs predict cross sections very close to the true value.

### 1.2.2 Cross Sections

Cross sections estimated by the distributed Gaussian processes are compared to cross sections from **Prospino** and **NLL-fast**. The estimators consist of 10 experts with 5000 training points each, trained on the MSSM-24 data. Note that these experts are smaller than those used in Sec. 1.2.1, where each expert used 8000 training points. This is due to the size of saved models, which will be discussed in Sec. 1.3. In Fig. 1.4 the cross sections for the processes  $\tilde{d}_L\tilde{d}_L$  and  $\tilde{d}_L\tilde{u}_L$  are



**Figure 1.3:** Distribution of the relative deviance  $\varepsilon$  as a function of the logarithm of the total cross section for all squark pair production processes for NLL-fast and Gaussian processes for mSUGRA data. The 'true' values are the values generated by Prospino. 10 experts with 5000 training points each were trained on MSSM-24 data for each process  $\tilde{q}_i \tilde{q}_j$ .

shown as a function of  $m_{\tilde{d}_L}$ . The masses  $m_{\tilde{d}_L}$  and  $m_{\tilde{u}_L}$  are in the range [200, 2500] GeV, with the approximate mass splitting of

$$m_{\tilde{d}_L}^2 - m_{\tilde{u}_L}^2 \approx m_W^2, \quad (1.3)$$

for  $m_W = 80$  GeV. All other squark masses are held at 1000 GeV, and the gluino mass is  $m_{\tilde{g}} = 500$  GeV. The DGP prediction is very close to the cross sections from Prospino, with a slightly lower prediction for large  $m_{\tilde{d}_L}$ . The DGP predicts lower cross sections than those calculated Prospino, but the estimated values are within the uncertainty band. In addition, the uncertainty in the DGP prediction is very small, with the plots in Fig. 1.4 showing  $50\sigma$  uncertainty bands.

The uncertainty in the Prospino calculation comes from the renormalization scale dependence. Cross sections are calculated for twice the renormalization scale, and half the renormalization scale, to see how scale dependent the cross sections are. As discussed in Sec. ??, the scale dependence is reduced with the addition of higher order terms to the cross section, so this is also a way of estimating the order of magnitude of higher order terms (in this case, next-to-next-to-leading order).

The cross sections calculated by NLL-fast for the MSSM-24 are very far from the true values, and only coincide with Prospino for  $m_{\tilde{d}_L} = 1000$  GeV, where the masses are in fact degenerate. In addition, the uncertainty from NLL-fast is



Tool	Computation time [s]
Prospino	1711.96
NLL-fast	0.00739
Distributed Gaussian Processes	28.03

**Table 1.1:** Computation times for 1 parameter point for all 36 squark pair production processes.

very large. The uncertainty from **NLL-fast** includes the uncertainty from scale dependence, from the pdf's and from  $\alpha_s$ .

Cross sections are also calculated as a function of the gluino mass,  $m_{\tilde{g}}$ , and shown in Fig. 1.6. All squark masses are here held at 1000 GeV, and the gluino mass is in the range [200, 2400] GeV. In this case the DGP prediction and cross sections from **Prospino** and **NLL-fast** all coincide, but DGP gives an uncertainty that is considerably lower than **Prospino** and **NLL-fast**.

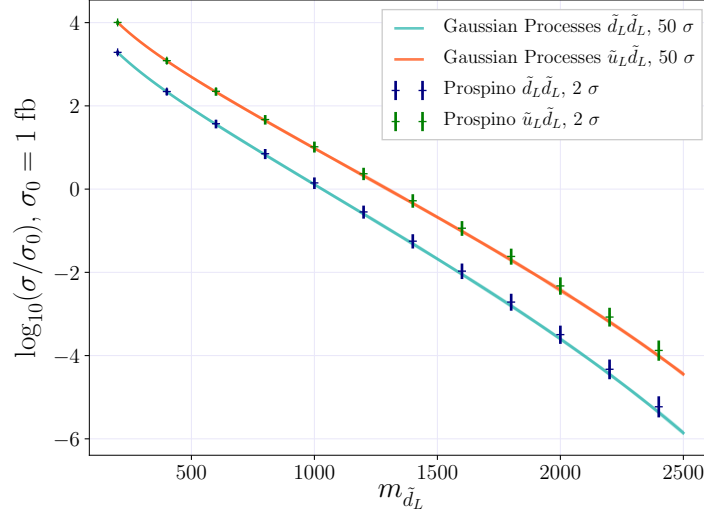
In Tab. 1.1 the computation times for all 36 process cross sections for a single parameter points are shown for **Prospino** and distributed Gaussian processes. For **Prospino** 3 cross sections are calculated for each process, with scales 0.5, 1.0 and 2.0, to include the scale uncertainty. The time for **NLL-fast** is also shown for one parameter point. Note that the distributed Gaussian prediction was not performed in parallel, meaning that the computation time could optimally be reduced to  $28.03 \text{ s} / 36 = 0.78 \text{ s}$ , if each of the 36 processes is performed on a separate core. **NLL-fast** is considerably faster than both DGP and **Prospino**, but has a very large error relative to the **Prospino** computation, as seen in Fig. 1.5.

