Impact of statistical fluctuations in the Monte Carlo on template fits

Matteo Defranchis[1], Jan Kieseler[2]

[1]Deutsches Elektronen-Synchrotron (DESY), [2]CERN

Introduction

Template fits to final state distributions are an essential tool in high energy physics analyses as they allow to constrain systematic uncertainties using the data. Such templates describe the variation of shape and normalisation of the distributions with respect to the uncertainties (nuisance parameters). This dependence is often derived by simulation and **the impact of statistical fluctuations in the Monte Carlo can lead to unphysical constraints of the nuisance parameters** and therefore to underestimating systematic uncertainties.

Here we illustrate a fully-consistent method to estimate the impact of this effect on template fit results. As an example we will take a CMS analysis where a template fit to final state distributions is performed to extract the $t\bar{t}$ cross section and the top quark MC mass in the final state with an electron and a muon. The analysis is currently being reviewed internally by CMS.

The method

In order to assess the impact of statistical fluctuations in the Monte Carlo, **toy experiments are performed where each bin in the templates is smeared according to a Poisson distribution** based on the effective number of MC entries in each bin.

Toy experiments are created **simultaneously for all the nuisance parameters while consistently taking into account the correlation between templates**. The template dependencies are then re-derived and the fit to the data points is repeated. The spread of the best fit values quantifies the effect of statistical fluctuations in the MC and is added as an additional uncertainty to the final result.

