Influence of Likelihood cut

April 29, 2015

Problem: Minimum of V_R not found at zero as expected ...

⇒ Forcing the -ln($\mathcal{L}$) to have a minimum between $-0.1/-0.05$ and $0.1/0.05$ can help!

1 Comparison between correct, wrong and un-matched jet combinations

The number of events in each of these categories is given in the following table:

Table 1: Grouping of the different jet-matching types for 10 000 ttbar semi-muonic (+) events.

Correctly matched	Wrongly matched	Unmatched
13 608	15 345	34 176
21.56 %	24.31 %	54.14 %

The top mass distributions for each of the categories are given in Figure 1. These distributions can give an idea which m_{top} value should be retrieved from the MadWeight calculations.

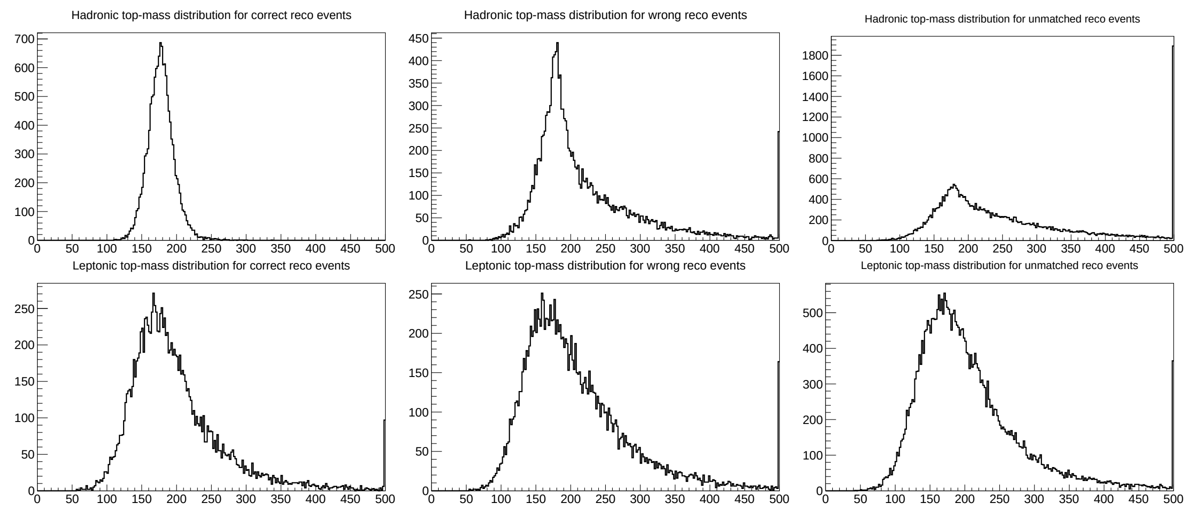


Figure 1: Distributions for the hadronically (upper) and leptonically (lower) decaying top quark mass for the correctly matched, wrongly matched and unmatched jet combinations, respectively.

1.1 Inefficiency of MadWeight depends on category-type ...

The number of remaining events which have been used for the measurements discussed further in this Chapter are given in Table 2. The dependence of the MadWeight inefficiency on the considered category can be understood from the large number of wrong event topologies in the wrongly matched and unmatched category.

Table 2: Number of events for each of the four considered categories succesfully calculated by MadWeight. The number of failing events, for which a weight equal to 0.0 has been returned or for which one of the considered configurations is missing, is especially significant for the category of unmatched reco-level events.

Category	Generator-level events	Reco-level events		
		Correctly matched	Wrongly matched	Unmatched
Succesful events (m_{top})	10000	9982	9085	7538
Succesful events ($Re(V_R)$)	10000	9995	9376	8059

2 Measurement of top-quark mass using Matrix Element Method

In order to check the influence of the event selection, first the measurement of the top-quark mass has been performed. The distributions for the generator-level events can be found in Figure 2 and Figure 3 . Both the -ln($\mathcal{L}$) distribution and the obtained mass distributions show the expected behavior.

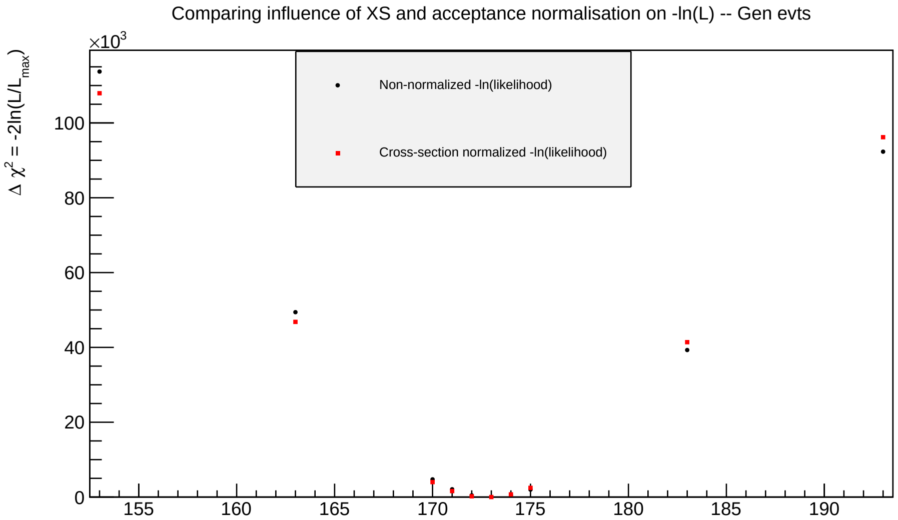


Figure 2: -ln($\mathcal{L}$) distribution for 10000 generator-level $t\bar{t}$ semi-mu (+) events. The minimum of the distribution seems to be located around the value used for simulating these events, namely 172.5 GeV.

