Wavelet analysis as a multiresolution method for new particle searches.

Margherita Spalla, Università di Pisa.

Summary

- The problem of searching small signals at unknown mass in high energy physics.
  - Review of standard approach.
- Investigation of a new analysis method (wavelet analysis).
  - Definition
  - Inspection of properties
  - Data
  - Comparison with other methods

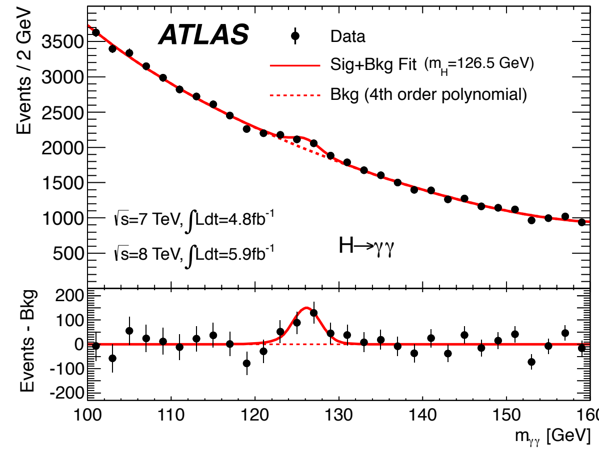
Context

- The search of new physics phenomena is one of the most relevant topic in recent years high energy physics.
  - Direct search at accelerators (LHC) requires an enormous effort from the worldwide physic community.
  - Up to now, no evidence of new physics have been found at LHC.

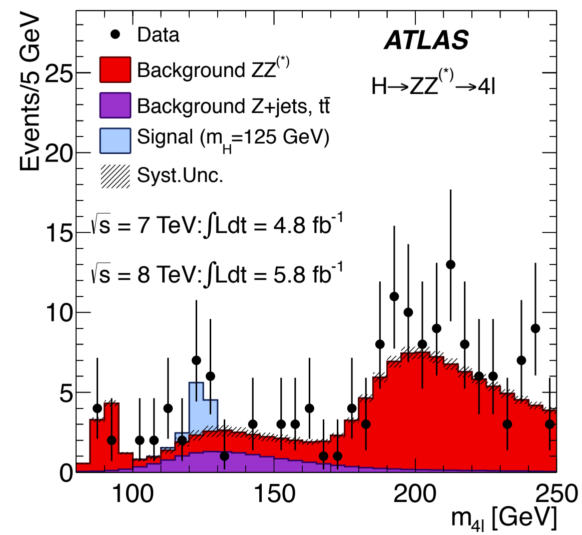
- The problem of searching a new particle in LHC-like conditions:
  - The signal is very small.
  - The new particle's mass is unknown.
  - The background is in general (very) huge.

Particle measurement: a reminder

- A decaying particle is detectable as a resonance in the invariant mass spectrum of it's decay products.
  - Both it's mass and cross section are determined from the mass spectrum.



- In the scenario we're considering, such a signal should be distinguished among a background orders of magnitude huger than the signal itself.
  - Background shape can be simulated with MonteCarlo (MC) or fitted to the data themselves.
  - Any excess with respect to background is then considered.
  - Standard statistical tools are applied to determine if a given excess is due to statistical fluctuation or is the evidence of a new particle.



Hypothesis test

- Check if the data are incompatible with a given hypothesis (*null hypothesis* H_0).
  - This is done computing the probability (*p-value*) of finding an excess equal or greater than the one in data.
  - A small *p-value* indicates that the null hypothesis can be rejected (i.e. the excess is a signal)

- More complex extensions of the hypothesis test have been developed for the case in which there is no information on the mass of eventual signals (multiple hypothesis test, Bump Hunter).

- In most cases, this kind on analysis relies on background modeling.

Alternative method: wavelet analysis

- Wavelet analysis was developed to detect localized structures in time series, it is based on *wavelet transform*.
  - Wavelet transform is an evolution of Fourier transform, substituting the plane wave function with a local complex function Ψ(ξ).

$$\hat{x}(\omega) = \int x(t) \underbrace{e^{-i\omega t}} dt \quad \Rightarrow \quad W(t,s) = \int x(t') \underbrace{\psi^* \left(\frac{t-t'}{s} \right)} dt'$$

- It can be applied to the analysis of any random variable *m* of density *f(m)*.
  - Here, *f(m)* is the invariant mass spectrum.

