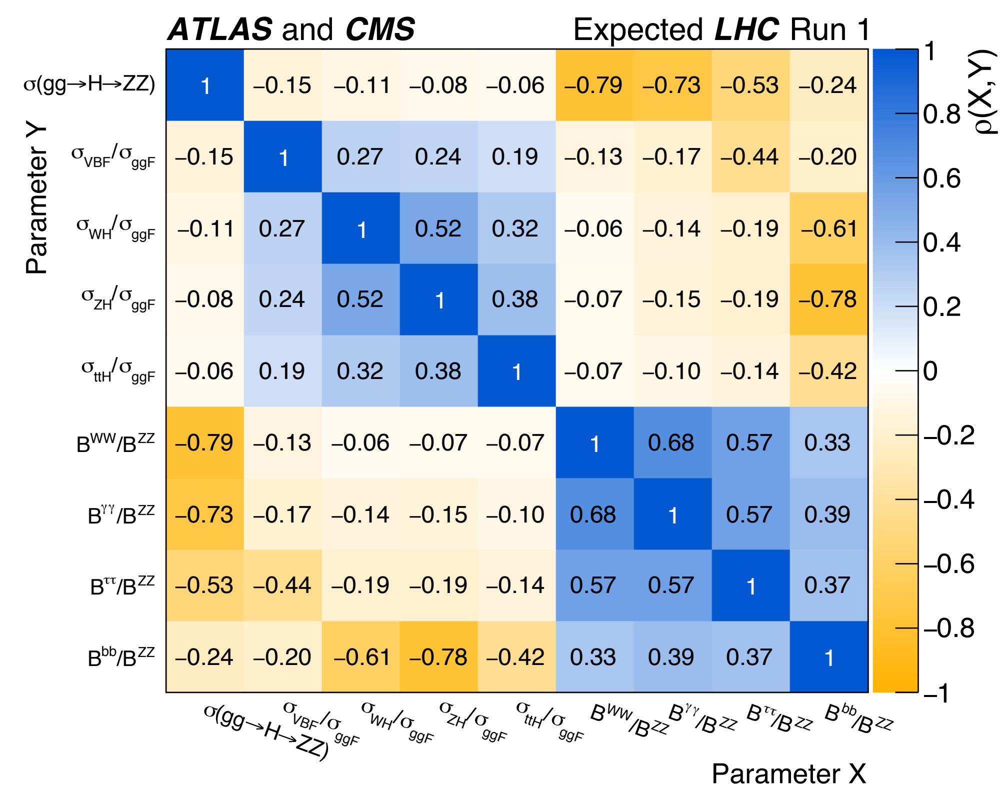
Taylor expanding the likelihood

A. Gilbert

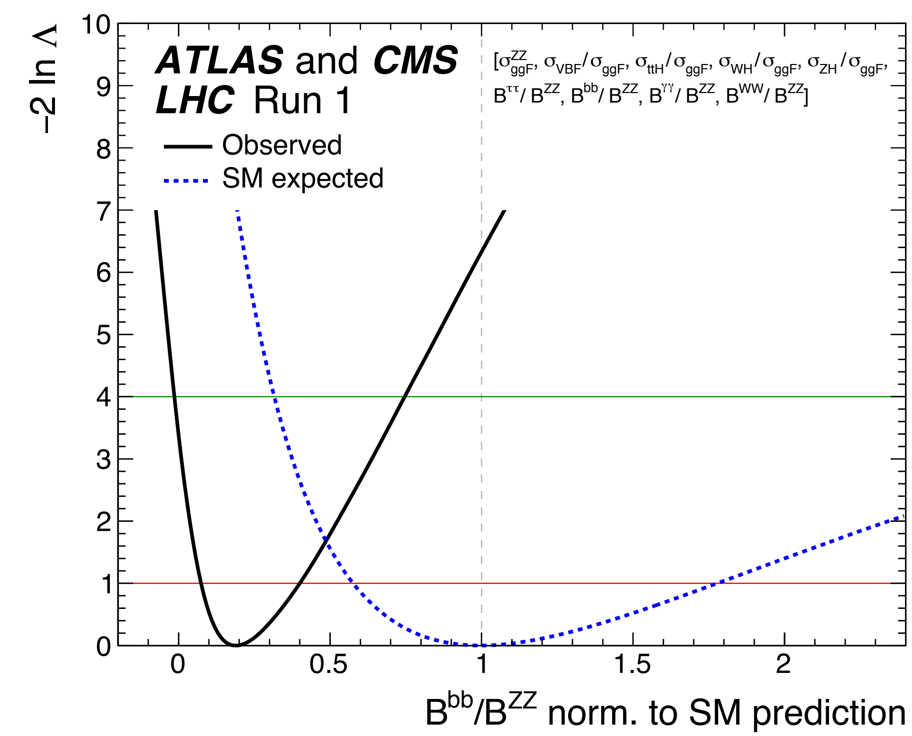
WG2 meeting | 23 November 2018

Background

- In the CMS+ATLAS Run-1 combination intended that results in the "generic" models would be used for recasting constraints to different models by the wider community (e.g. HiggsSignals)
  - Known that giving only the 1D NLL scans not sufficient
  - Correlation matrices also made public



- Use the covariance matrix to define a multi-variate Gaussian PDF as a proxy for the full likelihood
  - However found not to work well in several models - we are sensitive to the non-Gaussian nature of the NLL around the minimum (i.e. non-parabolic NLL scans



Taylor expansion

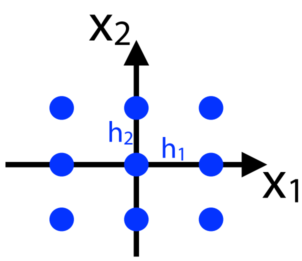
- Difficult to motivate an a-priori parametrisation of the NLL around the minimum
- Instead, Taylor expand the NLL function in the POIs
  - Parametrising with the covariance matrix is already equivalent to expanding up to second order
- General Taylor expansion in *d* variables:

$$T(x_1,\ldots,x_d) = \sum_{n_1=0}^{\infty}\cdots\sum_{n_d=0}^{\infty}\frac{(x_1-a_1)^{n_1}\cdots(x_d-a_d)^{n_d}}{n_1!\cdots n_d!}\left(\frac{\partial^{n_1+\cdots+n_d}f}{\partial x_1^{n_1}\cdots\partial x_d^{n_d}}\right)(a_1,\ldots,a_d)$$

- To calculate terms of the expansion up to arbitrary N need a way to calculate arbitrary mixed partial derivatives
  - Finite difference formulae available (and can be calculated) for arbitrary derivatives wrt. one parameter:

$$f''(\hat{x}) \approx \frac{f(\hat{x}-h) - 2\cdot f(\hat{x}) + f(\hat{x}+h)}{h^2}$$

  - Can then generate formulae for mixed derivatives recursively: e.g. $\frac{\partial f}{\partial x_1 \partial x_2} = \frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_2}\right)$



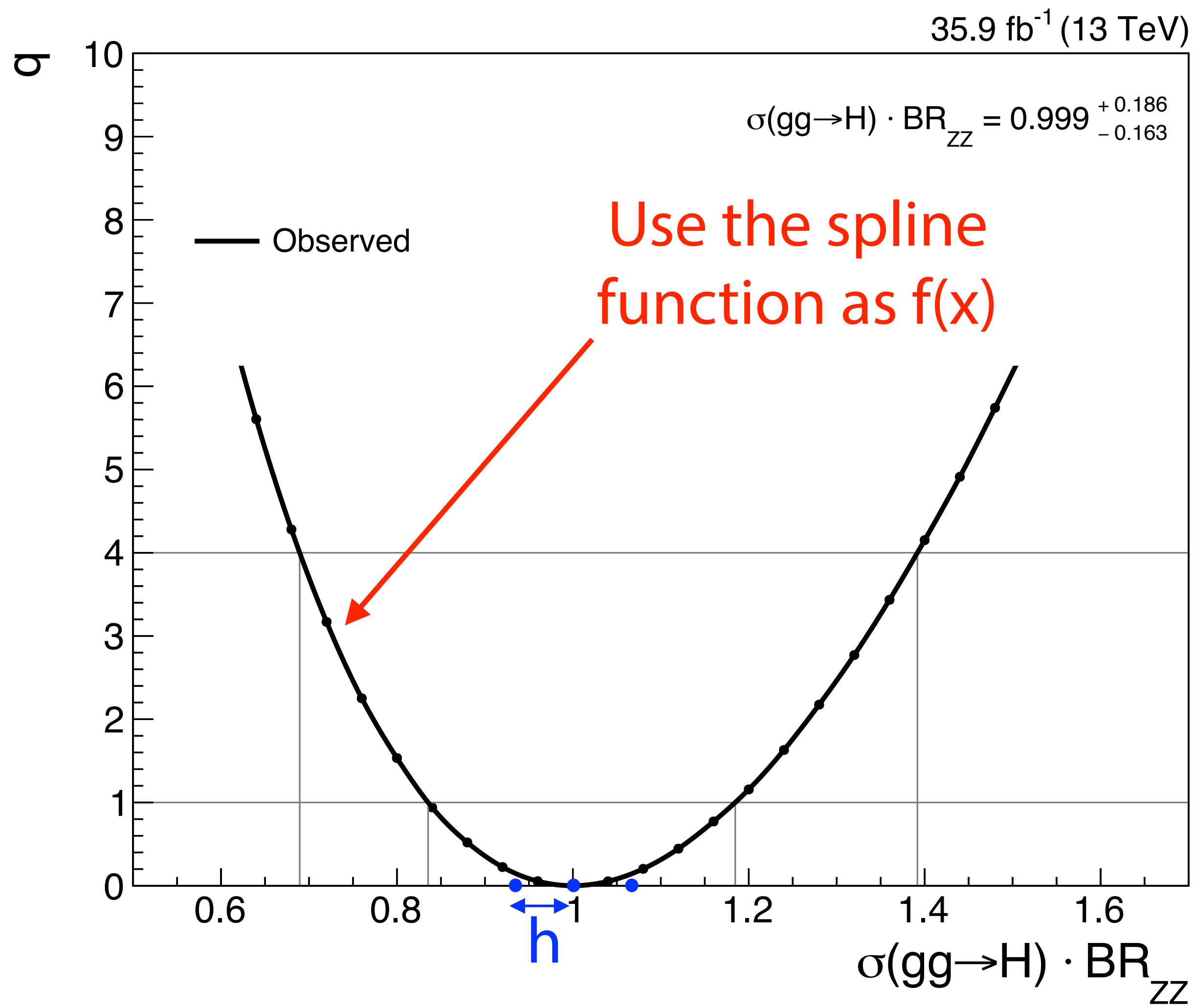
$$\frac{\partial f}{\partial x_1\partial x_2} \approx \frac{-\frac{\partial f}{\partial x_2}|_{\hat{x}_1-h_1} + 0\cdot\frac{\partial f}{\partial x_2}|_{\hat{x}_1} + \frac{\partial f}{\partial x_2}|_{\hat{x}_1+h_1}}{2h_1} = \frac{-\left[-f|_{\substack{\hat{x}_1-h_1\\ \hat{x}_2-h_2}} + 0\cdot f|_{\substack{\hat{x}_1-h_1\\ \hat{x}_2}} + f|_{\substack{\hat{x}_1-h_1\\ \hat{x}_2+h_2}}\right] + 0\cdot\left[-f|_{\substack{\hat{x}_1\\ \hat{x}_2-h_2}} + 0\cdot f|_{\substack{\hat{x}_1\\ \hat{x}_2}} + f|_{\substack{\hat{x}_1\\ \hat{x}_2+h_2}}\right] + \left[-f|_{\substack{\hat{x}_1+h_1\\ \hat{x}_2-h_2}} + 0\cdot f|_{\substack{\hat{x}_1+h_1\\ \hat{x}_2}} + f|_{\substack{\hat{x}_1+h_1\\ \hat{x}_2+h_2}}\right]}{4h_1h_2}$$

First tests

- Start with some simple proof of principle tests using the STXS-stage 0 model from the CMS 13 TeV combination (HIG-17-031)
- **Take a real NLL scan and treat this as a 1D function to expand**
  - Goal is to show we can accurately capture shape of the curve, and demonstrate the effect of different stencil choices
- Start with a 2nd order expansion. Calculate f''(x) using the simplest stencil:

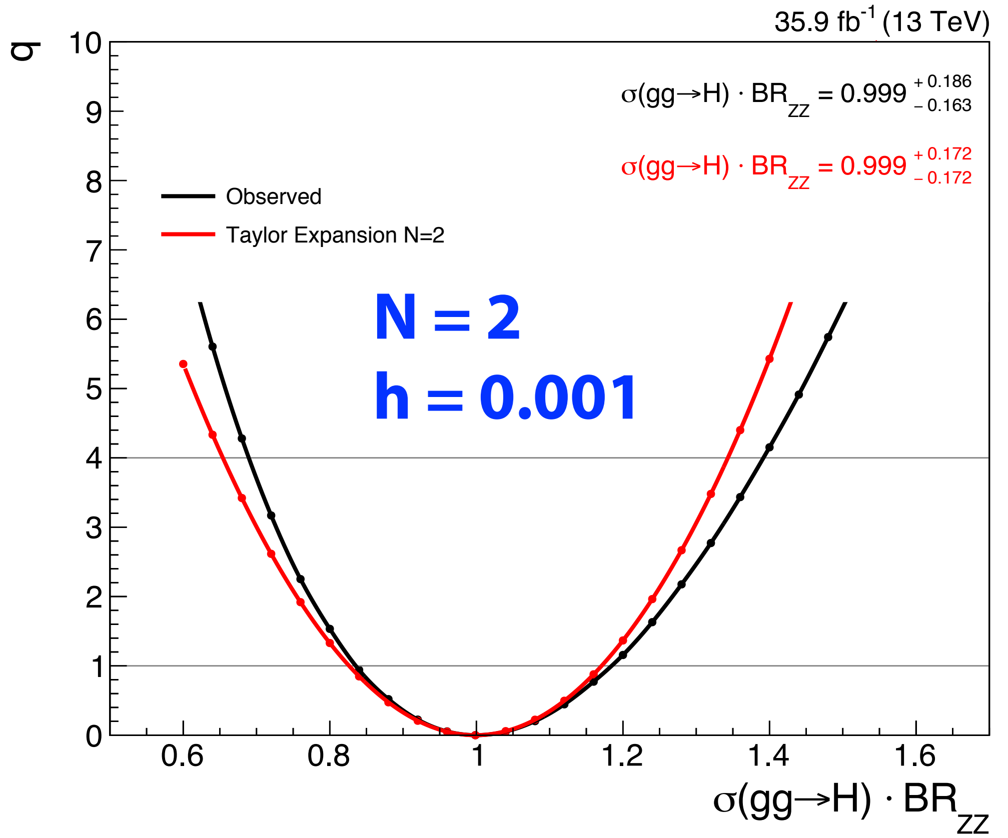
$$f''(\hat{x}) \approx \frac{f(\hat{x}-h) - 2 \cdot f(\hat{x}) + f(\hat{x}+h)}{h^2}$$

- This approaches f''(x) as h→0, so we could start with some small value of h

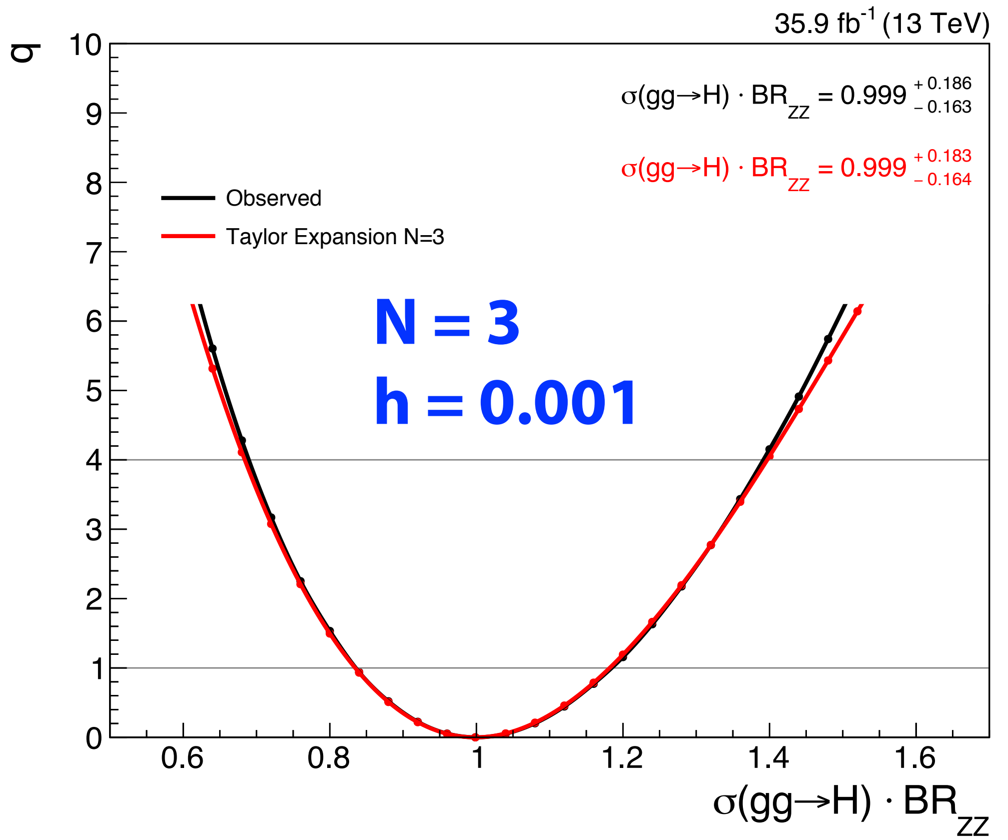


- Can express concisely as stencil points, {-h, 0, +h} and corresponding weights, {1, -2, 1}. Possible to generate weights for <u>arbitrary stencil points</u>

Expansion with small h

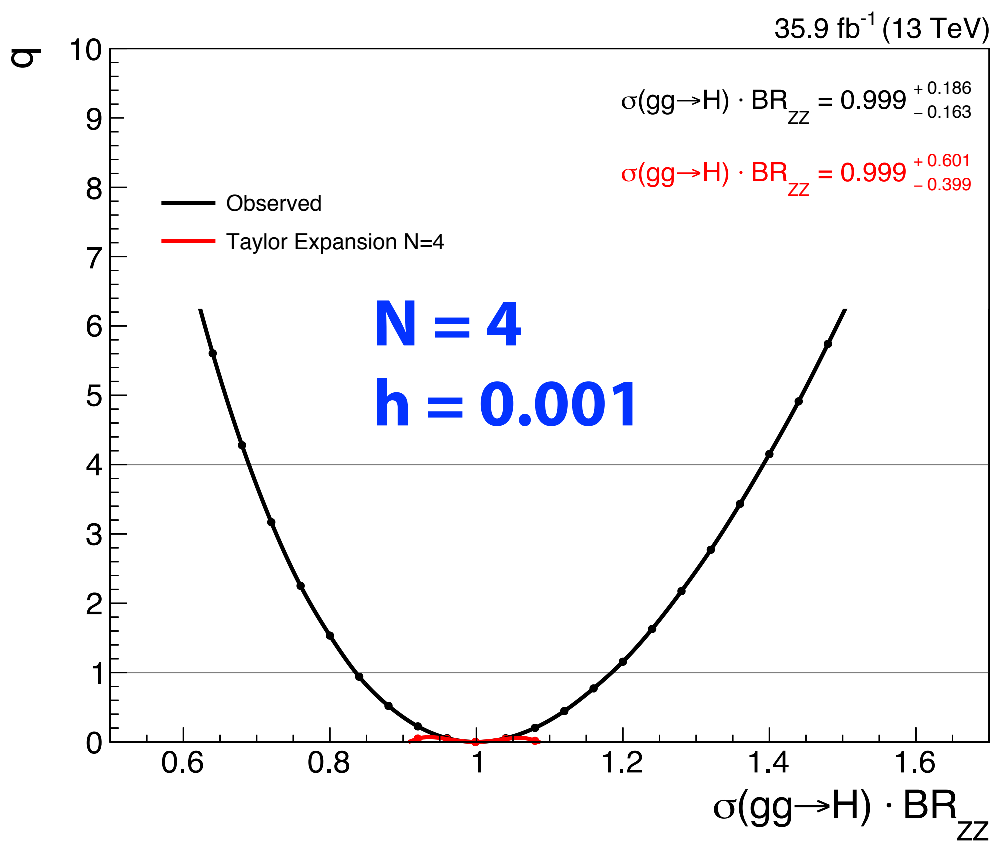


- Evaluating the expansion up to N=2 gives what we expect, a symmetrised version of the scan, where the 1σ interval is roughly the average of the actual asymmetric interval

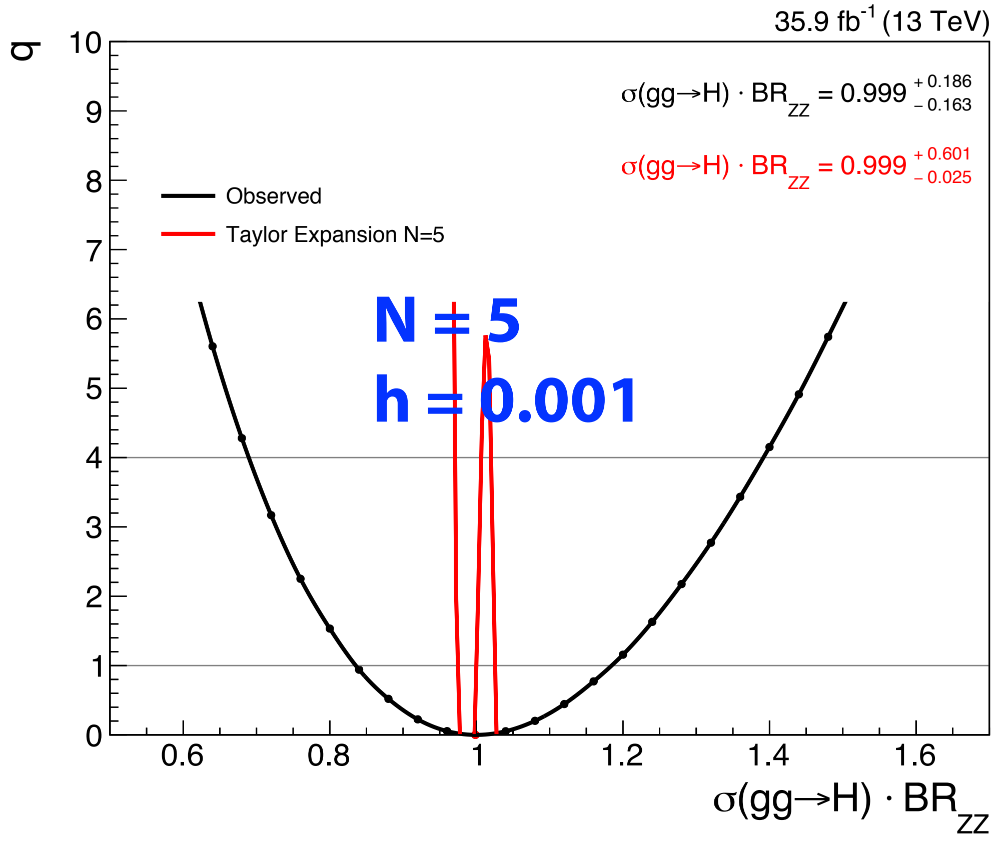


- For N=3 need to use at least a 4-point stencil. To keep it symmetric use a 5-point one:
  - {-h, -h/2, 0, +h/2, +h}
- This already shows much improved agreement, could call it "good enough" up to around the 2σ level.

