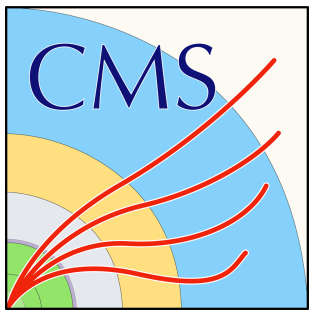
ggH stage 1.2: uncertainty prescription

Haider Abidi, Jonathon Langford

work performed within the fid/diff/STXS working group

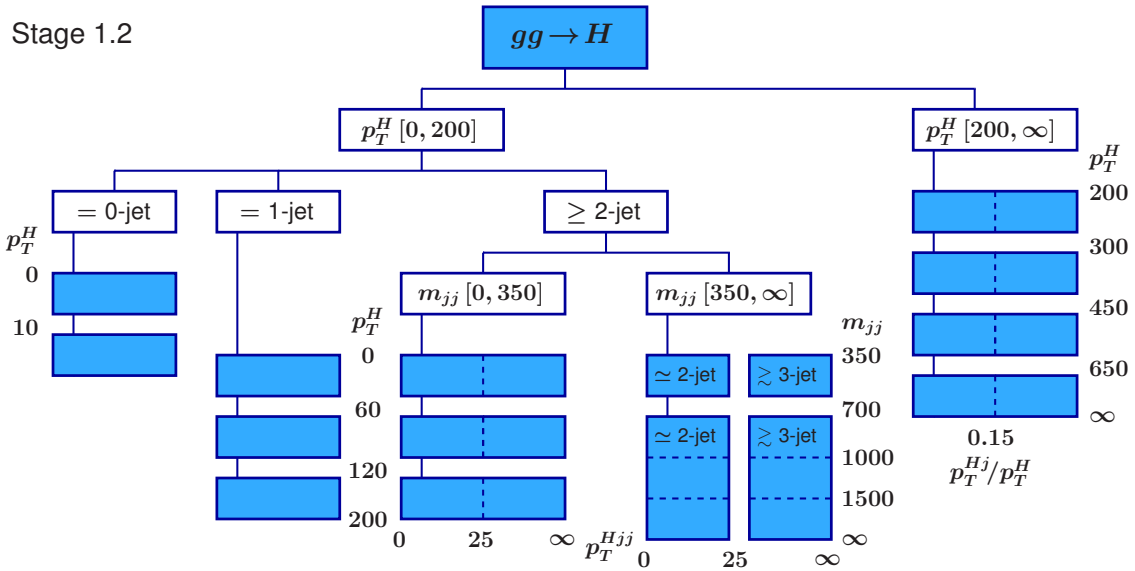
The 17th Workshop of the LHC Higgs WG
10 Nov 2020



Imperial College London

Introduction

- Many analyses targeting STXS stage 1.2 measurements + interpretations
- **Goal**: common prescription for stage 1.2 uncertainties
  - evolution of stage 1.0 scheme[1]: ▸ Back-up

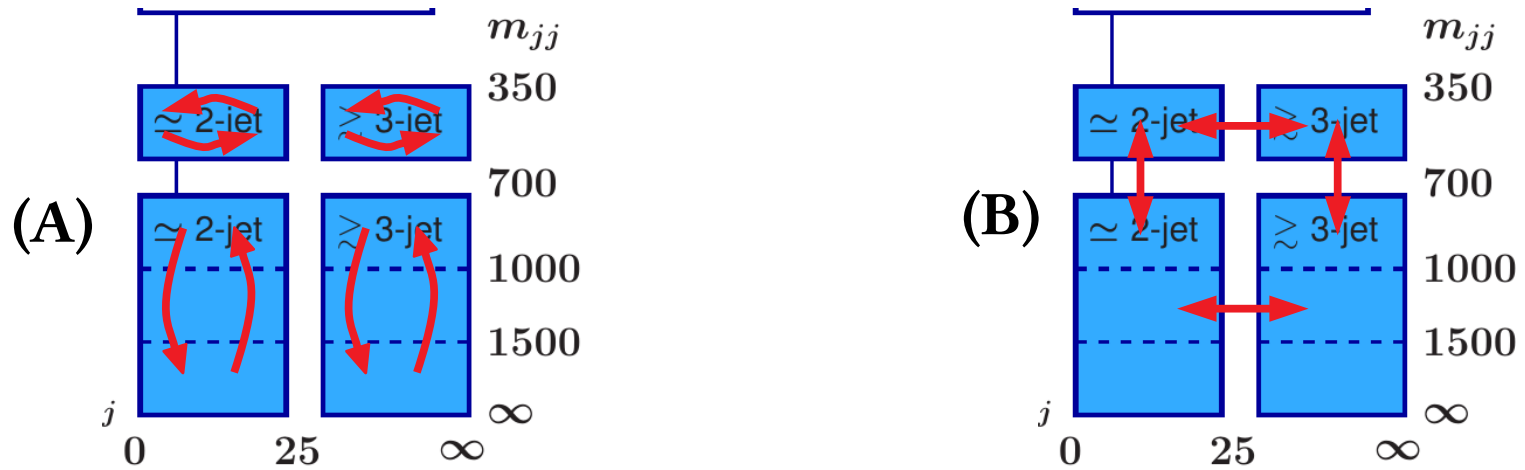


- **This talk**: summarise discussions in fid/diff/STXS working group
  - Proposal for stage 1.2 scheme: numbers are **still preliminary**
    $\Rightarrow$ methodology likely to require tweaking, do not take as final words!
  - Thoughts on how to use scheme
  - Open questions: limitations (residual shape effects), input from theory etc

[1] also known as 2017 scheme

Overview

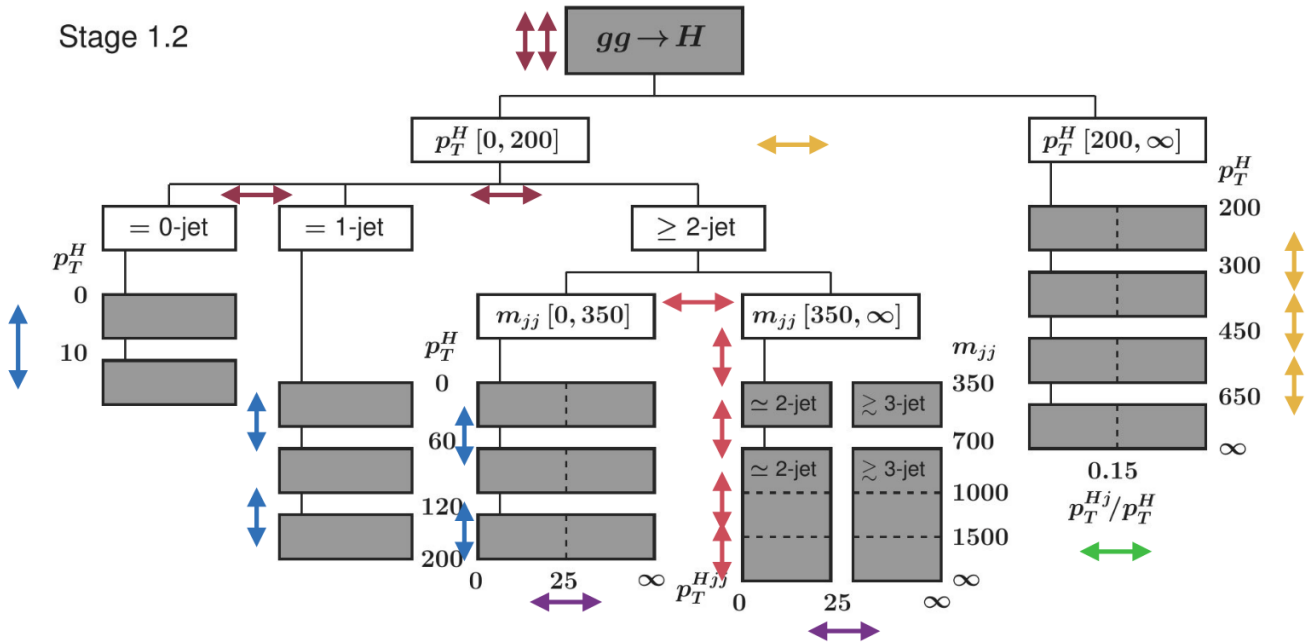
- There are two types of uncertainty that must be addressed...
- A) Uncertainty in the measurement: σ_i
  - **within-bin migrations** i.e. kinematic/shape effects
  - affects the estimated experimental acceptance: $N_{\text{obs}} = \sigma \cdot \text{BR} \cdot (\epsilon \cdot \mathcal{A}) \cdot \mathcal{L}$
  - visible only to the experiment
- B) Uncertainty in the interpretation: e.g. μ_i, κ, EFT
  - yield + **migrations across bin boundaries**
  - affects predicted cross section in a given bin
  - to be used by theory community (test new models) & in-house by experiments



Overview (2)

(A): within-bin migrations, (B): across-bin migrations

- Unfortunately, distinction/interplay between (A) and (B) is not simple!
  - when we merge bins in the measurement: (B) $\Rightarrow$ (A)
  - increase splitting: reduce (A), increase (B)
- Define uncertainty scheme at the **granularity of the dashed boundaries**
  - nuisance params (NP) to account for migrations across all boundaries: $\{\Delta^i\}$
  - dashed boundaries provide handle for important within-bin migrations
  - **Summary**: defines (B), whilst capturing some of (A)



