ggF Stage 1.2 Uncertainty scheme

Frank Tackmann, Matthias Kerner, Stephen Jones, Micheal Spira, Bernhard Mistlberger, Stephen Philip Jones
Jonathan Langford, Ed Scott, Andrea Massironi
Nicolas Berger, Robin Hayes, Hui-Chi Lin, Yun-Ju Lu, Roberto Di Nardo, Tae Park, Laurelle Veloce, Hongtao Yang, Haider Abidi
Oct 5, 2021

**BROOKHAVEN**
NATIONAL LABORATORY

Introduction

- STXS measurements are very common in the Higgs group + **provide a convenient way to parametrize the uncertainties**
  - STXS recommendations have **two parts** - bin definition + associated uncertainty
- For Stage 1 scheme, had both - ggF Stage 1.2 only has a preliminary systematic scheme
  - **Tons of work put into the defining a new scheme - ~ year long collaboration with people from ATLAS, CMS & theorists all involved**
  - **Collaborations across various LHC XS G!**
- Results documented @ Link, plan to release a document after getting community feedback
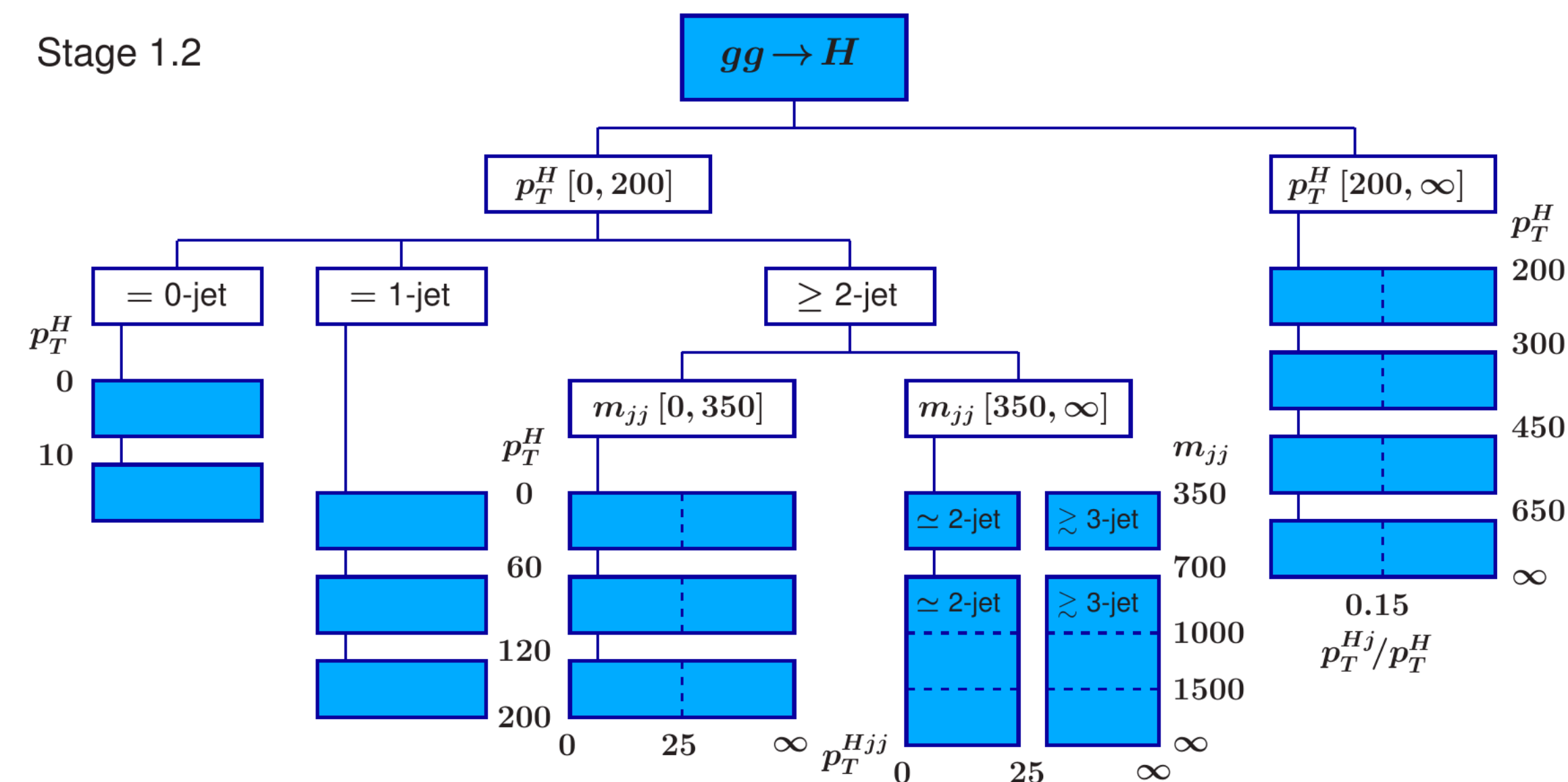
Uncertainty scheme

- Couple of **key ingredients** that need to be defined for such a scheme:

  - **Common default MC** - ATLAS: PP8+NNLOPS & CMS: MG5+NNLOPS

    - Talks on harmonizing this even further for Run 3

  - **List of NPs to parametrize the uncertainty**

    - All stakeholders need to agree on this - defines how to factorize the systematics

    - Many meetings within the LHC XS WG - Finalized a common scheme @ Link

    - Overview for completeness - Has been agreed to it by all parties

  - The method to evaluate systematics across bin boundaries - final numbers

    - Different ways to evaluate the numbers

    - ~ year long collaboration on the methodology - an evolution of the ST method settled as the main choice

  - Systematics impacting the acceptance - shapes within an STXS bin

    - Largely agreed to leave this up to each analysis as there are too many possibilities
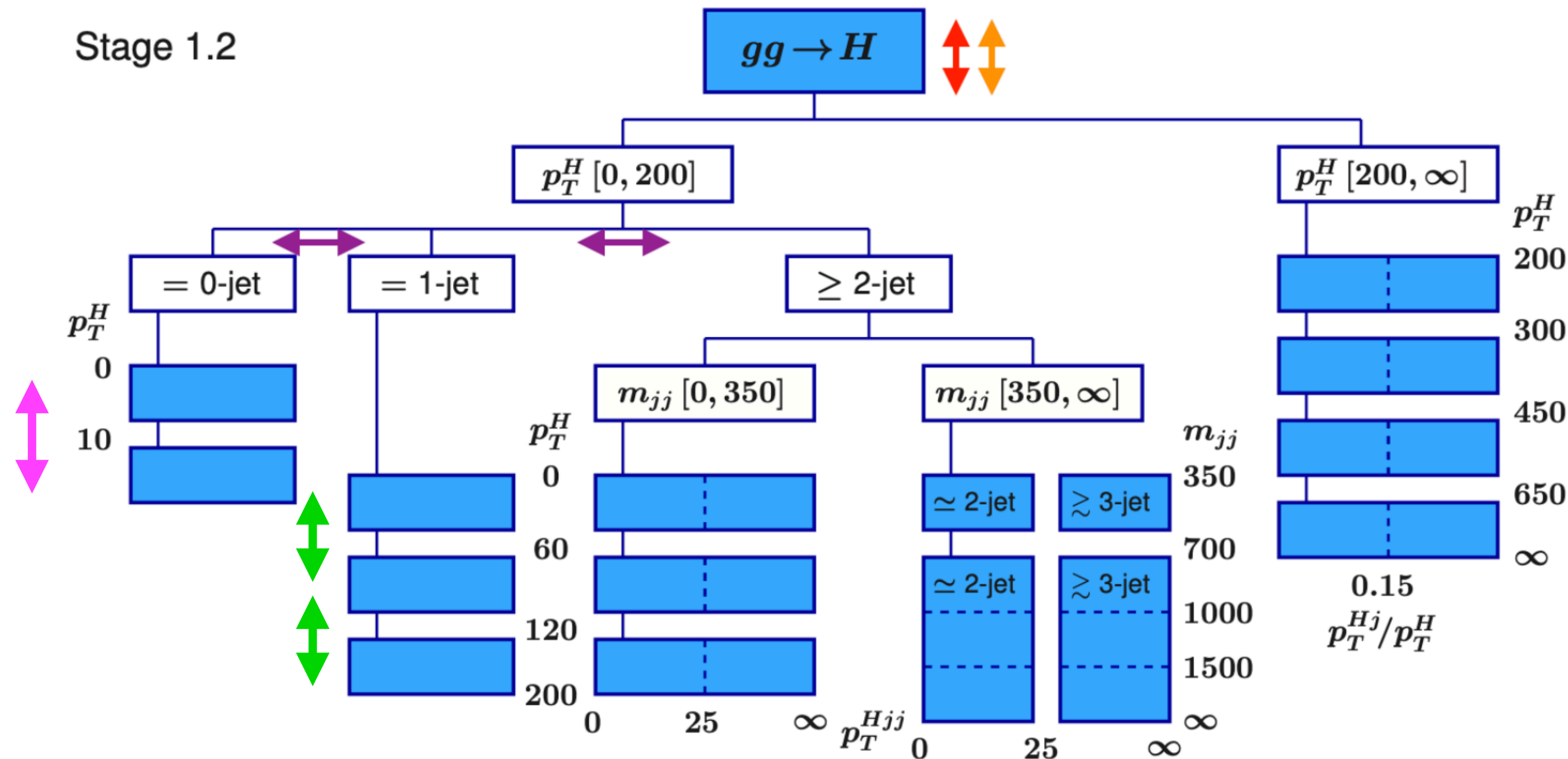
    - However, have a proposal on how to cover for this

Stage 1.2
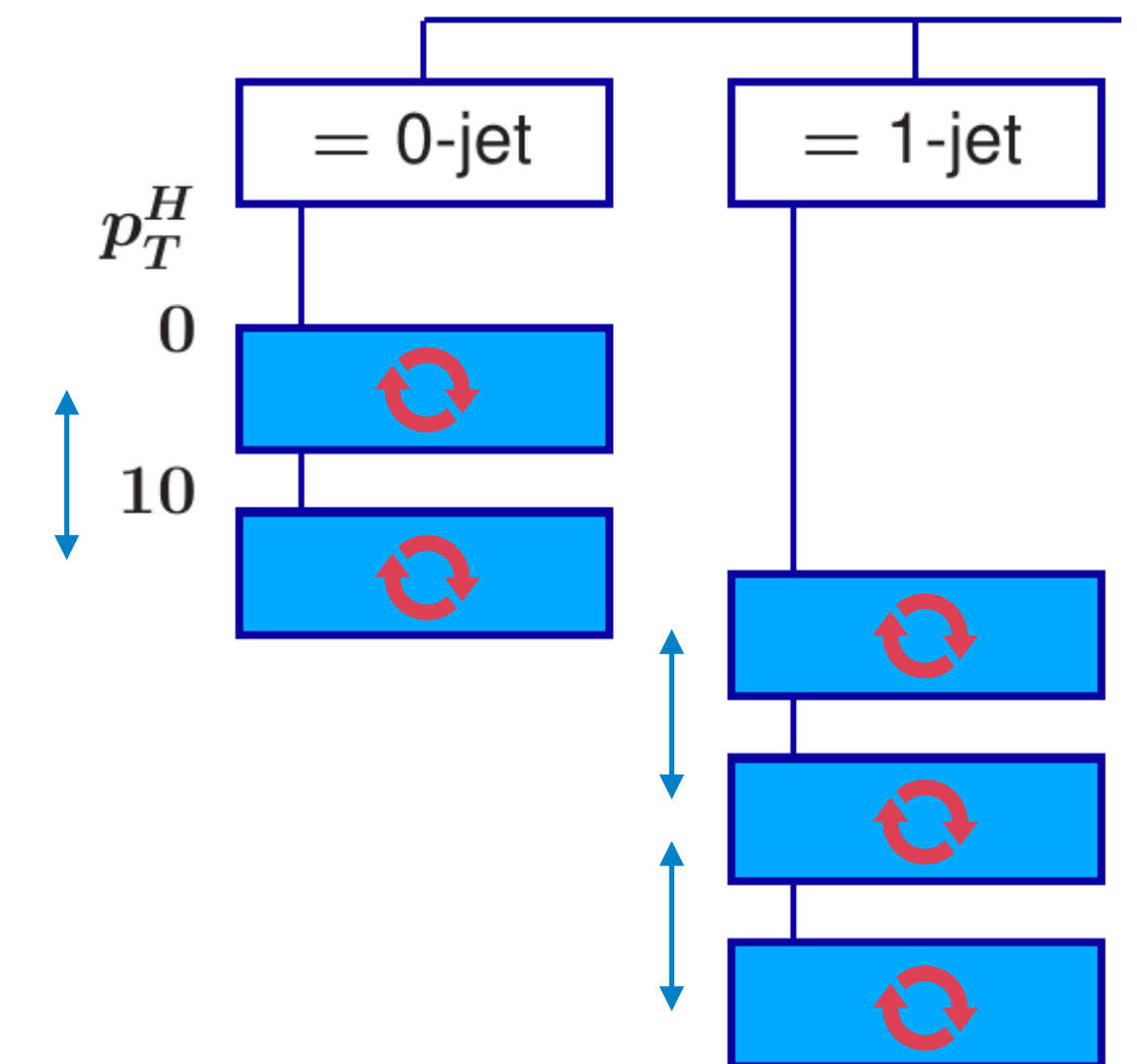
Parameter Scheme

NP scheme

- Final decision from discussions documented @ Link
  - This is an evolution of the Stage 1 theory scheme with many common parts taken directly
- Overall, 18 NPs have been decided upon to parameterize the uncertainties

