Estimating the SM background for supersymmetry searches: challenges and methods

Geert-Jan Besjes (On behalf of the ATLAS collaboration)
(Nijmegen/Nikhef)

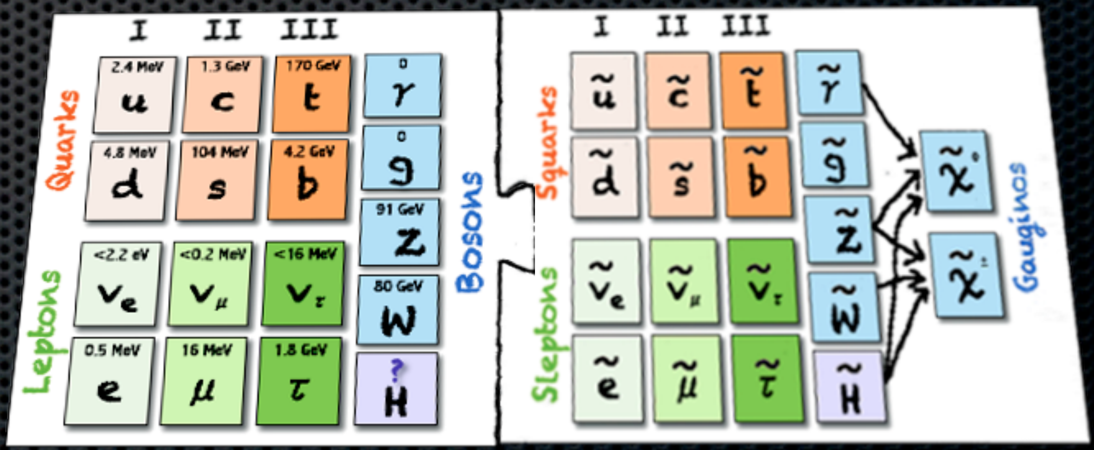
23 April 2013

Overview

- Supersymmetry
- SUSY at the LHC
- Reducible backgrounds
- Irreducible backgrounds
- Reweighting
- Conclusion