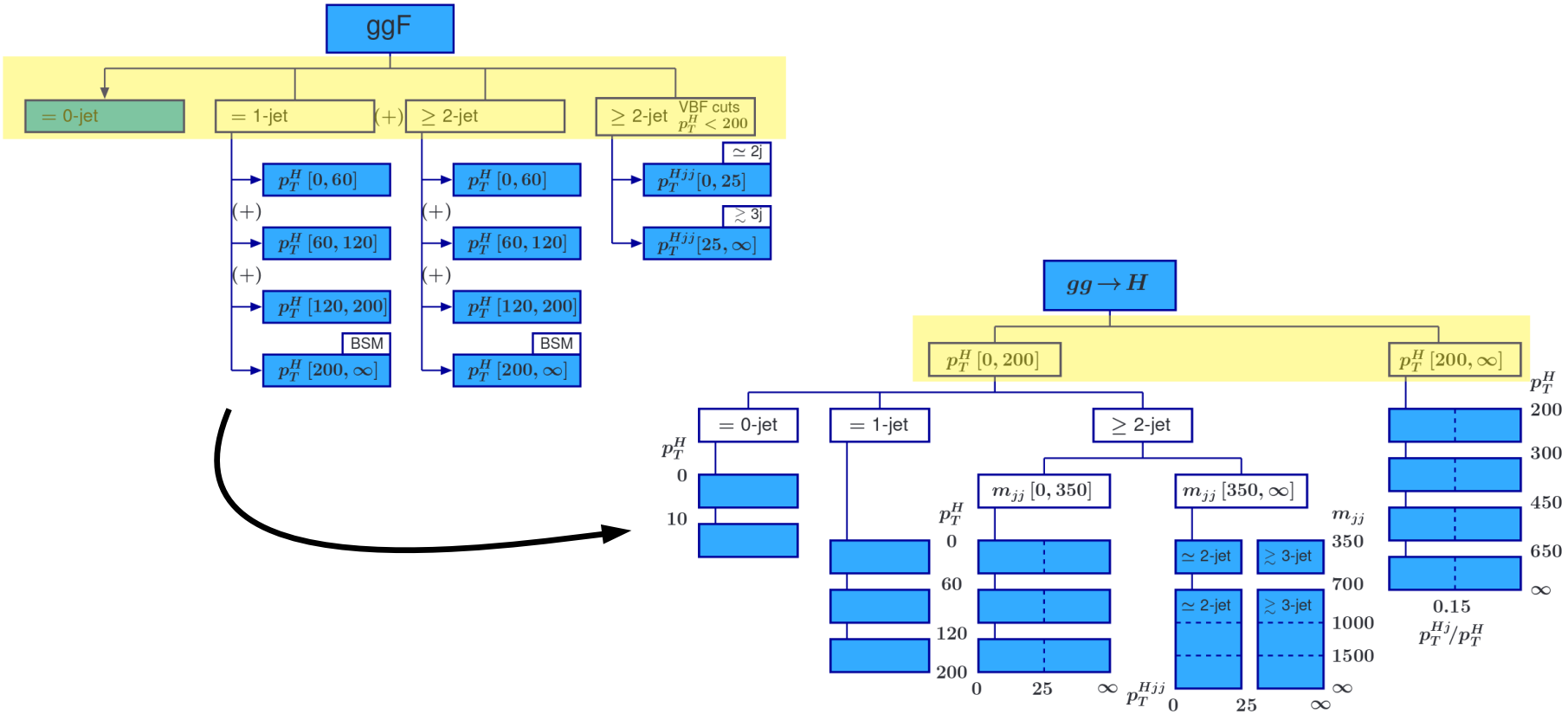
Uncertainty scheme: proposal

- Evolution of the stage 1.0 scheme
- $\{\Delta^i\}$: account for migrations across all boundaries i.e. dashed + solid
  - merged bins: type (A), split bins: type (B)
- Jet bin uncertainties: according to BLPTW scheme (YR4, official values)
  - as was used for stage 1.0 scheme
  - two absolute/normalisation uncertainties: Δ^{yield}, $\Delta^{\text{resummation}}$

    $\Rightarrow$ includes effect of inclusive N3LO QCD scale uncertainty: do not add on top!
  - two migration uncertainties: $\Delta^{njet=1}$, $\Delta^{njet=2}$
  - evaluated per N_{jet} bin
- Remaining $\mathcal{O}(20)$ NPs for migrations across boundaries
  - one for each boundary/splitting
  - anti-correlated across boundary $\Rightarrow$ cancel in sum (i.e. $\sigma^{\text{tot}}_{\text{ggH}} = \text{const.}$)
  - **Note**: any numbers for p_T^H, m_{jj} etc are still preliminary i.e. **not final**

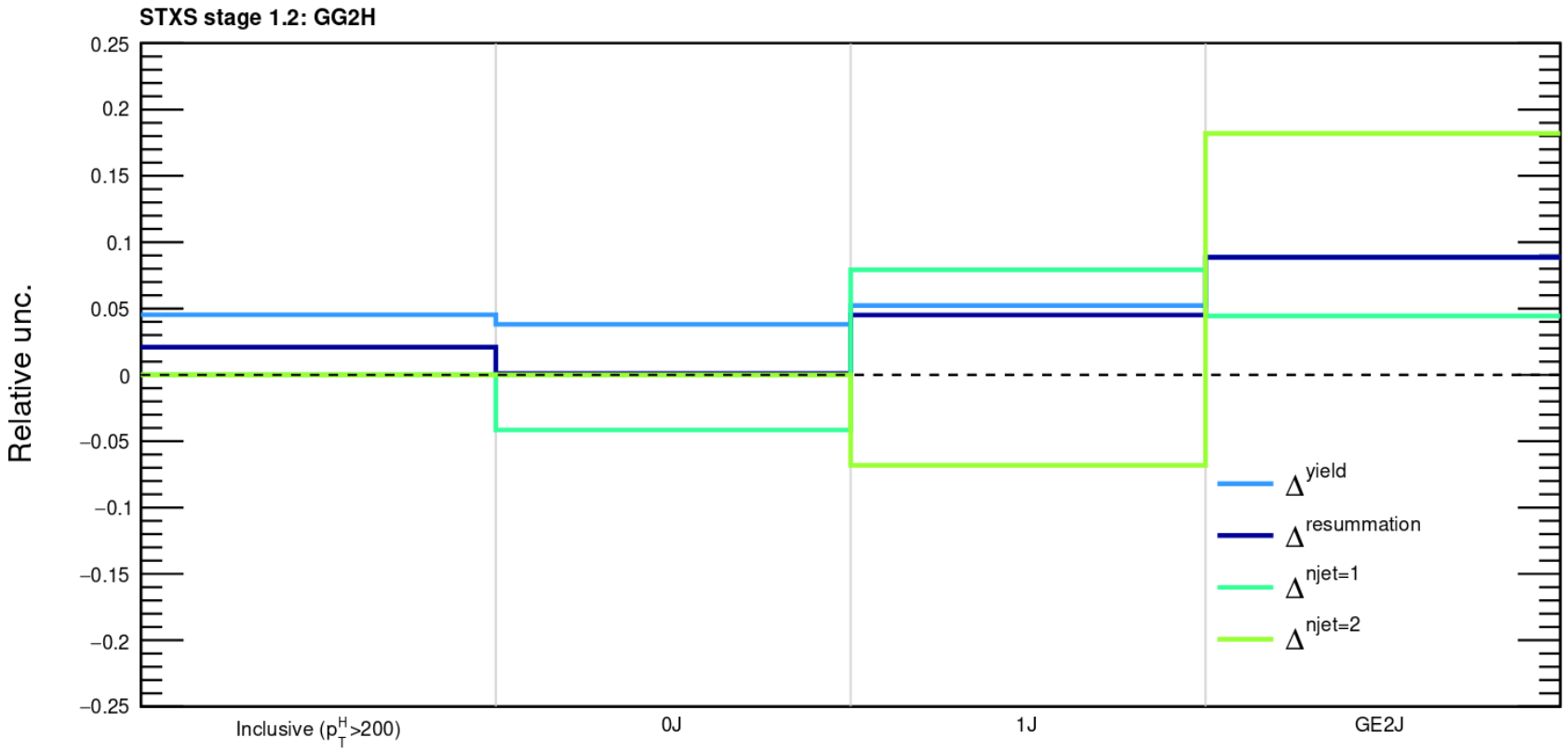
Yield, resummation and N_{jet} migrations

- Δ^{yield}, $\Delta^{\text{resummation}}$, $\Delta^{njet=1}$, $\Delta^{njet=2}$
- Re-use BLPTW calculations (YR4) as in stage 1.0 scheme: Back-up
- Stage 1.0 $\rightarrow$ 1.2: high p_T^H region now inclusive in N_{jet}



Yield, resummation and N_{jet} migrations

- Use **inclusive** N_{jet} bin in BLPTW scheme for $p_T^H > 200$ GeV:
  - $(\Delta^{\text{yield}}, \Delta^{\text{resummation}}) = (4.6\%, 2.0\%)$
  - $\Delta^{njet=1}$, $\Delta^{njet=2}$: migrations set to 0 for this region
- For 0J, 1J and GE2J regions use same numbers as used for stage 1.0 scheme
  - justified as contribution from high p_T^H region small: $\sim$0.5% of total XS



High p_T^H region

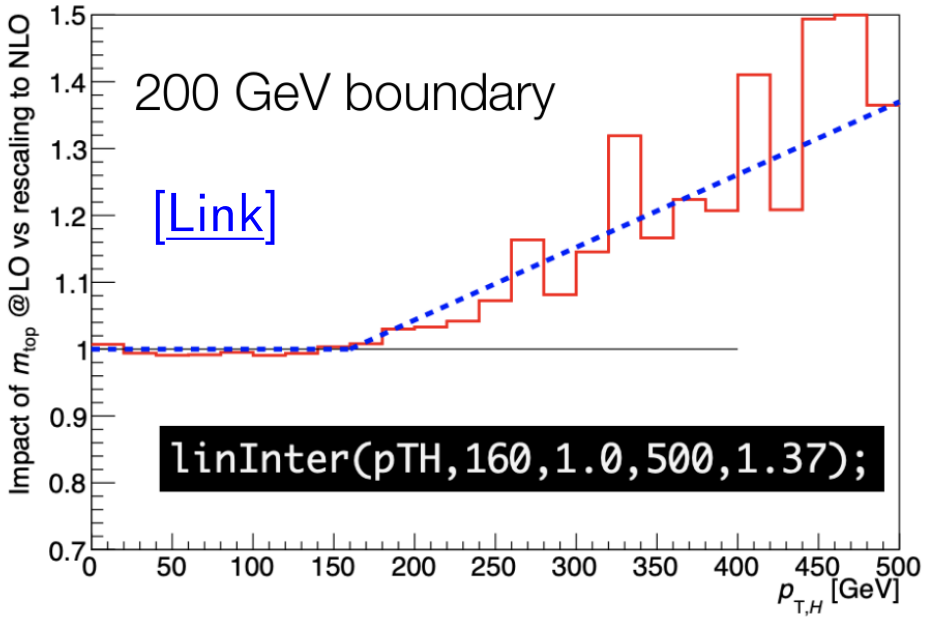
- We are now becoming sensitive to this region: important for BSM
  - require accurate prediction of uncertainty in $p_T^H > 200$ GeV bins