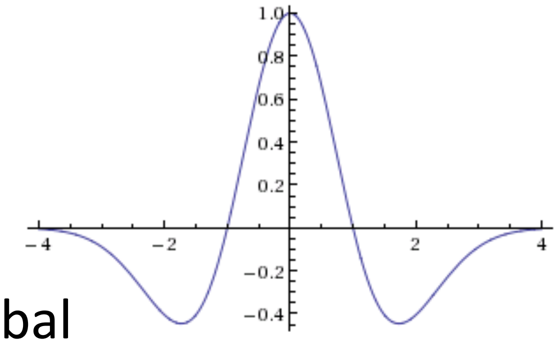
- It is applied in a variety of fields:
  - Denoising tool (e.g. in gravitational waves experiments)
  - Data compression (JPEG standard)
  - Analysis of quasi periodic phenomena (geophysics, metereology)
  - Detection of weak light sources in photon counting detection images

Wavelet analysis: an introducton

- The wavelet analysis is a multiscale method: it allows to separate structures of different dimension in mass.

- Wavelet transform (continuous):
  - Here, ψ is the *Mexican Hat* (DoG) function.
  - It can be any local function with zero mean.

$$W(m,s) = \int f(m')\psi^*\left(\frac{m-m'}{s}\right) dm'$$



- Varying *m* and the *scale s*, *W(m,s)* gives a global picture of *f(m)* features.

$$s_j = s_0 2^{j\delta j}, \qquad j = 0, 1, ..., J$$

- In practice, *f(m)* is substituted by the mass histogram.

$$W(m = n \cdot \delta m, s) = \sum_{n'=0}^{N-1} x_{n'} \cdot \psi^*\left(\frac{(n'-n)\delta m}{s}\right)$$

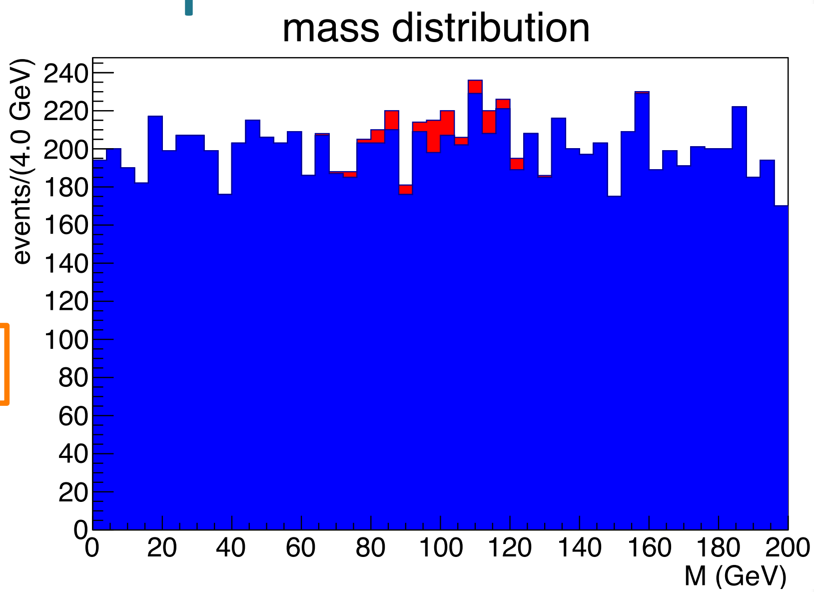
Wavelet analysis: an example

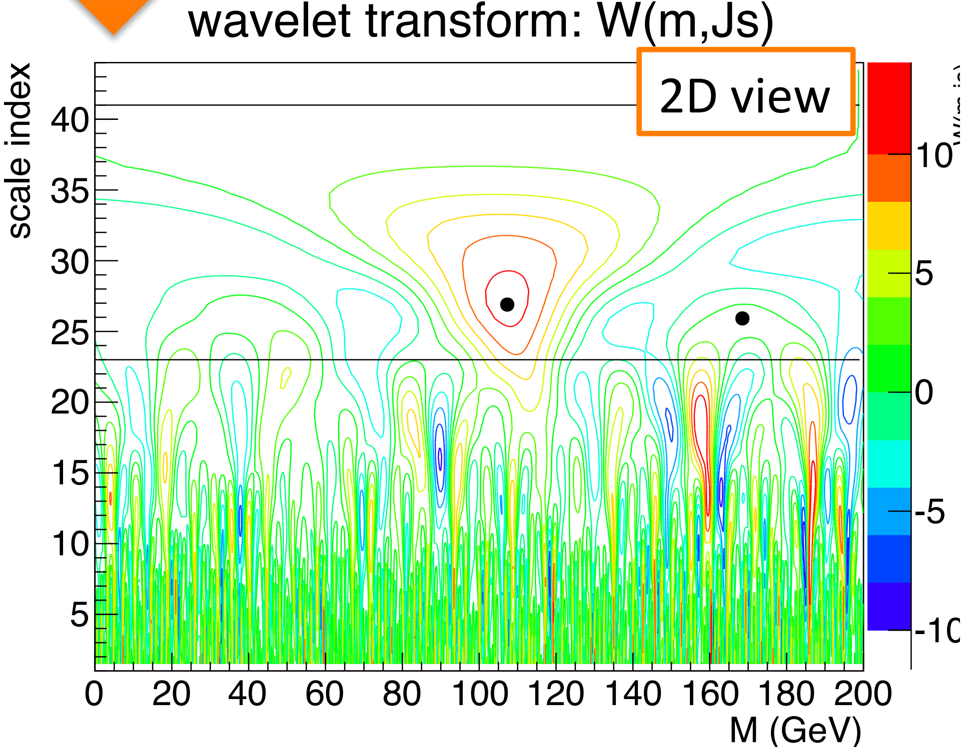
Flat background: 10000 events.