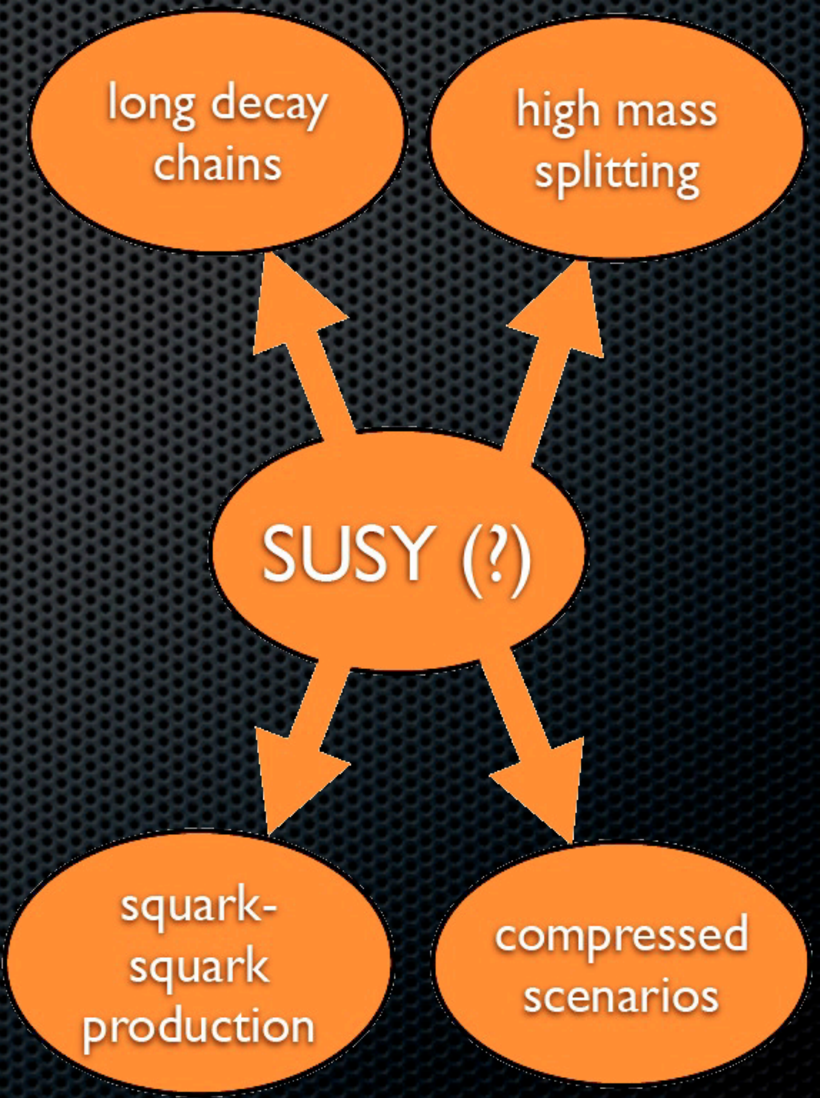
Supersymmetry

- Supersymmetry (SUSY) is a new **symmetry between bosons and fermions**
- Every Standard Model particle gets a **partner particle**
- Naturally **solves the hierarchy problem** in the Standard Model
- Could provide **solutions to other problems**:
  - dark matter candidate
  - gauge unifications
  - ...
- **Searches** for supersymmetric particles **ongoing at the LHC**!



SUSY at the LHC

- If SUSY exists at the TeV scale, **we expect to observe a strong production** of squarks and gluinos at the LHC
- **Sparticles will decay in a cascade** producing quarks and gluons, appearing as jets, and possibly leptons
- **Assuming R-parity conservation**, the decays leads to production of the lightest SUSY particle (LSP) and thus E_T^{miss}
- A **typical signature for SUSY production** at the LHC, assuming R-parity conservation, is jets + E_T^{miss} (+leptons)
- All affected by a wide range of Standard Model backgrounds!

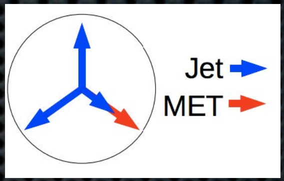


Reducible backgrounds

- **Broad range of signatures**, each affected by one or more of the following
- **Fake E_T^{miss}**:
  - **Multijets**
    - jet smearing method (0-lepton inclusive, sbottom 0-lepton)
    - templates (multijets analysis)
  - **Z->ll+jets**: jet smearing method (2l stop, $Z+E_T^{miss}$)
- **Fake leptons**: matrix method (any multilepton analysis, is main bkg)
- **Charge misidentification** (2l same-sign)

- Also see the ATLAS SUSY talks by E. Romero Adam (inclusive searches), P. Pani (direct 3rd gen.) and C. Deluca (gluino-mediated 3rd gen.), as well as M. King (R-parity violating signatures) in the EW session later this afternoon

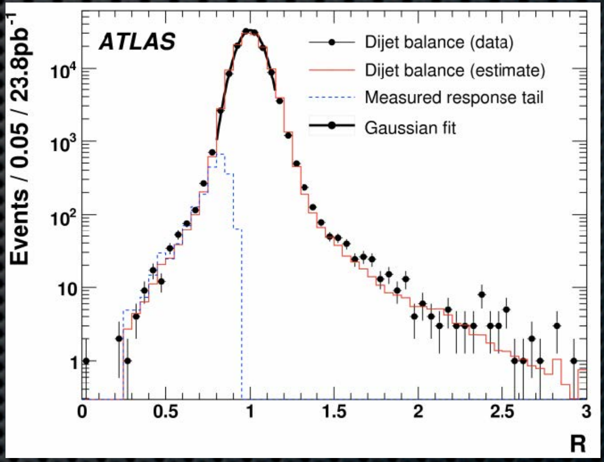
Fake E_T^{miss}: jet smearing

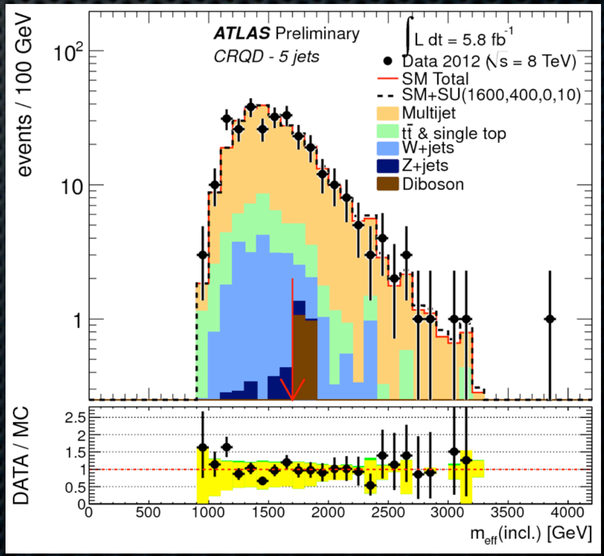


- **Jets can be mis-measured**, introducing large **fake E_T^{miss}**
- Relevant for **multijet processes and Z->ll** (high cross-section, no real E_T^{miss})
- **Not enough high E_T^{miss} events** to constrain background, aim to generate such events from ones at low E_T^{miss}
- Derive a **jet response function** using MC