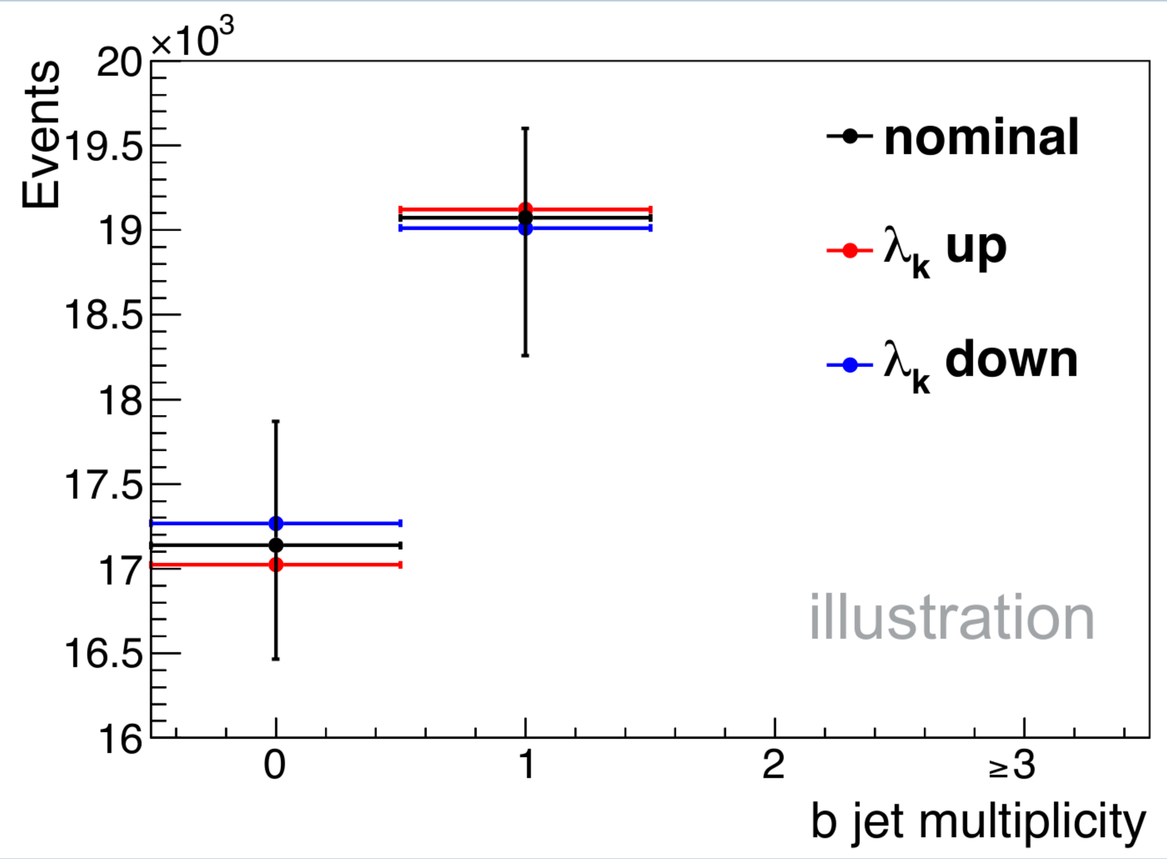
Conclusions

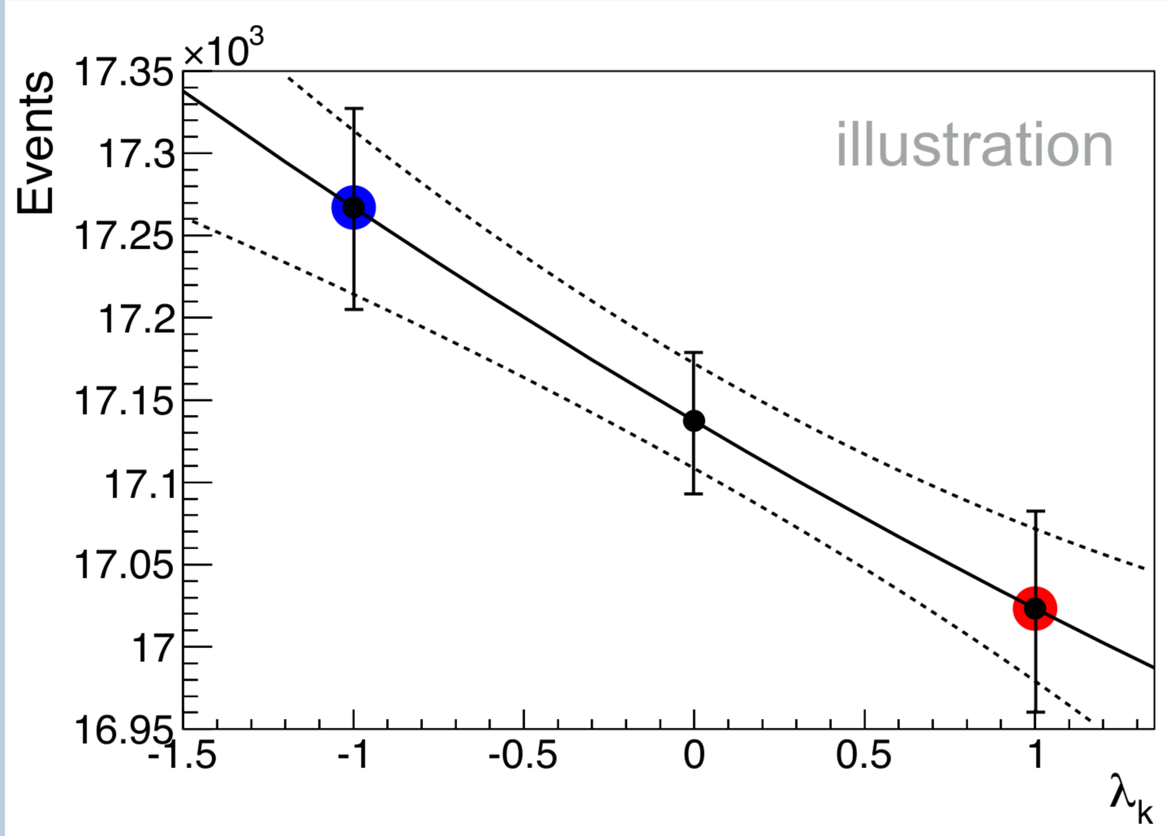
We have shown that statistical fluctuations in the Monte Carlo can play a significant role in template fits, using a realistic state-of-the-art analysis example.