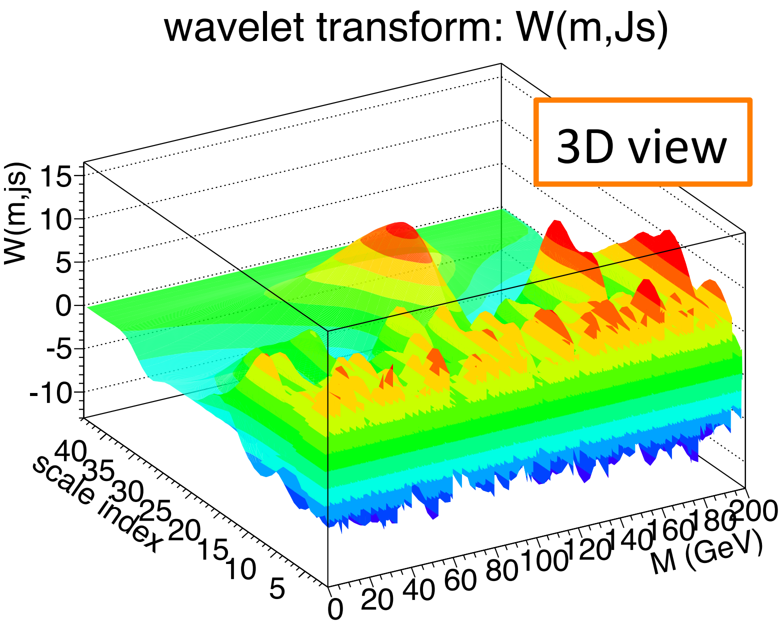
Gaussian signal: 100 events,
mean μ=100 GeV, width σ=15 GeV.