Expansion with small h

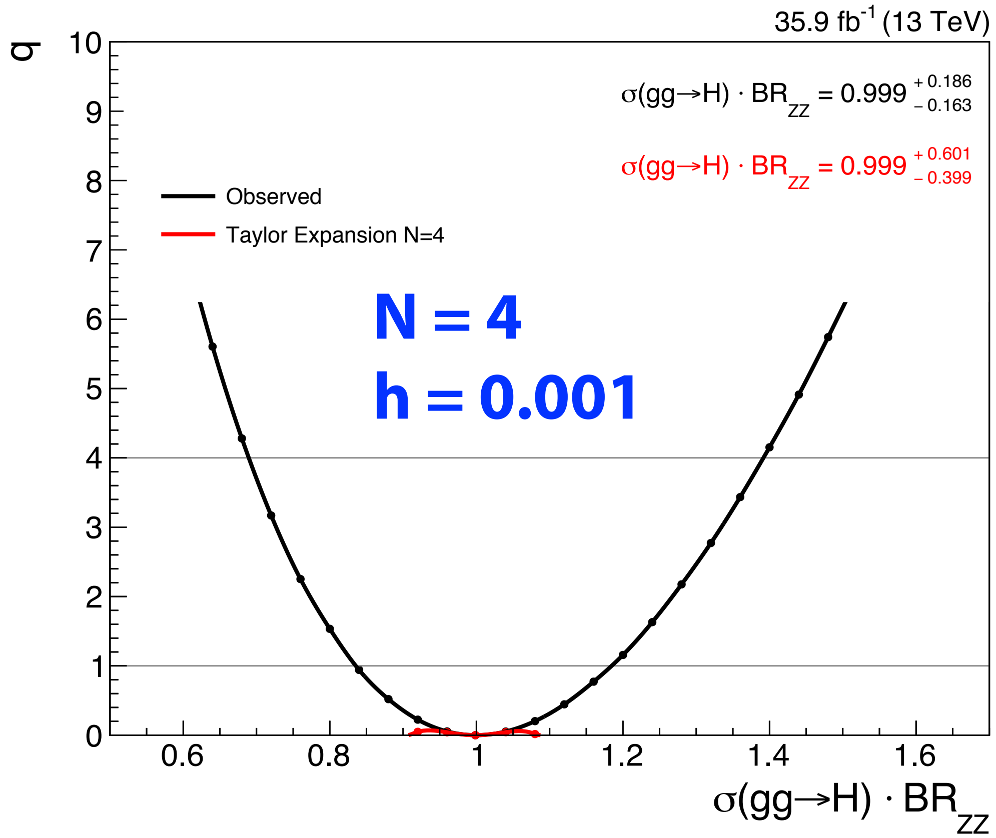


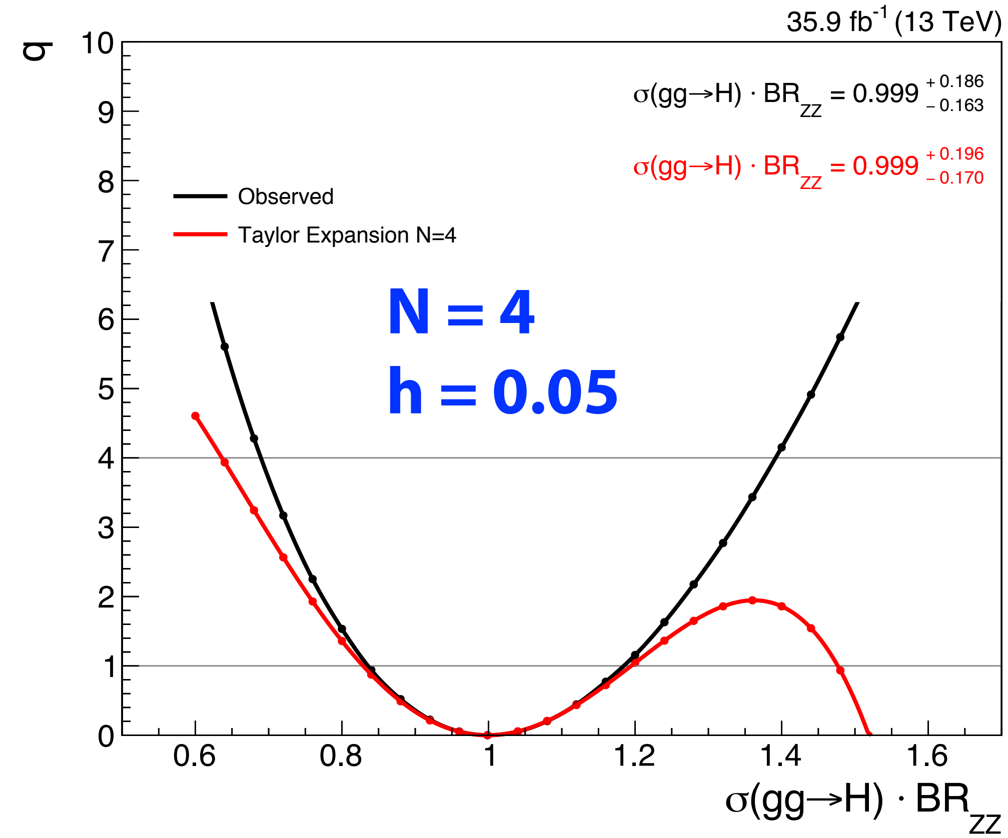
- At N=4 and above it fails: the expansion quickly diverges from the true value of the function.

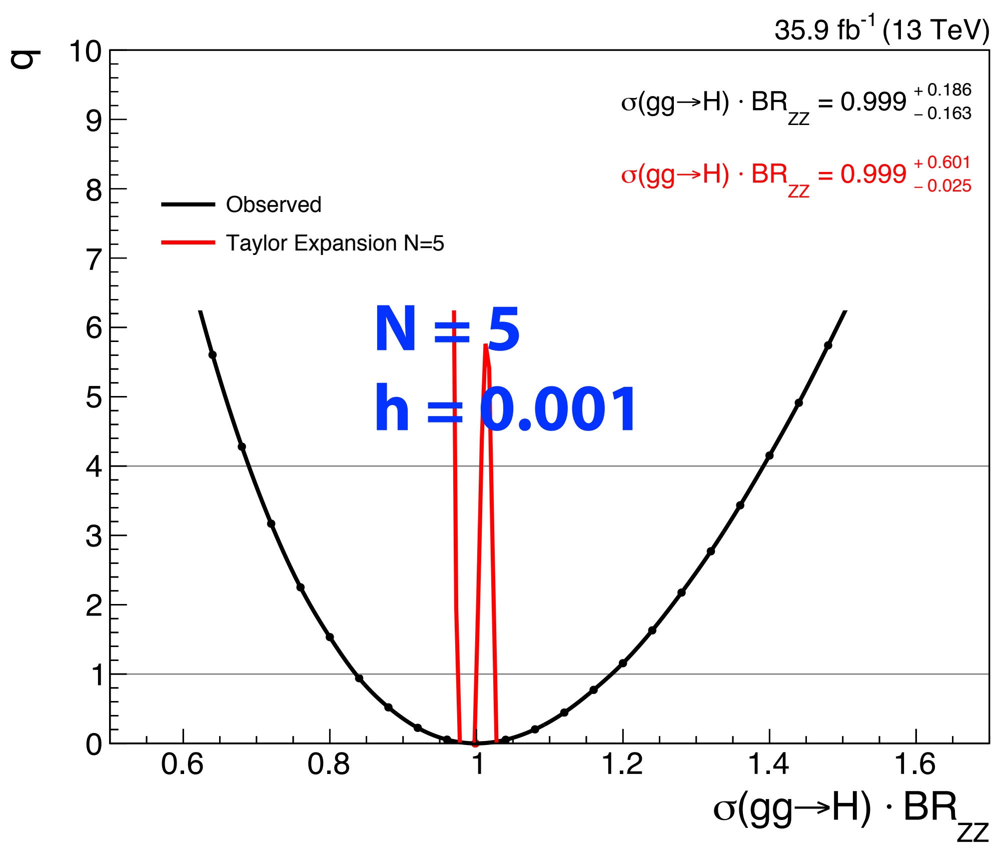


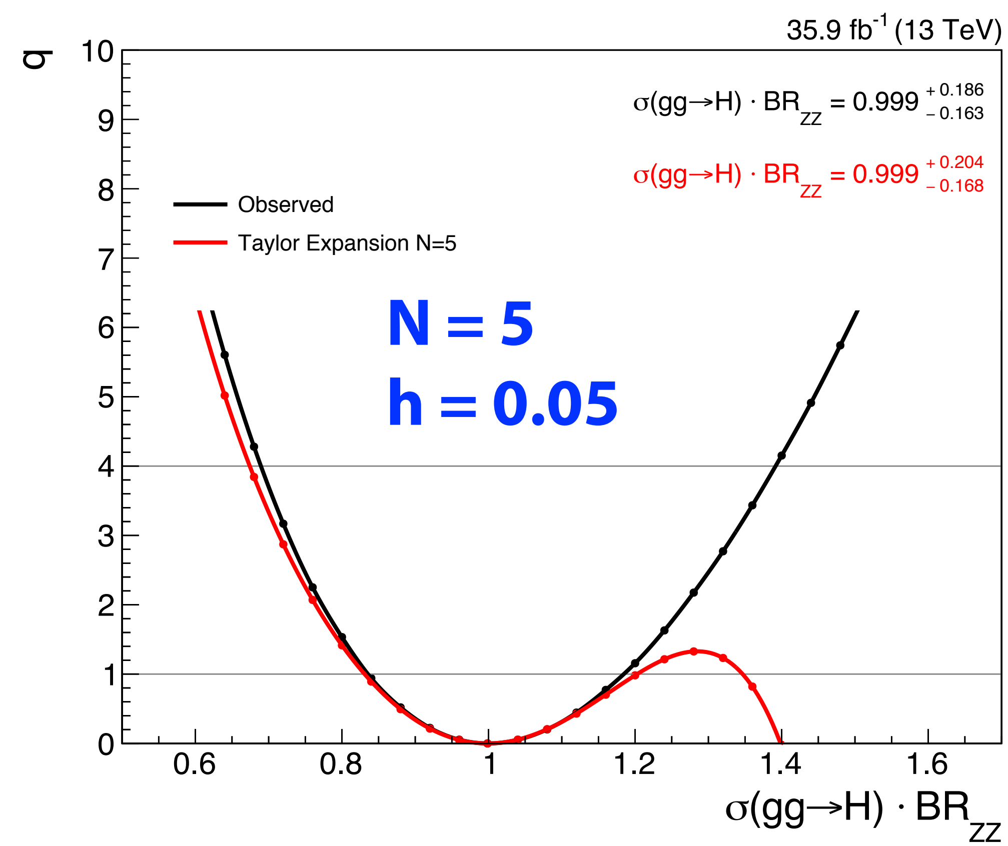
- At this point could just use the N=3 expansion. Going to higher orders (and accuracy) requires a different approach. Observed that by increasing h we can influence the range beyond which the expansion diverges.