Overall yield and jet migration - 4 NPs
1 NP for overall fixed-order effects
1 NP for overall resummation effects
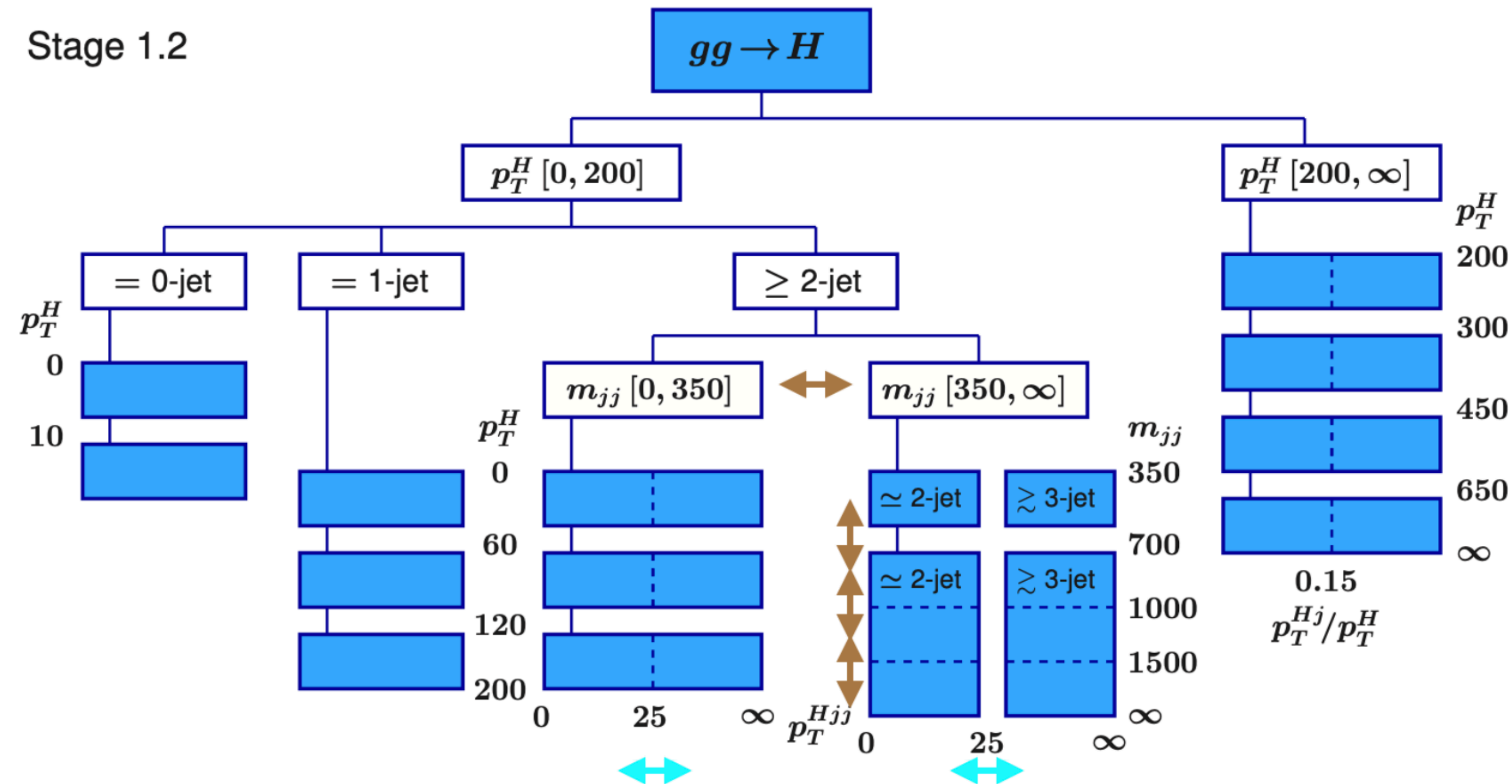1 NP for 0-1 jet bin migration
1 NP for 1-2 jet bin migration

pTH migrations - 3 NPs
1 NP for p_T^H migration in 0 jet
2 NP for p_T^H migration at the 60/120 boundary
Correlated across 1-2 jet bins
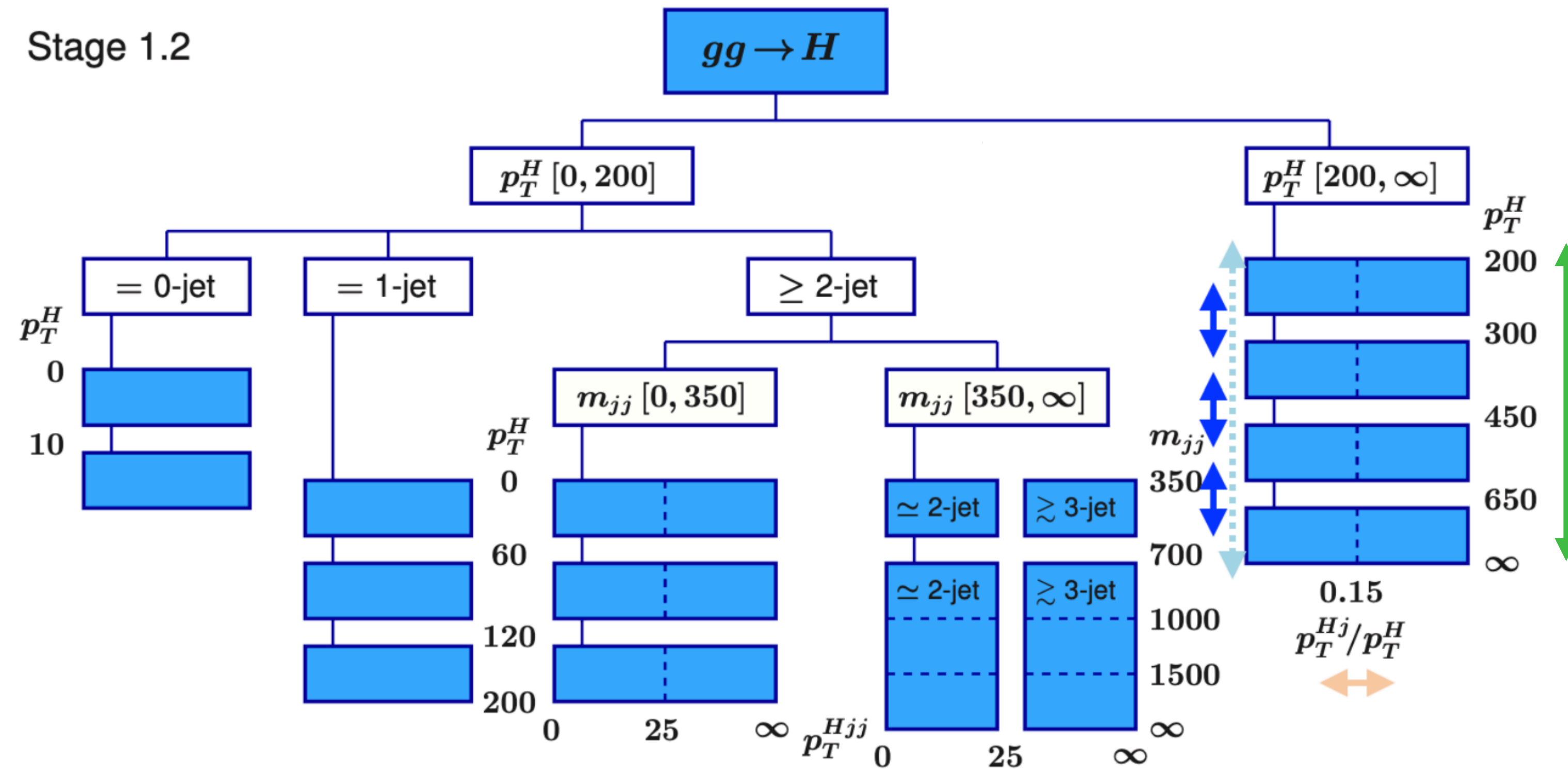
No change from Stage 1
for these NPs

NP scheme

m_{JJ} and $p_T{}^{Hjj}$ migration - 5 NPs
4 NP for new m_{JJ} bins
1 NP for $p_T{}^{Hjj}$ variations

New m_{JJ} bins added

$p_T{}^{Hjj}$ sys now impacts the full
2J phase space

NP scheme

Stage 1.2

High p_T^H migration - 6 NPs
1 NP for overall XS variation
3 NP for new p_T^H bins
1 NP for p_T^{Hj}/p_T^H variations
1 NP for top mass scheme

Previously one NP for full region
Now take dedicated state of art
predictions and factorize out the
various effects

Uncertainty evaluation

Long range ST method

- Many ways to evaluate - all involve **varying** the muR & muF scales and using the XS variations
- Build upon the ST method to remove some of its limitations - **LR ST method collaboratively developed**
  - Evaluate the yield variations inclusively and replace with better calculation if available
  - Distribute the migration sys across all 'higher' bins
- Leads to **double counting** - if we apply the same method in $m_{JJ}>350$ and $m_{JJ}>700$, double counting in the upper region
  - Introduce ρ scaling param to prevent this - no clear way to estimate this correlation theoretically
  - Nominal choice of $\rho = 0.5$ chosen to ensure that total variation is ~ equal to the scale variations in that bin

Bin definitions
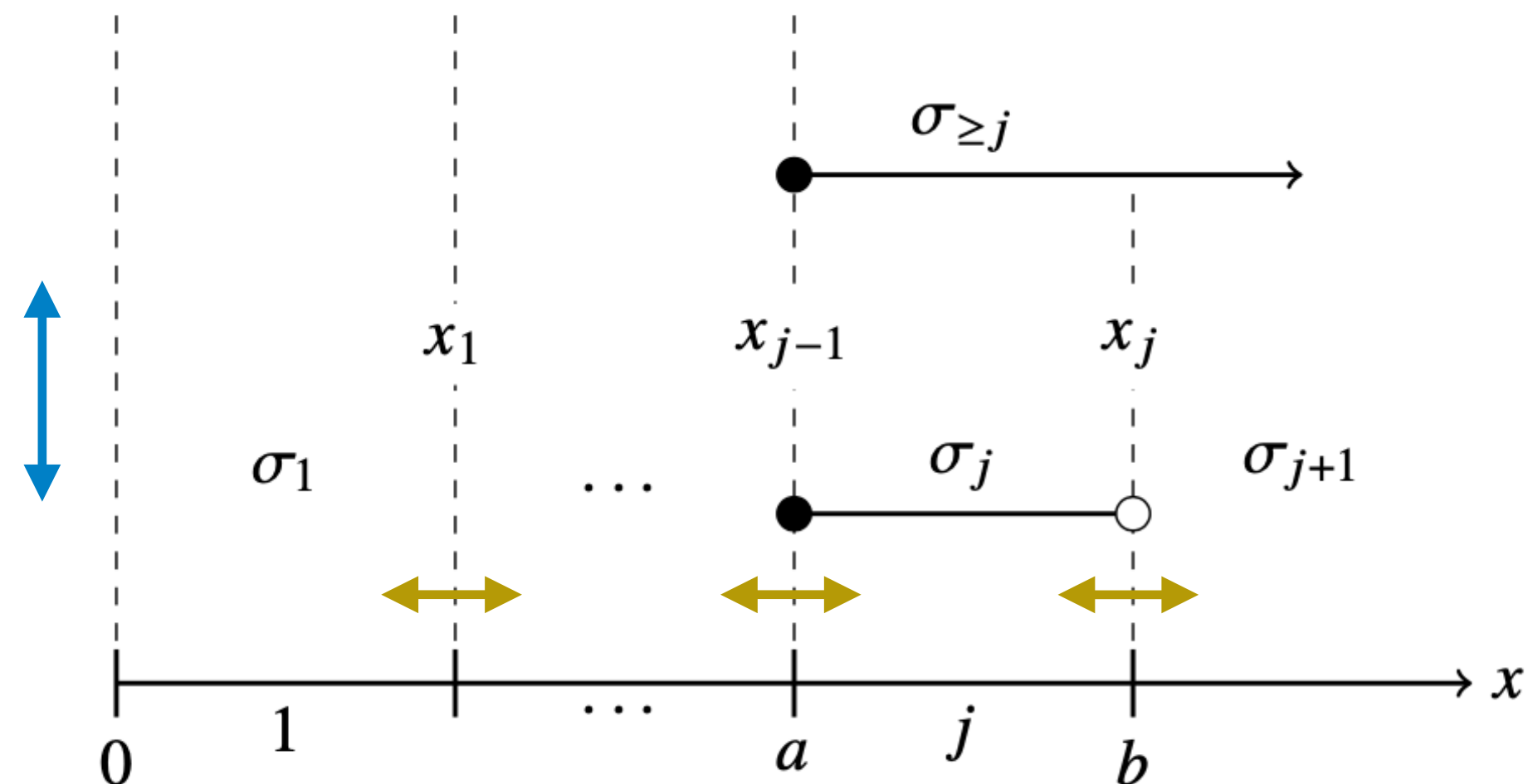
Long range ST method

Take the max scale variation inclusive region and apply as yield NP for all bins

$$\theta^y(j)\left\{\delta^y = \max|\Delta_\mu|/\sigma \ \ \text{(Replaced by state-of-the-art number when available)},\right.$$