Invariant mass spectrum, small binning

Wavelet transform



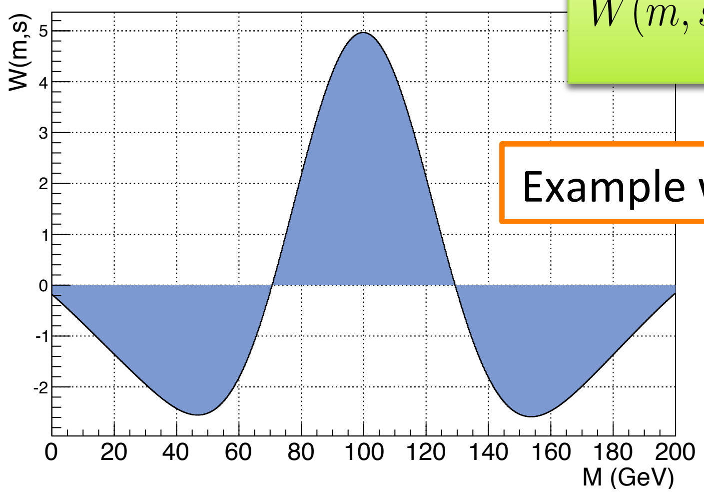




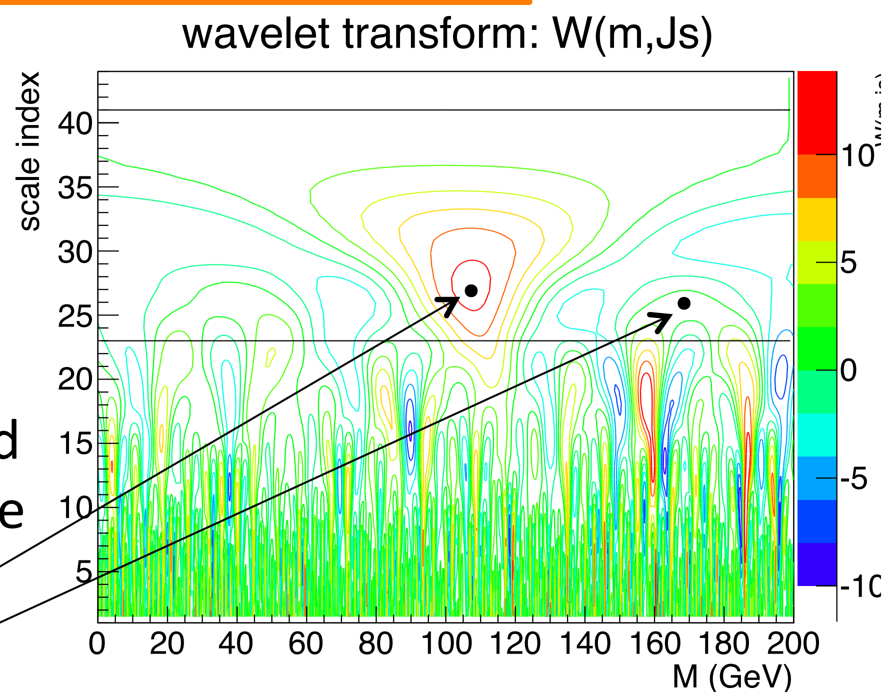
Signal expectation

- The wavelet transform of a gaussian signal is expected to be a bell-shaped function of *m*, peaking at the signal mass.
  - The peak height can be an estimator of the number of signal events N_{ev}.

$$W(m,s) \;=\; A(s,\sigma)\cdot\delta m\cdot N_{ev}\cdot\left(1-\frac{(n\delta m-\mu)^2}{\sigma^2+s^2}\right)e^{-\frac{(n\delta m-\mu)^2}{2(\sigma^2+s^2)}}$$



Example without background: $W(m,j_s=29)$

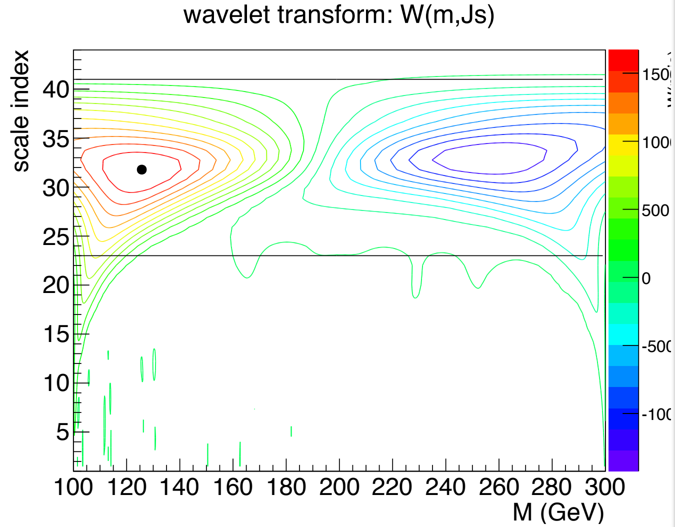


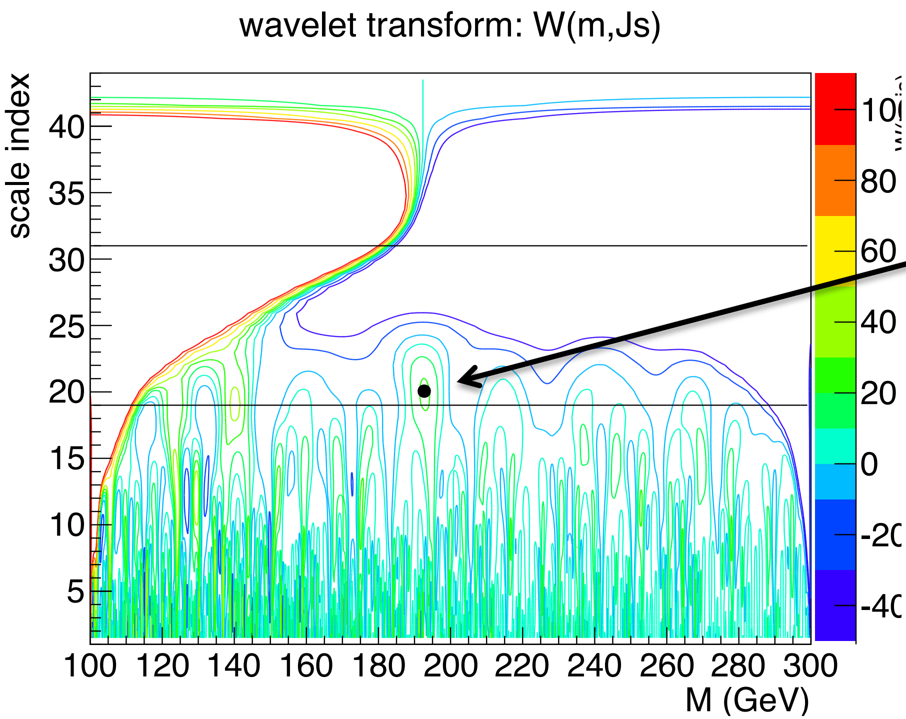
- To reject background, only wavelet transform maxima found in an appropriate scale region are considered.