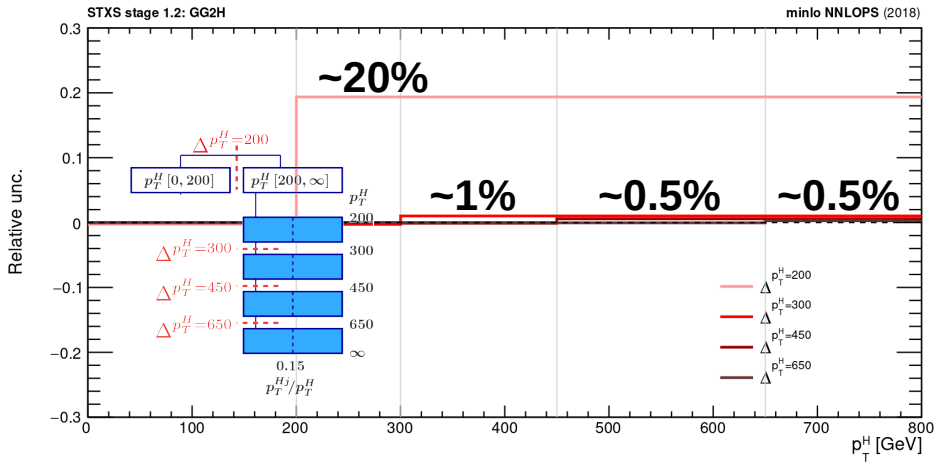
- Stage 1.0 scheme:
  - NP for treatment of top mass, m_t $\Rightarrow$ finite-vs-infinite
  - **not** a migration, affects overall yield
  - derived setting $m_t = \infty$: unphysical
  - over-estimate? input from theory needed!

- Dedicated predictions in this region?
  - as in [arXiv:2005.07762]
  - (re)configure for STXS boundaries $\Rightarrow$ extend down to $p_T^H > 300$ GeV?

- Add 4 NP for high p_T^H migrations
  - derived using minlo hj sample ▸ Back-up
  - pick (μ_R,μ_F) with largest variation

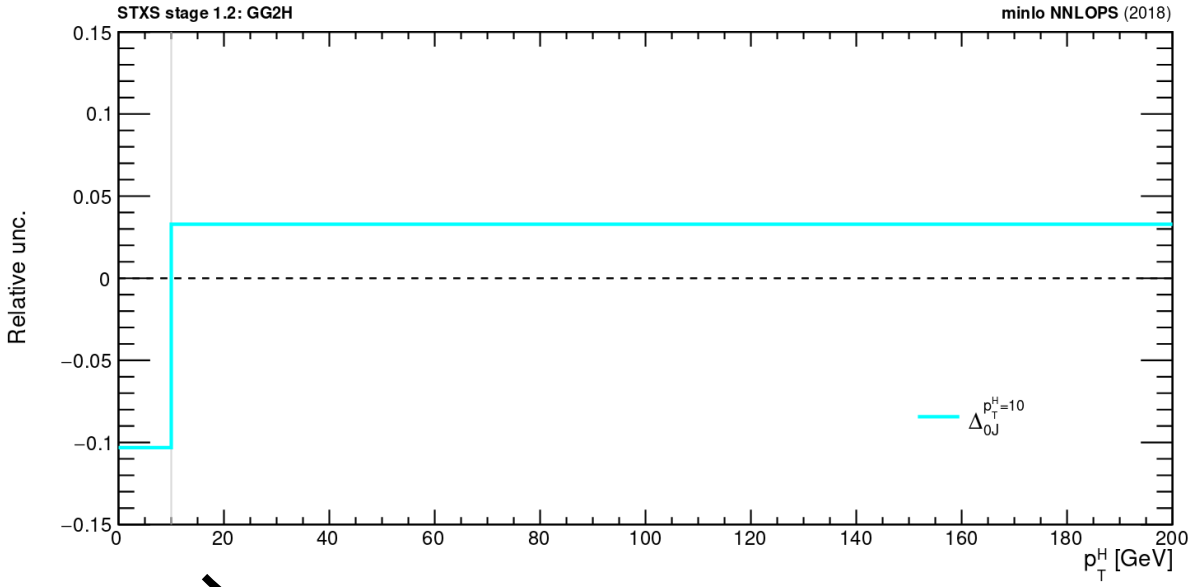
- **To do**: calculate p_T^{Hj}/p_T^H migrations

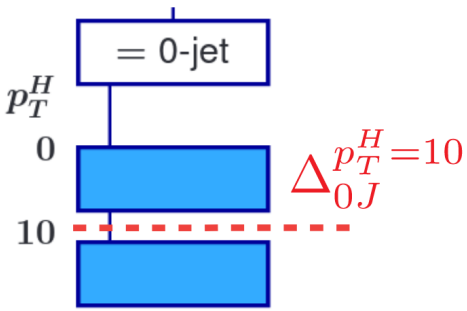


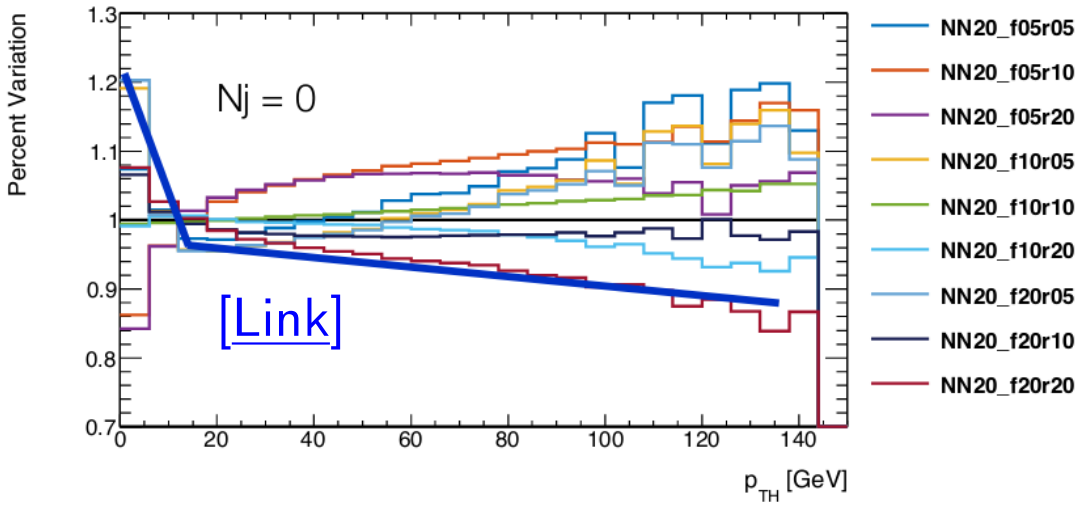


0J p_T^H migration

- Single NP: exclusive to 0J bins (defined by subscript: Δ_j)



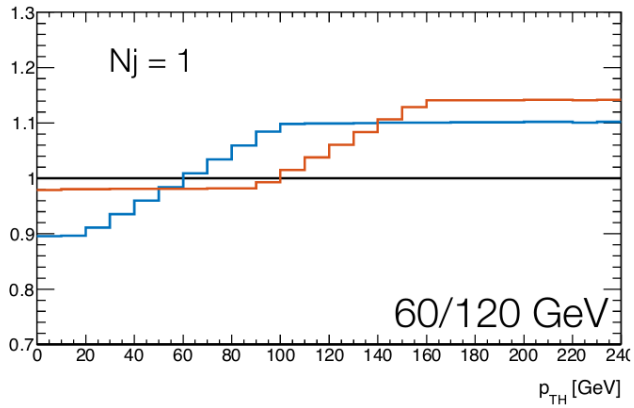




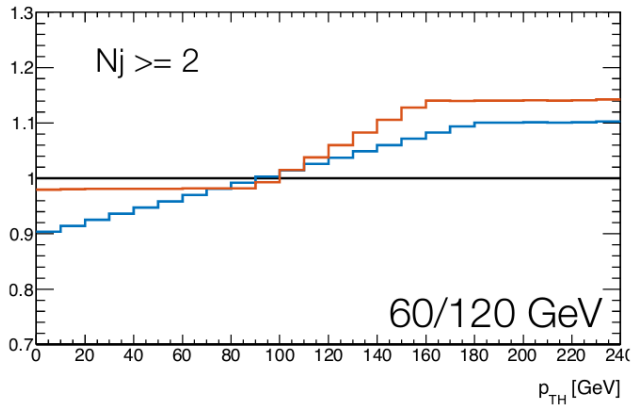
- Presented and 'approved' at a LHCHXSWG meeting
  - parametrized $\Delta \to \Delta(p_T^H)$
  - envelope of scale variations

p_T^H migrations: 1J + GE2J

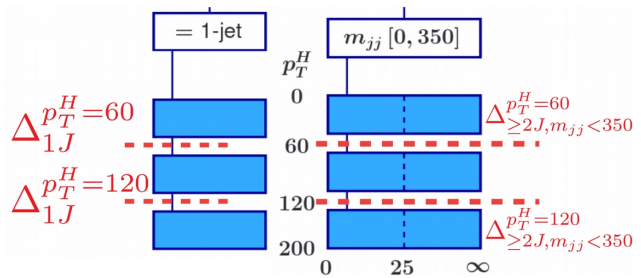
- Stage 1.0 scheme: 2NPs for migrations across 60, 120 GeV boundaries
  - smoothed at transition to avoid unphysical kink, correlated for 1J and GE2J