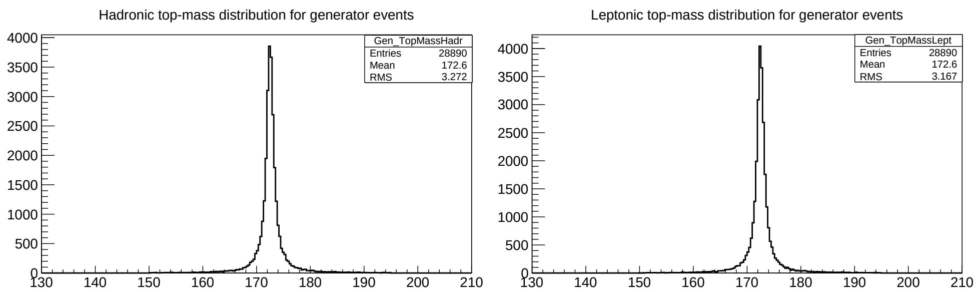


Figure 3: Distributions for the hadronically (left) and leptonically (right) decaying top quark for generator-level events.

Also for the generator-level events the -ln($\mathcal{L}$) distributions, given in Figure 4, can be studied. The -ln($\mathcal{L}$) distribution for the correctly matched jet combinations is the only one which follows the distribution obtained for the generator-level events. The two other distributions show a significant deviation of the position of the minimum indicating an important bias introduced by the applied event selection.

Fitting the -ln($\mathcal{L}$) distributions of the generator-level events and the three categories of reco-level events with a 2^{nd} degree polynomial gives a measurement of the m_{top} mass. The results of this fit is given in Table 3. Comparing the different m_{top} measurements for each of the categories and for the

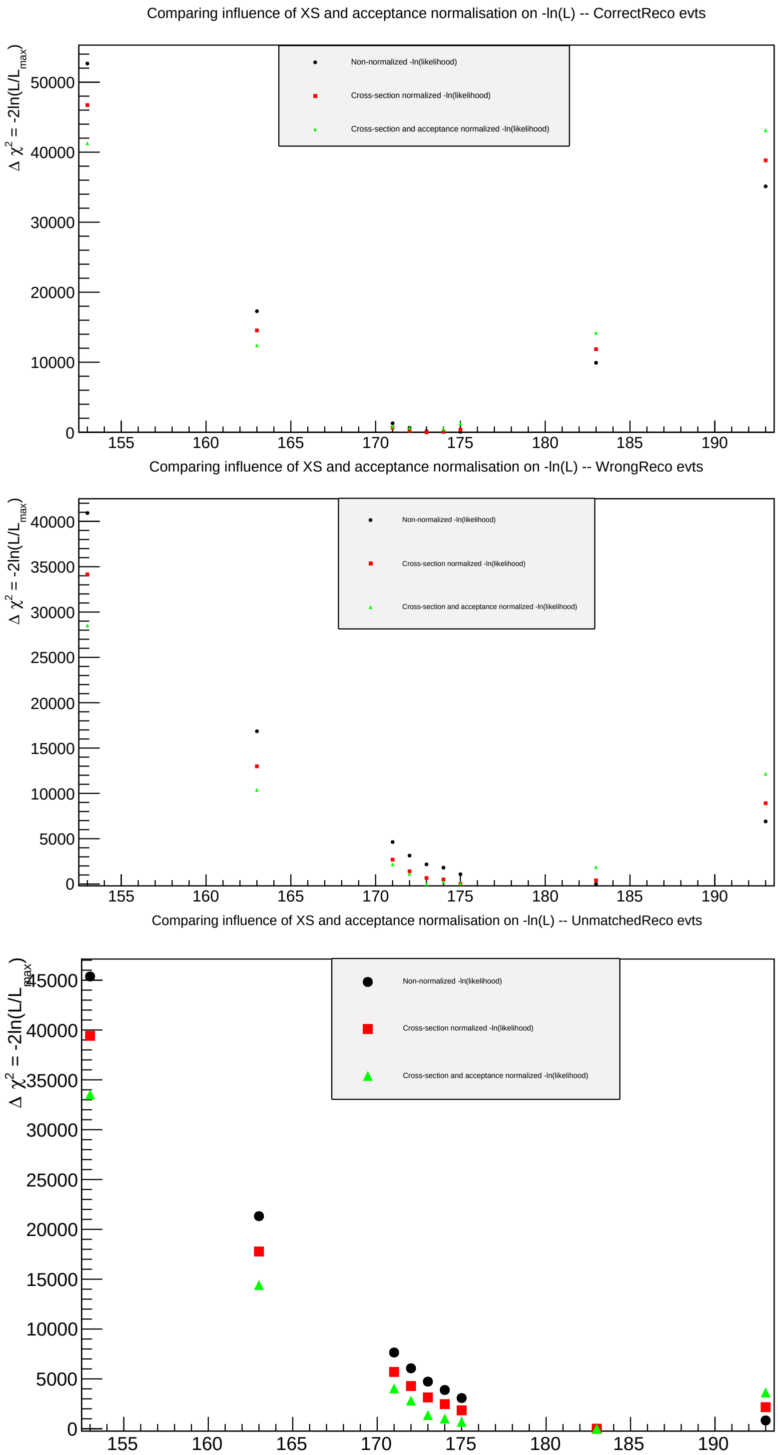


Figure 4: -ln($\mathcal{L}$) distributions for 10000 reco-level $t\bar{t}$ semi-mu (+) events, respectively correctly matched, wrongly matched and unmatched jet combinations. The position of the minimum for the correctly matched jet combinations still corresponds with the value used for simulating these events, while the wrongly matched or unmatched jet combinations significantly distort the agreement with the expected minimum position.

different normalisations applied clearly shows that applying both the cross-section normalisation and the acceptance normalisation significantly improves the measurement of the top-quark mass and brings it closer to the expected value.
Current results are given without uncertainties.

Table 3: Obtained measurement for m_{top} for the different categories considered. Fit on the obtained -ln($\mathcal{L}$) has been done using a 2nd degree polynomial.