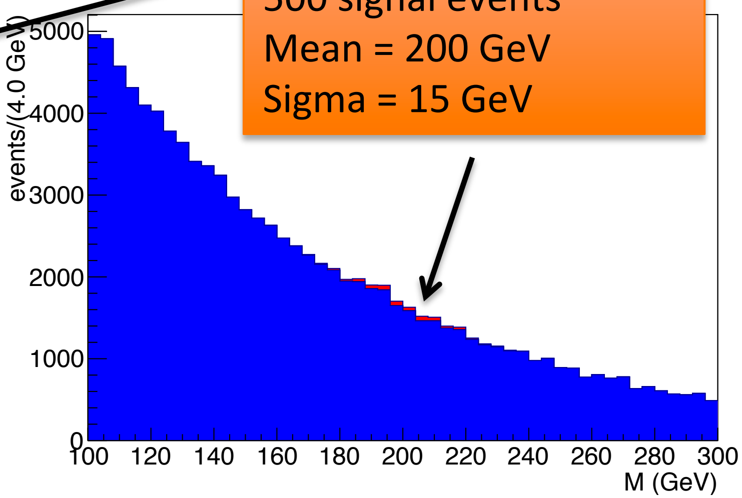
$$j_s \in [24, 40]$$

Non flat background

- A non flat background strongly affects the wavelet transform, making difficult to identify the signal.
  - With an appropriate choice of acceptable scales region and contour levels, a not too small signal can still be visible.
  - The method performances are strongly reduced.