In the measurement we considered, this leads to an additional uncertainty on the top quark mass of 350 MeV, which is a sizable contribution to the total uncertainty. The effect on the cross section is found to be small.

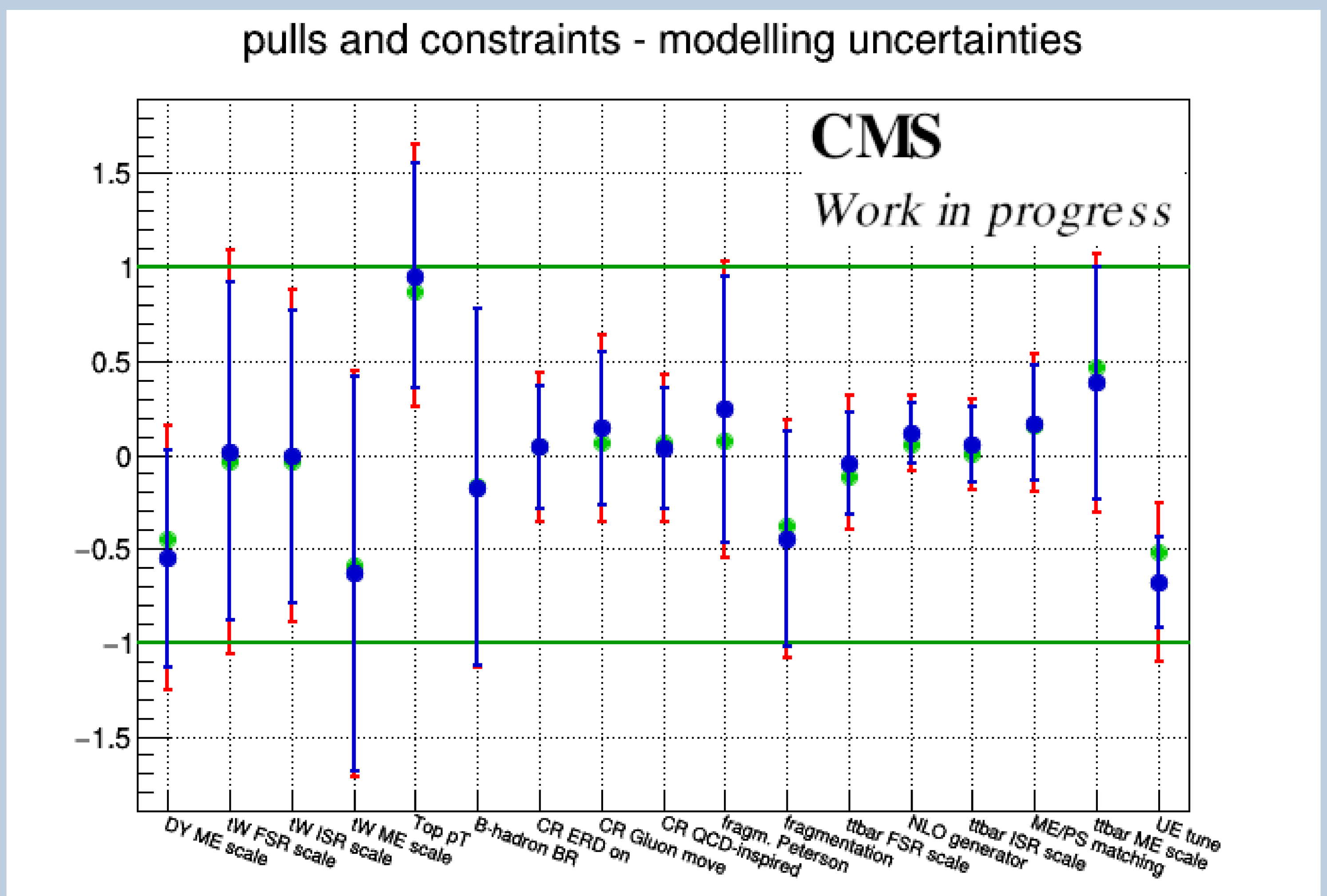
Template dependencies and statistical correlation

The template fit requires a continuous dependence of the template w.r.t. the nuisance parameter, which is typically derived from three points: "nominal", "up" and "down".