$$R = \frac{p_{\mathrm{T}}(\mathrm{reco})}{p_{\mathrm{T}}(\mathrm{true})}$$

- The tail in this response function comes from mis-measured jets (mercedes events)
- Corrected using in-situ measurements of QCD dijets and multijets samples
- With the response function, **smear jets** with low E_T^{miss} in real data
- **Obtain events with a large fake E_T^{miss}**
- Validate this estimation in a **dedicated control region**





ATLAS-CONF-2012-109

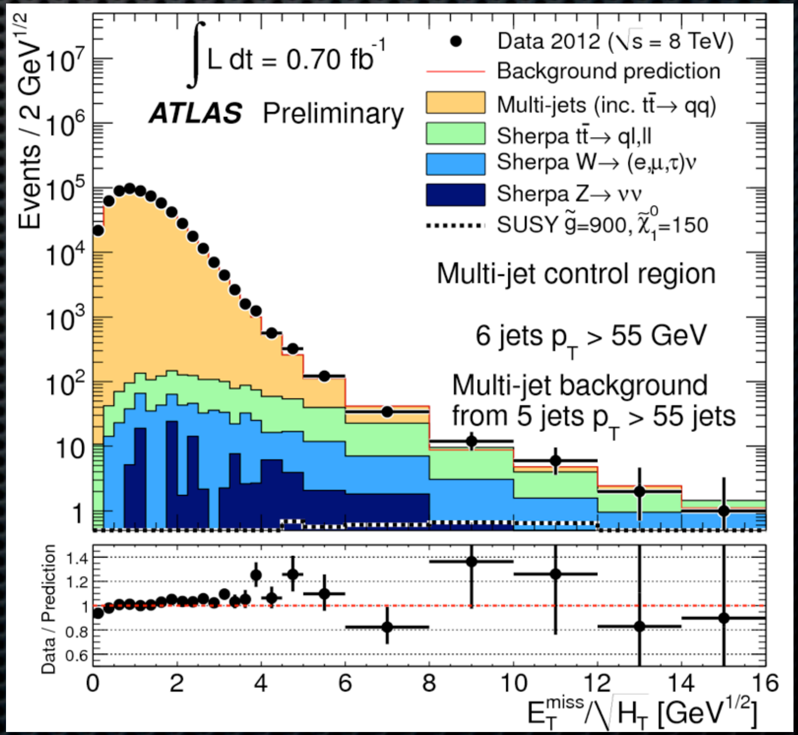
Fake E_T^{miss}: templates

- E_T^{miss} is dominated by **jet-mismeasurement**
- $E_T^{miss}/\sqrt{H_T}$ shape does not depend on jet multiplicity
- **Take a template** of $E_T^{miss}/\sqrt{H_T}$ **at low E_T^{miss}**
- **Reweigh data at high E_T^{miss}** using this template:

$$N^{SR}_{N_{jet}\geq j} = N^{CR}_{N_{jet}\geq j}\frac{N^{SR}_{N_{jet}=i}}{N^{CR}_{N_{jet}=i}}$$

- Method can be **validated using**:
  - variations in shape at lower jet multiplicities than those used in the SRs
  - at low $E_T^{miss}/\sqrt{H_T}$ in the SRs

$$H_T = \sum_{i\in \text{jets}} |\mathbf{p}_T^{(i)}|$$



ATLAS-CONF-2012-103

Fake leptons: matrix method

- Fake leptons can arise **from non-prompt leptons and photon conversion**
- General approach is **based on loose/tight matrix method**
- **Define a "loose"** region with preselected leptons and a **"tight" region** (signal)
- Solve the set of equations