	Applied normalisation on -ln($\mathcal{L}$)		
	None (GeV)	Cross-section (GeV)	Acceptance (GeV)
Generator level	$173.014^{+0.0442}_{-0.0442}$	$172.783^{+0.0441}_{-0.0441}$	X
Reco-level, correctly matched	175.10	173.84	172.58
Reco-level, wrongly matched	181.00	179.00	177.00
Reco-level, unmatched	185.40	183.80	181.80

2.1 Improvement of top-quark mass measurement by applying cuts on -ln($\mathcal{L}$)

In order to reduce the influence of the event selection the effect of requiring the second derivative to be positive (= -ln($\mathcal{L}$) is parabola with a minimum) is studied.

The results are summarized in Tables 7 and 8: first the efficiency of this cut has been given by showing the percentage of remaining events after applying this cut on the different categories and afterwards the obtained m_{top} value after the cut is given. For the calculation of this second derivative 5 different points have been studied with the middle point the expected SM value. Hence a distinction can be made whether the second derivative of the inner three points, the second derivative of the two outer ones with the middle point or both of the two should be positive.

Table 4: Percentage of remaining events for the four considered categories and three possible second derivative requirements. The numbers given here have been found by applying the above-mentioned cut on the -ln($\mathcal{L}$) obtained by running MadWeight on 10000 $t\bar{t}$ semi-mu (+) events. The number of successfully calculated events by MadWeight have been given before in Table 2.

	Events remaining after requiring 2^{nd} derivative > 0		
	Inner (%)	*Outer* (%)	*Both* (%)
Generator level	89.87	94.75	88.91
Reco-level, correctly matched	84.32	75.22	71.27
Reco-level, wrongly matched	73.31	67.58	59.87
Reco-level, unmatched	70.60	66.40	57.38

Table 5: Measured top-quark mass for the four considered categories and the three possible second derivative requirements compared to the mass measured originally (see Table 3). The fit has been applied on the acceptance normalised (cross-section normalised) -ln($\mathcal{L}$) for reco-level (generator-level) events.

	Original m_{top}	m_{top} after requiring 2^{nd} derivative > 0		
		Inner (GeV)	*Outer* (GeV)	*Both* (GeV)
Generator level	$172.783^{+0.0441}_{-0.0441}$	$172.730^{+0.0450}_{-0.0450}$	$172.768^{+0.0434}_{-0.0434}$	$172.724^{+0.0447}_{-0.0447}$
Reco-level, correctly matched	$172.849^{+0.0697}_{-0.0697}$	$172.694^{+0.0639}_{-0.0639}$	$172.919^{+0.0544}_{-0.0544}$	$172.860^{+0.0572}_{-0.0572}$
Reco-level, wrongly matched	$174.116^{+0.0664}_{-0.0664}$	$173.238^{+0.0518}_{-0.0518}$	$173.070^{+0.0353}_{-0.0353}$	$173.079^{+0.0412}_{-0.0412}$
Reco-level, unmatched	$175.056^{+0.0695}_{-0.0695}$	$173.548^{+0.0520}_{-0.0520}$	$173.260^{+0.0334}_{-0.0334}$	$173.268^{+0.0389}_{-0.0389}$

From this the large gain of requiring the -ln($\mathcal{L}$) to have a minimum around the Standard Model value is clearly visible. And it doesn't seem to introduce an additional bias such that it can definitely be applied on the anomalous couplings measurement as well!

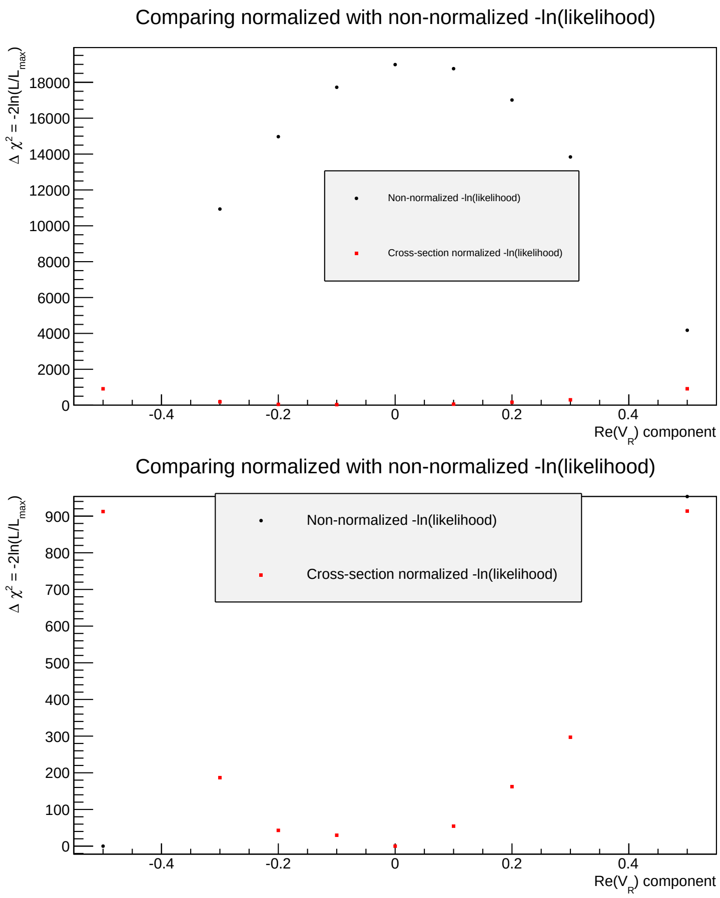
Only point which should still be considered (and which is probably more imporant for RVR measurement) is how the outer points are distributed with respect to the inner five points which are used for the fit ... This to have an idea whether the cut requirement results in weird-shaped events

3 Measurement of right-handed vector coupling, V_R, using Matrix Element Method

Applying exactly the same condition on the measurement of V_R!