$$\theta_{x_k}^{\text{mig}}(j)\begin{cases} 0 & : j < k \\ \rho \times \left(\delta_k^- \equiv -\max|\Delta_{\mu,\geq k+1}|/\sigma_k\right) & : j = k \\ \rho \times \left(\delta_k^+ \equiv +\max|\Delta_{\mu,\geq k+1}|/\sigma_{\geq k+1}\right) & : j \geq k+1 \end{cases}$$

Take the max scale variation in ≥ k+1 region and apply as migration NP between k and ≥ k+1 bins

Inclusive and Jet migrations

- No change from the Stage 1 scheme

- Use BLPTW method from the YR4

  - These end up being the 'yield' uncertainties when we evaluate other uncertainties

| Uncertainty [%]	jet bin		
	σ_0	σ_1	$\sigma_{\geq 2}$
θ_μ	3.8	5.2	7.9
θ_{res}	0.1	4.5	7.9
$\theta_{0/1}$	-4.2	7.9	3.9
$\theta_{1/2}$	-	-6.8	16.1
Total	5.6	12.5	19.9
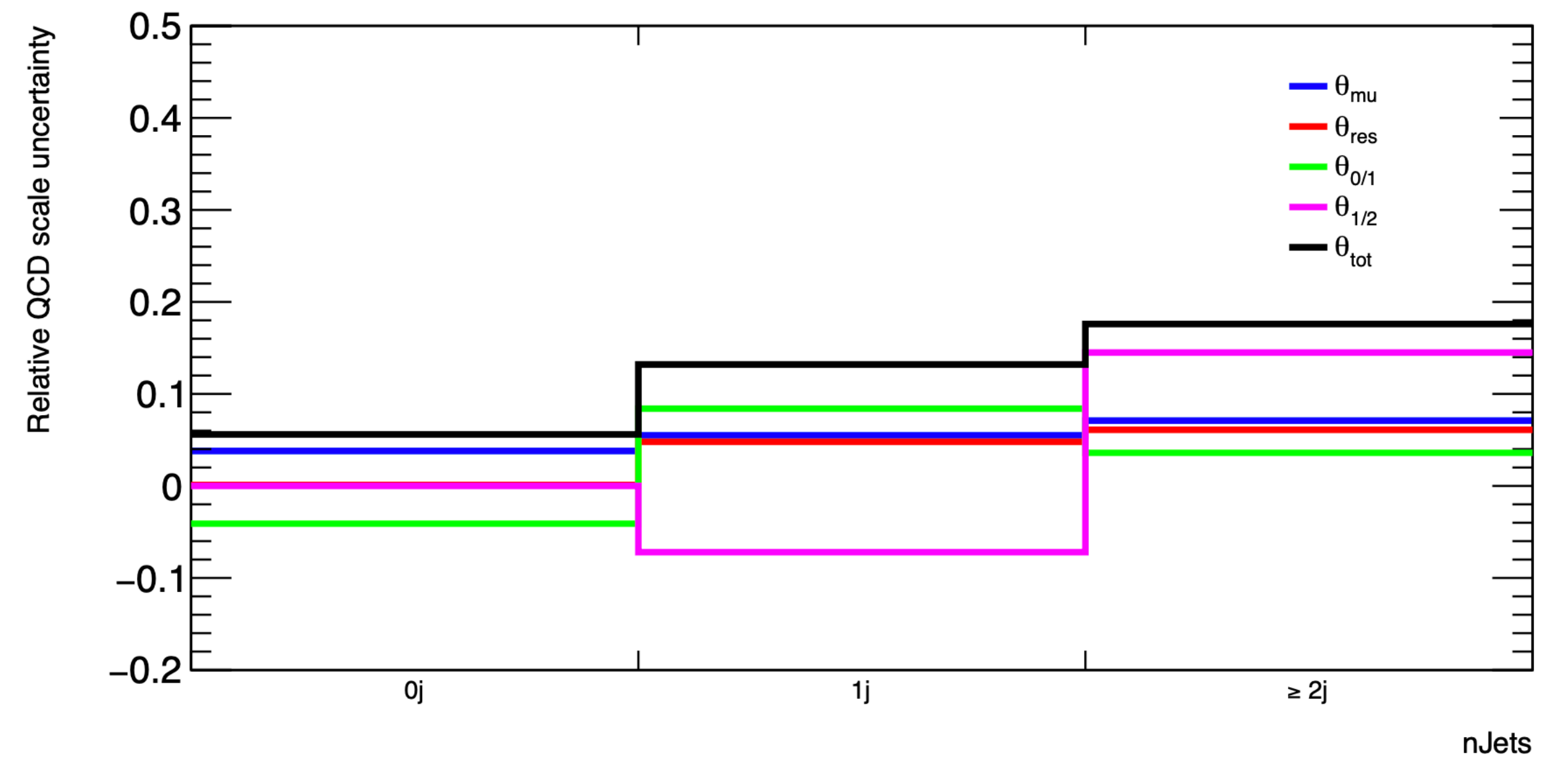
Low p_T^H region - 0 jet topology

- Scheme developed a few years ago - approved as part of the Stage 1.1 scheme - Link
  - Envelop of HNNLO NNLOPS/muR/muF scale taken as the sys
- No application of LR method - dominant effects are from low p_T resummation + only one bin boundary
  - **Care taken** to ensure that uncertainty in the region are inline with calculations
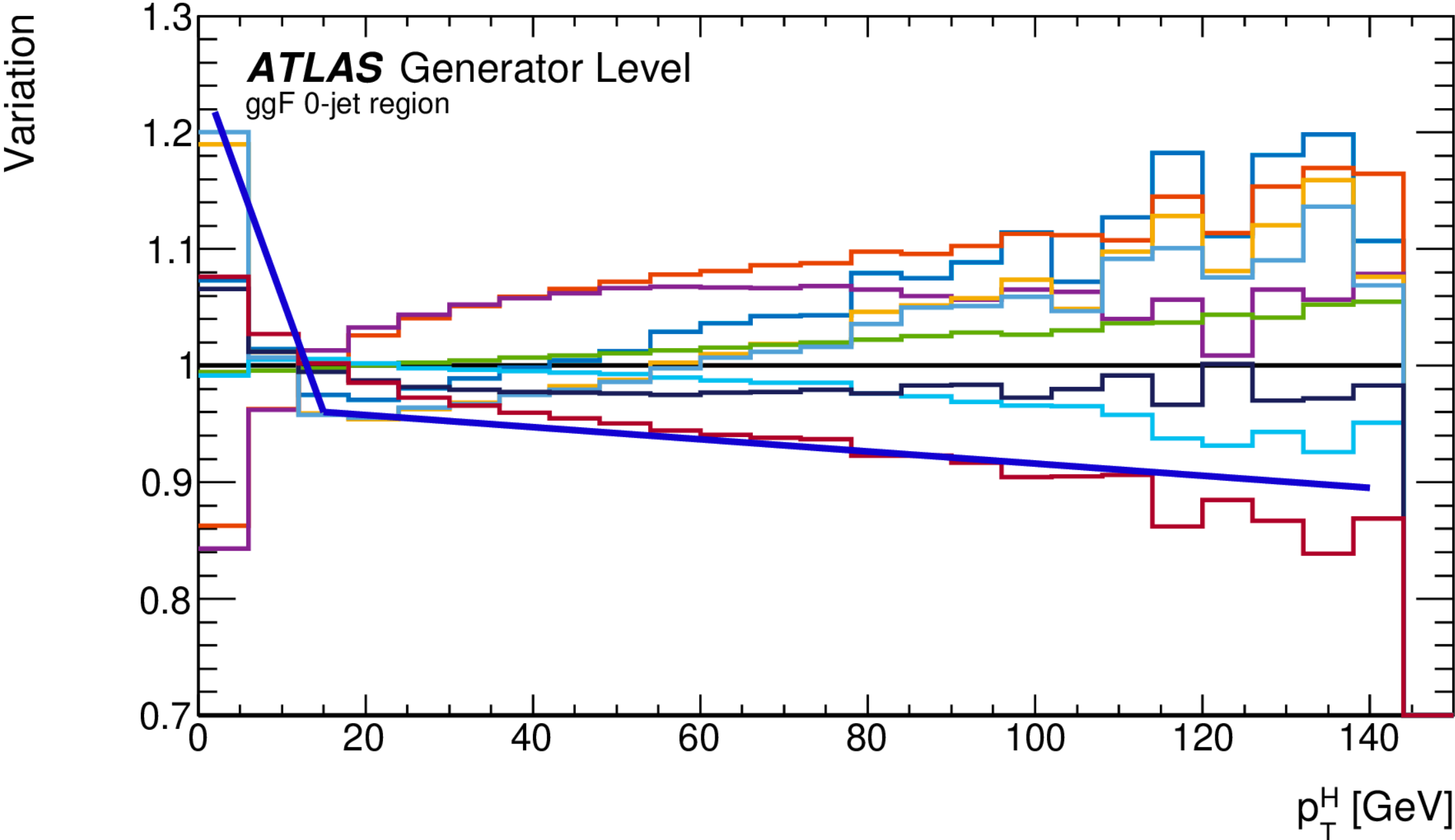
Uncertainty [%]	p_T^H [GeV] region in 0j		
	$\sigma_{=0j}$	$\sigma_{<10}$	$\sigma_{\geq 10}$
$\theta_{p_T^H=10}$		11.2	-3.6

Low p_T^H region - 1/2 jet topology

- Similar as 0 jet - but dominant uncertainty source **expected** to be covered by muR/muF/HNNLO variations

- Following a similar procedure as Stage 1 scheme - but applying LR ST method in the middle

  - Use the largest scale variation to define the input

- Also checked the scale variations in m_{jj} bins - consistent within the statistical error