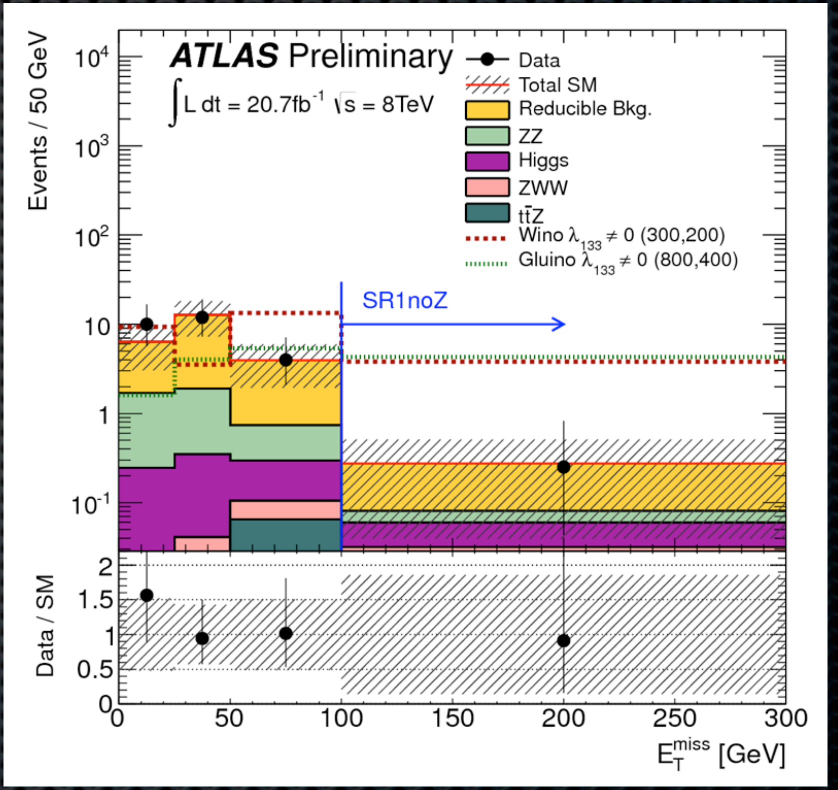
$$N^{\text{loose}} = N^{\text{loose}}_{\text{real}} + N^{\text{loose}}_{fake}$$

$$N^{\text{tight}} = \varepsilon_{\text{real}} N^{\text{loose}}_{\text{real}} + \varepsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}}$$

- The **number of fake leptons** is thus given by

Count how many

$$N^{\text{tight}}_{\text{fake}} = \frac{\varepsilon_{\text{fake}}}{\varepsilon_{\text{real}} - \varepsilon_{\text{fake}}} \left[N^{\text{loose}}_{\text{real}} \varepsilon_{\text{real}} - N^{\text{tight}} \right]$$



ATLAS-CONF-2013-036

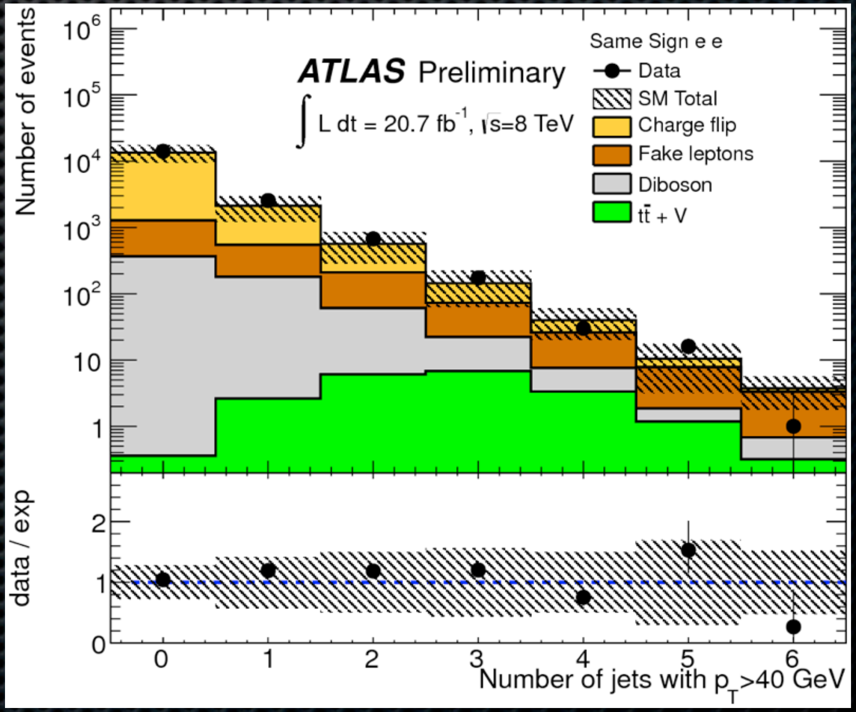
Measure independently from dedicated fake-dominated control region and using tag-and-probe method

