Investigation of Max likelihood fit and Expected Limit

Lisa Paspalaki

Expected Limit

- datacard
- Asymptotic approximation
- create a pseudo-data set where the data values are set equal to the expectation
- the data_obs distribution is blinded

To lnU or lnN ??

lnN:

- log normal → recommended choice for multiplicative corrections

lnU:

- log-uniform distribution **-->** ”set a large a-priori uncertainty on a given background and then rely on the correlation between channels to constrain it”
- this is done when pre fit normalization is not known

- **QCD_norm**: use the data driven ABCD to constrain by hand QCD_norm: lnN instead of lnU

- **ttbar_norm**: fit simultaneously all categories with ttbar control region
- lnN because we trust the MC prediction or lnU to let ttbar region constrain norm???

rate plus shape uncertainties

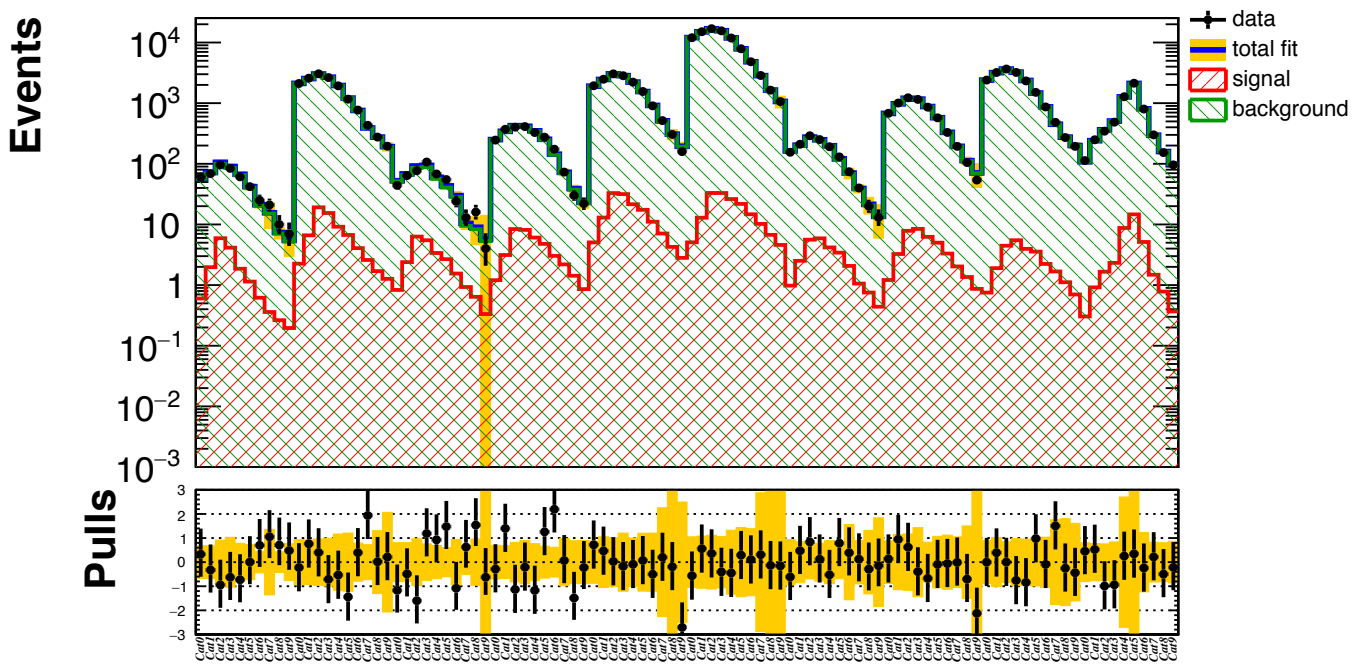
rate uncertainties

- affect the overall normalization
- nuisance: log-normal distribution (lnN)

shape uncertainties

- variations of expected event yields that are correlated among bins of a template
- variation of the underlying parameter by ±1 standard deviation
- template “morphing” procedure
- the normalization of a template is not conserved during morphing
- scale each template to the nominal