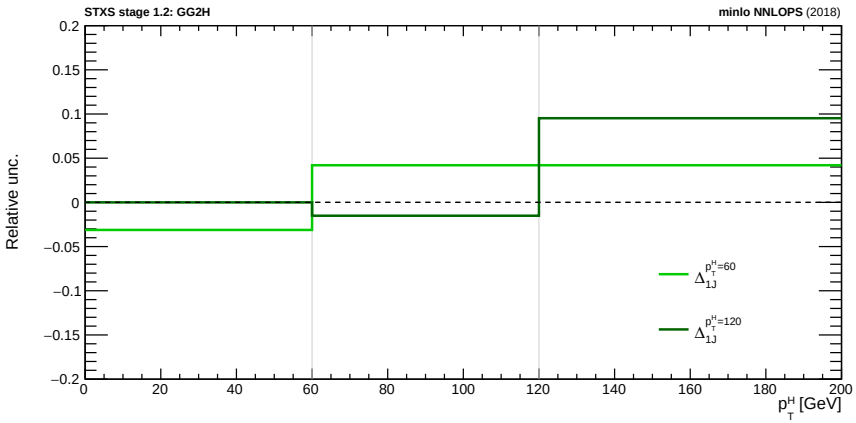


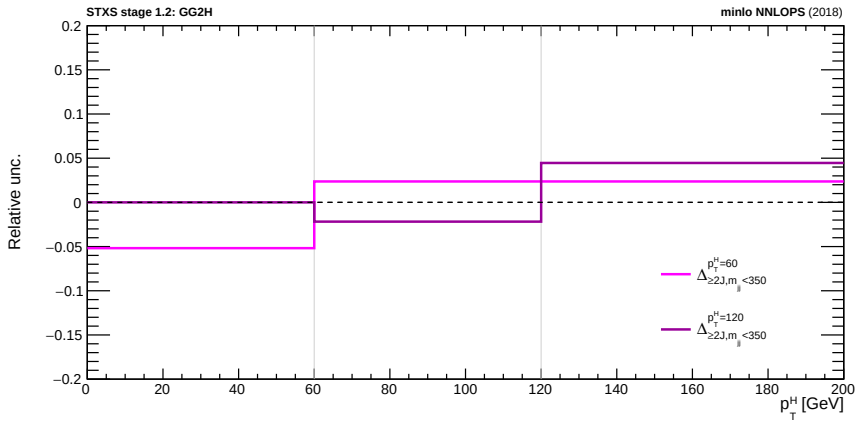
[Link]



- Check: correlation (1J/2J) still well motivated?
  - different phase space? → 4 independent NP
  - re-derive for GE2J with $m_{jj} < 350$ GeV cut?
  - final results should not depend on level of correlation



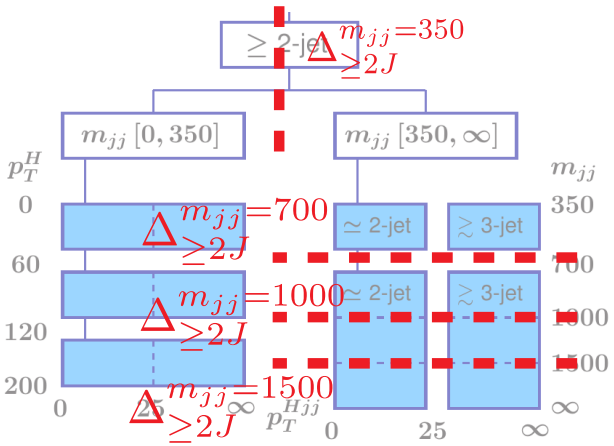




GE2J m_{jj} migrations

- **Sample**: MG5 FxFx (NLO)
  - 4NP: one for each m_{jj} boundary

	σ	σ(0,350)	σ(350,700)	σ(700,1000)	σ(1000,1500)	σ(1500,∞)
Δy	X	X	X	X	X	X
Δ350	0	-0.486%	2.033%	2.033%	2.033%	2.033%
Δ700	0	0.0%	-0.996%	2.251%	2.251%	2.251%
Δ1000	0	0.0%	0.0%	-1.623%	1.924%	1.924%
Δ1500	0	0.0%	0.0%	0.0%	-1.127%	2.190%
Total	0	0.486%	2.264%	3.440%	3.765%	4.207%



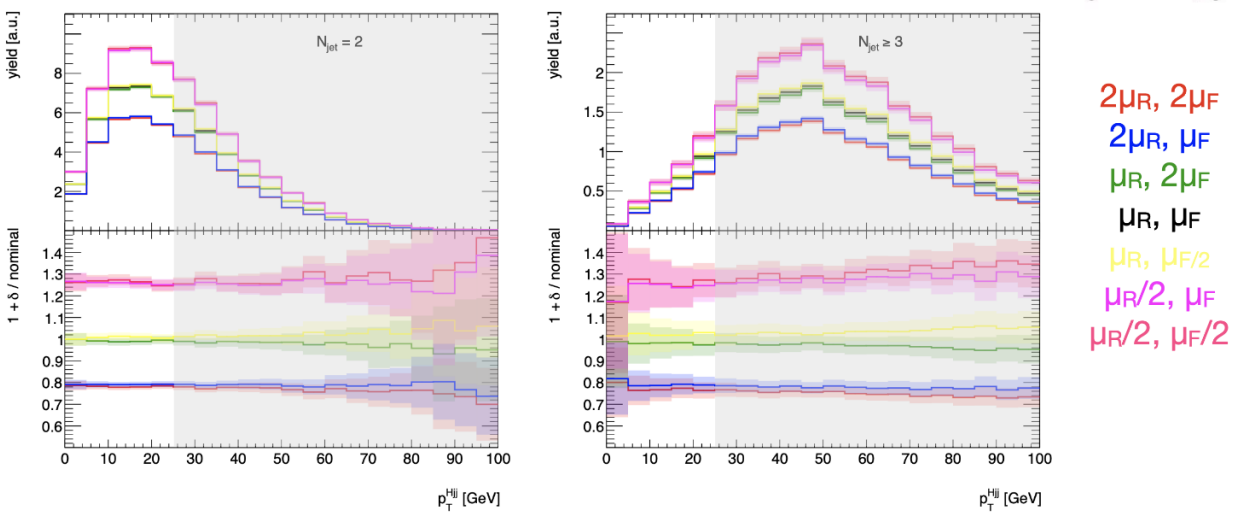
- Derived using ST-like method
  - migration uncertainties $\mathcal{O}$(0.5-2.5%)
- Currently flat numbers in each STXS bin
  - are shape effects (A) sufficiently captured by dashed boundaries?

GE2J p_T^{Hjj} migrations

Very challenging theoretically!

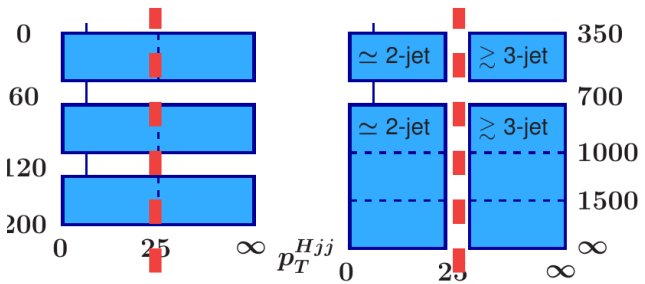
- p_T^{Hjj} boundary is used as stand-in for VBF-like topology vs H+3jet
  - $p_T^{Hjj} < 25$ GeV: (2-jet,3-jet) = (92%,8%)
  - $p_T^{Hjj} > 25$ GeV: (2-jet,3-jet) = (55%,45%)

} from MG5 FxFx sample



- Using standard ST-like approach:
  - large 'leakage' of 2J topology in the ≥3J region: 80% uncertainty in 2J
  - **require feedback from experts here**: availability of HJJJ@NLO?

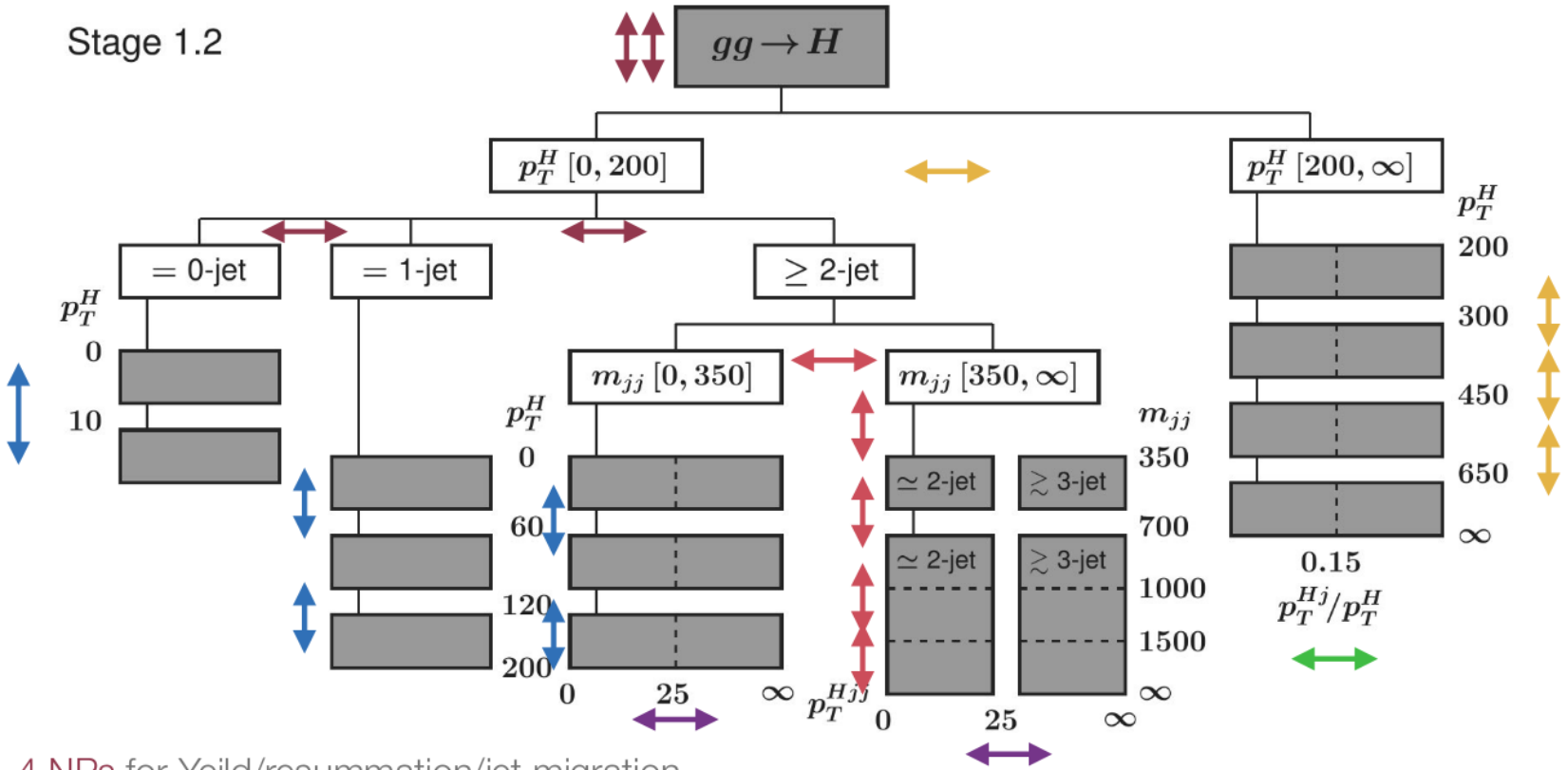
- Plan: use dedicated p_T^{Hjj} migrations when available
  - for time being: could use uncertainties from stage 1.0 scheme as educated guess?
  - $N_{\text{jet}} = 2 \leftrightarrow 3$ migrations for VBF topology: 20-30%
  - separate NP for each p_T^H / m_{jj} bin?