For 1 lepton, but can easily be extended to include more

Charge misidentification

- Relevant for the 2-lepton same-sign analysis
- Negligible for muons, **only applies to electrons**
- Possible since charge is measured in inner detector, and energy in calorimeter
- Due to **hard photon bremsstrahlung**, followed by a conversion
- The **"charge flip" rate is estimated in a Drell-Yan control sample**
- Applied in data with the same selections as the SRs, except that **the leading two leptons are required to have opposite sign charges**
- Flip **rate extracted using "aided" tag and probe** (tag from barrel, |η|<1.37)
- In each bin in of eta and p_T:

$$\varepsilon^{\text{flip}} = \frac{N_{\text{SS}}}{N_{\text{SS}} + 2N_{\text{OS}}}$$



ATLAS-CONF-2013-007

Irreducible backgrounds

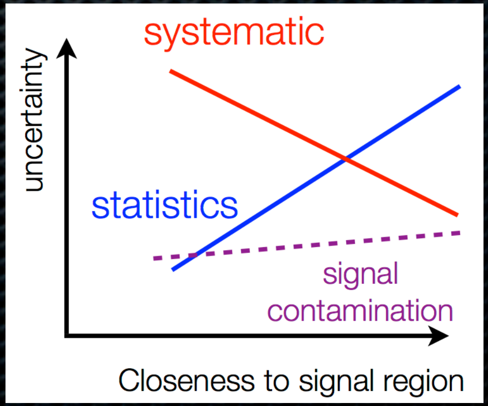
- Use a **semi-data-driven estimate, or pure MC samples**
- **Normalise MC in control regions** for the dominant background: e.g. ttbar, W, Z for the 0-lepton inclusive search
- **Pure MC for minor backgrounds** or ones too similar to signal (such as WW, or ttbar+V)

Semi-data-driven estimate

ATLAS-CONF-2012-109

CR	SR Background	CR process	CR selection
CRY	**$Z(\to \nu\nu)$ + jets**	γ + jets	Isolated photon
CRQ	**QCD jets**	QCD jets	Reversed $\Delta\Phi(j_i, E_T^{miss})$ and E_T^{miss}/m_{eff} cuts
1l CRW	**$W(\to l\nu)$ + jets**	$W(\to l\nu)$ + jets	1 lepton, b-veto, 30 GeV < $m_T(l, E_T^{miss})$ < 100 GeV
1l CRT	**ttbar and single-t**	ttbar->bbqq'lν	1 lepton, b-tag, 30 GeV < $m_T(l, E_T^{miss})$ < 100 GeV

- Example from 0-lepton inclusive search
- Control regions are used to **each constrain one particular SM process**
- Every SR has 4 CRs that **follow the cuts for that SR as close as possible**
- Dedicated **validation regions** with low signal contamination **to test the MC prediction**



Fitting irreducible backgrounds

- **Likelihood given by the product of Poisson distributions** in all the control regions and the signal region, and a term constraining all systematics:

$$L(\mathbf{n}|\mu, \mathbf{s}, \mathbf{b}, \theta) = P_{\rm SR} \times P_{\rm CRW} \times P_{\rm CRT} \times P_{\rm CRY} \times P_{\rm CRQ} \times C_{\rm cyst}$$

- Poisson distribution determined by **expected number of events λ_i and observed number of events n_i**
- The expected number of events depends on the expected signal, the expected background, a fully free signal scale parameter and nuisance parameters

$$\lambda_i(\boldsymbol{\mu}, s_i, \mathbf{b}, \boldsymbol{\theta}) = s_i(\theta) \cdot \mu_s + \sum_j^{W,Z,t\bar{t},\rm QCD} b_{i,j}(\boldsymbol{\theta}) \cdot \mu_j + b_i^{single-t}(\boldsymbol{\theta}) + b_i^{VV}(\boldsymbol{\theta})$$

- **Assuming the shape of the distribution does not change** from a control region to a signal region, **the number of events is given by**