In the case of the V_R measurement the influence of the cross-section is even important to obtain the correct minimum for generator-level events. This can be seen in Figure 5 which first shows the non-normalized -ln($\mathcal{L}$) and then the one after applying this normalisation.

Figure 5: -ln($\mathcal{L}$) distribution for generator-level events

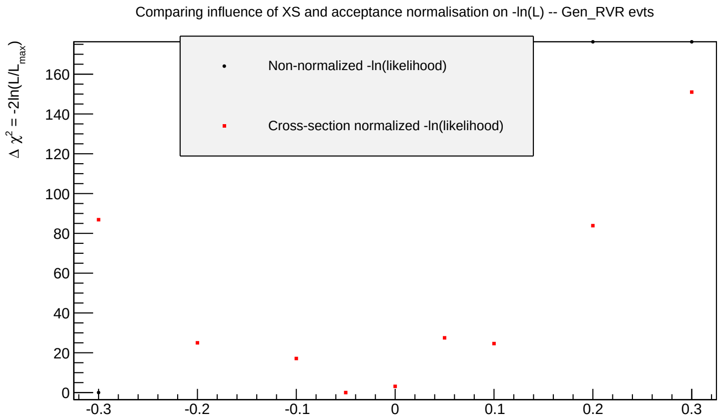


The same -ln($\mathcal{L}$) has also been obtained when the range of the V_R component is restricted between -0.3 and 0.3 because the value of ± 0.5 is too far from the expected value when taking into account the existing uncertainties on the value! This is given in Figure 6.

However it seems that the added points on ± 0.05 (for having enough data-points around the minimum) actually distort the smooth curve ... *Maybe this is caused by the influence of the XS which actually varies more than would be expected based on the uncertainty given by MadGraph ...*

Similar distributions (zoomed on normalized -ln($\mathcal{L}$)) for the reco-level events when considering the narrow $Re(V_R)$ range can be found in Figure 7. From these can be concluded that the influence of requiring the -ln($\mathcal{L}$) to have a minimum will probably have a very large effect... Since the shape currently looks a little bit random it seems to suggest that quite a lot of events will be discarded when applying this requirement!

Figure 6: -ln($\mathcal{L}$) distribution for generator-level events (narrow range of $Re(V_R)$)



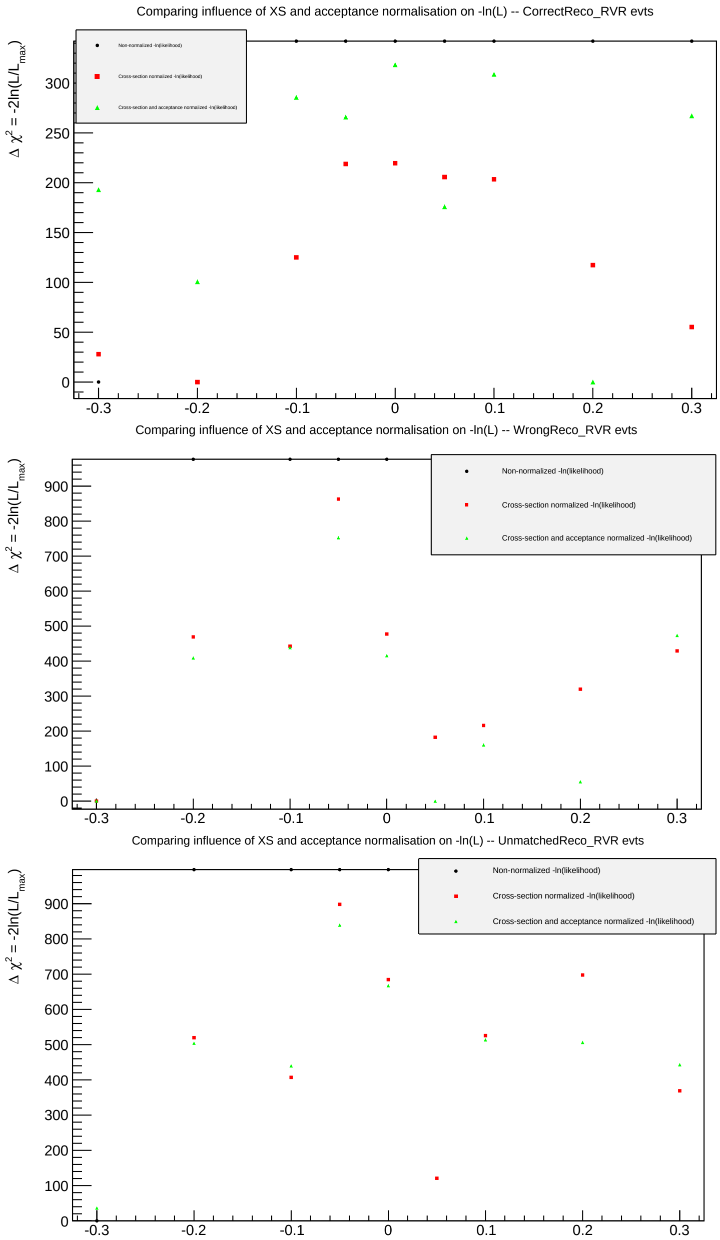


Figure 7: -ln($\mathcal{L}$) distributions for reco-level events. (narrow range of $Re(V_R)$)

3.1 Improvement when applying cuts on -ln($\mathcal{L}$)

Since it seems completely ridiculous to fit the reco-level -ln($\mathcal{L}$) distributions with a quadratic function, only an attempt to calculate the $Re(V_R)$ value using generator-level events has been done. This both for the full and narrow range of the coefficient as shown in Table 6.
Before any conclusion can be made from this table, the uncertainties on this $Re(V_R)$ value should be added ...