Increasing h



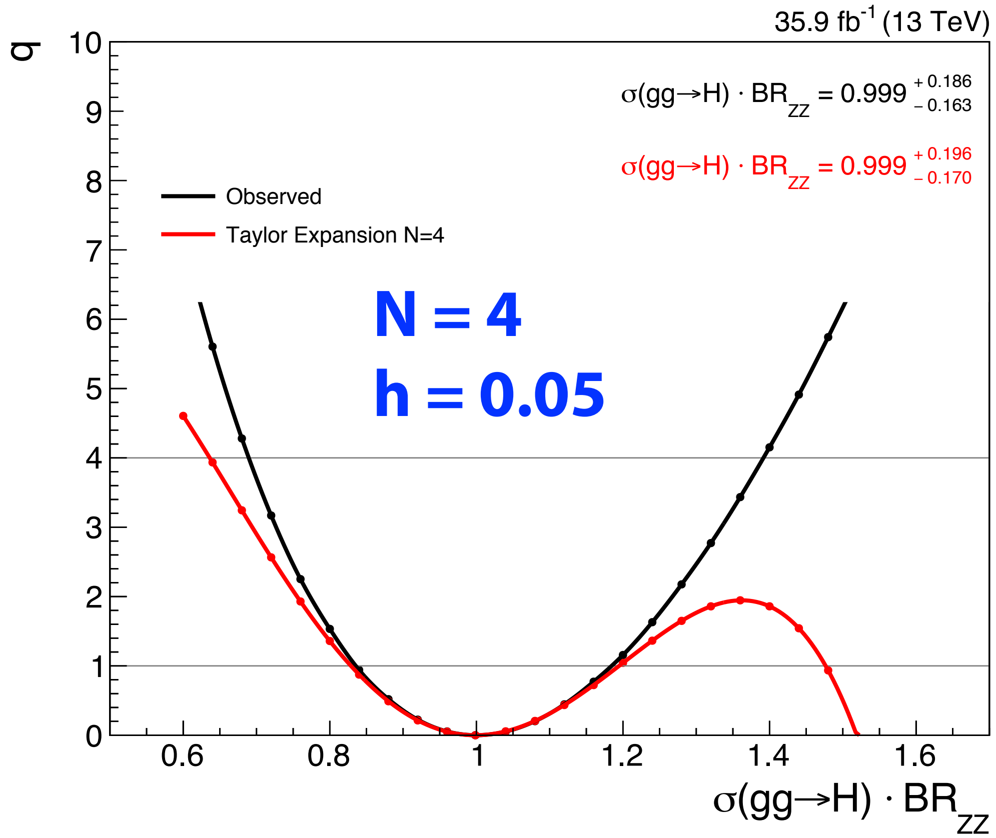


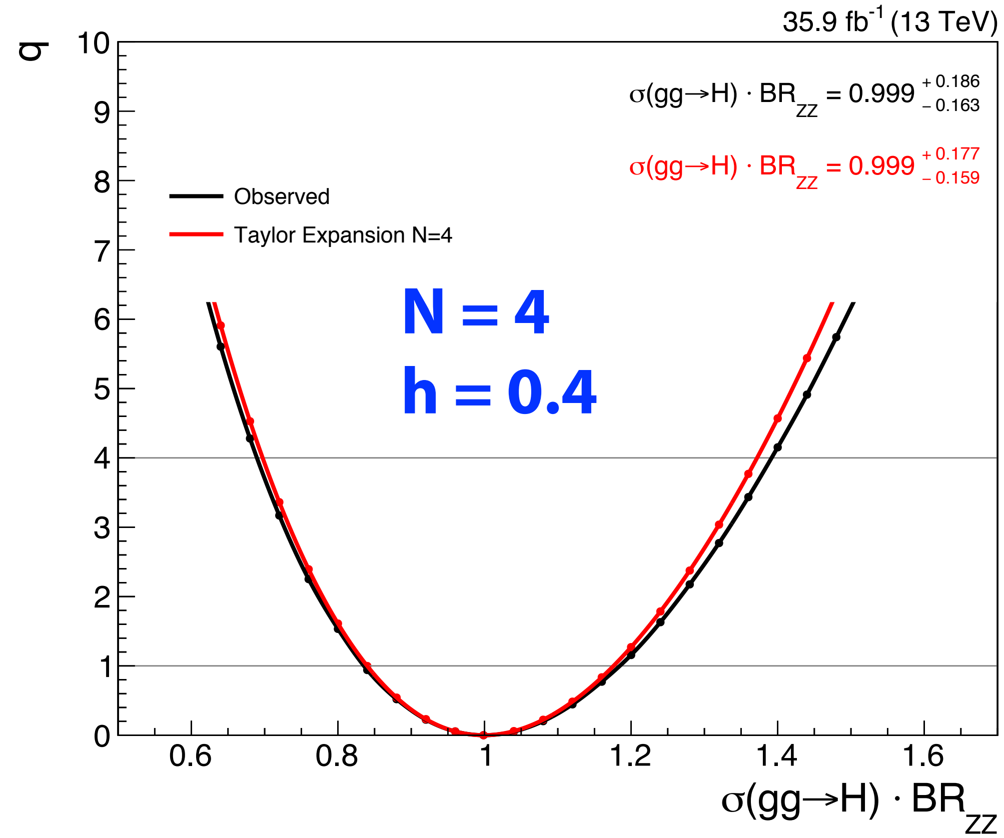




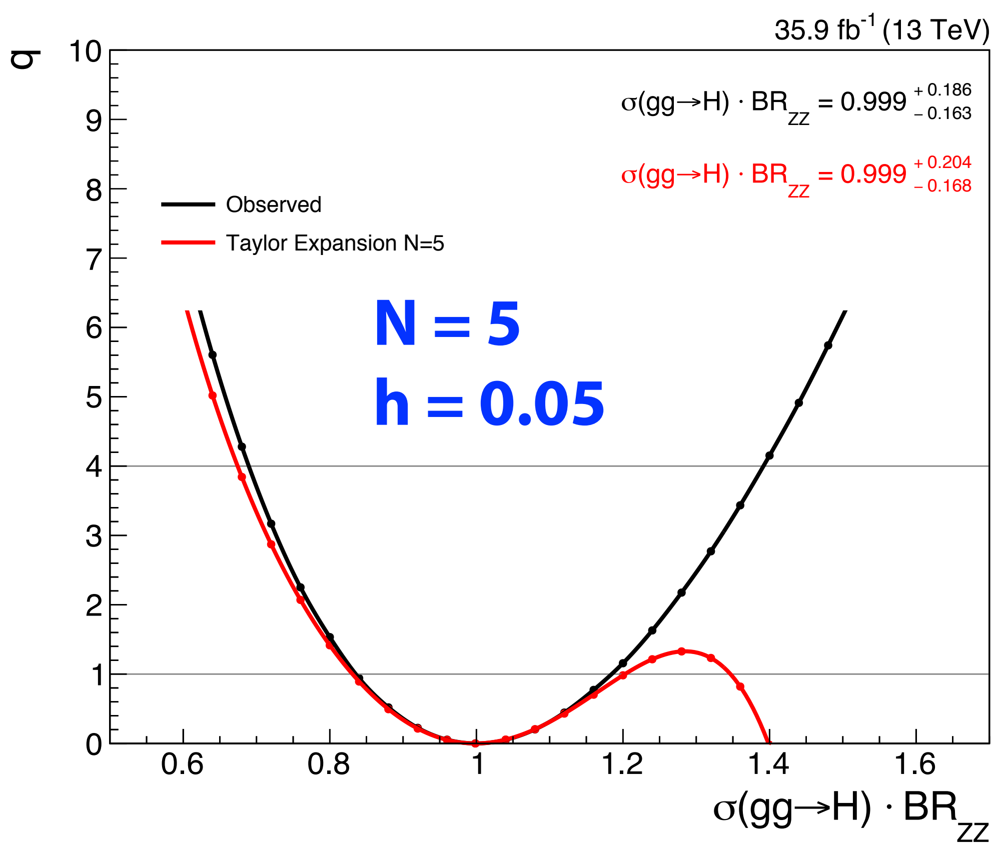
- Increasing h to 0.05 increases the range in which it converges, but still not enough.
- Turns out matching h to the range we want a good approximation in works well.

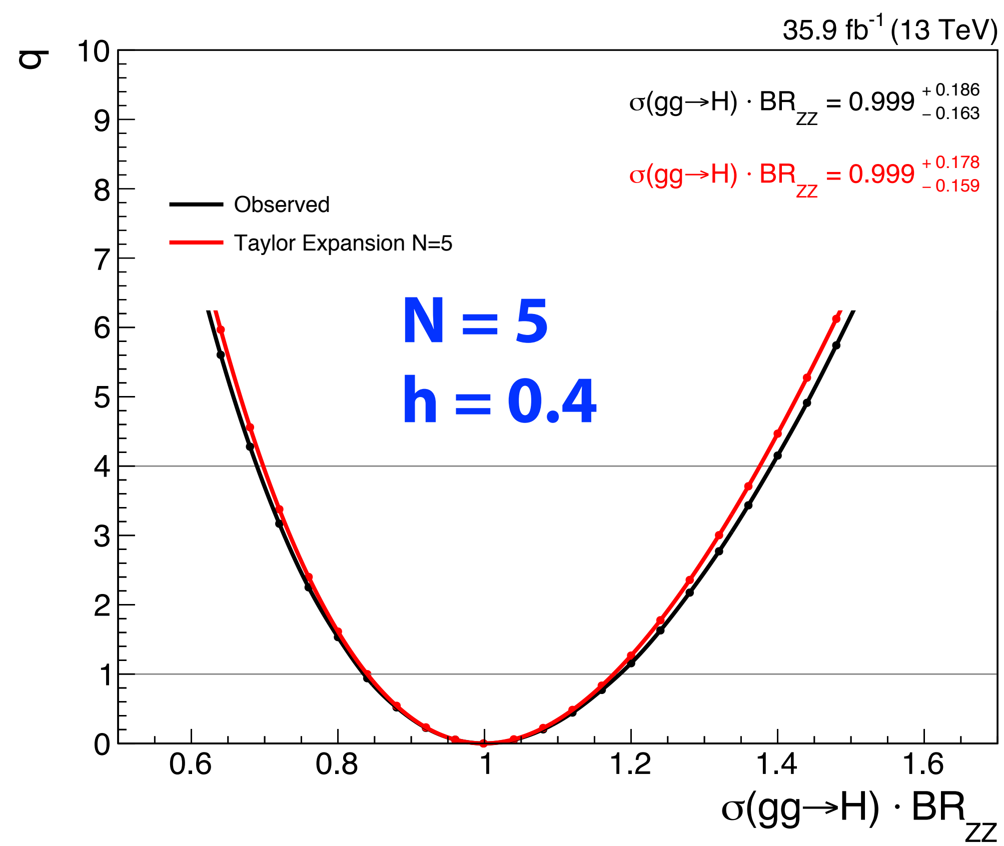
Increasing h



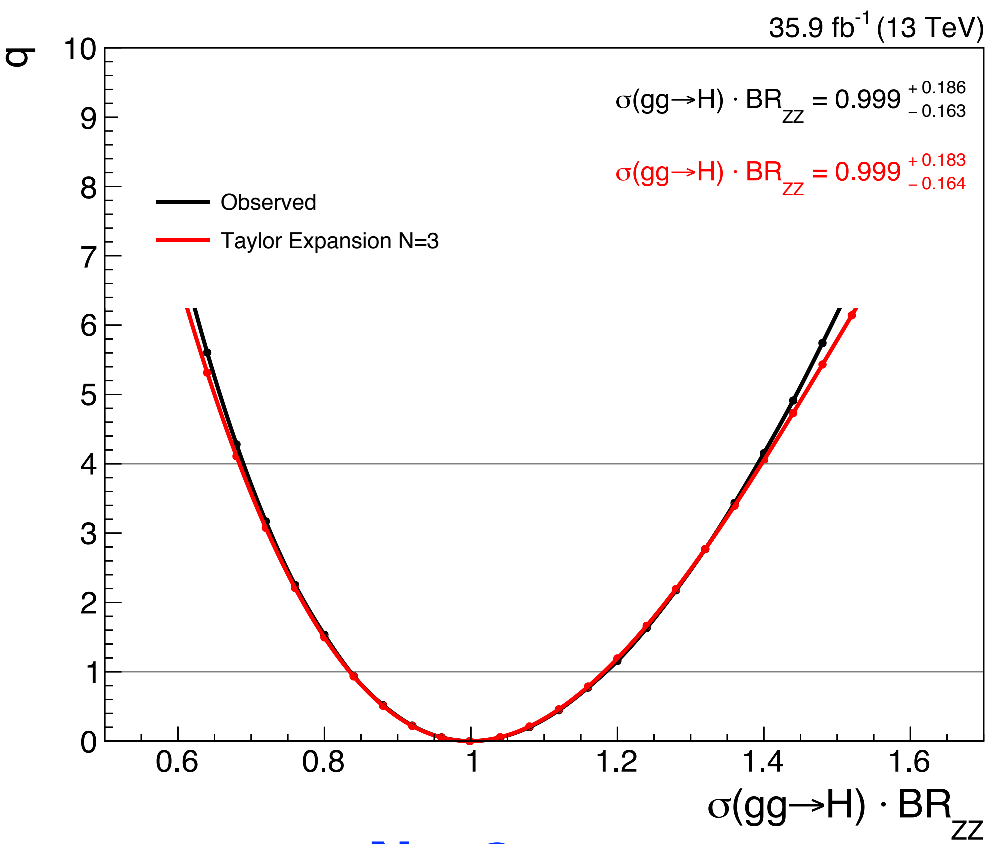


Divergence problem solved, but not more accurate than the N=3, h=0.001 expansion. To understand go back and look at lower orders...