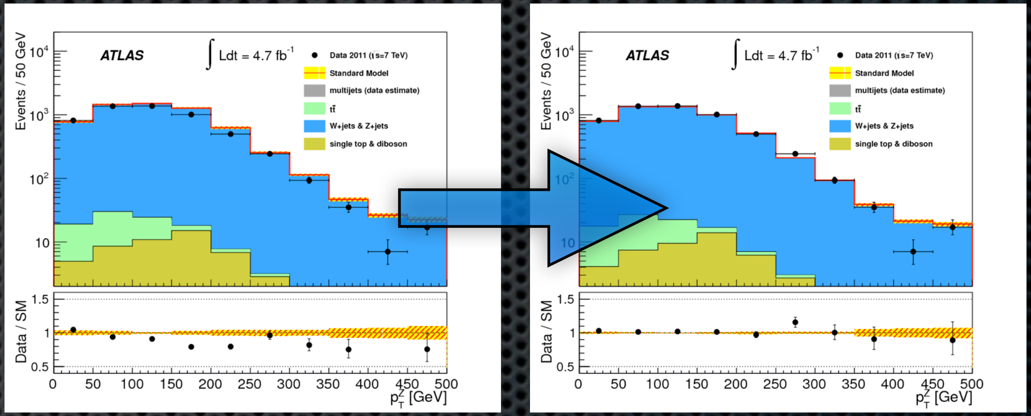
$$N(\rm SR, scaled) = N(\rm CR, obs) \times \left[\frac{N(\rm SR, unscaled)}{N(\rm CR, unscaled)}\right]$$

- The **systematics** considered to estimate the **error made extrapolating** are both **theoretical** (scale dependence, generator modelling, ISR) and **experimental** (JES, JER, pileup) errors (see backup)

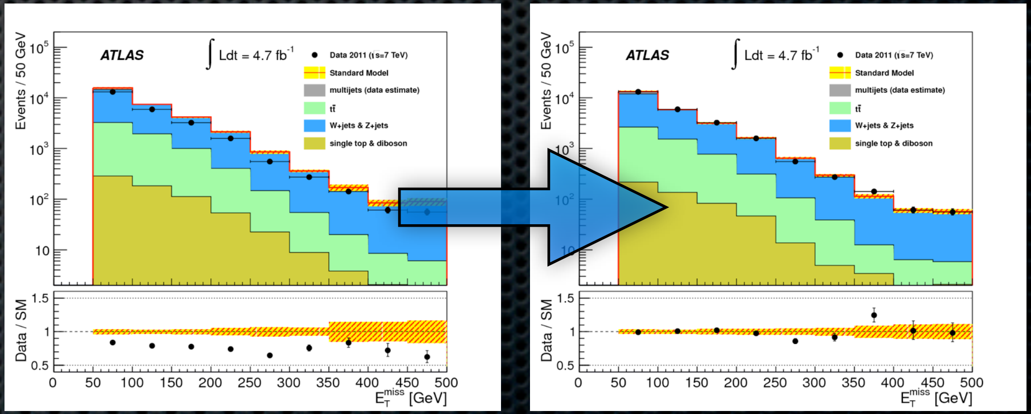
Reweighting

- In certain analyses (e.g. 1-lepton stop search), some MC **backgrounds can be found to be too hard** compared to data
- Alternative generators are used, spectra are checked to **identify the issue** (e.g. Sherpa overestimating high p_T for W+jets)
- **Derive reweighting factors** for several subsets of the background (e.g. categories for various types of W decays coming from ttbar)
- **Apply these factors** to the Monte Carlo after the fit **to match data**

From 2011 1-2 leptons + >=2-4 jets + E_T^{miss} **arXiv:1208.4688**



Z p_T



W E_T^{miss}

Conclusions

- Have **shown how to estimate SM backgrounds** to SUSY analyses
- We discriminate between irreducible and reducible backgrounds

- **Reducible backgrounds** are treated using **dedicated methods**
- These can be validated through **validation regions and closure tests**

- **Irreducible backgrounds** are dominant backgrounds in the signal region
- Such backgrounds are **fitted in dedicated control regions**
- **Or taken from pure MC**, if they are small and/or too signal-like

- (Un?)fortunately, all our **data is described well** by the Standard Model

Backup

Uncertainties

Experimental uncertainties

- Trigger efficiency
- Jet energy scale and resolution
- Lepton energy scale and efficiency
- E_T^{miss} soft component
- b-tagging
- Luminosity
- pileup modelling

Theory uncertainties

- Generator modelling (μ_F,μ_R, ME/PS matching, α_s scale choice when possible; otherwise compare generators)
- Parton shower uncertainties (typically compare Pythia and Herwig)
- PDF choice