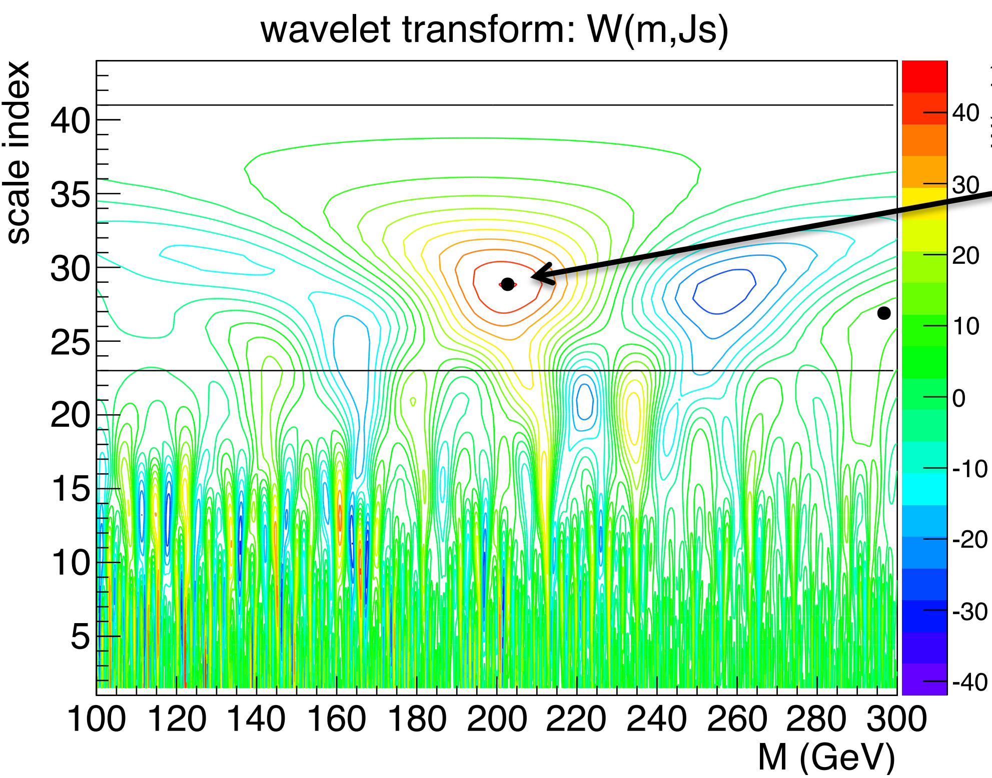




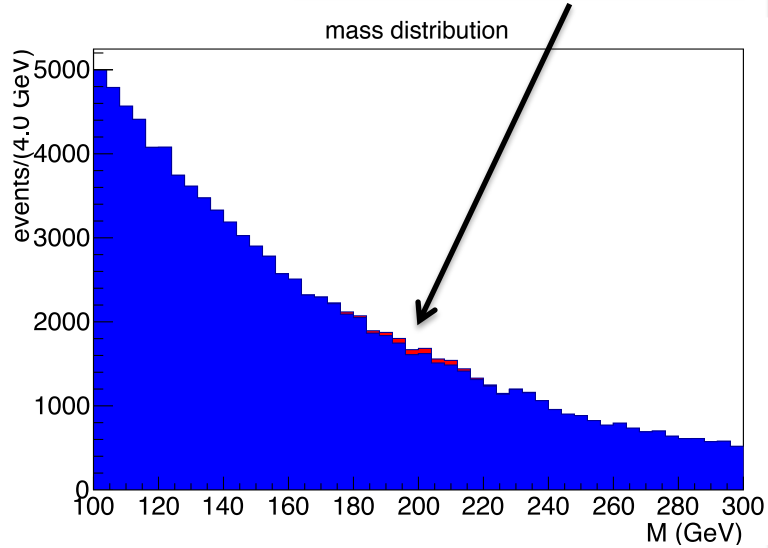
10^5 background events
500 signal events
Mean = 200 GeV
Sigma = 15 GeV

Non flat background (2)

- A more efficient solution is to fit the background shape and subtract the fit result from real data.
- Since the wavelet analysis is sensitive to any small structure, fit quality is a very delicate point.
  - Simulated samples represent the ideal case in which the background is a pure exponential, this is not the case in real data.



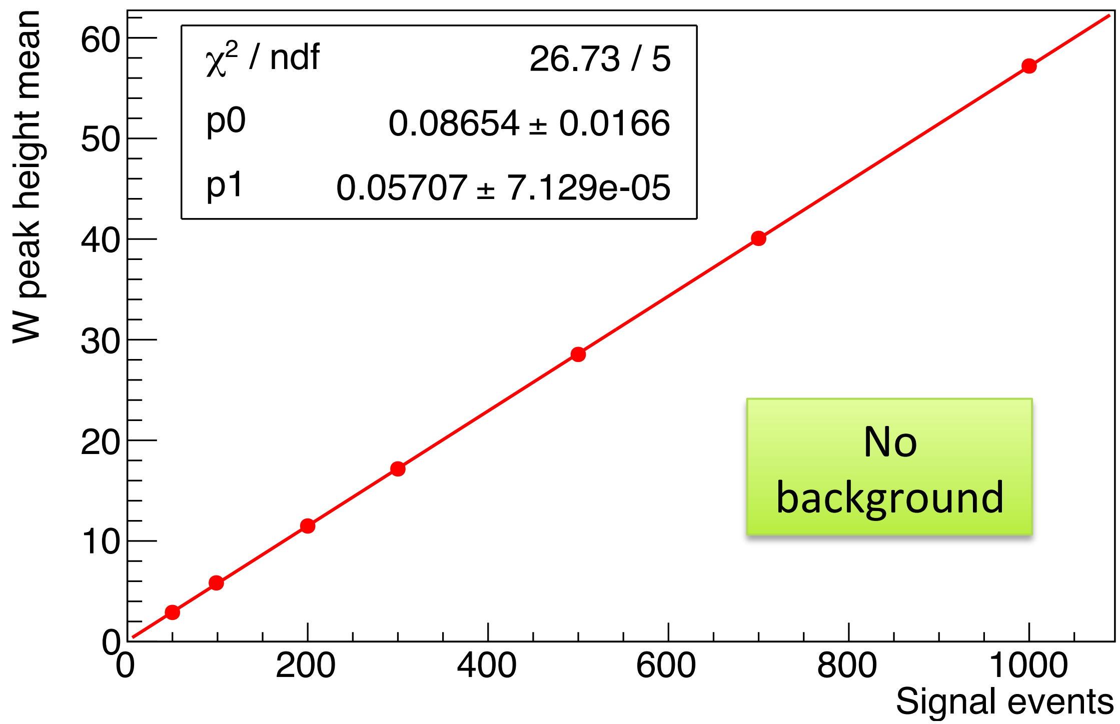
10^5 background events
500 signal events
Mean = 200 GeV
Sigma = 15 GeV



Testing the method: W(m,s) vs Number of signal events

- Background samples with a gaussian signal (μ=100 GeV σ=15 GeV) have been generated using toy MonteCarlos.
- The height of *W(m,s)* peaks has been measured varying the number of signal events (*S*).
  - Only peaks found inside the acceptance region [μ-σ;μ+σ] are considered.