  - Assume for now, p_T^H uncertainty is **independent** of m_{jj}
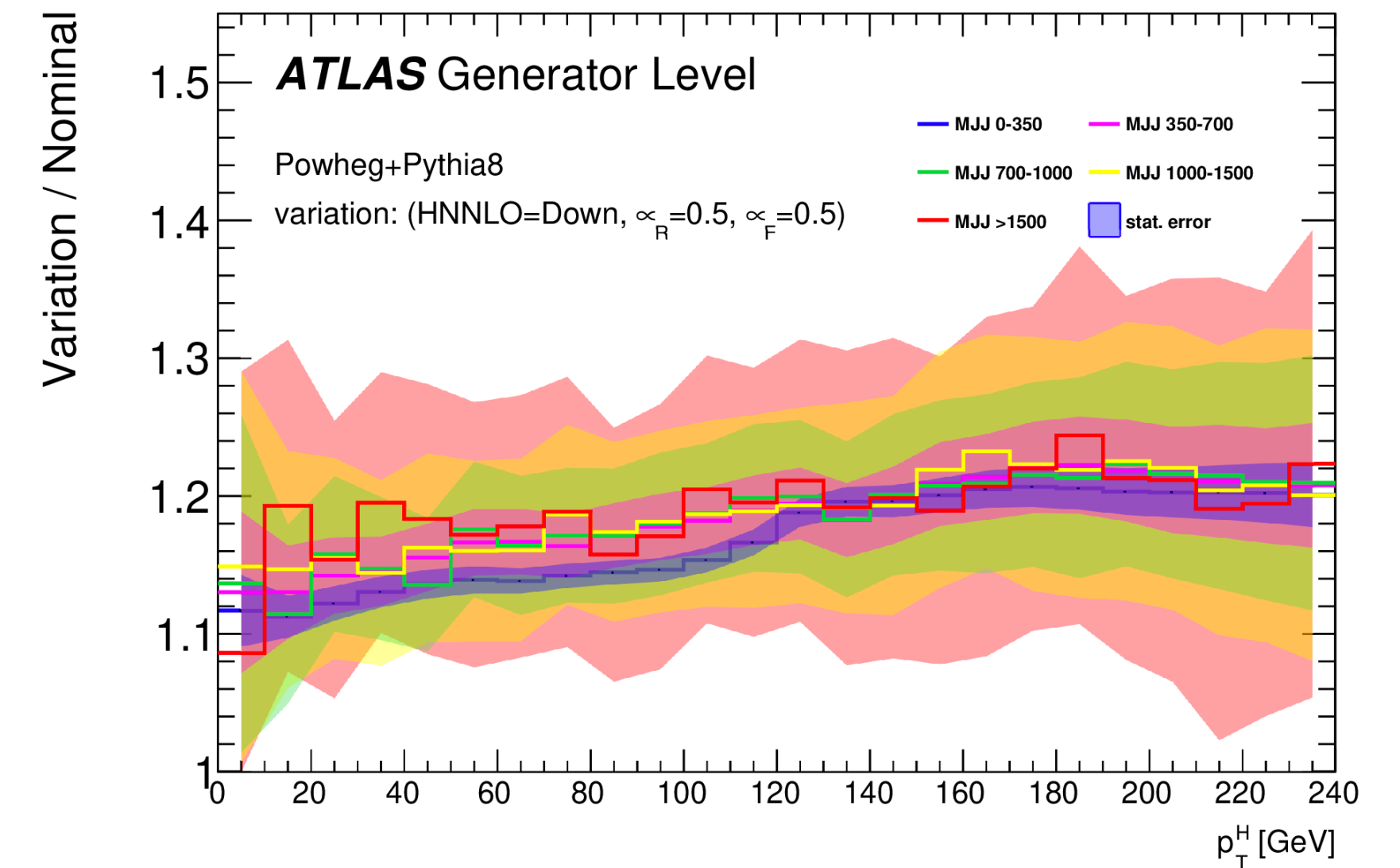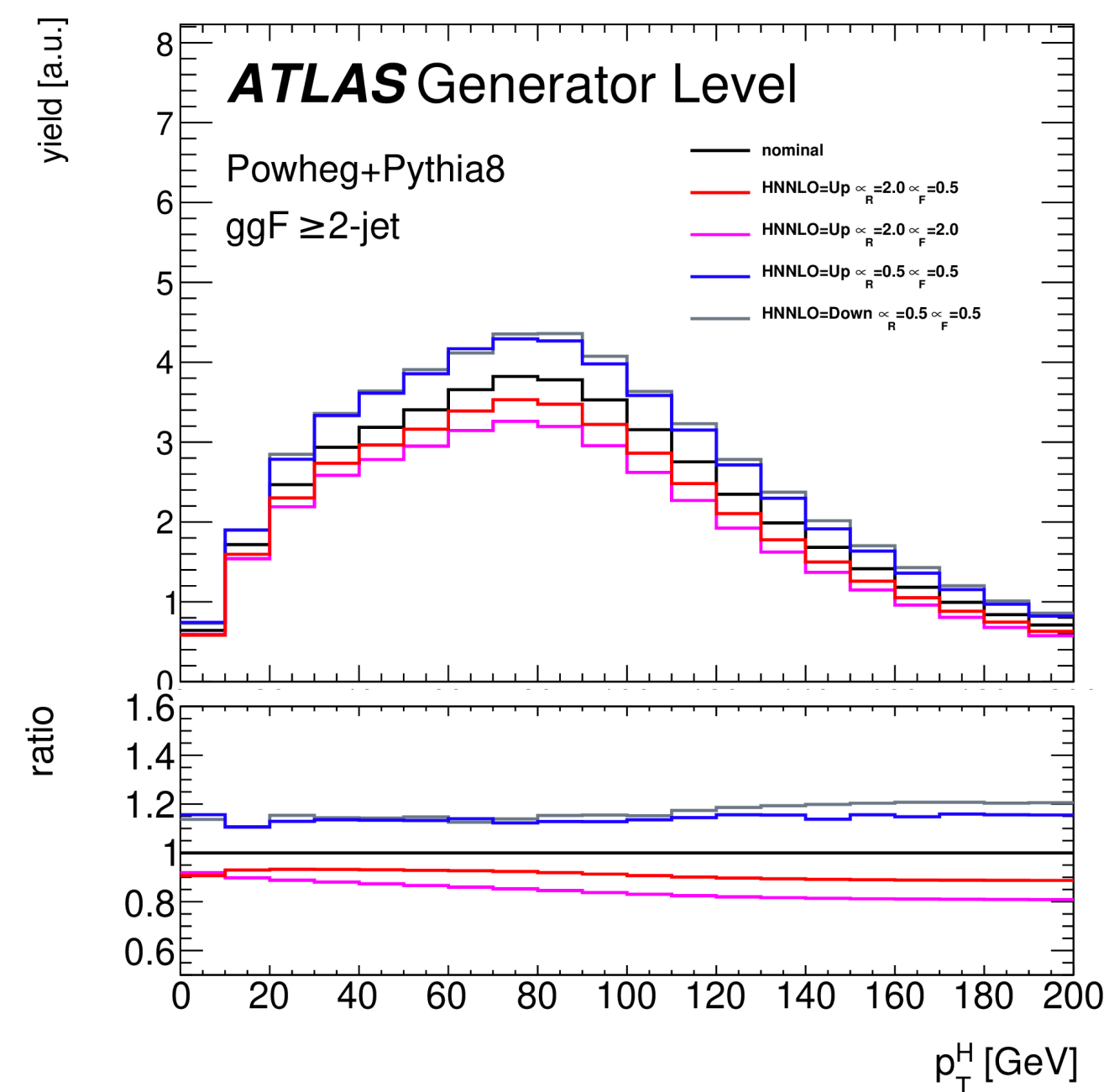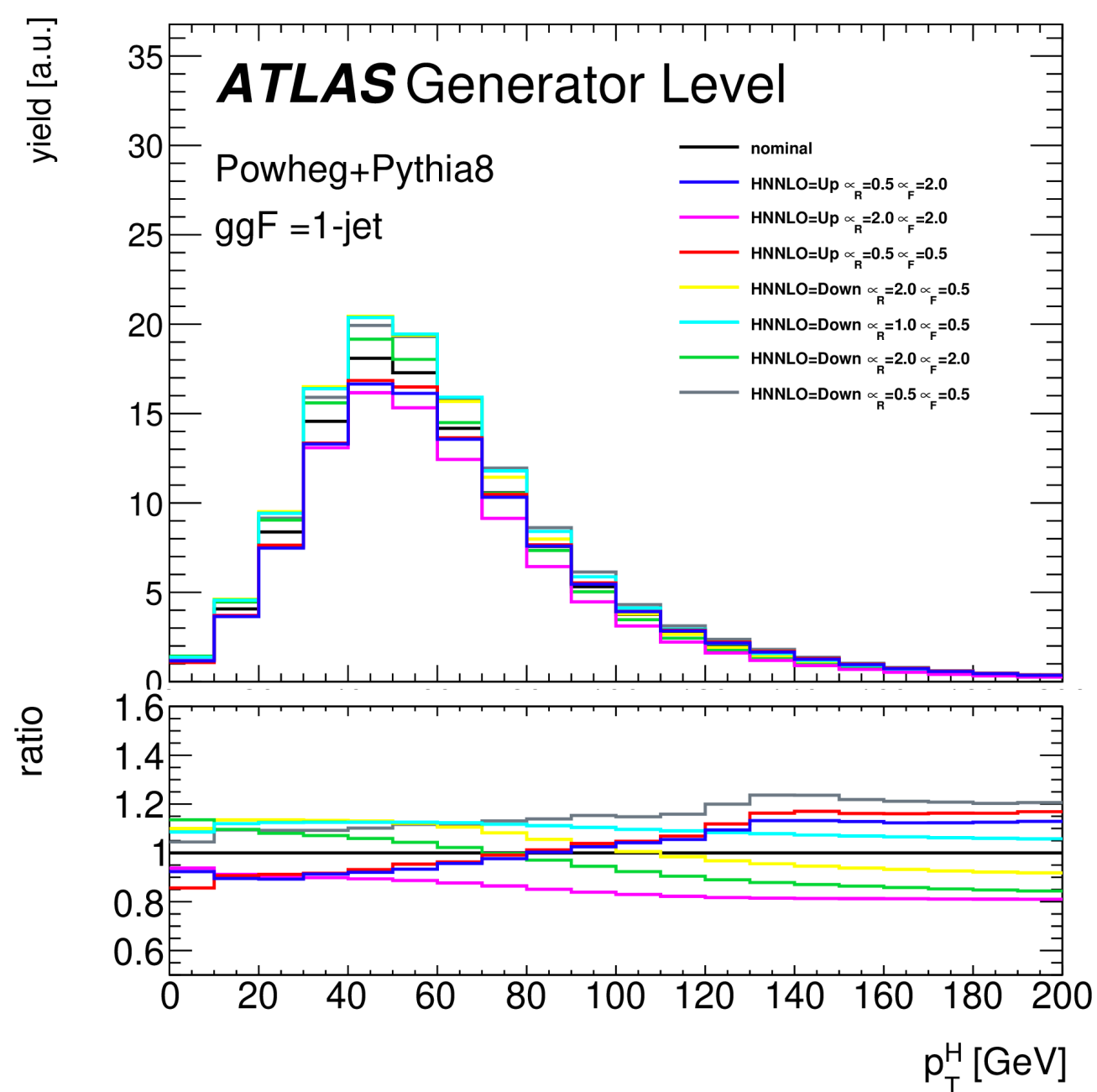
Low p$_T^H$ region - 1/2 jet topology

- **Derive** the results in ≥ 1 region

  - Since p$_T^H$ shape changes in jet regions, apply a different **smoothing** function to distribute the XS impact evenly across

- Only place where smoothing is applied - p$_T^H$ is typically correlated with acceptance effects due to analysis selection

  - Smoothing allows to get the impact of these acceptance effects

  - Non-trivial to parameterize other variables once p$_T^H$ has been - **tackle in the next iteration** of the scheme

| Uncertainty [%]	Powheg NNLOPS			
	$\sigma_{\geq 1}$	p_T^H [GeV] region		
		$\sigma_{[0,60)}$	$\sigma_{[60,120)}$	$\sigma_{[120,200)}$
θ_y	13.1	13.1	13.1	13.1
$\theta_{p_T^H=60}$		-8.1	+7.6	+7.6
$\theta_{p_T^H=120}$			-2.9	+10.3
Total	13.1	15.4	15.5	18.3
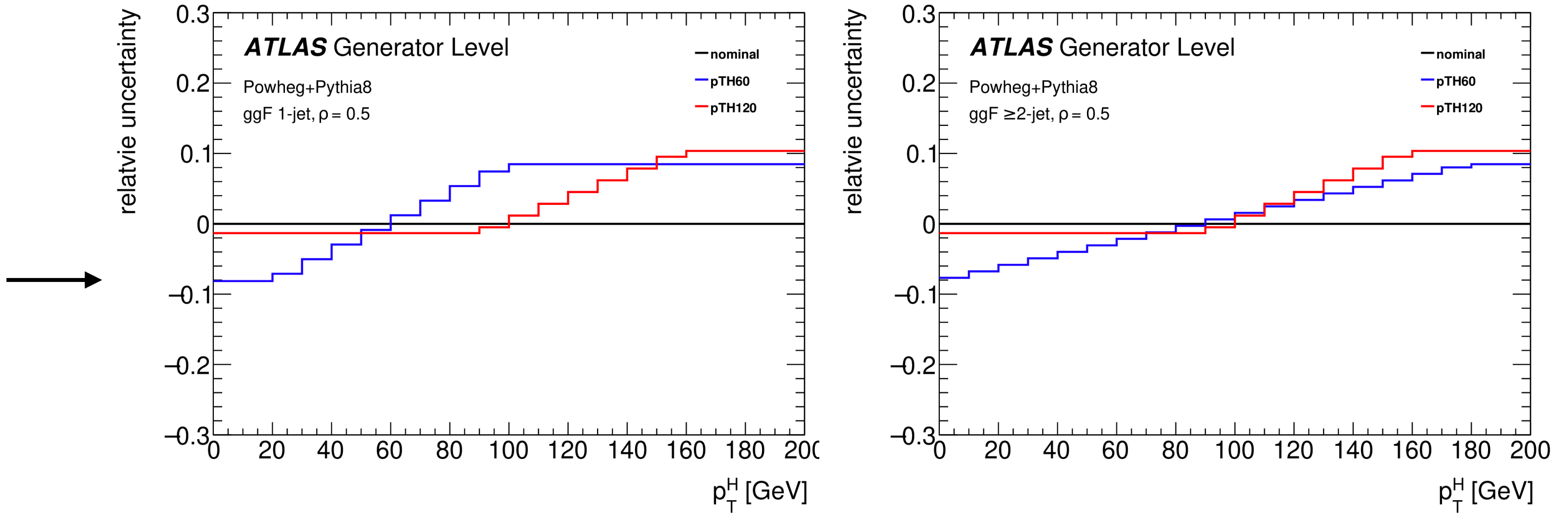
2 Jet region - m_{jj}