Table 6: Obtained measurement for $Re(V_R)$ for the two ranges considered for generator-level events. Again the fit on the obtained -ln($\mathcal{L}$) has been done using a 2nd degree polynomial. The full $Re(V_R)$ generator-level measurement is done using 20 000 events while the narrow one only uses 10 000 events.

	Applied normalisation on -ln($\mathcal{L}$)	
	None	Cross-section
Generator level (full $Re(V_R)$)		
Generator level (narrow $Re(V_R)$)		

Applying the -ln($\mathcal{L}$) -level event selection will be rather important for measuring the anomalous couplings coefficient. As expected, the number of events which gets excluded by this requirement is higher than for the m_{top} measurement.
From this it seems that applying a cut on the second derivative of $(-0.05/0.0/0.05)$ actually cuts away a higher percentage of events. However if this has a positive influence on the obtained measurement (*with uncertainties*) it can be interesting to apply this cut and only have a pure sample of signal events!

Table 7: Percentage of remaining events for the four considered categories and three possible second derivative requirements. The numbers given here have been found by applying the above-mentioned cut on the -ln($\mathcal{L}$) obtained by running MadWeight on 10000 $t\bar{t}$ semi-mu (+) events. The number of successfully calculated events by MadWeight have been given before in Table 2.

	Events remaining after requiring 2^{nd} derivative > 0		
	Inner (%)	*Outer* (%)	*Both* (%)
Generator level (full $Re(V_R)$)	56.27	59.74	48.51
Generator level (narrow $Re(V_R)$)	52.01	56.31	40.40
Reco-level, correctly matched	30.99	54.08	24.84
Reco-level, wrongly matched	37.13	38.85	28.53
Reco-level, unmatched	34.42	42.36	26.27

Table 8: The fit has been applied on the acceptance normalised (cross-section normalised) -ln($\mathcal{L}$) for reco-level (generator-level) events.

	Original $Re(V_R)$	$Re(V_R)$ after requiring 2^{nd} derivative > 0		
		Inner	*Outer*	*Both*
Generator level (full $Re(V_R)$, 20k Evts)	-0.05648 ± 0.020715	0.00703 ± 0.009937	0.00744 ± 0.008011	0.00808 ± 0.008944
Generator level (full $Re(V_R)$, 10k Evts)	x	0.00846 ± 0.013902	0.00704 ± 0.011307	0.00819 ± 0.012585
Generator level (narrow $Re(V_R)$)	-0.02990 ± 0.026524	0.00318 ± 0.012949	0.00470 ± 0.007933	0.00455 ± 0.009982
Reco-level, correctly matched	0.01389 ± 0.017817	0.00024 ± 0.010231	0.00433 ± 0.005542	0.00181 ± 0.007454
Reco-level, wrongly matched	0.12500 ± 0.000186	0.04724 ± 0.006188	0.01089 ± 0.002932	0.02162 ± 0.003857
Reco-level, unmatched	0.12500 ± 0.000304	0.01531 ± 0.006617	0.00551 ± 0.002625	0.00687 ± 0.003568

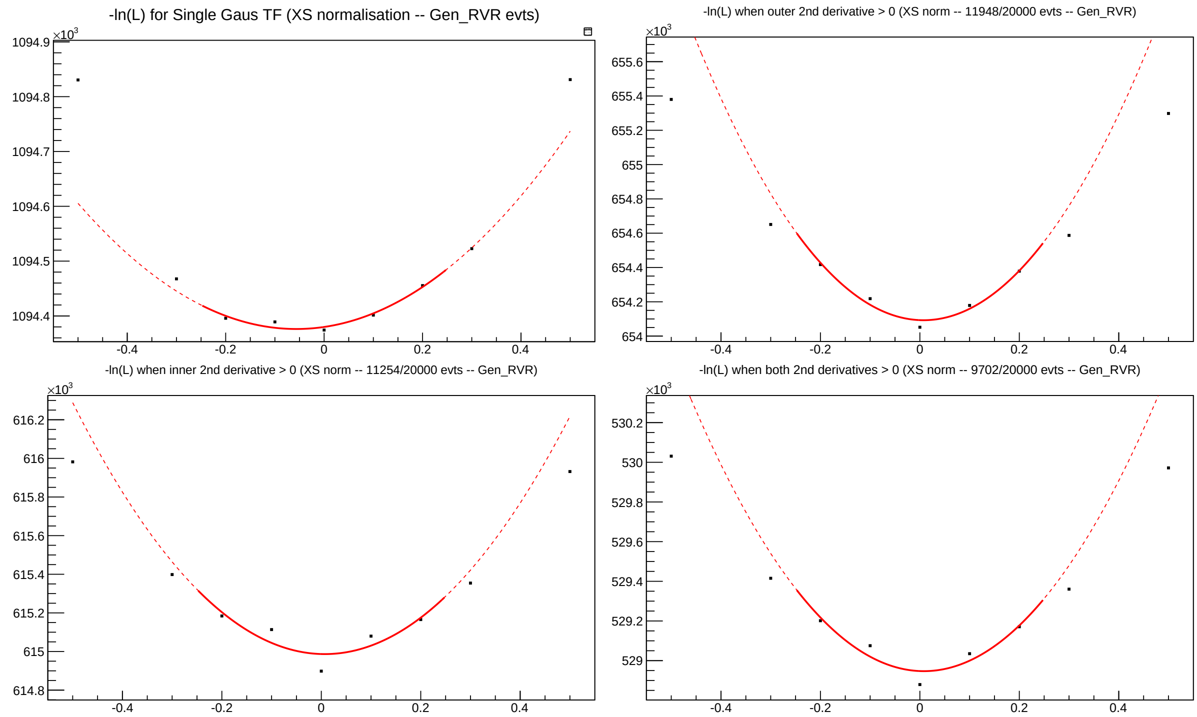


Figure 8: -ln($\mathcal{L}$) distributions for 20000 generator level events (looking at wide $Re(V_R)$ range, -0.5 to 0.5)