"Up" and "down" variations are determined either by event weights or by comparing to an alternative MC sample. In the first case, the statistical fluctuations are correlated to the nominal sample and the dependency can be derived exactly. In the second, the fluctuations are independent, resulting in a **large uncertainty in the dependence of each bin on the nuisance parameters**. This effect is typically not taken into account by standard procedures to estimate statistical uncertainty in the MC and **can lead to a significant underestimation of the uncertainty**.



blue dots: best fit value

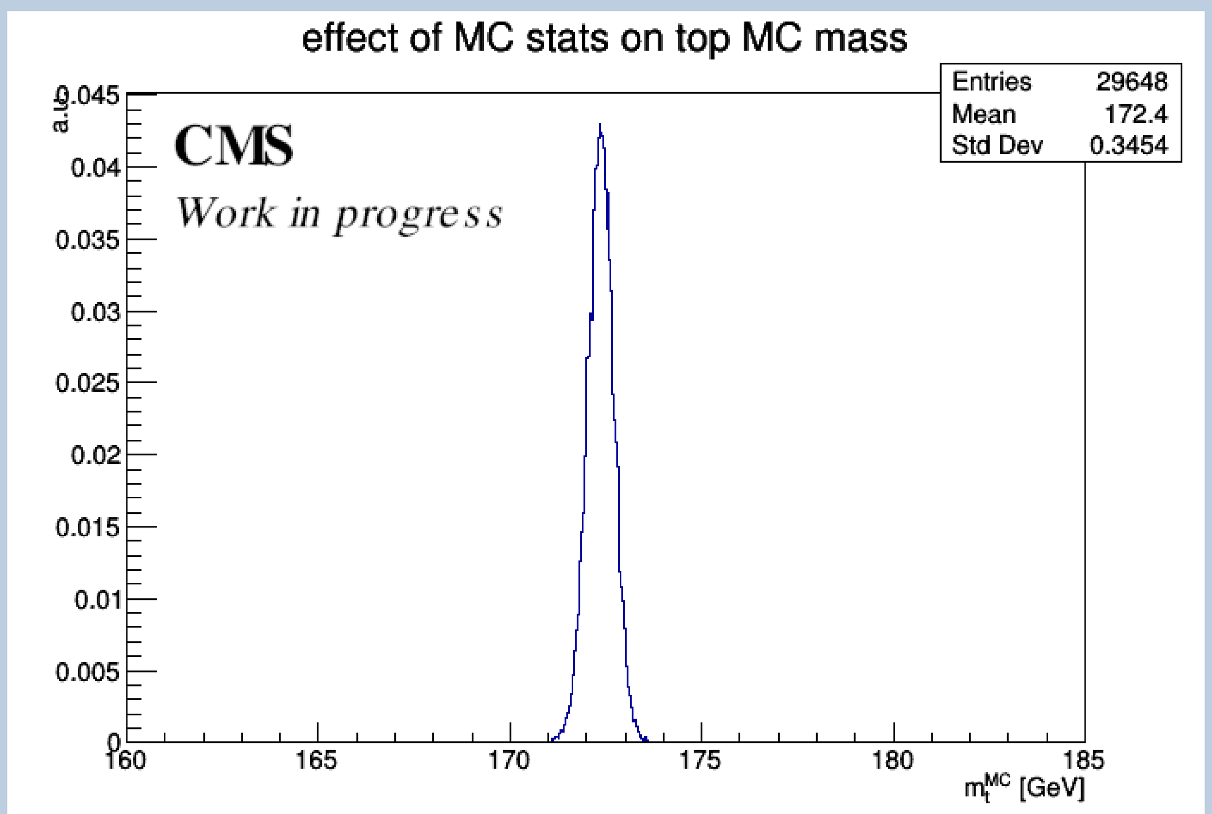
green dots: central value from toys (mean of distribution)