Proposal summary

(A): within-bin migrations, (B): across-bin migrations

- Defines (B) for interpretations, whilst capturing some of (A)



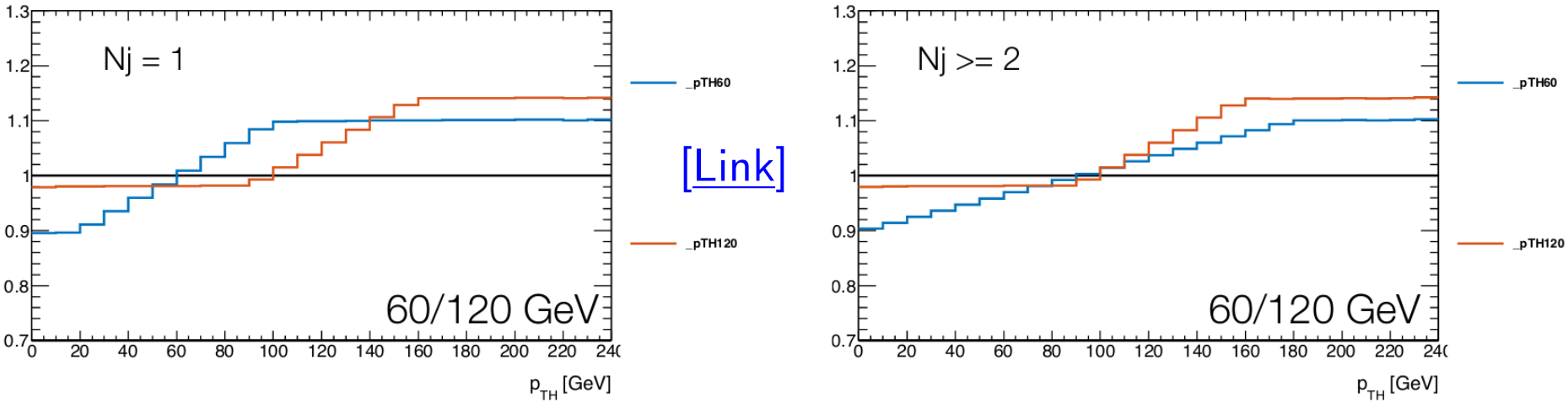
- 4 NPs for Yeild/resummation/jet migration
- 3/5 NPs for pTH migration below 200 GeV (extra if de-correlate 1/2 jet)
- 5 NPs for mJJ migration
- 4 NPs for High pTH migration
- 1/4 NPs for pTHj migration in high pTH region (depending on the correlation with pTH)
- 1/2/7 NPs for pTHjj migration (depending on the correlation with pTH/mJJ)

Limitations

- Assumes within-bin migrations (A) are mostly covered by dashed boundaries
  - (as well as UE/PS/PDF uncertainties)
- In most cases, these assumptions are valid
  - this is how/why the dashed bin boundaries were defined!
  - i.e. choose observable which accounts for majority of shape variation (p_T^{Hjj})
  - (in the future dashed boundaries may become solid)
- In some cases this validity breaks down: **residual shape effects**
  - shape effects not limited to dashed-boundary-observable e.g. p_T^{Hjj}
  - using BDT/DNN: reco phase space can be very different from truth-level
  - if large, firstly **recommend variable for (dashed) boundary in future!** For now...
- Want prescription on how to check effects and incorporate uncertainties:
  - so far left to experiments: decide necessary treatment on analysis-by-analysis level
  - may become a problem when combining ATLAS and CMS
  - two options...

Treatment of residual shape effects (1)

- When possible can **bake shape effects into NPs** of uncertainty scheme
  - e.g. p_T^H Δ in stage 1.0 scheme: $\Delta \rightarrow \Delta(p_T^H)$
  - shape inspired by scale variations



- Pros:
  - removes un-physical kinks in spectrum $\Rightarrow$ smooth transitions at bin boundaries
  - enables correlation of shape effects with across-bin migrations (B)
- Cons:
  - still only accounting for one observable dependence (e.g. p_T^H)
- **Be careful**: must factor out effect on absolute yield for each measured bin

Treatment of residual shape effects (2)

- Use **scale variations** (μ_R,μ_F) keeping normalisation of bin constant
  - calculate event migrations across analysis categories
  - ∼same result as baked in shape effects: smoothing
  - **Key**: normalisation before any event selection (including out-of-acceptance)
  - more detail in Back-up

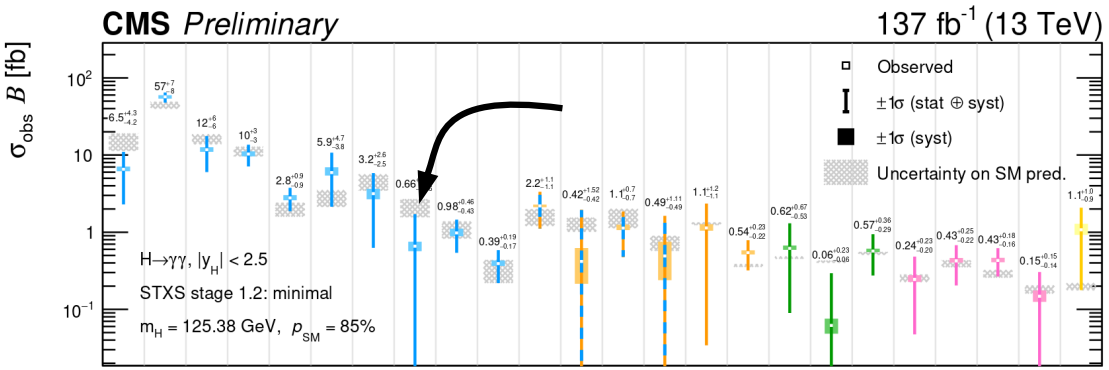
- Pros:
  - migrations are continuous, in full kinematic phase space e.g.
    ⇒ better covers migrations in BDT/DNN score distributions
    ⇒ therefore migrations in/out-of experimental acceptance

- Cons:
  - (μ_R,μ_F) variations are not always appropriate e.g. p_T^{Hjj} migrations in H+1J sample
    ⇒ include additional NP for badly modelled variables
    ⇒ e.g. 20-30% uncertainty for $N_{\text{jet}} = 2 \rightarrow 3$ Slide 11

- **Be careful**: do not double count!
  - migrations across dashed boundaries: ideally do (μ_R,μ_F) vars in each dashed bin
  - not to be used in addition to "baked in shape effects"

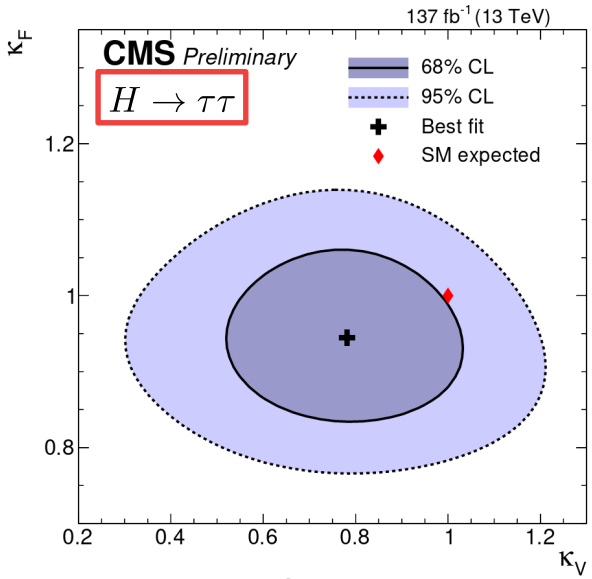
Thoughts on how to use (1)

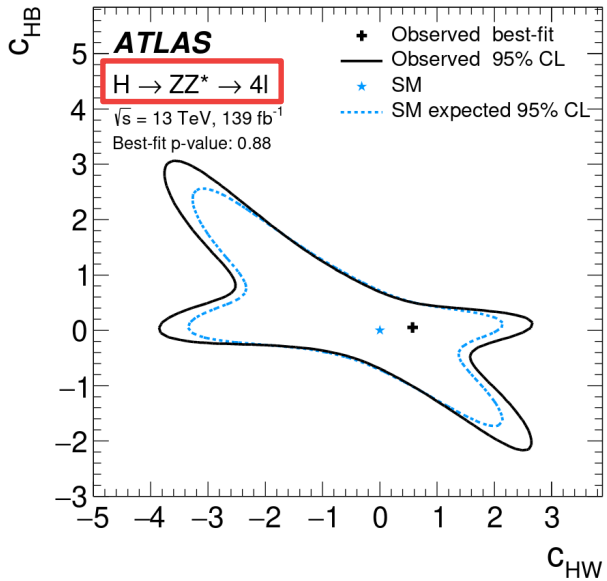
- Try to deliver final scheme as quickly as possible
  - we encourage to include new/updated NP
  - can benefit ongoing Run 2 analyses + combination
  - at the very least sets foundations for future measurements
- For the measurement of STXS bins: σ_i
  - dealing with type (A) uncertainties
  - define bin merging scheme, with choice driven by sensitivity of analysis
  - **crucial**: total cross section of any measured bin (merged or not) is constant
    $\Rightarrow$ you must factor out the absolute effect of each NP
    $\Rightarrow$ absolute variation is then attributed to SM prediction: see grey bands
  - if necessary, include treatment of **residual shape effects**
    $\Rightarrow$ shape effects baked into NP vs scale variations inside each bin
    $\Rightarrow$ avoid double counting!