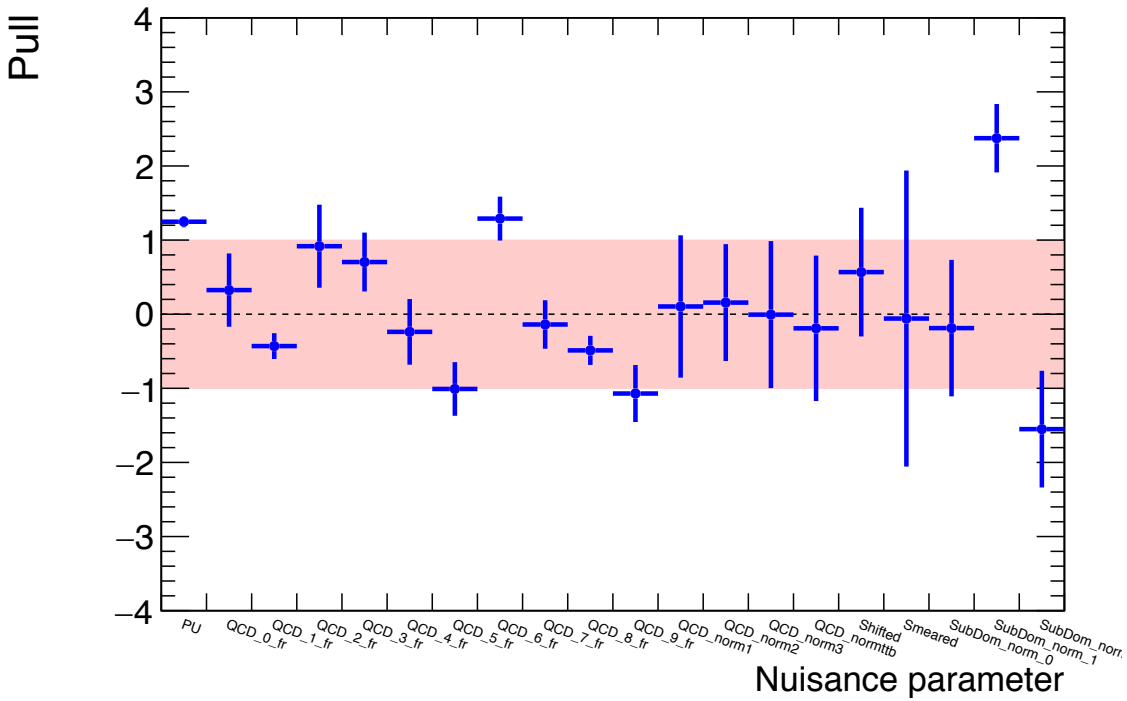
Post fit distributions



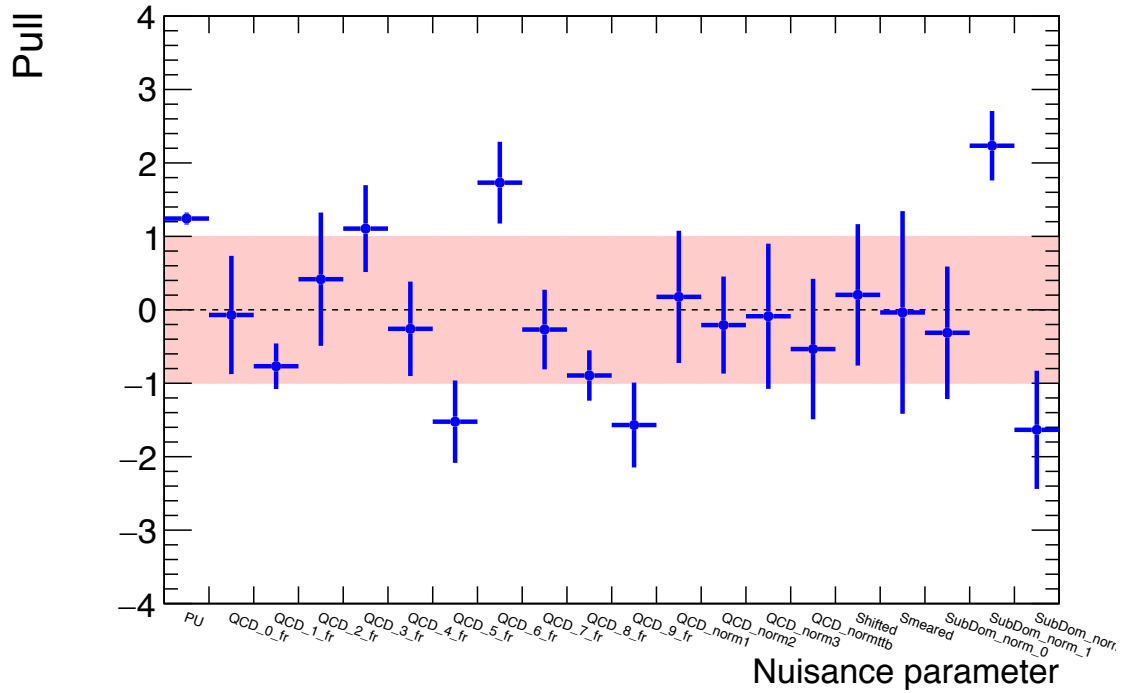
- binned max likelihood fit on the M_{SD} of the Higgs candidate
- signal and the control regions fitted to the data simultaneously
- data_obs template is the actual data distribution
- is it still **blinded**?