## 1.3 The Optimal Model

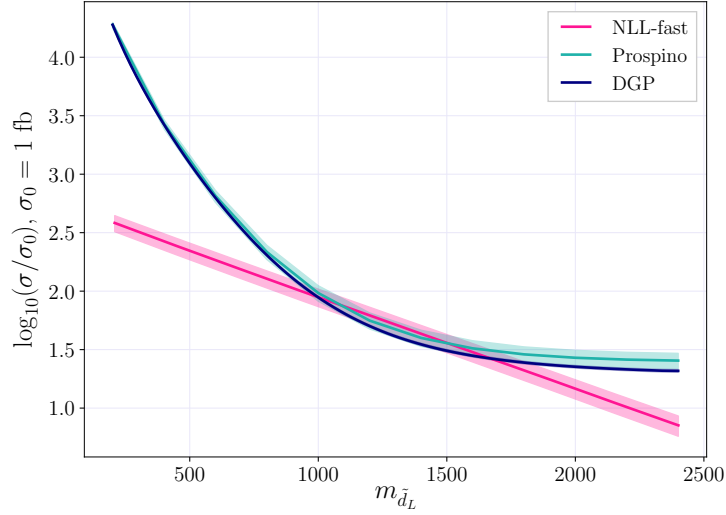
The distributed Gaussian process predictions using 10 experts with 8000 training points each are very good. Unfortunately, the Gaussian process models take up a lot of space when stored. In addition, larger DGP models take longer to predict values. In this section the model sizes and computation times are discussed, to find the optimal model as a compromise between prediction quality, size and computation time.

### Training data

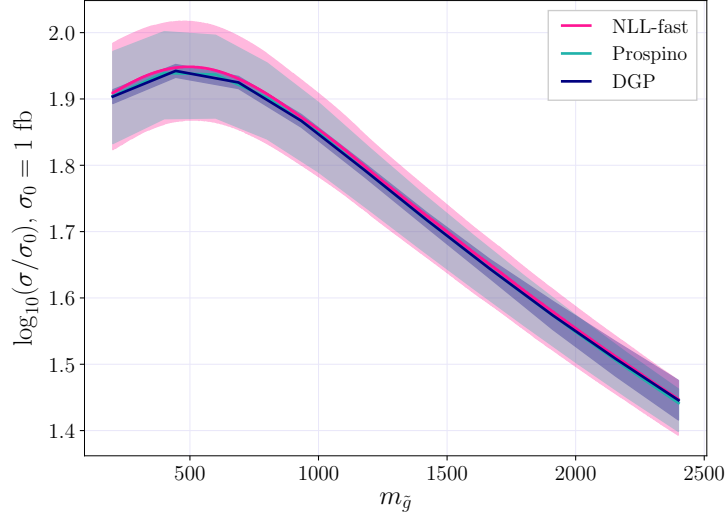
The saved models are very large. A single Gaussian process model trained with 2000 training points takes up 31 MB. Since one model is needed per process, this means that a relatively small model with a single expert with 2000 training points will require  $36 \times 31 \text{ MB} \approx 1.09 \text{ GB}$ . As seen from Table 1.2, the model size scales approximately as  $n^2$ , so the model size puts limits on the number of



**Figure 1.4:** Cross sections for  $\tilde{d}_L \tilde{d}_L$  and  $\tilde{u}_L \tilde{d}_L$ , using  $m_{\tilde{d}_L} = [200, 2400]$  GeV and all other masses set to  $m_i = 1000$  GeV generated by prospino (crosses) and predicted by the GP (lines with errors). The GP models used are for  $\tilde{d}_L \tilde{d}_L$  and  $\tilde{d}_L \tilde{u}_L$ .



**Figure 1.5:** Total cross sections for all 36 processes as a function of the  $\tilde{d}_L$ -mass as calculated by Prospino and NLL-fast, and estimated by Gaussian processes. The masses  $m_{\tilde{d}_L}$  and  $m_{\tilde{u}_L}$  are varied from  $[200, 2400]$  GeV while all other squark masses are held at 1000 GeV, and  $m_{\tilde{g}} = 500$  GeV.



**Figure 1.6:** Total cross sections for all 36 processes as a function of the  $\tilde{g}$ -mass as calculated by Prospino and NLL-fast, and estimated by Gaussian processes. The mass  $m_g$  is varied  $[200, 2400]$  GeV while all squark masses are held at 1000 GeV.

Training data size	Model size $[MB]$
2000	31
3000	69
5000	191

**Table 1.2:** Size of saved GP models with 3 or 4 features and the Matérn kernel.

training points an optimal model should have. There is very little difference in model sizes for models with 3 and 4 features, and a difference of 29 bytes for all training sizes between using the RBF kernel and the Matérn kernel. Presumably this is the space needed to store  $\nu$ .

### 1.3.1 Times and Kernels

### 1.3.2 Model Size

Hvor stor er den ultimate modellen?

Sjekk om man kan zippe en pickla GP, eller andre settinger

Oevre grense fra hva man kan gi til andre (ikke stoerre enn 10GB)

Hvor stor er det den kan vre?

Naar er modellen saa stor at den er upraktisk?  
How large is the optimal model?

### **Independent Models**

In this project 36 Gaussian process models were trained, with 10 experts each. The effective number of models may be reduced, however, as same-flavour same-chirality processes have identical function expressions as functions of the squark mass. This reduces the 36 models to 32.