- Uncertainties for m_{jj} are a simple application of the LR method
  - Scale variation from FxFx used as it is NLO @ 2j
- Cross-checked results with NNLOPS, MG5 H+2J and Hjj MiNLO
- Ensured that migration uncertainties cancel out when applied to NNLOPS

Dijet mass, m_{jj} [GeV]
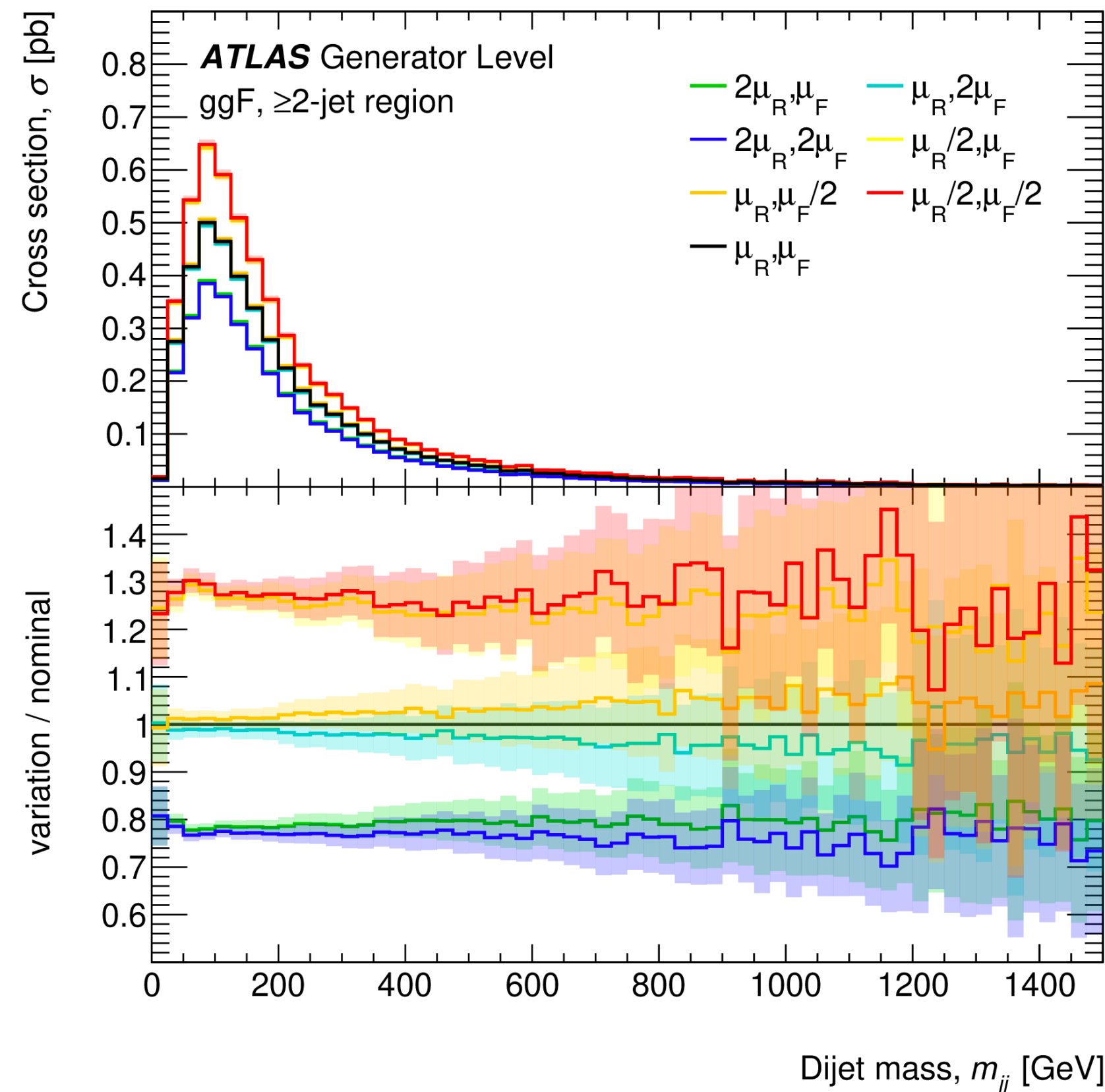
Uncertainty [%]	$H + 0, 1, 2j$ MG5_AMC@NLO (FxFx)					
			m_{jj} [GeV] region			
	$\sigma_{\geq 2j}$	$\sigma_{<350}$	$\sigma_{[350,700)}$	$\sigma_{[700,1000)}$	$\sigma_{[1000,1500)}$	$\sigma_{\geq 1500}$
θ_y	23.0	23.0	23.0	23.0	23.0	23.0
$\theta_{m_{jj}=350}$		-2.9	+11.8	+11.8	+11.8	+11.8
$\theta_{m_{jj}=700}$			-5.7	+12.4	+12.4	+12.4
$\theta_{m_{jj}=1000}$				-11.1	+12.6	+12.6
$\theta_{m_{jj}=1500}$					-6.8	+13.0
Total	23.0	23.2	26.5	30.7	32.0	33.9

- p_T^{Hjj} is an indirect probe for N_{jet} - however, there is significant leakage at the $p_T^{Hjj} = 25$ GeV boundary
  - Consistently found the same behaviour across generators
  - Need a better probe for $N_{jet} = 2 \leftrightarrow N_{jet} \geq 3$ migrations

- Leads to an **increase** in the apparent systematic in the lower bin
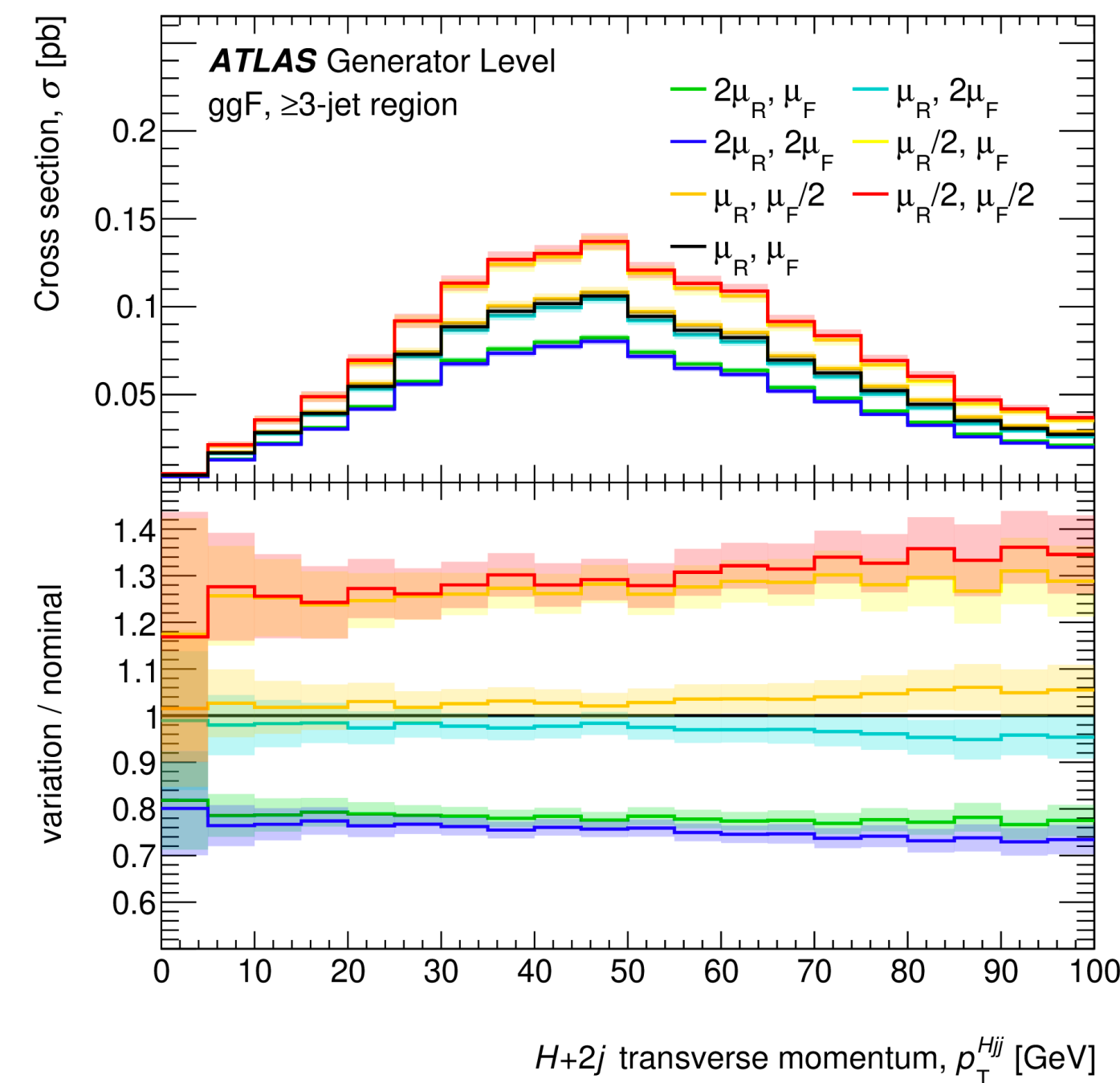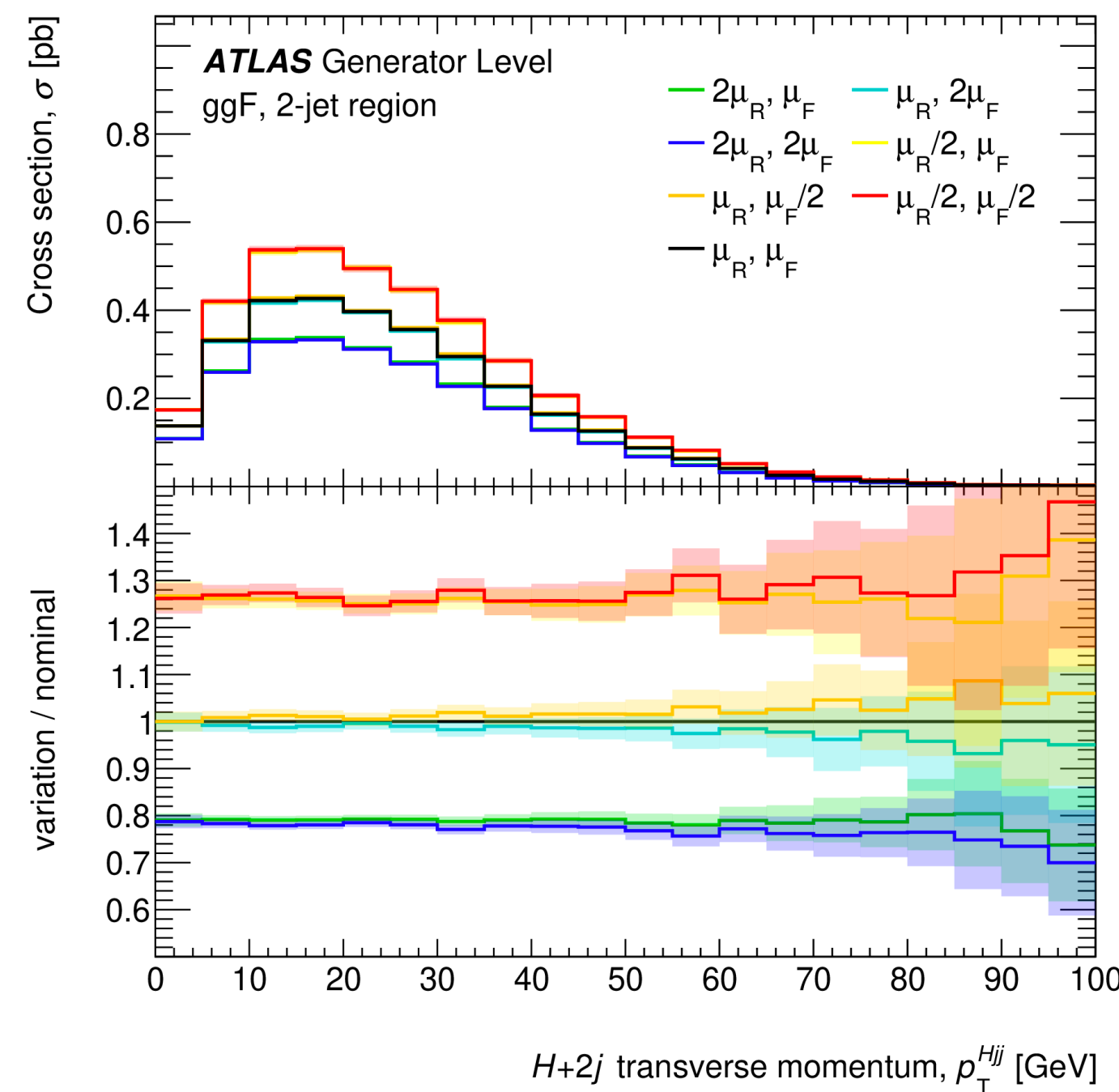
Composition [%]	$H + 0, 1, 2j$ MG5_AMC@NLO (FxFx)	
	p_T^{Hjj} [GeV] region	
	$\sigma_{<25}$	$\sigma_{\geq 25}$
$N_{jet} = 2$	91	9
$N_{jet} \geq 3$	55	45