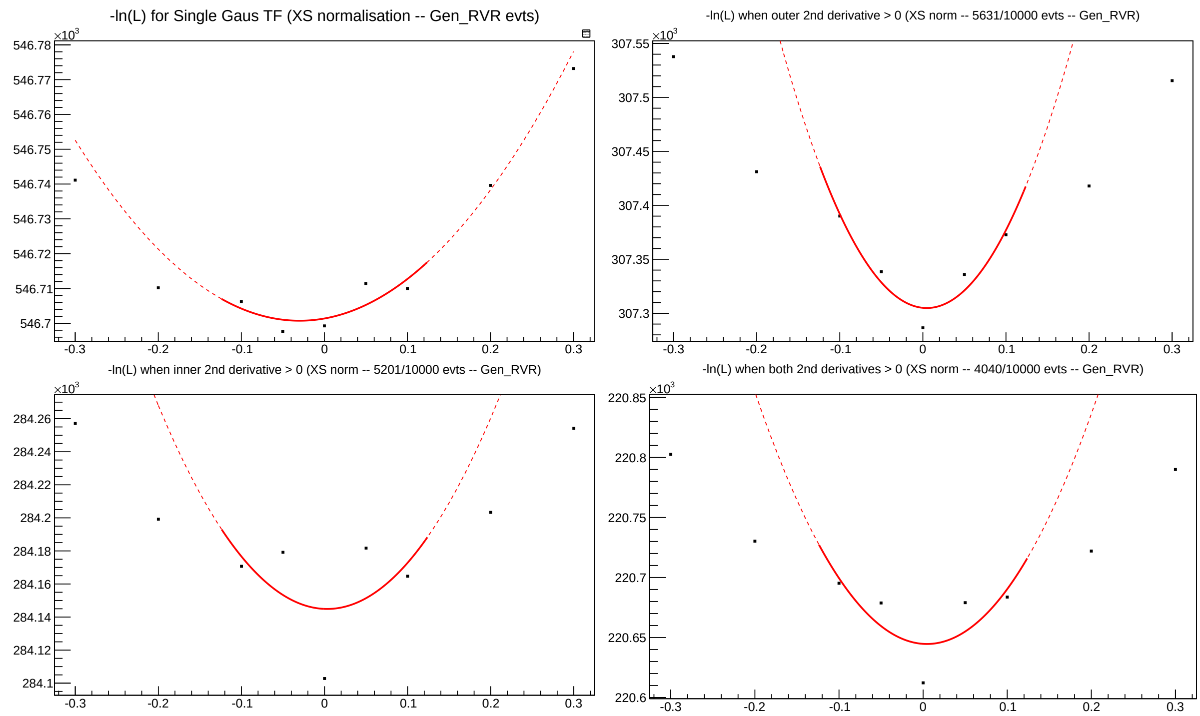


Figure 9: -ln($\mathcal{L}$) distributions for 10000 generator level events (looking at narrow $Re(V_R)$ range, -0.3 to 0.3)

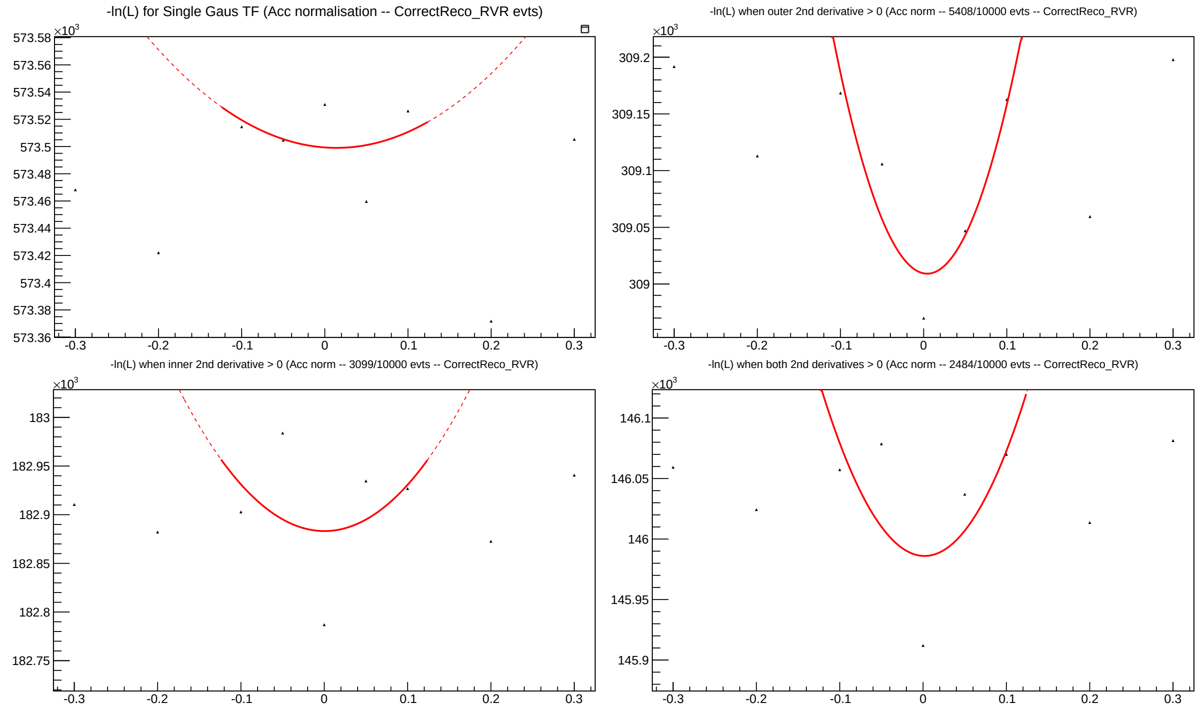


Figure 10: -ln($\mathcal{L}$) distributions for 10000 correctly matched reco-level events (looking at narrow $Re(V_R)$ range, -0.3 to 0.3)