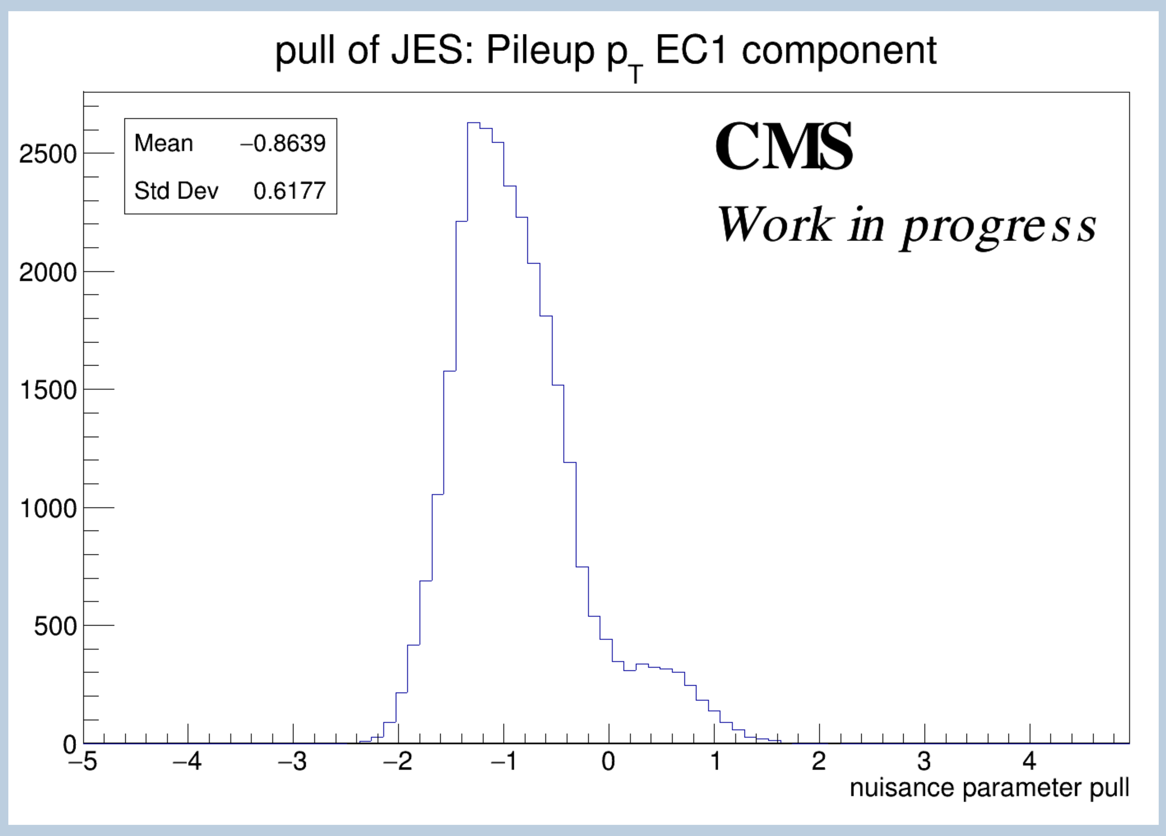
blue bands: fit constraint

red bands: additional uncertainty from toys (RMS of distribution)

Impact of Monte Carlo fluctuations on final results

Approximately 30 thousand toy experiments are performed to derive the distribution of the best fit values for the parameters of interest (m_t^{MC} and $\sigma_{t\bar{t}}$), both of which are found to be symmetric and with a quasi-Gaussian shape, with the central value corresponding to the best fit value of the nominal fit.





The RMS of the distributions is taken as an additional uncertainty:

top mass:	350 MeV	(sizable)
$t\bar{t}$ cross section:	9 pb	(small)