Thoughts on how to use (2)

- It is possible to include new scheme for future **interpretations**
  - in fact one of advantages of STXS: measurements remain long-term useful
  - accommodate improvements in theory predictions
- For the interpretation of STXS measurements: e.g. μ_i, κ, EFT
  - type (A) + type (B) i.e. include across-bin migrations + yield uncertainties
  - if performed in-house[1]: comes for free in a well defined scheme
  - if performed outside[2]: e.g. testing a new model using STXS measurements
    $\Rightarrow$ apply flat numbers of scheme to each measured bin
    $\Rightarrow$ drop relevant NP for merged bins as effect is **included in measurement**
    $\Rightarrow$ (assumes within each bin model behaves like SM, within type (A) uncertainty)





[1]meaning in-experiment, [2]meaning by theorists

Summary

- Summarised discussions between ATLAS + CMS + theory with respect to a ggH stage 1.2 uncertainty scheme
  - evolution of the previous stage 1.0 scheme
  - new/updated NP for stage 1.2 boundaries
  - thoughts on how to use: measurements vs interpretations
  - the numbers are only a proposal, not yet final

- A number of **open questions** remain...
  - treatment of high p_T^H region: dedicated predictions? treatment of m_t?
    $\Rightarrow$ + how to derive p_T^{Hj}/p_T^H migrations (dashed)?
  - treatment of VBF-like phase space: p_T^{Hjj} migrations, HJJJ@NLO?
  - correlations? e.g. p_T^H variations for 1J-vs-GE2J
  - dependence on other boundaries? e.g. are p_T^H migrations dependent on m_{jj} cut
  - treatment for residual shape effects? parametrized NPs vs scale variations

- Aiming to finalise prescription in near future $\Rightarrow$ documentation!

Back-Up Slides

Stage 1.0 scheme

- Also known as "2017" scheme: summarised in [slides]

Cross sections and fractional uncertainties

STXS	sig	stat	mu	res	mig01	mig12	VBF2j	VBF3j	pT60	pT120	qm_top	Tot
Incl	48.52	+/- 0.00	+4.6%	+2.1%	-0.0%	-0.0%	+0.3%	-0.0%	+0.0%	+0.2%	+0.2%	+5.1%
FWDH	4.27	+/- 0.01	+4.5%	+1.9%	-0.5%	-0.2%	+0.0%	+0.0%	-0.3%	-0.1%	+0.0%	+4.9%
VBF_J3V	0.27	+/- 0.00	+0.0%	+0.0%	+0.0%	+0.0%	+20.0%	-32.0%	-1.6%	+1.1%	+0.1%	+37.8%
VBF_J3	0.36	+/- 0.00	+0.0%	+0.0%	+0.0%	+0.0%	+20.0%	+23.5%	-0.2%	+2.5%	+0.2%	+31.0%
=0J	27.25	+/- 0.03	+3.8%	+0.1%	-4.1%	+0.0%	+0.0%	+0.0%	+0.0%	+0.0%	+0.0%	+5.6%
=1J_0-60	6.49	+/- 0.01	+5.2%	+4.5%	+7.9%	-6.8%	+0.0%	+0.0%	-4.8%	-1.6%	+0.0%	+13.5%
=1J_60-120	4.50	+/- 0.01	+5.2%	+4.5%	+7.9%	-6.8%	+0.0%	+0.0%	+4.8%	-0.9%	+0.0%	+13.4%
=1J_120-200	0.74	+/- 0.00	+5.2%	+4.5%	+7.9%	-6.8%	+0.0%	+0.0%	+10.0%	+10.1%	+0.5%	+18.9%
=1J_200->	0.15	+/- 0.00	+5.2%	+4.5%	+7.9%	-6.8%	+0.0%	+0.0%	+10.0%	+14.0%	+10.5%	+23.7%
>=2J_0-60	1.22	+/- 0.01	+8.9%	+8.9%	+4.4%	+18.2%	+0.0%	+0.0%	-5.9%	-1.6%	+0.0%	+23.3%
>=2J_60-120	1.86	+/- 0.01	+8.9%	+8.9%	+4.4%	+18.2%	+0.0%	+0.0%	-0.2%	-0.2%	+0.0%	+22.5%
>=2J_120-200	0.99	+/- 0.00	+8.9%	+8.9%	+4.4%	+18.2%	+0.0%	+0.0%	+6.6%	+10.6%	+0.6%	+25.8%
>=2J_200->	0.42	+/- 0.00	+8.9%	+8.9%	+4.4%	+18.2%	+0.0%	+0.0%	+10.0%	+14.0%	+11.8%	+30.7%
=0J	30.12	+/- 0.03	+3.8%	+0.1%	-4.1%	+0.0%	+0.0%	+0.0%	+0.0%	+0.0%	+0.0%	+5.6%
=1J	12.92	+/- 0.02	+5.2%	+4.5%	+7.9%	-6.8%	+0.0%	+0.0%	-0.1%	-0.4%	+0.2%	+12.5%
>=2J	5.47	+/- 0.01	+7.8%	+7.8%	+3.9%	+16.1%	+2.3%	-0.0%	+0.4%	+2.9%	+1.1%	+20.3%
>=1J 60-200	9.09	+/- 0.01	+6.2%	+5.8%	+6.4%	+1.9%	+0.9%	+0.1%	+4.2%	+1.7%	+0.1%	+11.8%
>=1J 120-200	1.96	+/- 0.01	+6.8%	+6.5%	+5.5%	+6.9%	+1.5%	+0.4%	+8.0%	+10.4%	+0.6%	+18.5%
>=1J >200	0.58	+/- 0.00	+7.9%	+7.7%	+5.4%	+11.6%	+0.0%	+0.0%	+10.0%	+14.0%	+11.4%	+26.7%
>=1J >60	9.68	+/- 0.01	+6.3%	+5.9%	+6.3%	+2.5%	+0.8%	+0.1%	+4.6%	+2.5%	+0.8%	+12.2%
>=1J >120	2.54	+/- 0.01	+7.0%	+6.8%	+5.5%	+8.0%	+1.2%	+0.3%	+8.4%	+11.2%	+3.0%	+19.9%
>=1	18.40	+/- 0.02	+6.0%	+5.5%	+6.7%	-0.0%	+0.7%	-0.0%	+0.0%	+0.5%	+0.4%	+10.6%

Treatment of residual shape effects (2)

- Posssible implementation: use $(\mu_R,\mu_F) = (0.5,0.5)$-$(2,2)$
  - most of the time this represents the max variation
  - enables correlation of uncertainties across analyses
- If phase space is close enough then could correlate across bins e.g.

for gluon fusion

ggH_scale_0jet for STXS bins GG2H_0J_PTH_0_10, GG2H_0J_PTH_GT10
ggH_scale_1jet_lowpt for STXS bins GG2H_1J_PTH_0_60, GG2H_1J_PTH_60_120, GG2H_1J_PTH_120_200
ggH_scale_2jet_lowpt for STXS bins GG2H_GE2J_MJJ_0_350_PTH_0_60, GG2H_GE2J_MJJ_0_350_PTH_60_120, GG2H_GE2J_MJJ_0_350_PTH_120_200
ggH_scale_highpt for STXS bins GG2H_PTH_200_300, GG2H_PTH_300_450
ggH_scale_very_highpt for STXS bins GG2H_PTH_450_650, GG2H_PTH_GT650
ggH_scale_vbf for STXS bins GG2H_GE2J_MJJ_350_700_PTH_0_200_PTHJJ_0_25, GG2H_GE2J_MJJ_350_700_PTH_0_200_PTHJJ_GT25, GG2H_GE2J_MJJ_GT700_PTH_0_200_PTHJJ_0_25, GG2H_GE2J_MJJ_GT700_PTH_0_200_PTHJJ_GT25

for VBF

vbf_scale_0jet for STXS bin QQ2HQQ_0J
vbf_scale_1jet for STXS bin QQ2HQQ_1J
vbf_scale_lowmjj for STXS bins QQ2HQQ_GE2J_MJJ_0_60, QQ2HQQ_GE2J_MJJ_60_120, QQ2HQQ_GE2J_MJJ_120_350
vbf_scale_highmjj_lowpt for STXS bins QQ2HQQ_GE2J_MJJ_350_700_PTH_0_200_PTHJJ_0_25, QQ2HQQ_GE2J_MJJ_350_700_PTH_0_200_PTHJJ_GT25, QQ2HQQ_GE2J_MJJ_GT700_PTH_0_200_PTHJJ_0_25, QQ2HQQ_GE2J_MJJ_GT700_PTH_0_200_PTHJJ_GT25
vbf_scale_highmjj_highpt for STXS bin QQ2HQQ_GE2J_MJJ_GT350_PTH_GT200