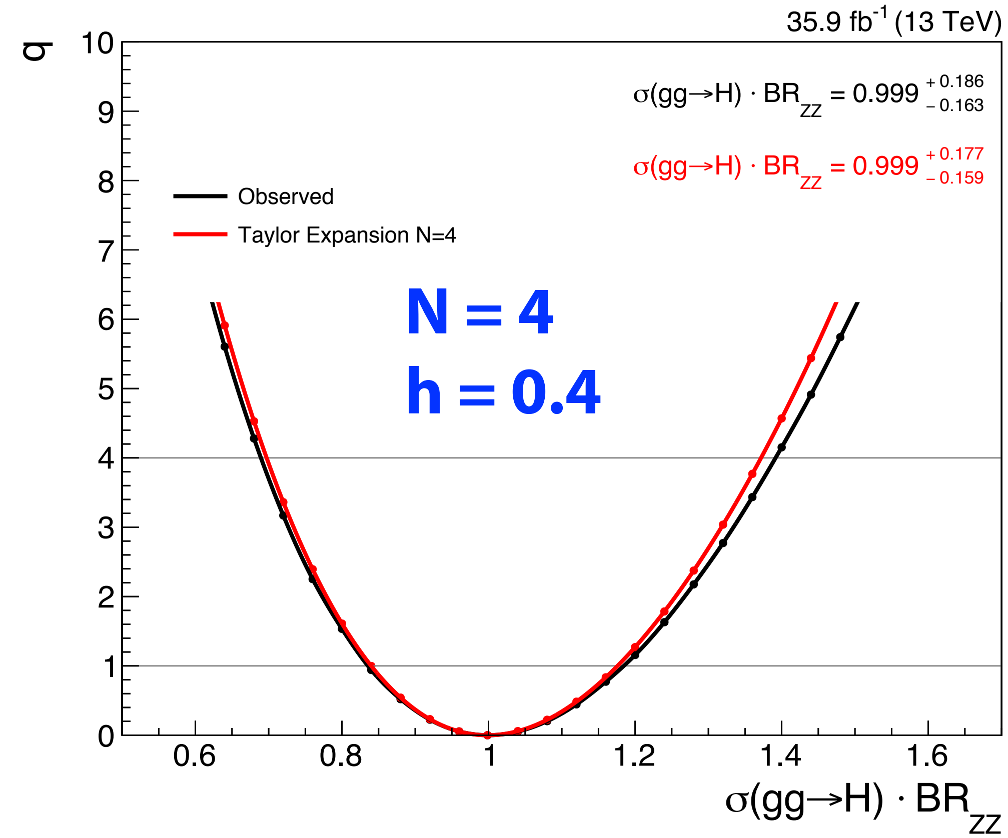


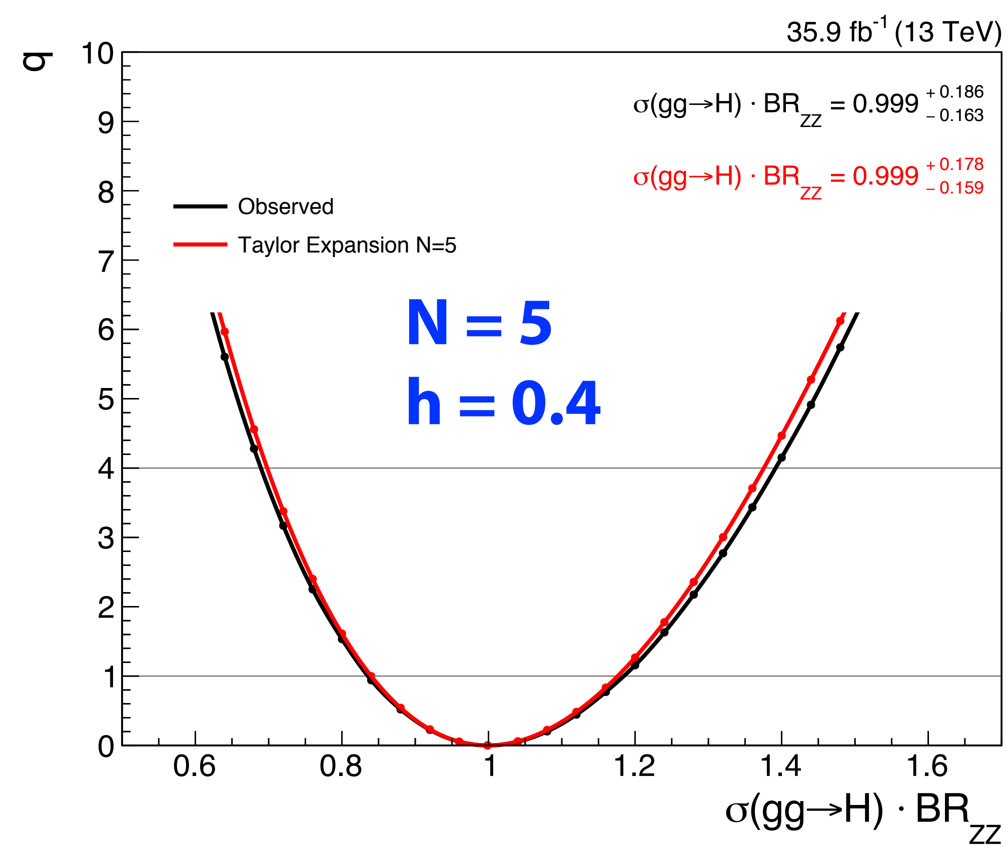


Increasing h



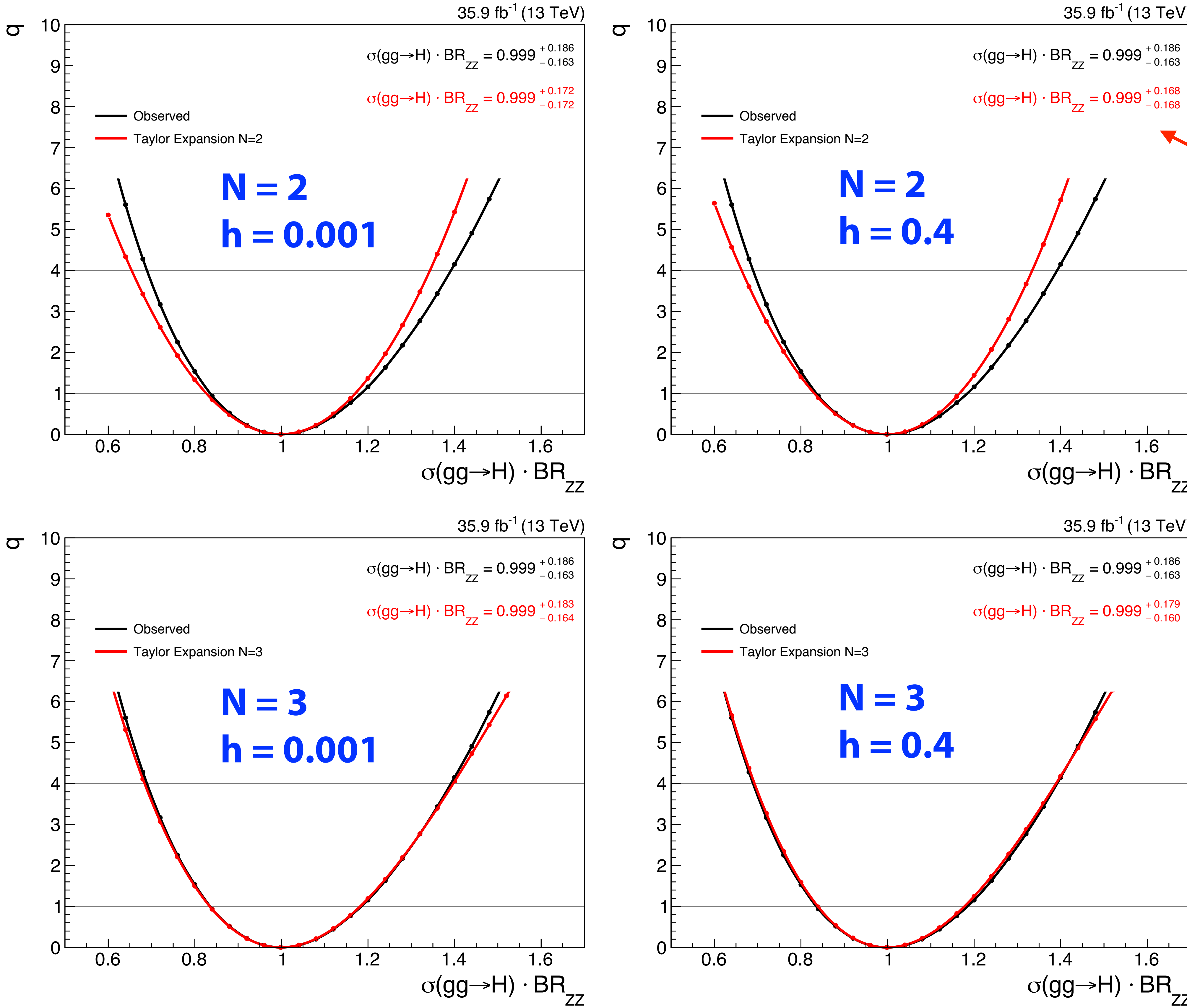
N = 3
h = 0.001





Divergence problem solved, but not more accurate than the N=3, h=0.001 expansion. To understand go back and look at lower orders...

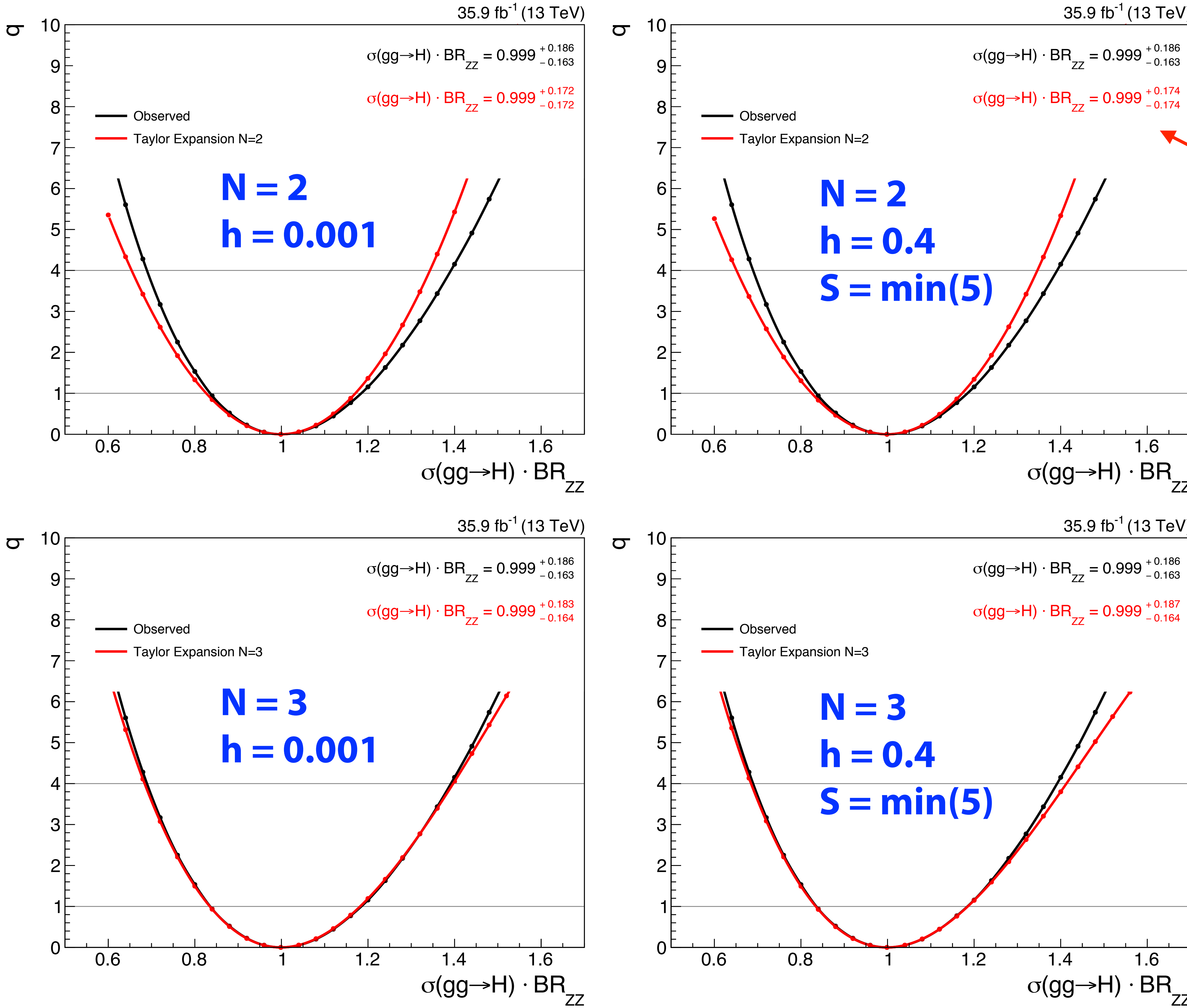
Increasing h



- Observe that 1σ uncertainties are less accurate with larger h

- With larger h pick up a larger bias from higher order contributions (consider how in approximating x^4 with $a \cdot x^2$, value of **a** depends strongly on chosen h)