2 Jet region - p_T^{Hjj}

- Due to NLO (FxFx) vs LO (NNLOPS) shape differences, the impact in the upper bin **increased** to O(30%) prevent a overall XS impact

- Cross-checked results with NNLOPS, MG5 H+2J and Hjj MiNLO

- ρ set to 1 as there is only one bin

Uncertainty [%]	$H + 0, 1, 2j$ MG5_AMC@NLO (FxFx)		
	p_T^{Hjj} [GeV] region		
	$\sigma_{\geq 2j}$	$\sigma_{<25}$	$\sigma_{\geq 25}$
θ^y	23.0	23.0	23.0
θ^{25}		-33.7	+30.0
Total	23.0	40.8	37.8

Increased from 23.8% to 30% to account for relative XS difference in NNLOPS wrt MG5 FxFx
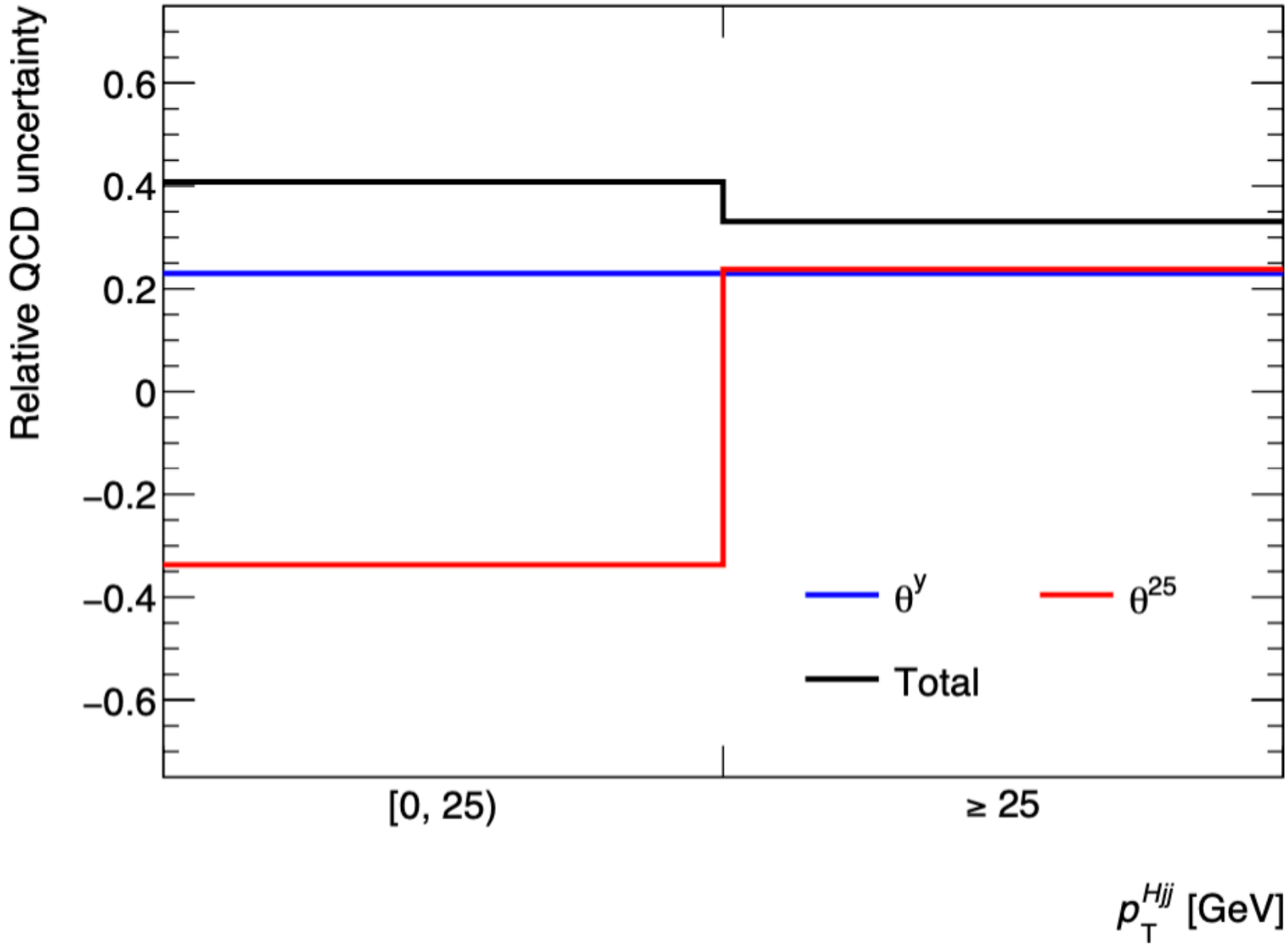
High p$_T^H$ region - Scale variation

- Significant improvements to the uncertainty scheme in p$_T^H$ > 200 GeV region

- Dedicated theoretical calculations and associated QCD scale uncertainty - 1802.00349

- Matthias Kerner & Stephen Jones extended and provided results in the needed binning

  - Ensured that these results are consistent with the NNLOPS results

- Very recently found that the top mass effect was overestimated - numbers will be updated ASAP
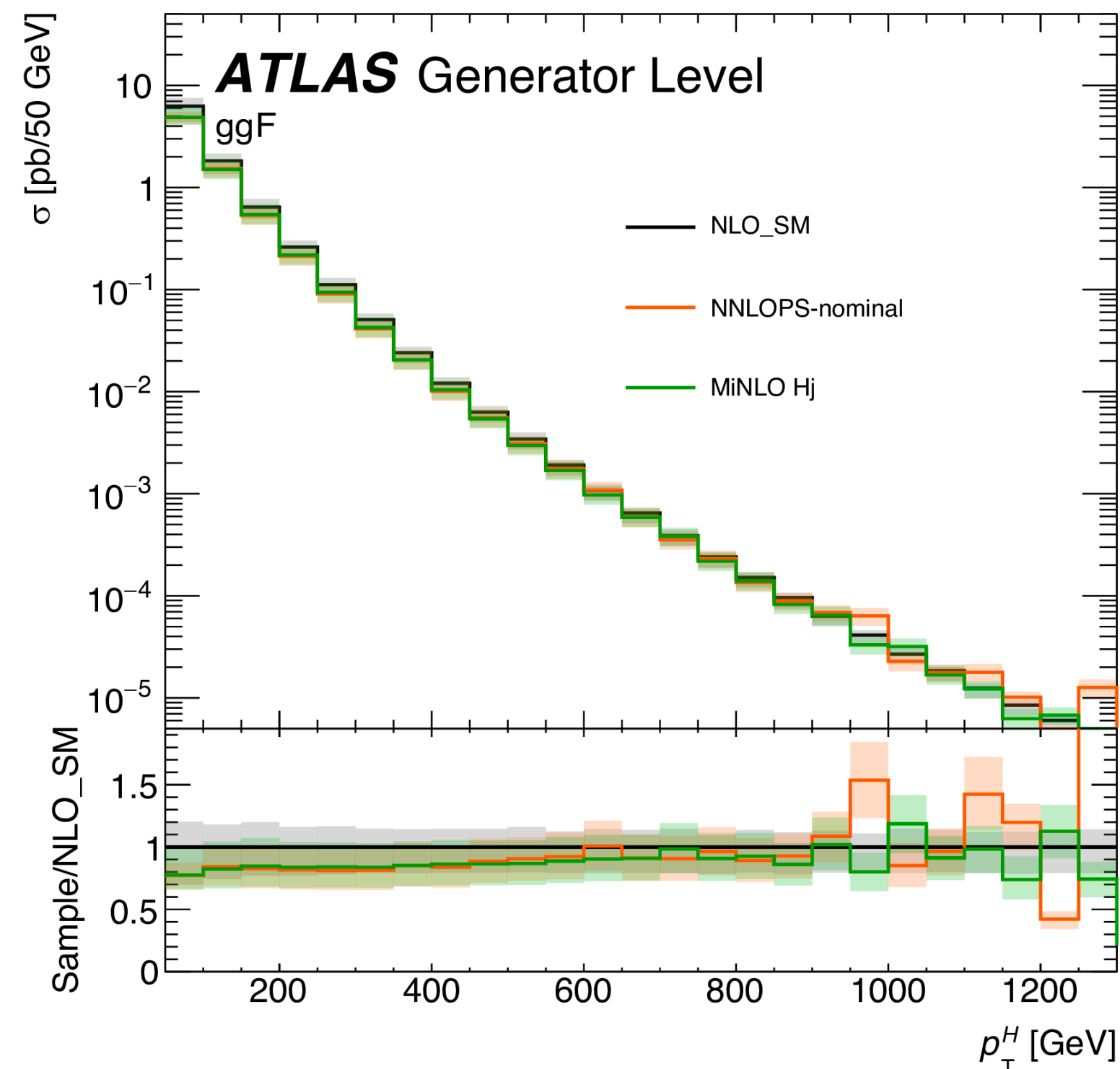
Comparison of central value
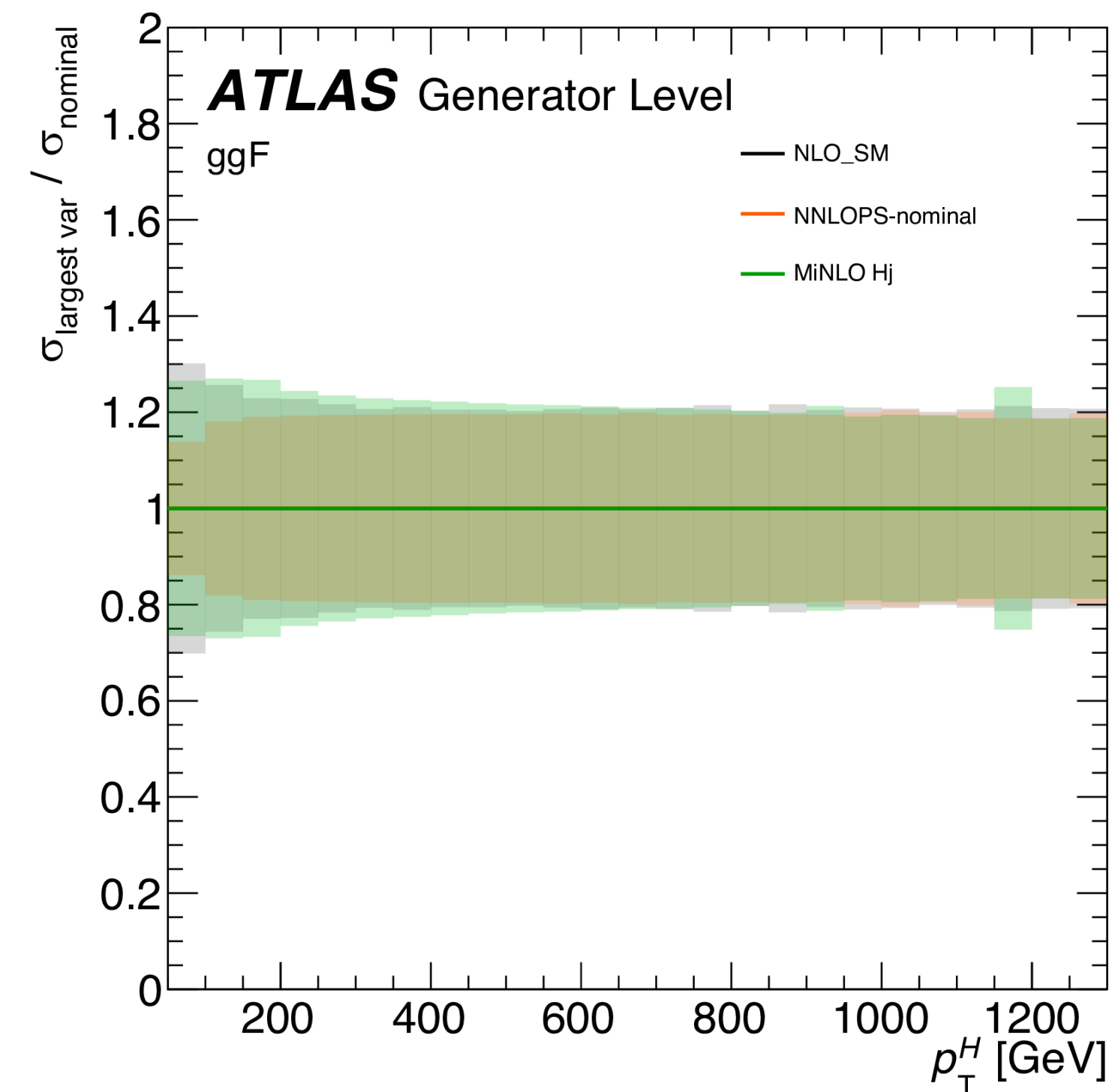
Comparison of uncertainty band