- A first check of the method has been done generating only signal events.

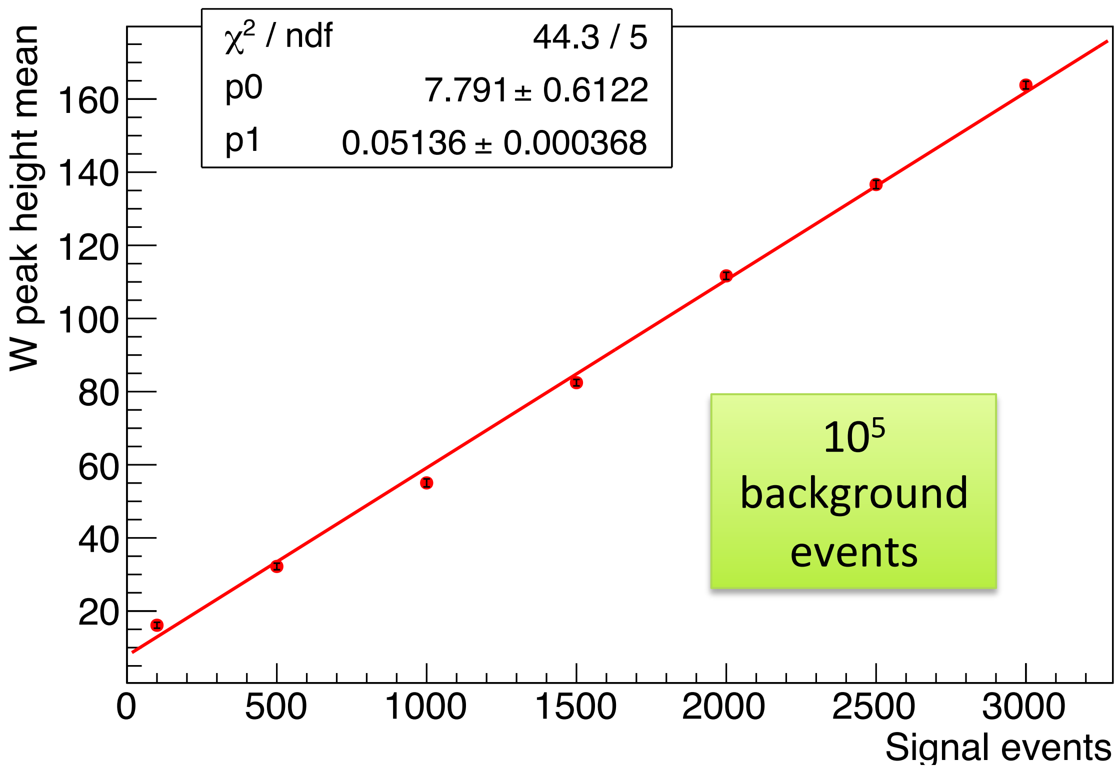


Testing the method: W(m,s) vs Number of signal events

- Adding 10^5 (exponential) background events *W* maxima height is still a linear function of *S* with good approximation.
  - The background has been fitted and subtracted before the wavelet transform.
  - The constant term has increased because of background.
  - The slope has variation smaller than percent.

Result without background:

p0	0.08654 ± 0.0166
p1	0.05707 ± 7.129e-05



Statistical treatment

- An hypothesis test is performed locally, evaluating *W(m,s)* distribution, fixed *m,s*.
  - x_n are Poisson variables: we assume gaussian approximation to be valid.
    - » The data arithmetic mean is subtracted before making the wavelet transform, then x_n should have zero mean.