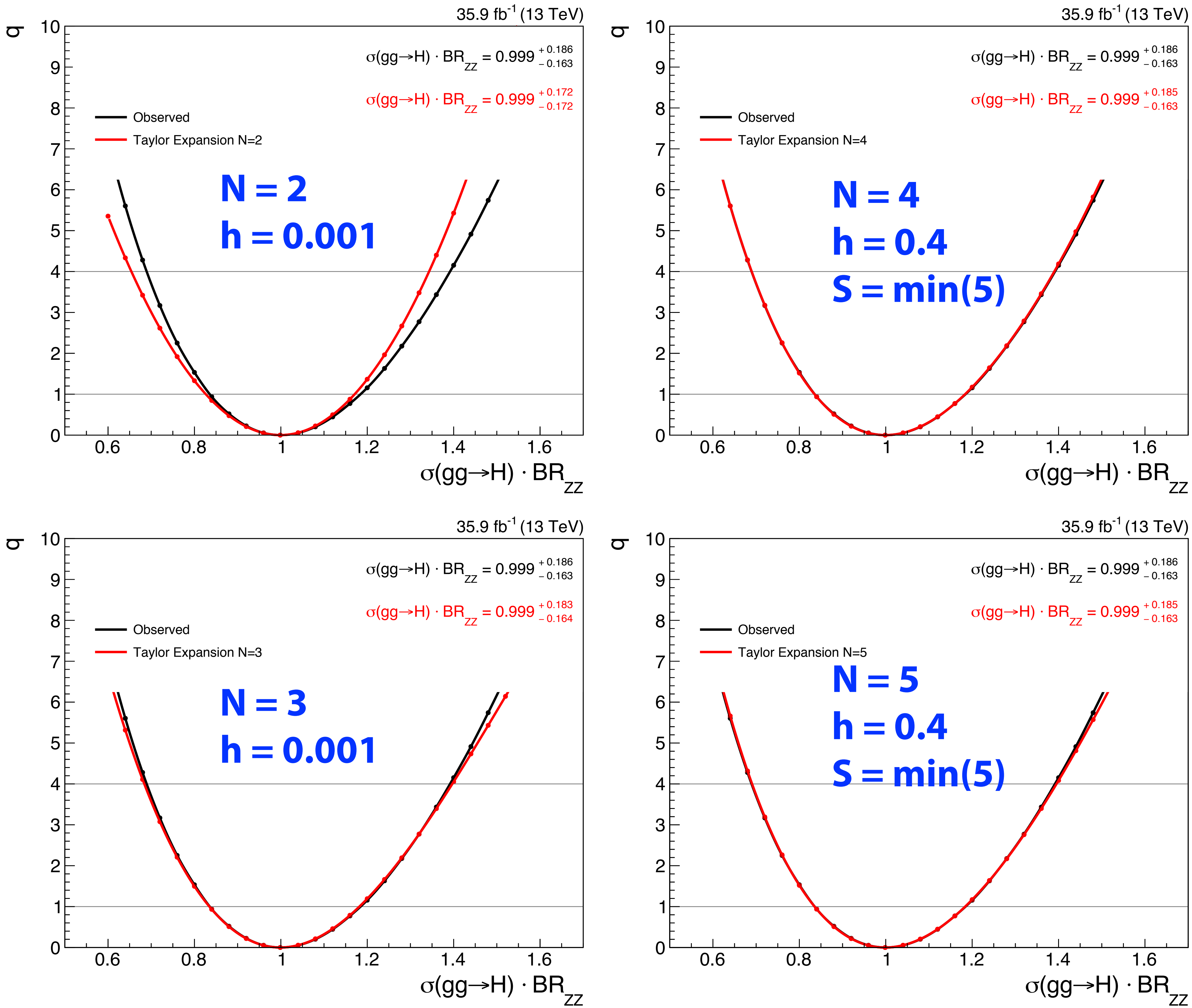
Increasing stencil size



- Now 1σ are much more accurate

- Using a 5-point stencil for N=2 term instead, now insensitive to any x^4 component

Increasing stencil size

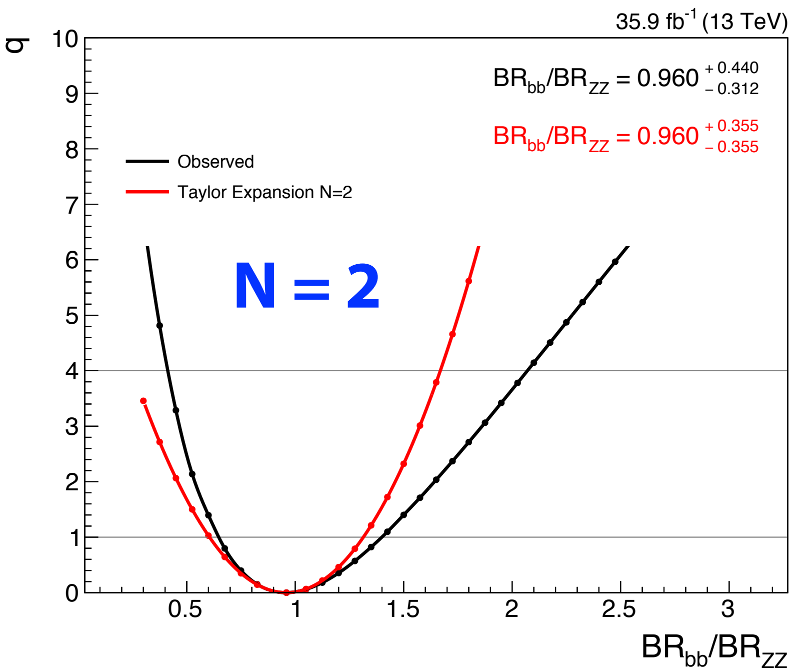


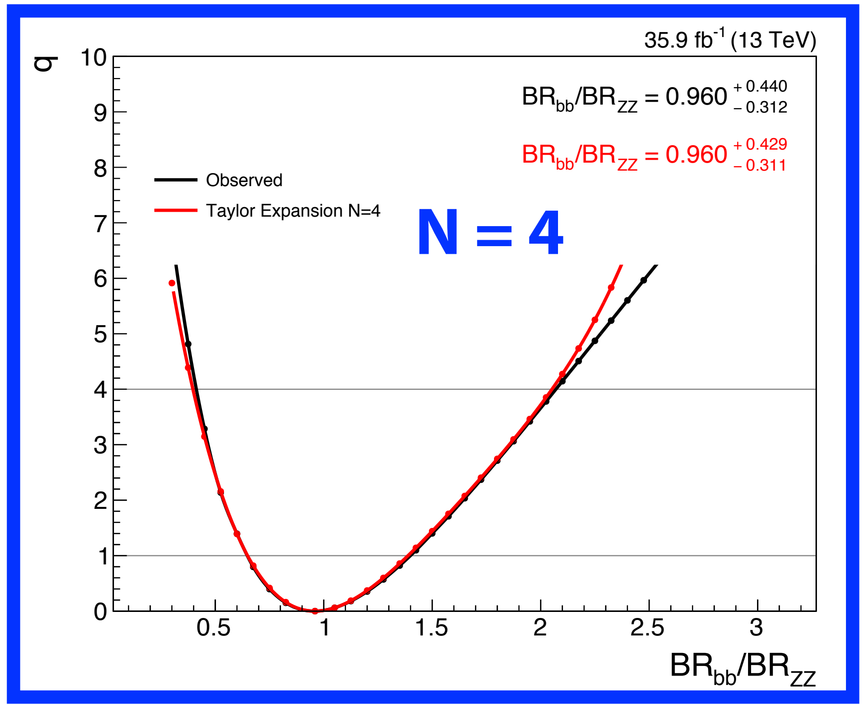
- N=4 with h=0.4 now even better than N=3 with h=001, at the cost of increasing stencil size for N=2

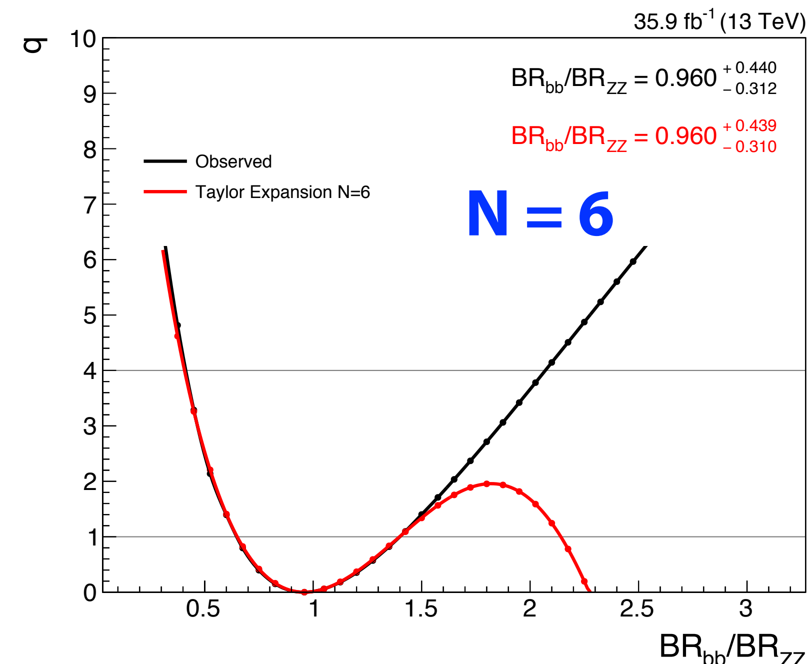
- N=5 starts to diverge more again, would need to increase N=3 stencil size from 5 to 7

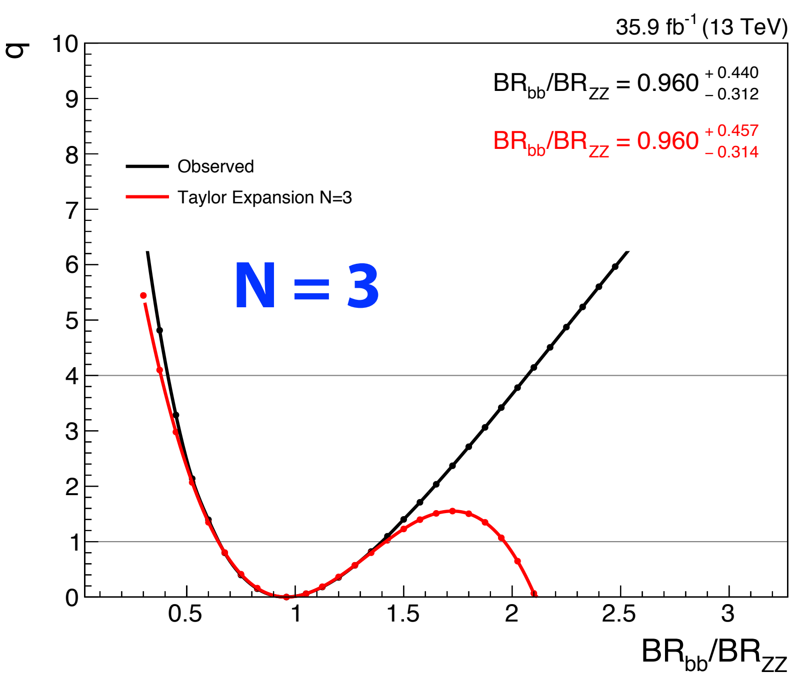
More challenging scans

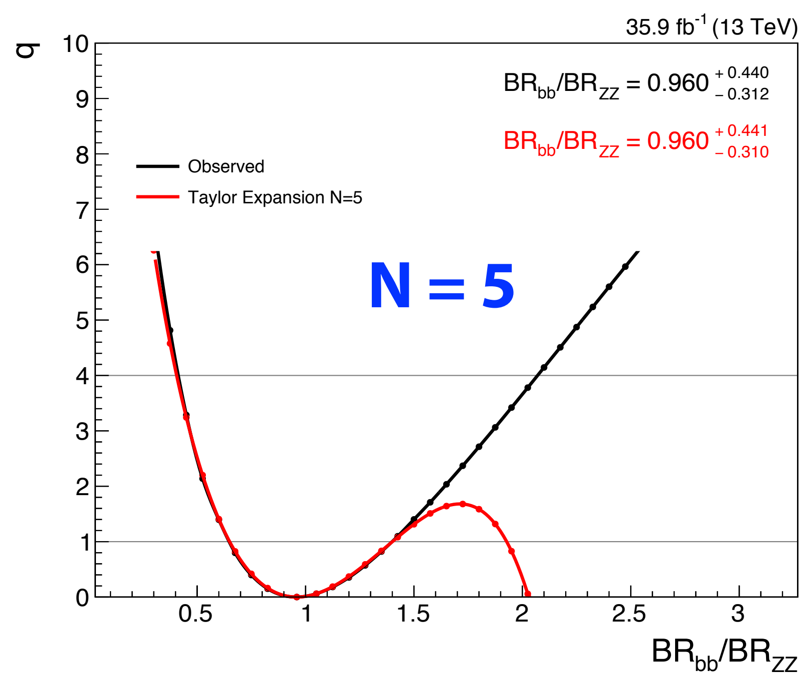
- Especially in models involving ratios of μ's, likelihood can be highly non-Gaussian:









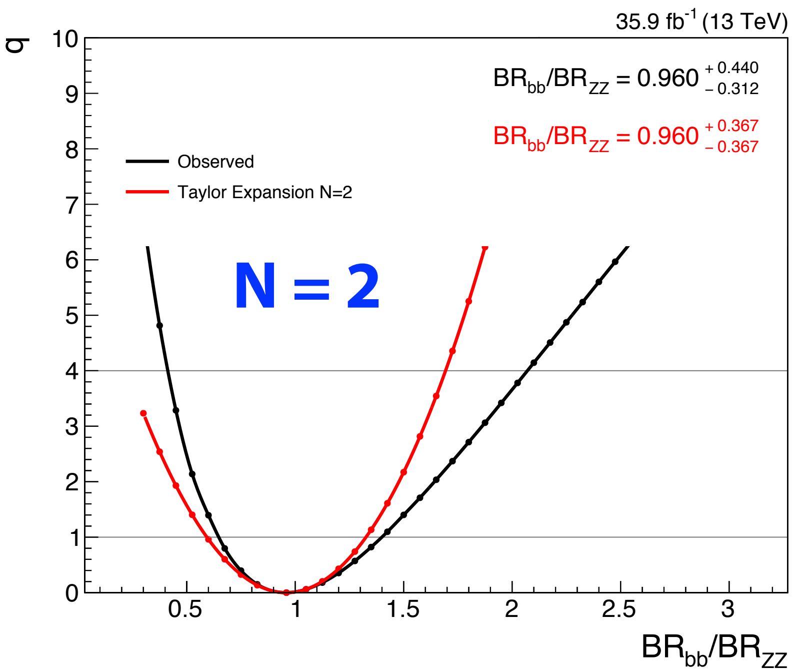


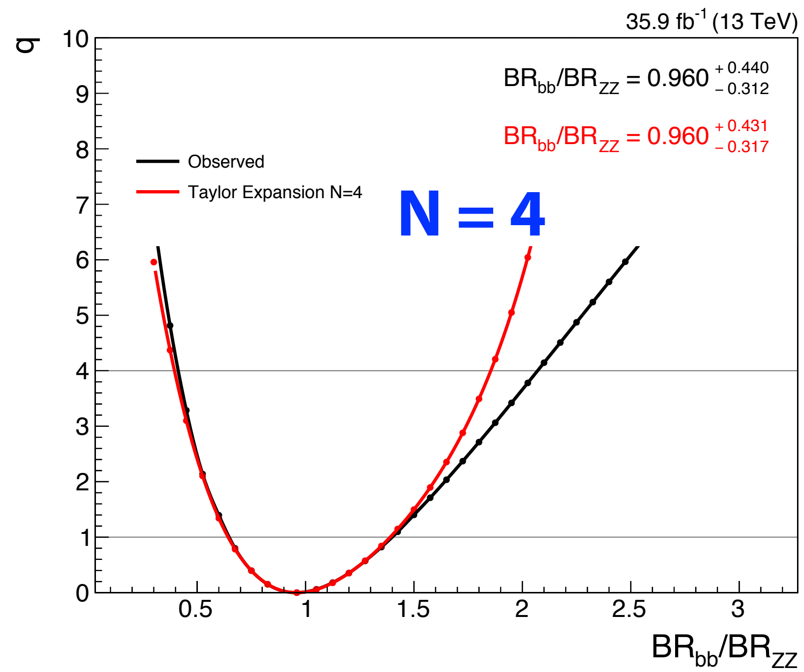
N = 2…6
h = [0.5, 2]
S = min(5)

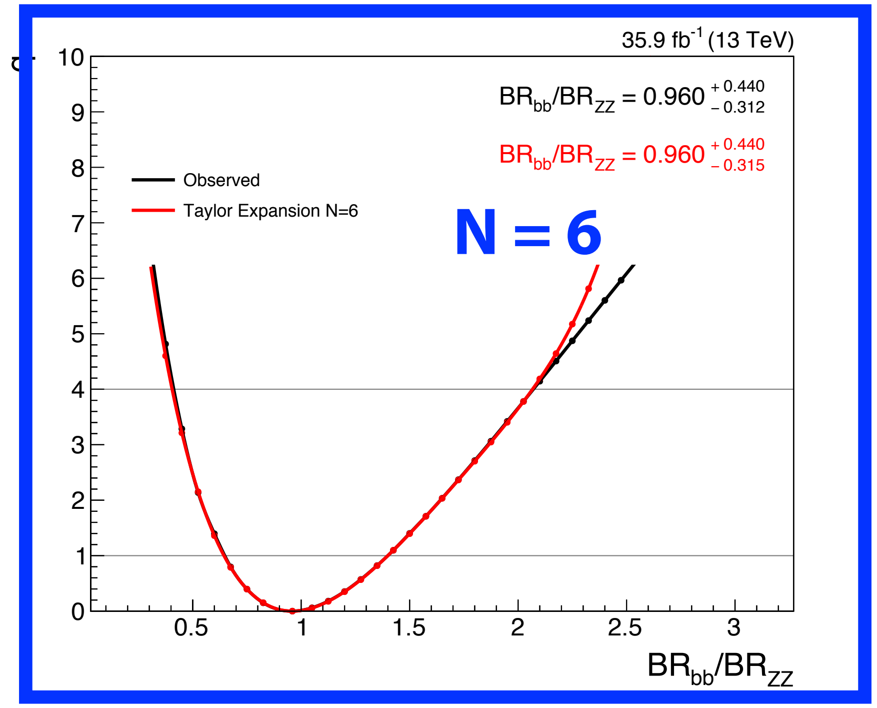
- For BR_{bb}/BR_{ZZ} convergence at N=4 is pretty good. Diverging going above this.

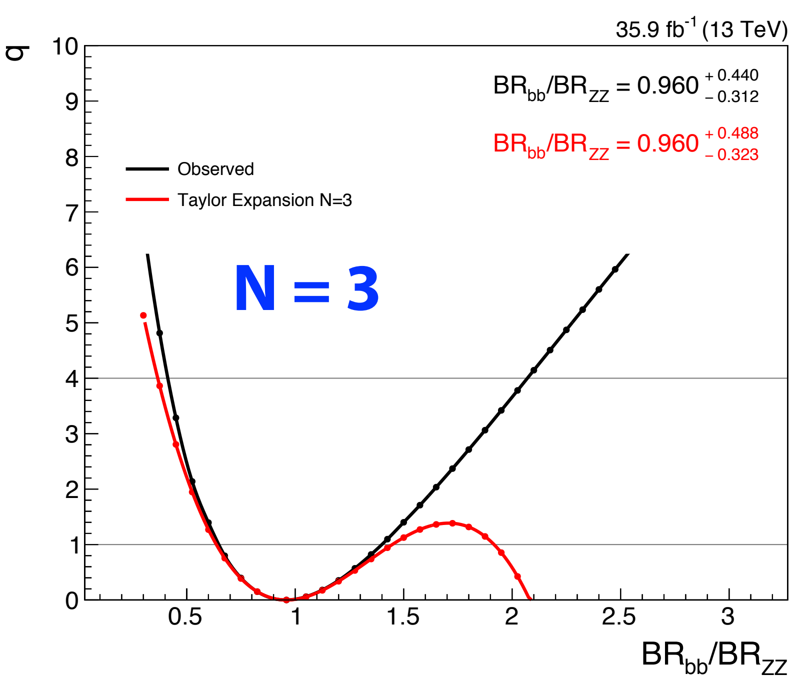
More challenging scans

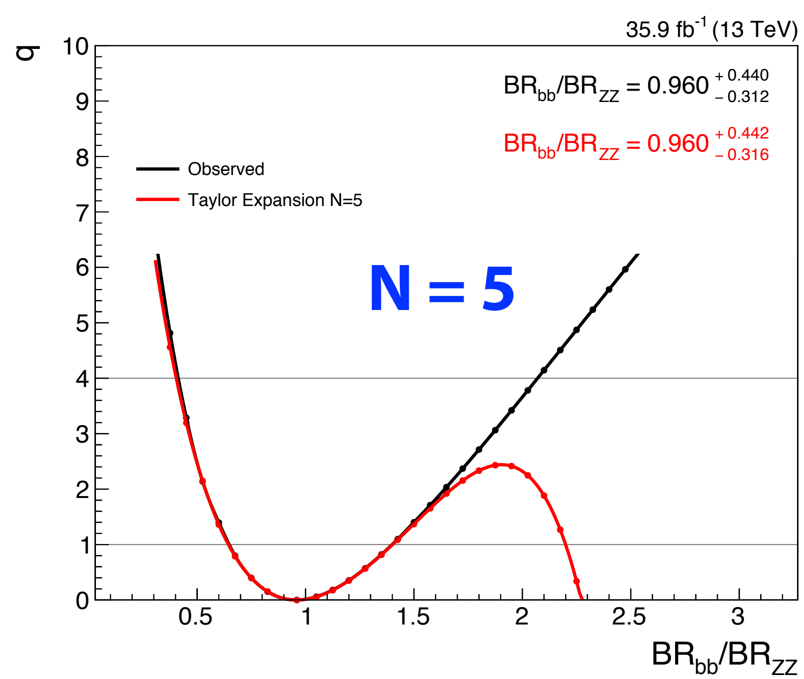
- Especially in models involving ratios of μ's, likelihood can be highly non-Gaussian:











N = 2…6
h = [0.5, 2]
S = min(7)

- Improved with minimum of 7-point stencils at each order.

Multivariate expansion

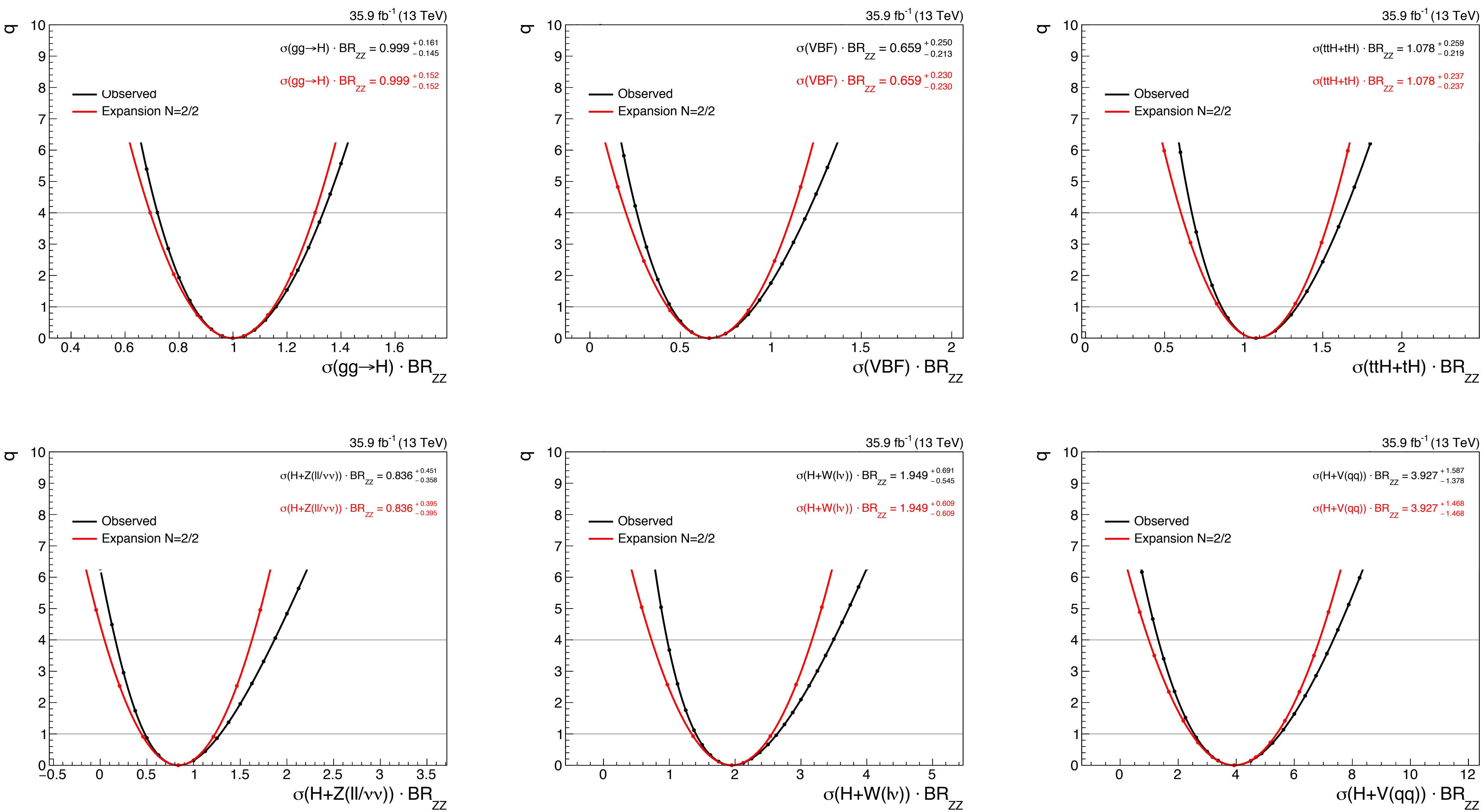
- **Expand the log-likelihood for the stage0 STXS model in all the POIs, with all other parameters frozen to the best-fit**
  - Can calculate expansion terms quickly without fitting
  - Compare profile likelihood scans fitting: a) full model b) Taylor expansion
- Have seen we need to go to at least N=4 for accurate scans in the 1D cases. Can automatically generate and evaluate all necessary terms, and recycle evaluation points in the POI parameter space wherever possible.
- However total number of evaluations still rather large [S=min(5)]:

Order	nEvals	nUniqueEvals	nActualUniqueEvals
2	1430	1233	1233
3	23430	17624	16391
4	272305	169115	151491

- Possible to reduce the number of evaluations at order N+1 by using the size of the terms at order N to predict which N+1 terms will be negligible (⇒ for a future presentation)

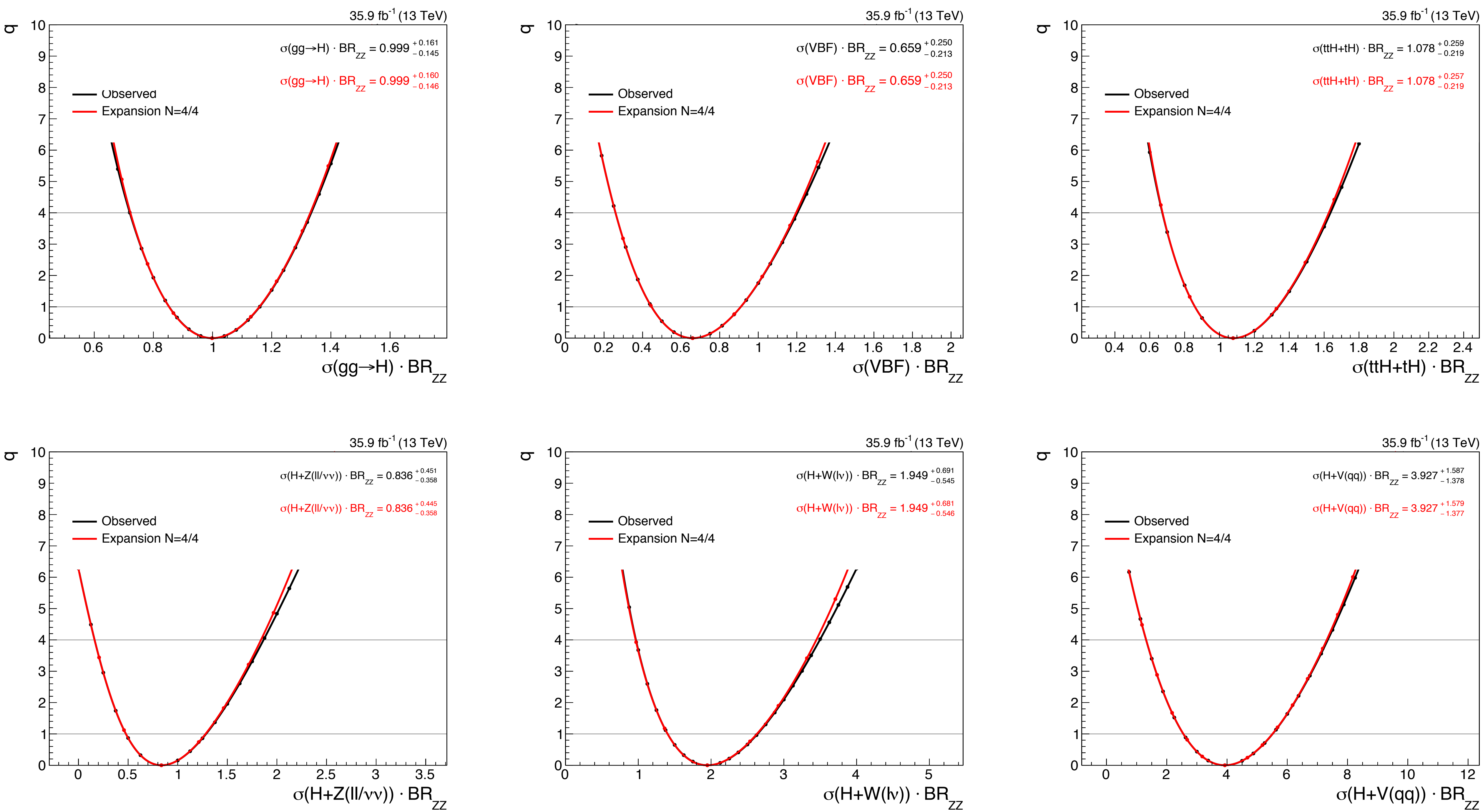
Order N=2

- At N=2 the expansion averages 1σ uncertainties pretty well (BR ratio parameters look similar)
- All stencils set to ~ 2σ range



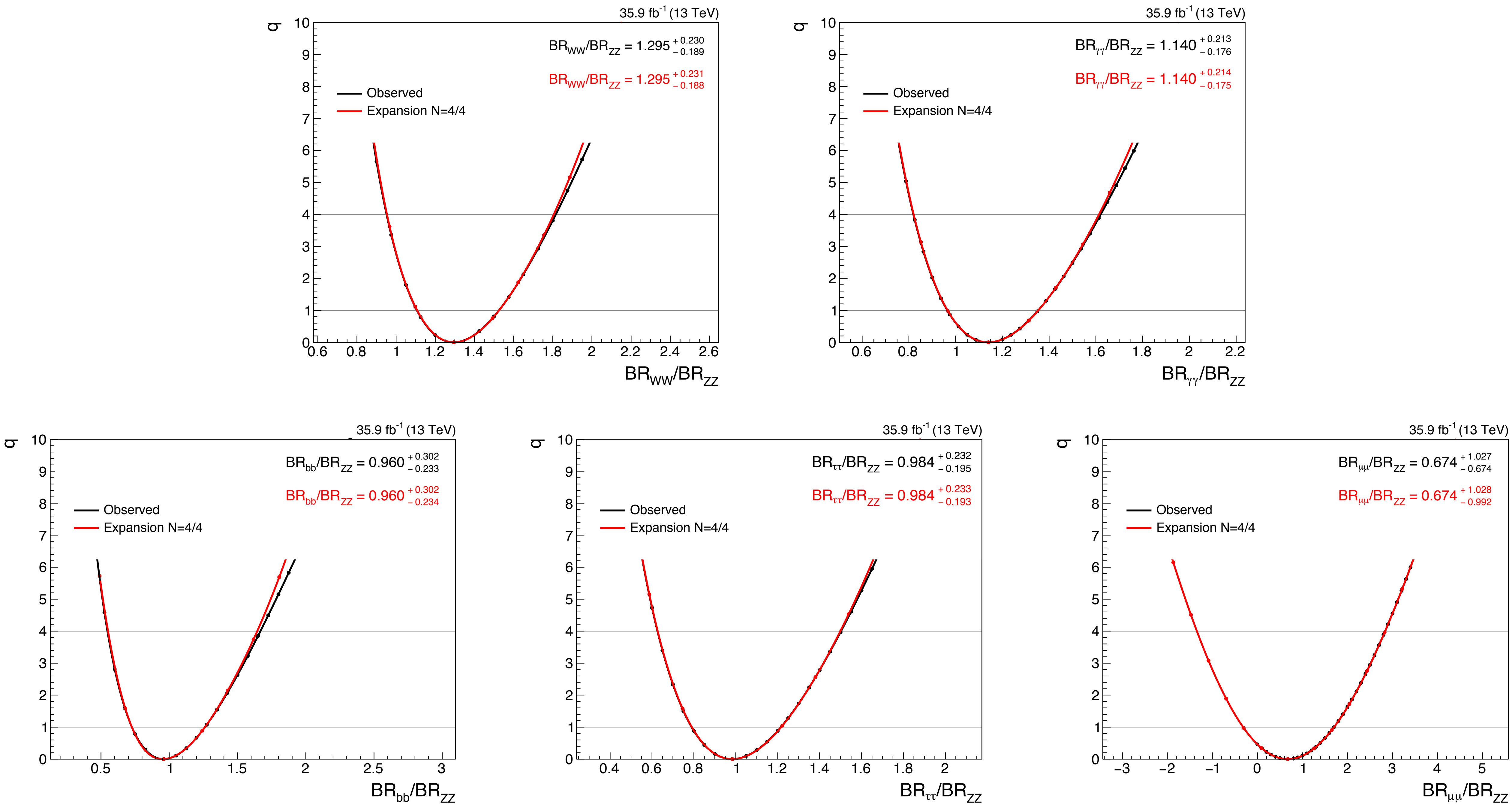
Order N=4

- Adding all terms up to N=4 gives good agreement



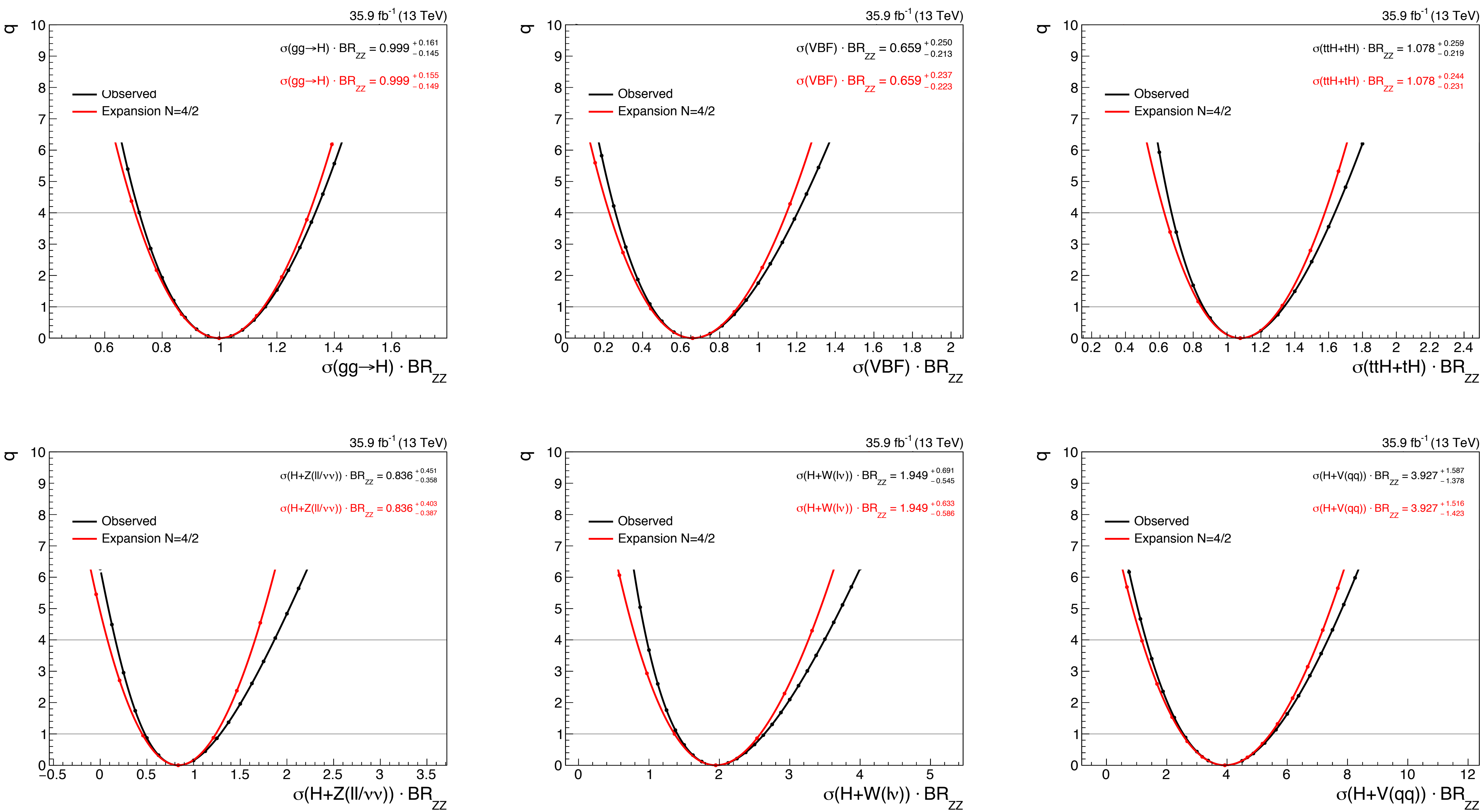
Order N=4

- Adding all terms up to N=4 gives good agreement



Order N=4, no mixed terms beyond N=2

- Shows that mixed terms with N > 2 are important



Summary

- Validated Taylor expansion approach in the Stage0 STXS model:
  - Individual 1D scans (one parameter expansions)
  - 11 parameter expansion of POIs with all other parameters frozen
- Best accuracy does require appropriate choice of stencil size and number of points
- Next steps:
  - Demonstrate on full fit with all nuisance parameters profiled
  - At higher orders will be useful to be able to predict negligible N+1 terms that can be dropped to reduce the number of fits
  - Test recasting to more constrained models

Order N=2, no mixed terms

- Dropping the mixed derivative terms illustrates that the correlations play an important role