High $p_T{}^H$ region - Scale variation

- Afterwards a **normal** application of the LR method
- In this case, yield migration is **kept** as a separate NP
  - BLPTW is not expected to cover this region

Uncertainty [%]	NLO_SM				
	p_T^H [GeV] region				
	$\sigma_{\geq 200}$	$\sigma_{[200,300)}$	$\sigma_{[300,450)}$	$\sigma_{[450,650)}$	$\sigma_{\geq 650}$
$\theta_{high-p_T^H}$	22.1	22.1	22.1	22.1	22.1
$\theta_{p_T^H=300}$	-	-2.8	10.4	10.4	10.4
$\theta_{p_T^H=450}$	-	-	-1.7	10.3	10.3
$\theta_{p_T^H=650}$	-	-	-	-1.4	10.4
Total	22.1	22.2	24.4	26.5	28.4
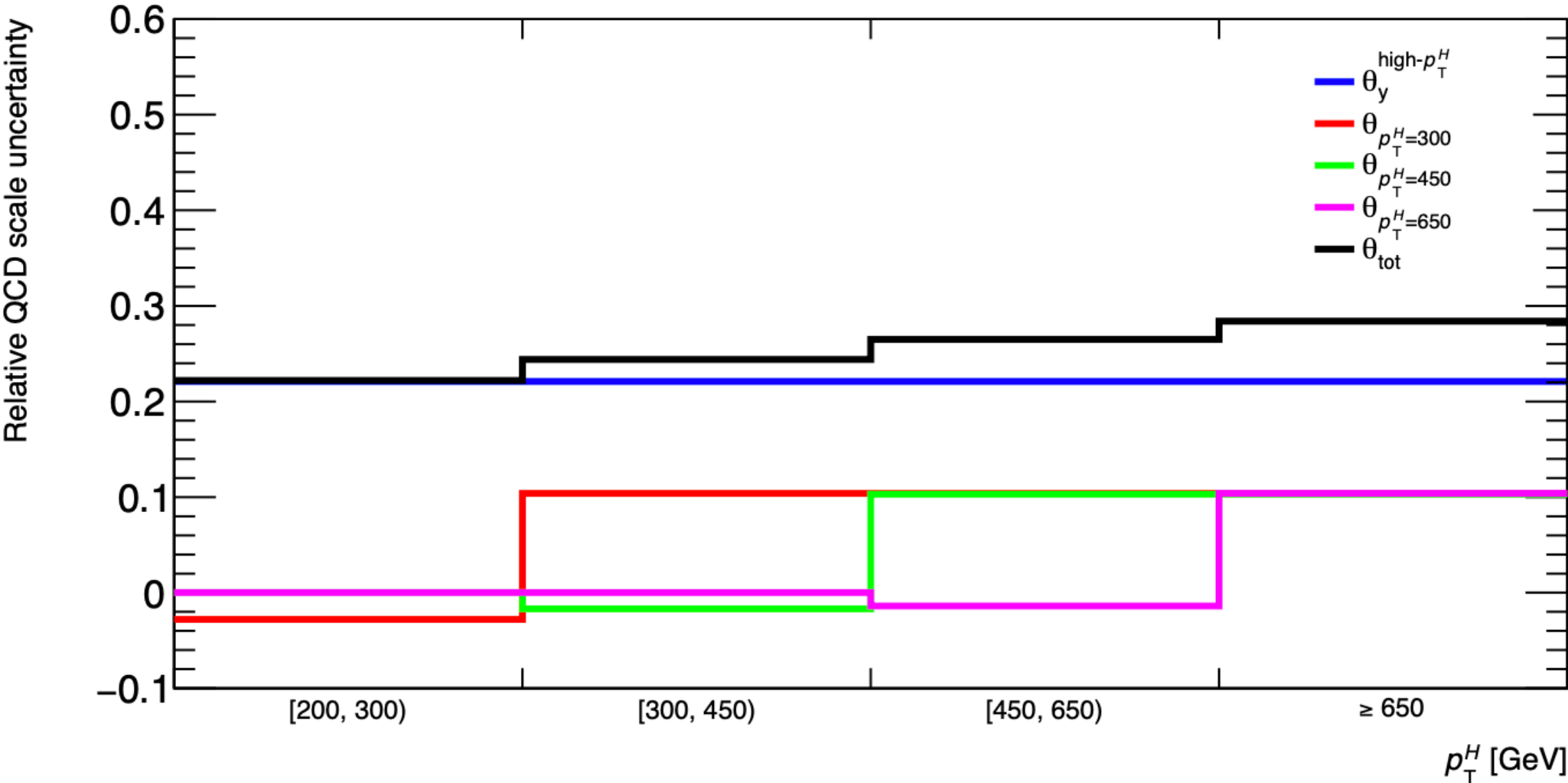
High p_T^H region - p_T^{Hj}/p_T^H

- p_T^{Hj}/p_T^H to account for **nJet migration** - similar to p_T^{Hjj}

  - **Checked** to ensure cut at 0.15 is a good probe for this effect

- ρ set to 1 as there is only one bin

- **Results** cross-checked with MG FxFx sample

Uncertainty [%]	NNLOPS		
	p_T^{Hj}/p_T^H region		
	$\sigma_{\geq 0}$	$\sigma_{[0,0.15)}$	$\sigma_{\geq 0.15}$
θ_y	20.7	20.7	20.7
$\theta_{p_T^{Hj}/p_T^H = 0.15}$	-	-51.0	18.1
Total	20.7	55.0	27.4
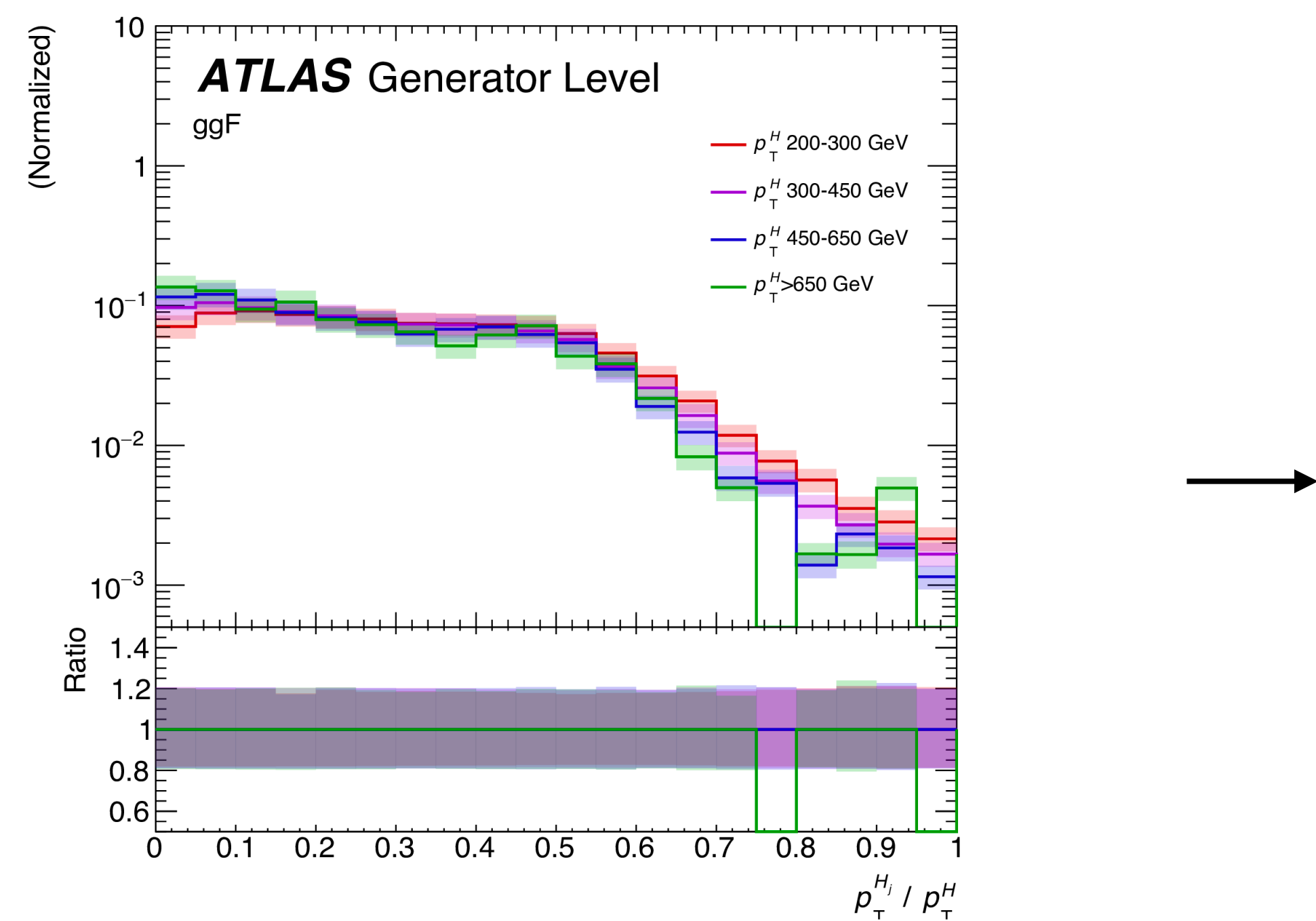
pTHj/pTH variations are not pTH dependant

High p$_T$^{H} region - top mass scheme

- Various calculations show that top mass scheme can lead to **different** prediction in the high p$_T$^{H} region - source of uncertainty