Pulls for QCD fraction

20% uncertainty



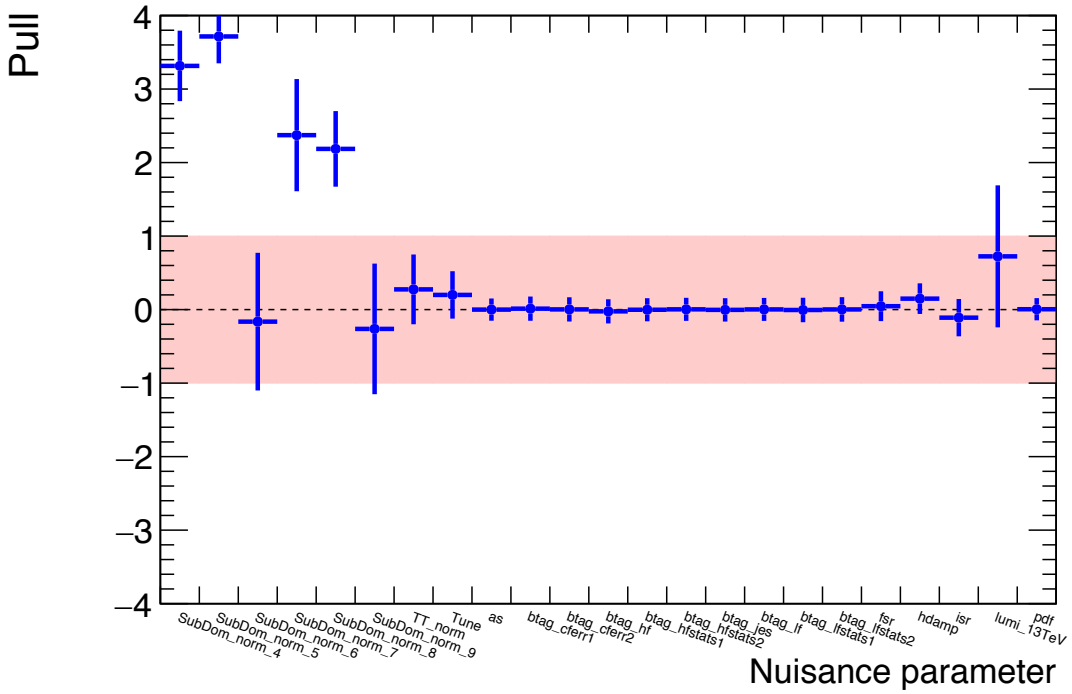
10% uncertainty



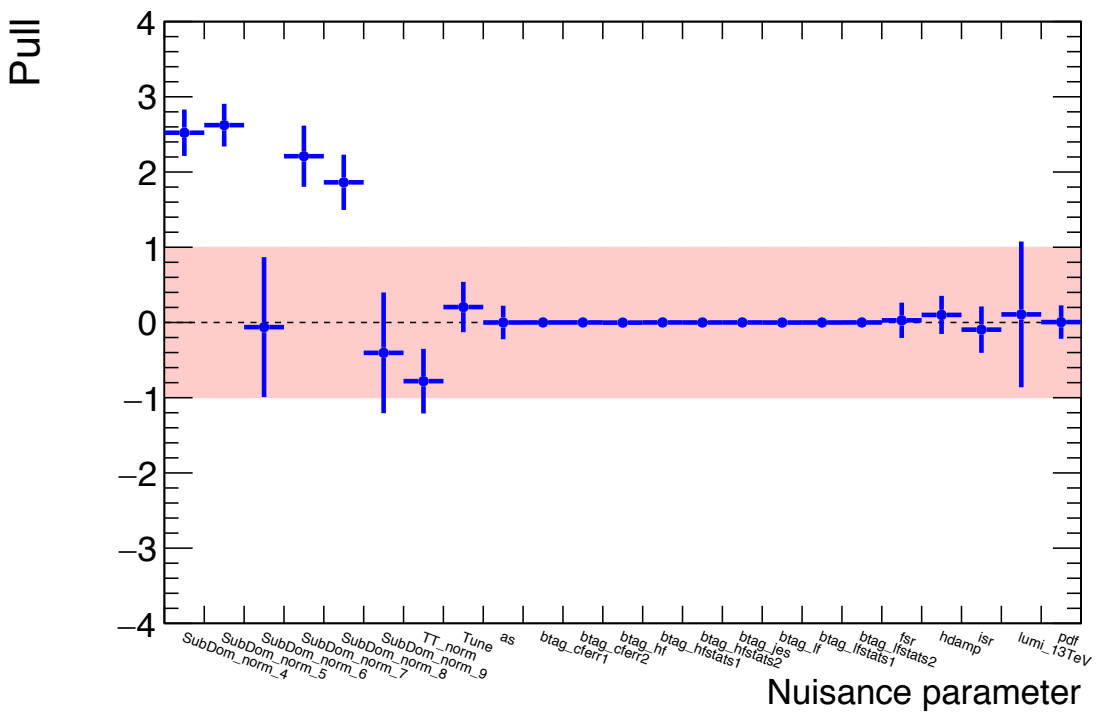
- if we increase the pre fit uncertainty on the QCD fraction the uncertainties are constrained by the fit
- expected limit is sensitive to this uncertainty: (10,7—12,..)