for V(qq)H

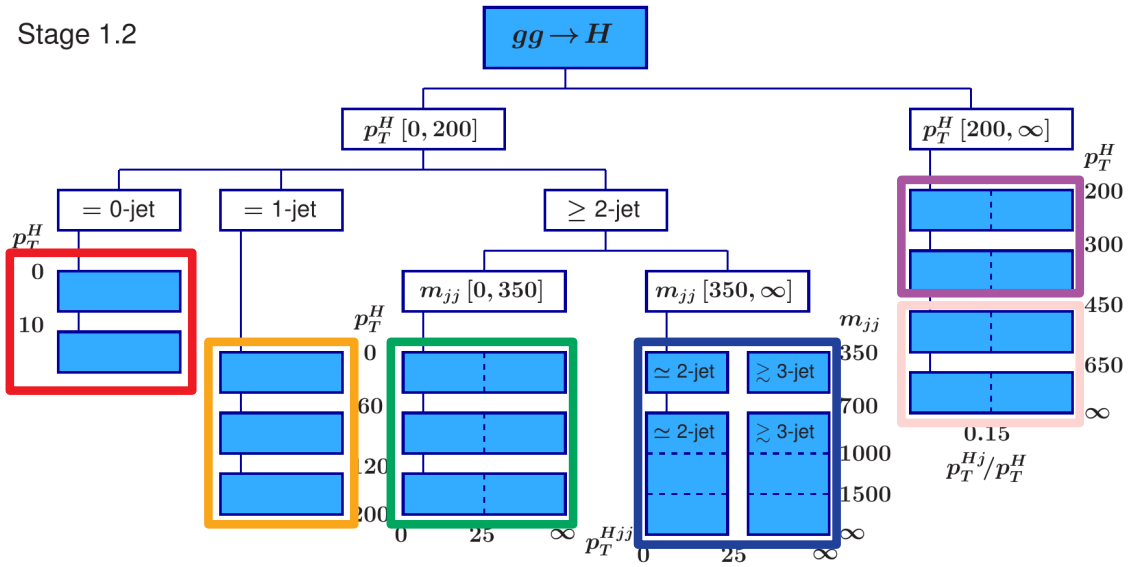
VH_scale_0jet for STXS bin QQ2HQQ_0J
VH_scale_1jet for STXS bin QQ2HQQ_1J
VH_scale_lowmjj for STXS bins QQ2HQQ_GE2J_MJJ_0_60, QQ2HQQ_GE2J_MJJ_60_120, QQ2H(
VH_scale_highmjj_lowpt for STXS bins QQ2HQQ_GE2J_MJJ_350_700_PTH_0_200_PTHJJ_0_
QQ2HQQ_GE2J_MJJ_GT700_PTH_0_200_PTHJJ_GT25
VH_scale_highmjj_highpt for STXS bin QQ2HQQ_GE2J_MJJ_GT350_PTH_GT200

for V(lep)H

WH_scale_lowpt for STXS bins QQ2HLNU_PTV_0_75, QQ2HLNU_PTV_75_150, QQ2HLNU_PTV_1
WH_scale_highpt for STXS bin QQ2HLNU_PTV_GT250
ZH_scale_lowpt for STXS bins QQ2HLL_PTV_0_75, QQ2HLL_PTV_75_150, QQ2HLL_PTV_150_
ZH_scale_highpt for STXS bin QQ2HLL_PTV_GT250, GG2HLL_PTV_GT250

for ttH

ttH_scale_lowpt for STXS bins TTH_PTH_0_60, TTH_PTH_60_120, TTH_PTH_120_200, TTH_
ttH_scale_highpt for STXS bin TTH_PTH_GT300



BLPTW: jet-bin uncertainties

$p_T^{\text{cut}} = 30$ GeV	σ/pb	Δ_μ	Δ_φ	$\Delta_{\text{cut}}^{0/1}$	$\Delta_{\text{cut}}^{1/2}$	total pert. unc.
$\sigma_{\geq 0}$	47.41±2.40	4.6%	2.0%	-	-	5.1%
σ_0	29.51±1.65	3.8%	0.1%	4.1%	-	5.6%
$\sigma_{\geq 1}$	17.90±1.88	6.0%	5.2%	6.8%	-	10.5%
σ_1	11.94±1.58	5.5%	4.8%	8.4%	7.2%	13.2%
$\sigma_{\geq 2}$	5.96±1.05	7.1%	6.1%	3.6%	14.5%	17.6%