- Calculations with MSbar and pole mass only available at LO - Micheal Spira

- Calculations of other processes show ~ 2x reduction in the difference at NLO

  - Take **half** of the variation for Higgs p$_T$^{H} as a systematic variation

$\longrightarrow$
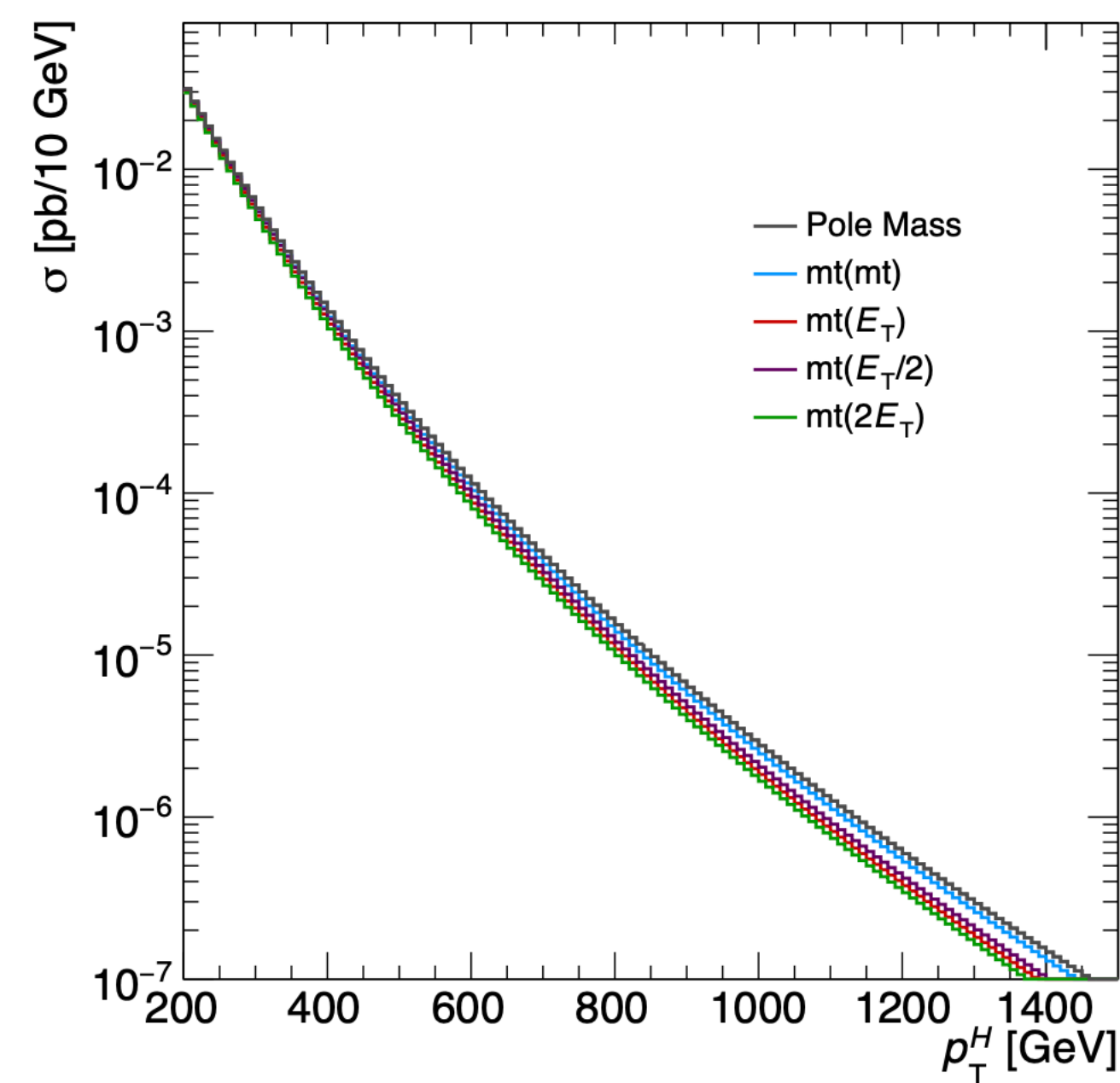
		LO finite top mass samples			
Uncertainty [%]	$\parallel$	p_T^H [GeV] region			
	$\sigma_{\geq 200}$	$\sigma_{[200,300)}$	$\sigma_{[300,450)}$	$\sigma_{[450,650)}$	$\sigma_{\geq 650}$
$\theta_{m_{\text{top}}}$ $\parallel$	3.7	2.8	6.4	10.3	14.1

Acceptance effects + uncovered variables

Acceptance effect + other variables

- Almost all uncertainty numbers are **applied** flat in a STXS bin

  - If analysis selection shapes the acceptance or ML algorithm uses the variable, the uncertainty will **factorize**

  - Many other QCD sensitive variables not **covered** by this scheme

  - This leads to an **underestimation**

- Way around it - **provide** one scale variation (e.g. muF = muR = 0.5) as part of the implementation of the scheme

- To avoid **'significant'** double counting, **normalize** scale variation in STXS bin to remove overall XS

  - To decide: if this variation should be one overall NP or one NP per bin or somewhere in between

- This proposal has some still has **double counting**

  - If NP is pulled/constrained/ranked highly - ask analysis to do detailed checks & make decisions on an case-by-case level

  - Leave this up to the collaboration to define how to implement this

Conclusion

- **Significant work** has been invested in defining the uncertainty scheme for the ggF Stage 1.2 STXS scheme
  - Collaborative effort between ATLAS, CMS and theorists!

- Complete version of both the nuisance parameter scheme and the associated numbers available!
  - Results documented @ Link which can be implemented by the analyses
- Plan to document these results in a note in the future!

- **Potential improvements** to the scheme in the future
  - Correlated effects across multiple dimensions
  - Smooth parameterization across variables
  - Any other feedback from the community!

Backup

Evaluating the uncertainty

- Many ways - all involve **varying** the renormalization & factorization scale and using the XS variations in some NP scheme

- **Our current go-to** is the ST method - One NP for overall yield variation and other NP for migration between categories
  - Removes the accidental cancellation of scale variations

Bin definitions
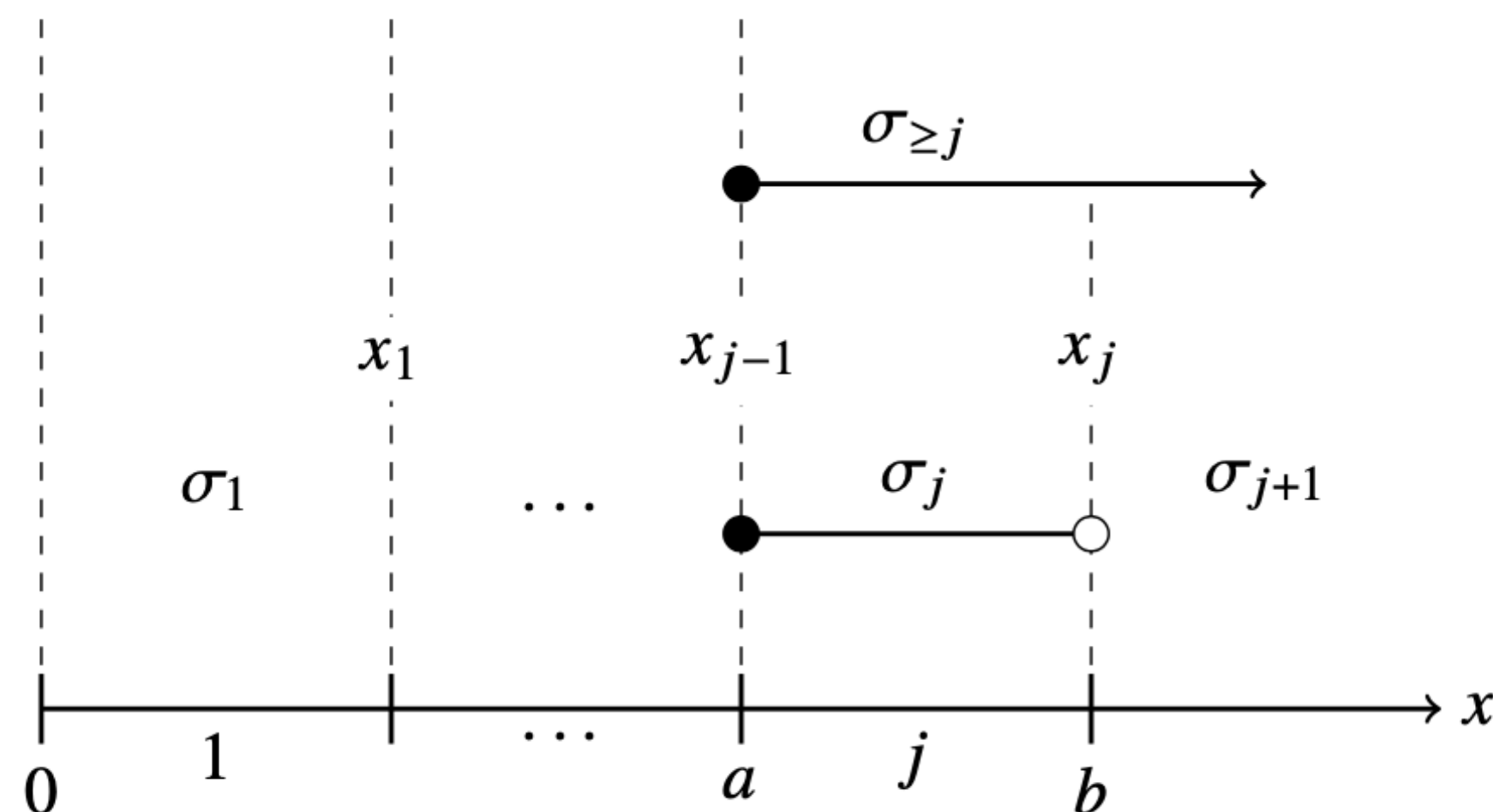
ST method

$$\theta^y(j) \begin{cases} 0 & : j < k \\ +\max |\Delta_{\mu, \geq k}|/\sigma_j & : j = k \\ 0 & : j \geq k+1 \end{cases}$$

Take the max scale variation in $\geq$ k region and apply as yield NP in k bin

$$\theta^{\mathrm{mig}}_{x_k}(j) \begin{cases} 0 & : j < k \\ -\max |\Delta_{\mu, \geq k+1}|/\sigma_j & : j = k \\ +\max |\Delta_{\mu, \geq k+1}|/\sigma_j & : j = k+1 \\ 0 & : j \geq k+2 \end{cases}$$

Take the max scale variation in $\geq$ k+1 region and apply as migration NP in k and k+1 bin

- But this method **breaks down** in the case of many or small bin width - unphysical blow up of uncertainty if XS is small

- For continuous variables, like pTH, it makes no sense that migration will be only between two neighbouring bins

Long range ST method

θ^y as the max scale variation
in the inclusive x region - $\delta_y = \max(\Delta_\mu)/\sigma$

θ_a^{mig} as the max scale variation
in the x > a region. First bin as the negative of
this value as the uncertainty
$\delta_a^+ = \max(\Delta_{x>a})/\sigma_{x>a}$
$\delta_a^- = -\max(\Delta_{x>a})/\sigma_{x<a}$

θ_b^{mig} as the max scale variation
in the x > b region. Second bin as the negative
of this value, with no sys applied to the first bin
$\delta_b^+ = \max(\Delta_{x>b})/\sigma_{x>b}$
$\delta_b^- = -\max(\Delta_{x>b})/\sigma_{a<x<b}$

	Kinematic observable x		
Nuisance parameter	$0 \leq x \leq a$	$a \leq x \leq b$	$x \geq b$
θ^y	$+\delta_y$	$+\delta_y$	$+\delta_y$
θ_a^{mig}	$\rho\delta_a^-$	$\rho\delta_a^+$	$\rho\delta_a^+$
θ_b^{mig}		$\rho\delta_b^-$	$\rho\delta_b^+$

ρ chosen to remove the overlap
between x > a and x > b uncertainty values

Samples used for theory sys

Generated Samples						
Name	Description	Generator	Shower	PDF	Variations	Other Notes
Powheg NNLOPS	H+0j@NNLO	Powheg	Pythia 8, AZNLO tune	PDF4LHC15_nlo_30_pdfas	(μ_R, μ_F) variations (7-point NLO, 3-point NNLO)	NNLOPS reweighting Rescaled for quark mass effects
MiNLO HJ	H+1j@NLO	Powheg	Pythia 8, AZNLO tune	PDF4LHC15_nlo_30_pdfas	7-point (μ_R, μ_F) variations	Rescaled for quark mass effects bornktmin=200
MiNLO HJJ	H+2j@NLO	Powheg	Pythia 8, AZNLO tune	PDF4LHC15_nlo_30_pdfas	7-point (μ_R, μ_F) variations	Rescaled quark mass effects
H+1j MG5_aMC@NLO	H+1j@NLO	MG5_aMC@NLO	Pythia 8, AZNLO tune	NNPDF30_nlo_as_0118	7-point (μ_R, μ_F) variations	HC_NLO_X0-heft model $m_{top} = \inf, m_b = 0$
H+0,1,2j MG5_aMC@NLO (FxFx)	H+0,1,2j@NLO	MG5_aMC@NLO	Pythia 8, AZNLO tune	NNPDF30_nlo_as_0118	7-point (μ_R, μ_F) variations	FxFx merging merging scale = 30 GeV HC_NLO_X0-heft model $m_{top} = \inf, m_b = 0$

Calculations						
Name	Description	Reference	PDF	Variations		Other Notes
NLO_SM	H+1j@NLO with finite top mass	arXiv:1802.00349	PDF4LHC15_nnlo_mc	7-point variations of (μ_R, μ_F) around $E_T = \sqrt{m_H^2 + p_{T,H}^2}$		mt = 173.05 GeV $\geq$ 1 jet with $p_T > 30$ GeV
Pole mass (and other top mass variations)	H+1j@LO with finite top mass	arXiv:2003.01700, arXiv:1811.05692 arXiv:2003.03227, arXiv:2008.11626	-	$\overline{MS}$ mass variations mt(mt), mt(E_T/2), mt(E_T), mt($2E_T$), with $E_T = \sqrt{m_H^2 + p_{T,H}^2}$		