Pulls for Subdominant bkg

50% flat uncertainty

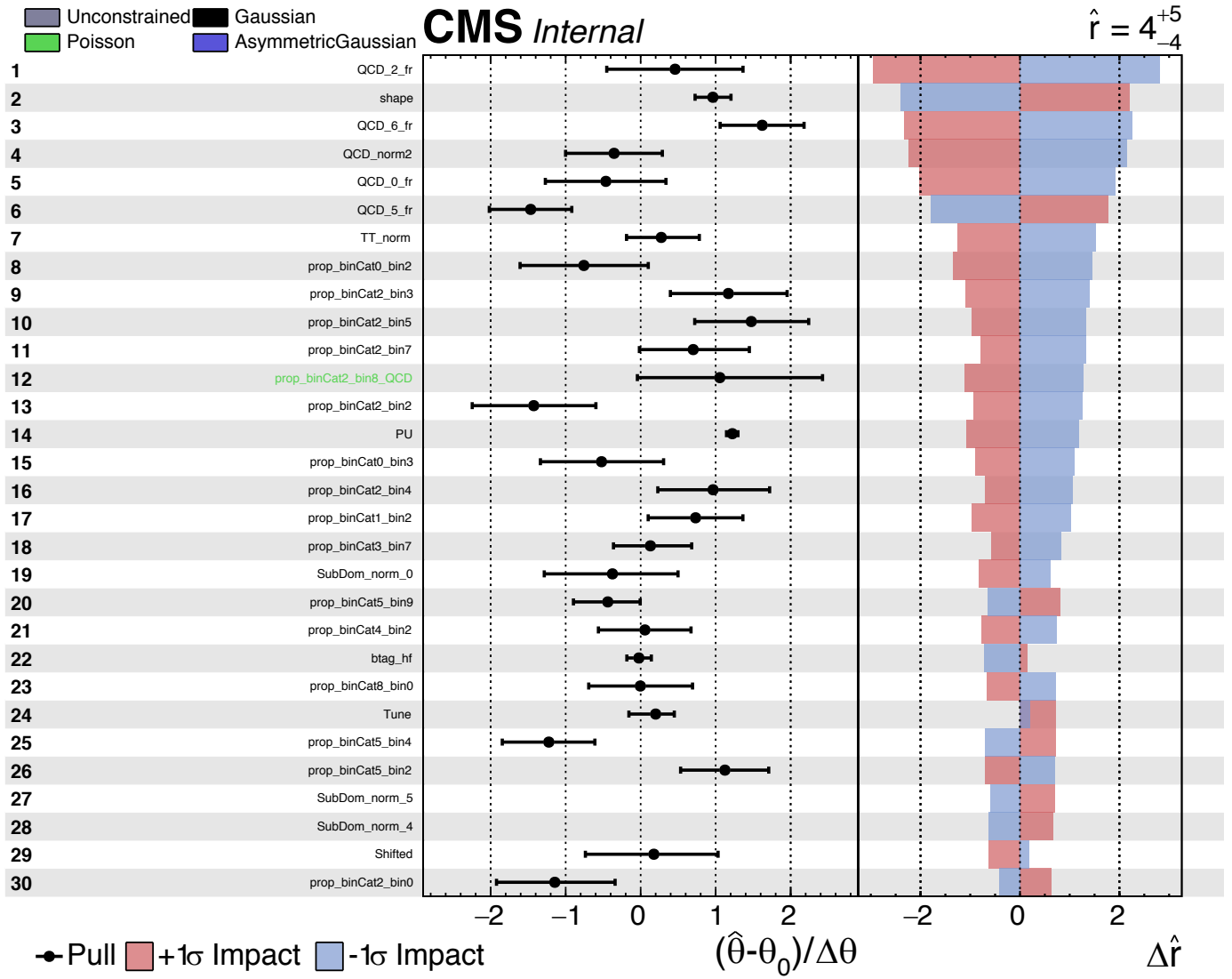


100% flat uncertainty



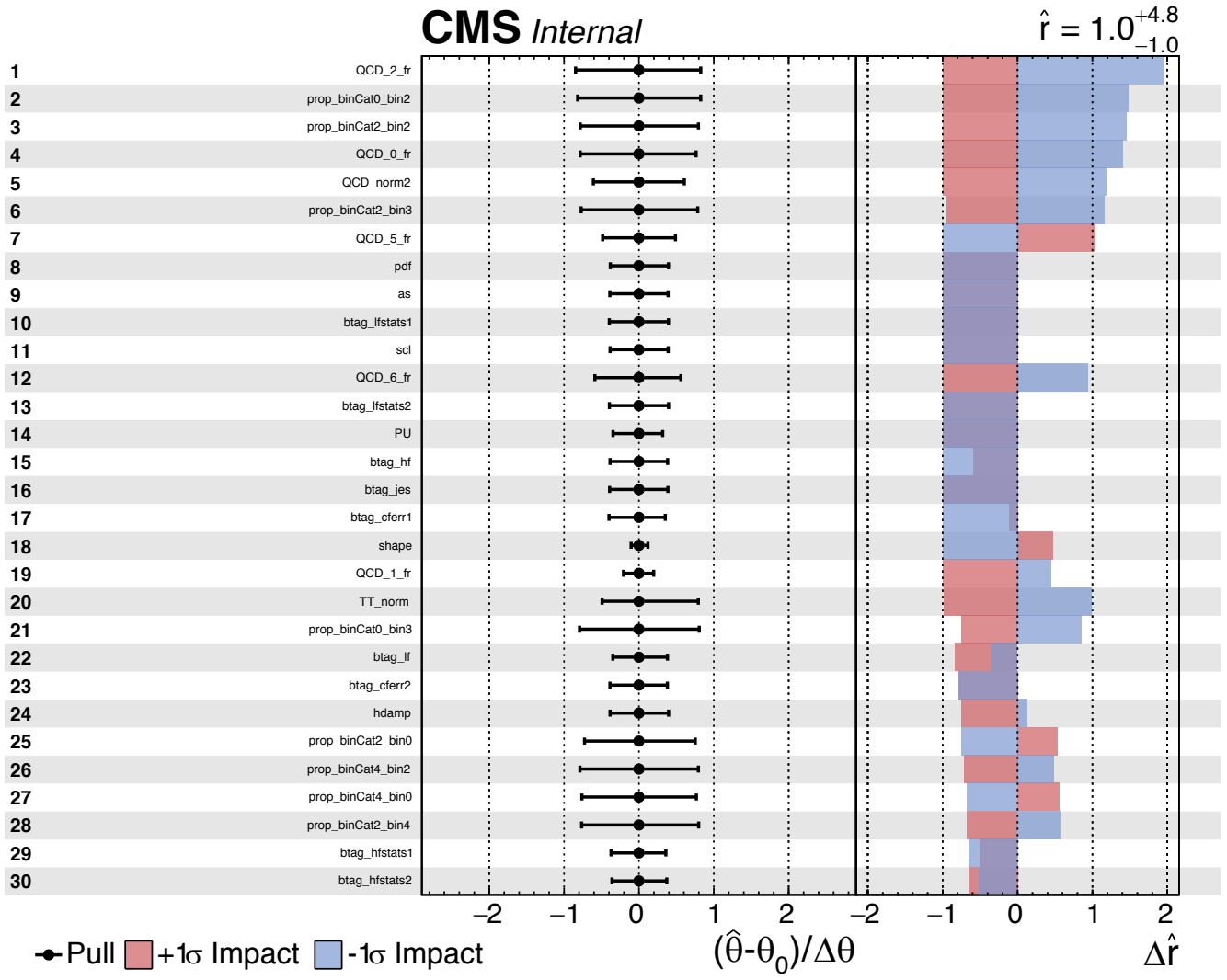
- not enough statistics on the templates
- even with 100% pre fit uncertainty the fit does not constrain this uncertainty

Impact of pulls



- **Post- fit values** of the nuisance parameters and their impacts on the extracted signal strength
- The pulls of the nuisance parameters (black markers with error bars) are computed as the differences between the best- fit values **θ** and the pre- fit values **θ0**, and then divided by the pre- fit uncertainties **Δθ**
- The impact $\boldsymbol{\Delta\hat{r}}$ for each nuisance parameter is computed as the difference between the best- fit signal strength value of the nominal fit, $\hat{r}$, and the best- fit value obtained when the nuisance parameter under investigation is fixed to its best- fit value plus/minus its post- fit uncertainty.

Impact of the Pulls for an Asimov dataset



Asimov dataset:
r is fixed to 1