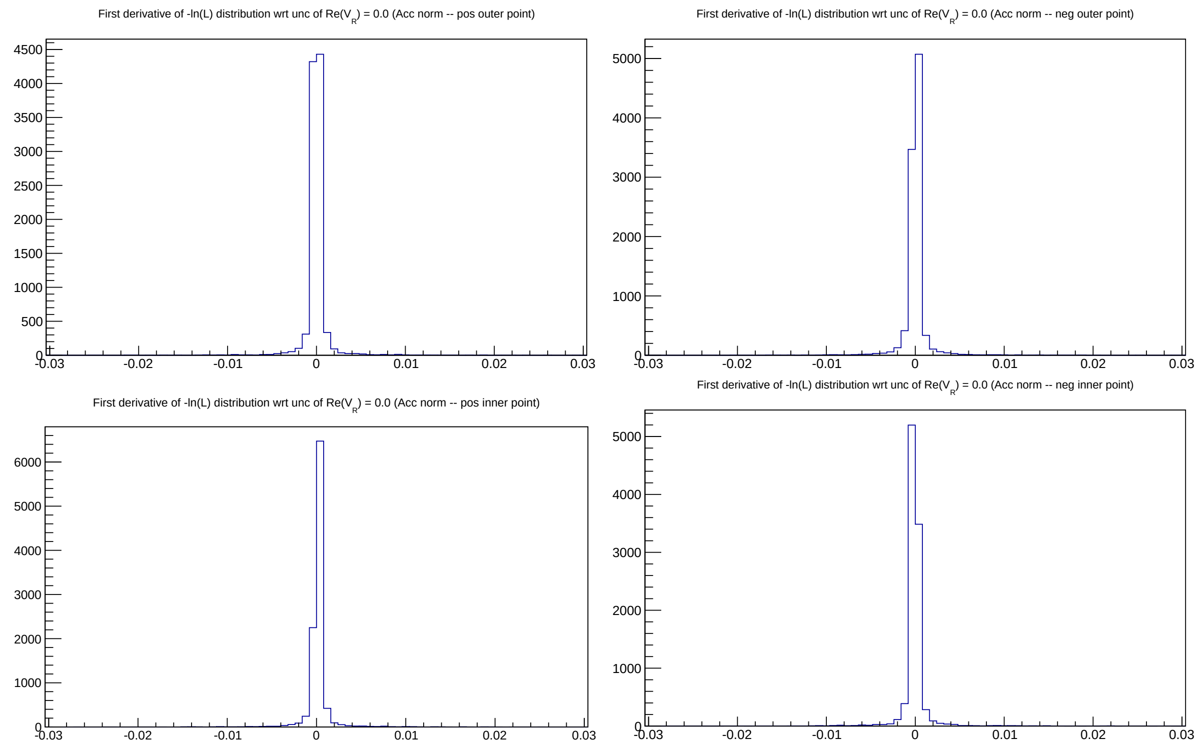


Figure 11: Relative derivative distribution for correctly matched reco events ((-ln($\mathcal{L}$) (pos/neg) - -ln($\mathcal{L}$) (min))/σ(-ln($\mathcal{L}$) (min)))

3.2 Likelihood distributions

3.3 Relative derivative distribution

4 Dependence on used cross-section values ...

The shape of the -ln($\mathcal{L}$) distribution is heavily influenced by the cross-section values used for the XS and Acc normalisation.

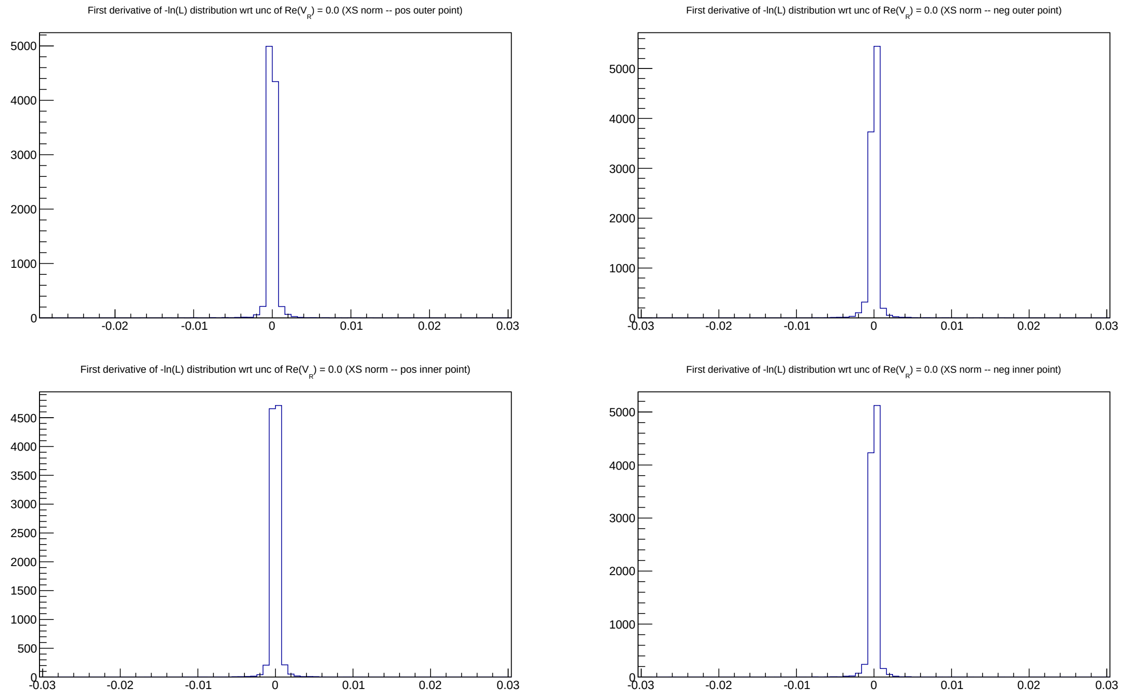


Figure 12: Relative derivative distribution generator level events ((-ln($\mathcal{L}$) (pos/neg) - -ln($\mathcal{L}$) (min))/σ(-ln($\mathcal{L}$) (min)))

For the acceptance normalisation the percentages can either be calculated using the original MadGraph files and apply the event selection using MadAnalysis or otherwise by calculating the cross-sections directly using the cuts in MadGraph.
Unfortunately both methods differ a little bit but are probably consistent within the statistical fluctuations. This because the difference in efficiency between the differen $Re(V_R)$ values is very low ... This is not the case for the top mass measurements. Figure 13 shows the event selection efficiency for both methods.