QCD uncertainty split into 4 independent sources

normalization

resummation

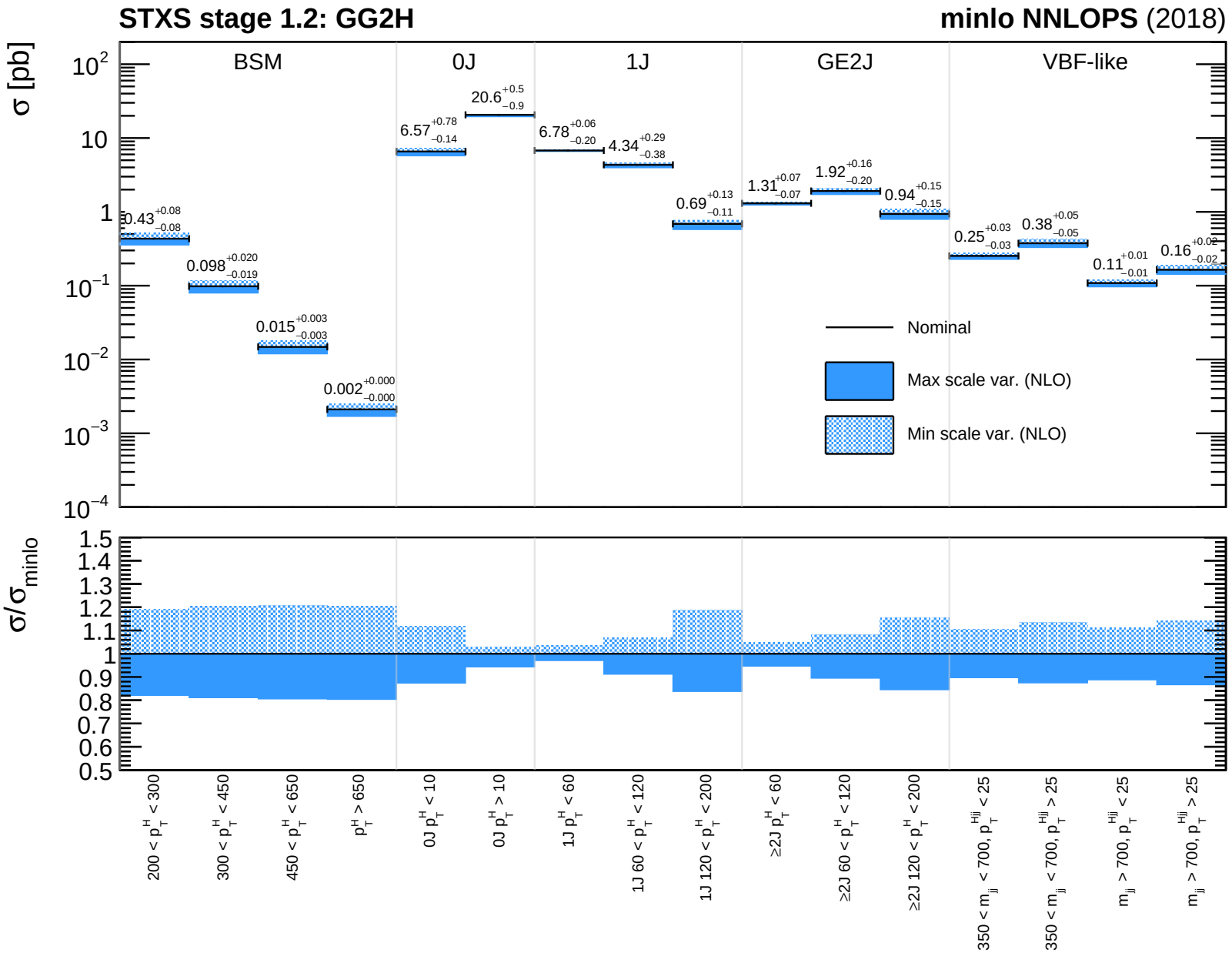
0↔1 jet migration

1↔2 jet migration

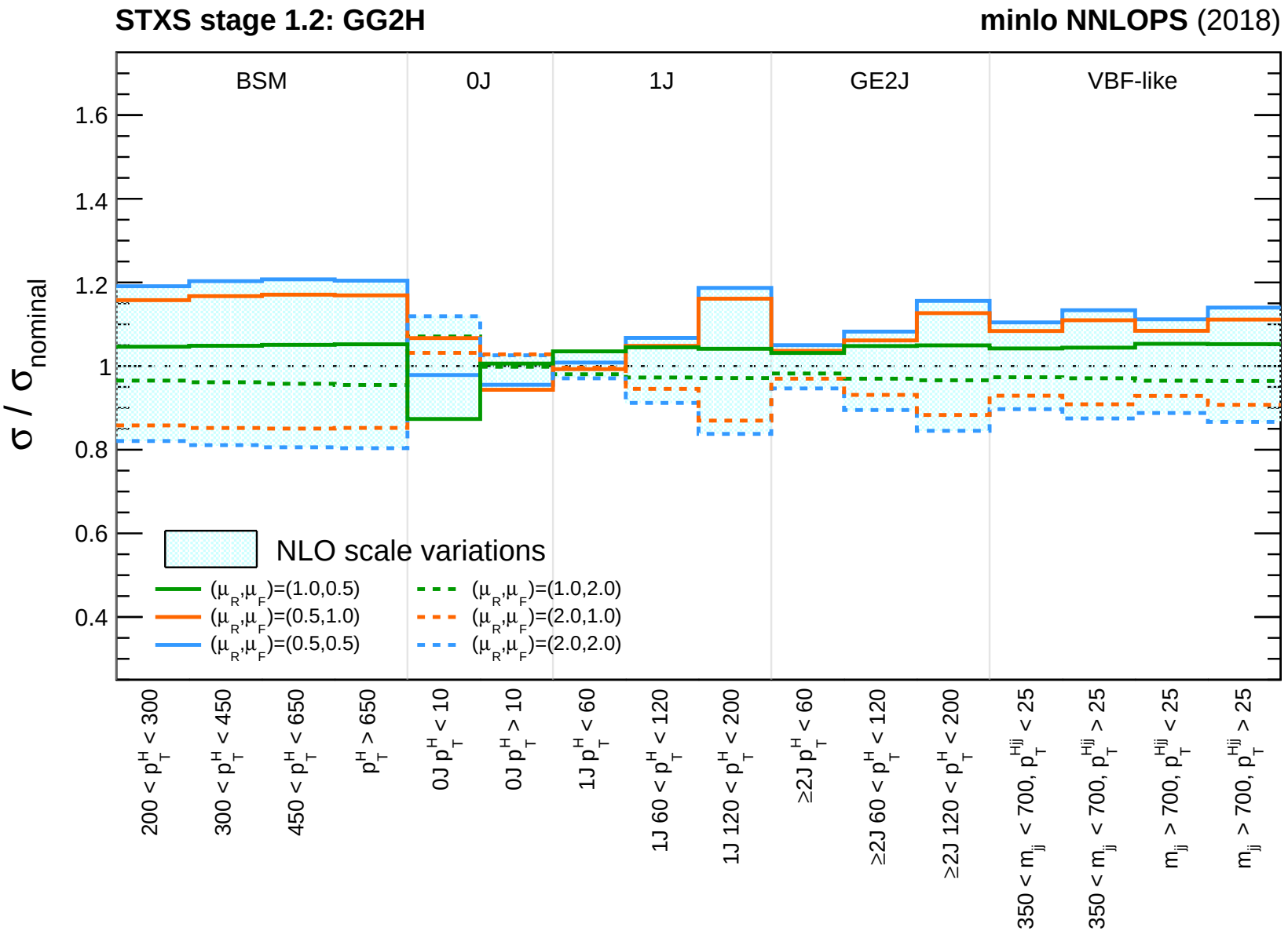
MINLO HJ sample

- For a number of migration unc used **MINLO HJ sample**: $m_H = 125.0$ GeV
  - NLO + re-summed calculation $\Rightarrow$ (almost) NNLO inclusively
  - NNLOPS (by construction)
  - [arXiv:1309.0017]
- Inclusive sample with no selection applied
- Normalised to $\sigma_{\mathrm{ggH}}^{\mathrm{N3LO}} = 48.58$ pb
- Consider scale variations (μ_R,μ_F) @ NLO and NNLO
  - NLO: (0.5,0.5), (0.5,1), (1,0.5), (1,2), (2,1), (2,2)
  - NNLO: (0.5,0.5), (2,2)
  - do **not** consider varying scales of both orders (NLO & NNLO) simultaneously

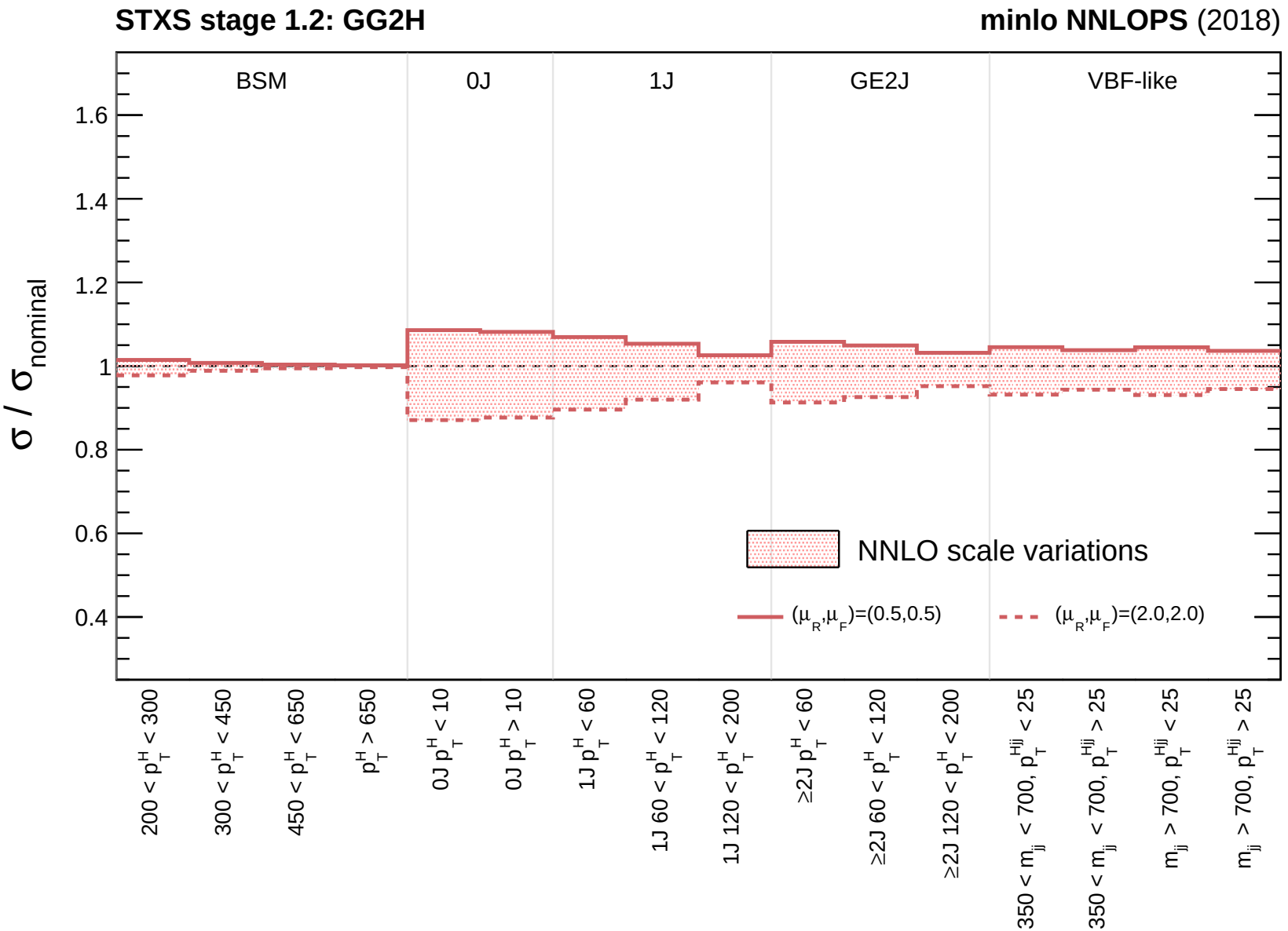
Nominal cross sections: stage 1.2



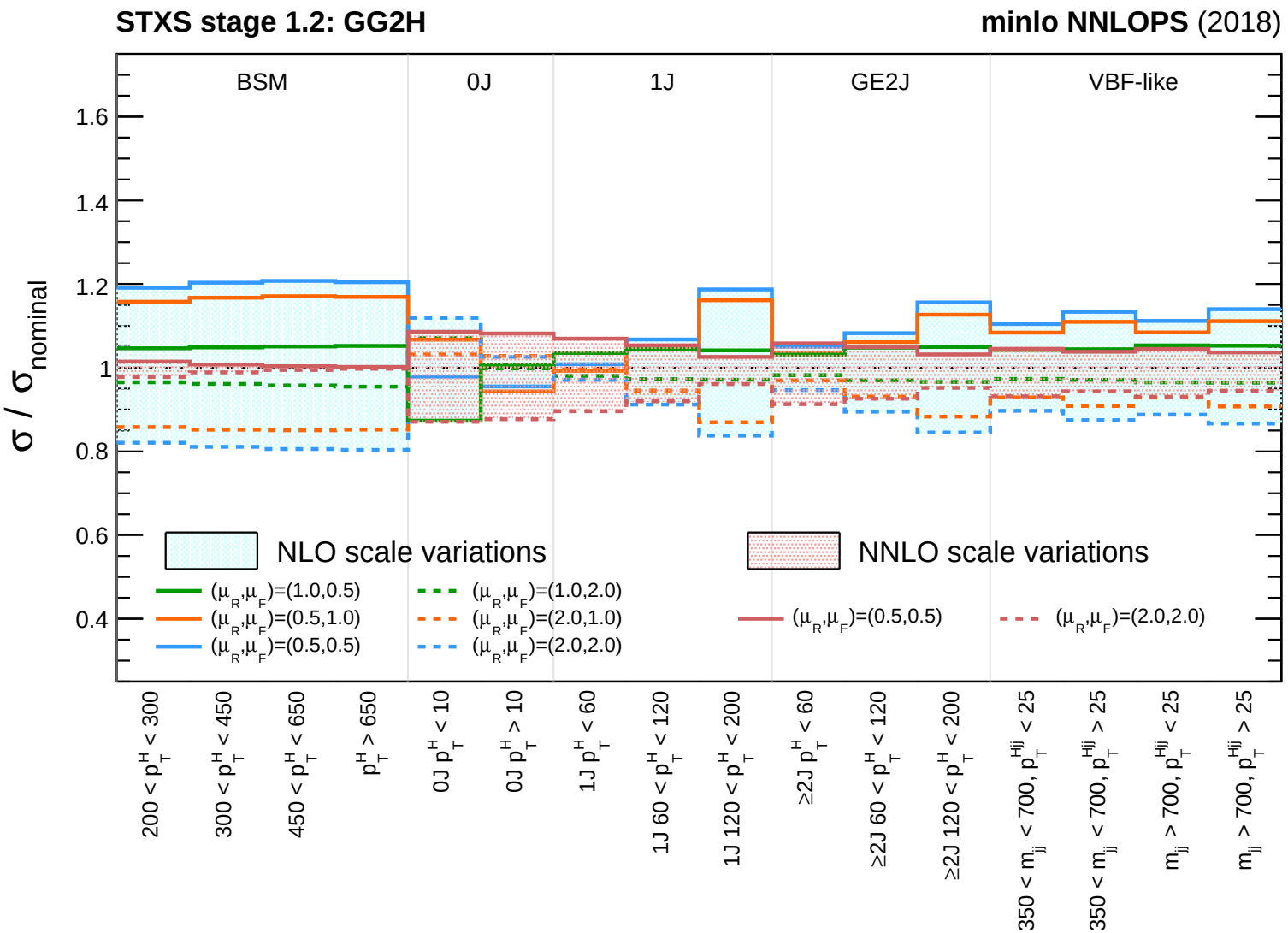
Scale variations @ NLO



Scale variations @ NNLO



Scale variations (all)



- **Aim**: define uncertainty scheme to cover max/min scale variations