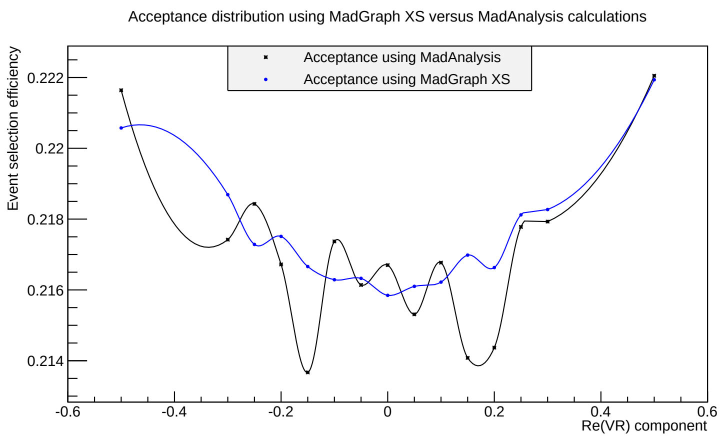


Figure 13: Event selection efficiency using either MadAnalysis or MadGraph.

5 -ln($\mathcal{L}$) variations on limited range

Looking at the $Re(V_R)$ interval of (-0.1, -0.09, -0.08, -0.07, -0.06, -0.05, -0.04, -0.03, -0.02, -0.01, 0.0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1) can help to see how the -ln($\mathcal{L}$) distribution fluctuates around the expected minimum. This interval corresonds very approximately to a 5σ interval using the

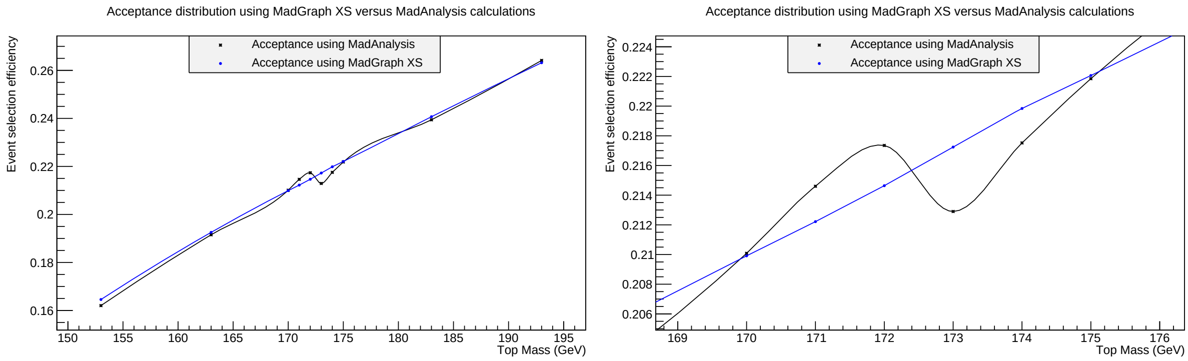


Figure 14: Event selection efficiency using either MadAnalysis or MadGraph.

preliminary results given above.
As can be seen from Figure 14 there is a significant influence from the acceptance on the -ln($\mathcal{L}$) distribution. This because the less narrow range is calculated using the MadAnalysis efficiencies while the narrow range used the MadGraph cross-sections.
However there can still be decided that the -ln($\mathcal{L}$) distribution fluctuates quite heavily inside the range of interest ...

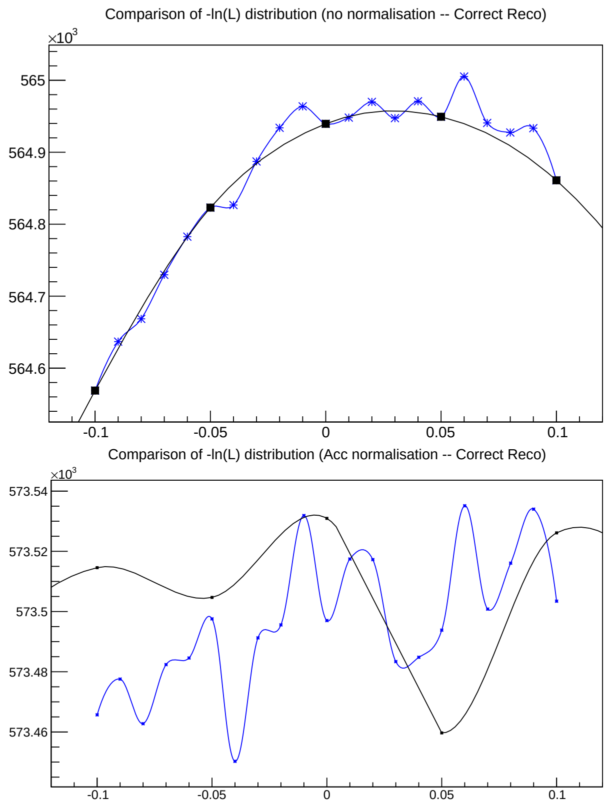


Figure 15: Variation of the -ln($\mathcal{L}$) distribution within a limited range. The obtained distribution on a less narrow range is also added in order to compare the -ln($\mathcal{L}$) distributions for different MadWeight calculations and get an idea of the importance of the event selection efficiency.