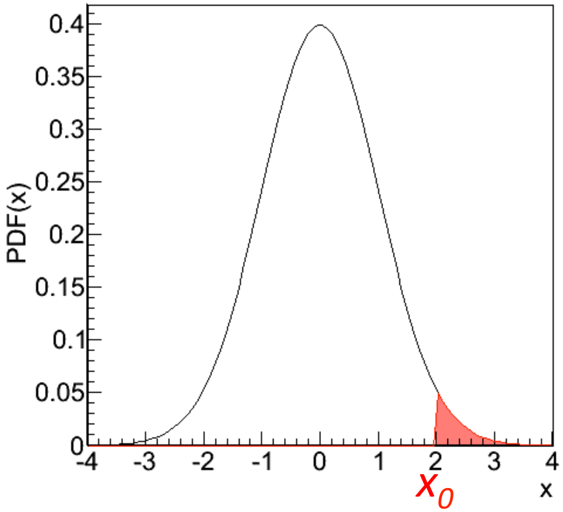
$$W(m,s) = \sum_{n'=0}^{N-1} x_{n'} \cdot \psi^* \left(\frac{(n'-n)\delta m}{s} \right) \quad \Rightarrow \quad W(m,s) \sim N(0, \sigma_{m,s})$$

$$\sigma^2_{(m,s)} = Var(W(m,s)) = \sum_{n'=0}^{N-1} x_{n'} \cdot |c_{n'}(m,s)|^2$$

- The *p-value* is computed given the value of $W(m,s)/\sigma_{m,s}$.

$$p\text{-}value = \int_{x_0}^{\infty} N(0,1)dx$$

$$x_0 = W(m,s)/\sigma_{m,s}$$

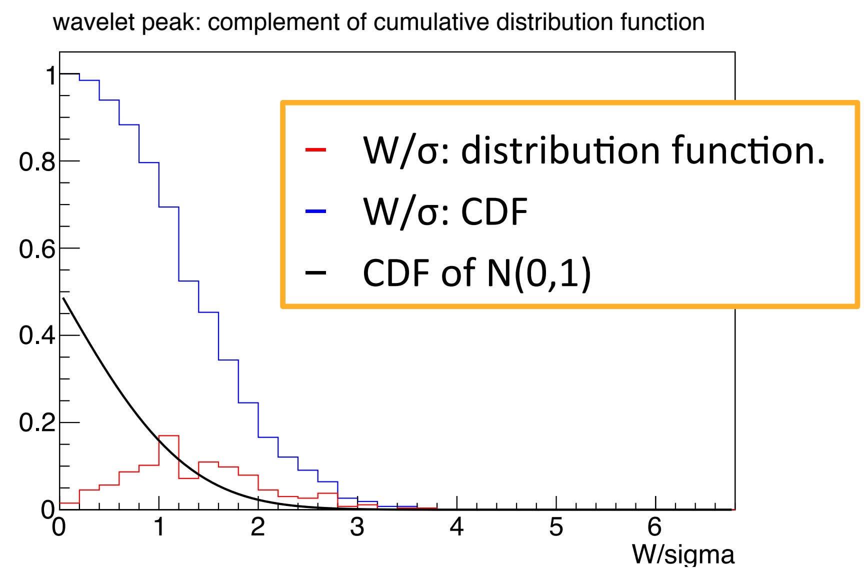


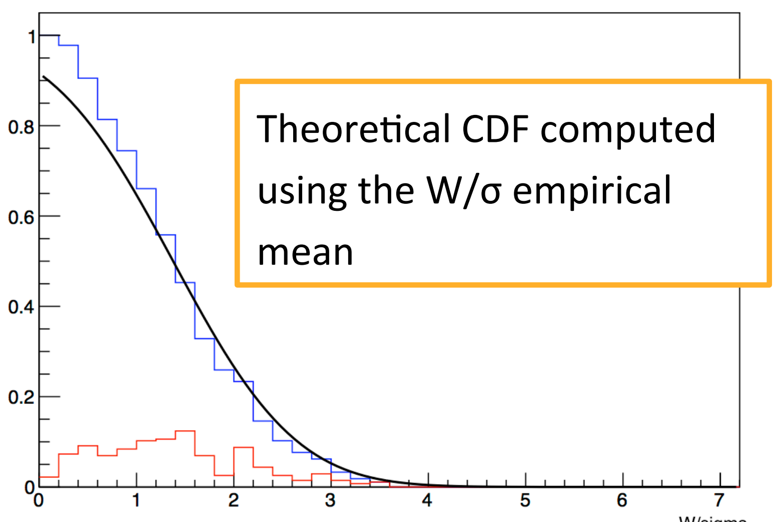
Statistical treatment: empirical check

- We have checked the assumption $W(m,s)/\sigma_{m,s} \sim N(0,1)$ computing empirically the cumulative distribution function (CDF) for $W(m,s)/\sigma_{m,s}$ using toy MonteCarlos with no signal added.

$$\mathrm{CDF}((\mathrm{W}/\sigma)) = \sum_{(W/\sigma)' > (W/\sigma)} (W/\sigma)'$$

- The distribution mean is nonzero: probably because we are considering *W* maxima, not single bins.





- Empirically, the *W/σ* mean is independent on the number of background events.

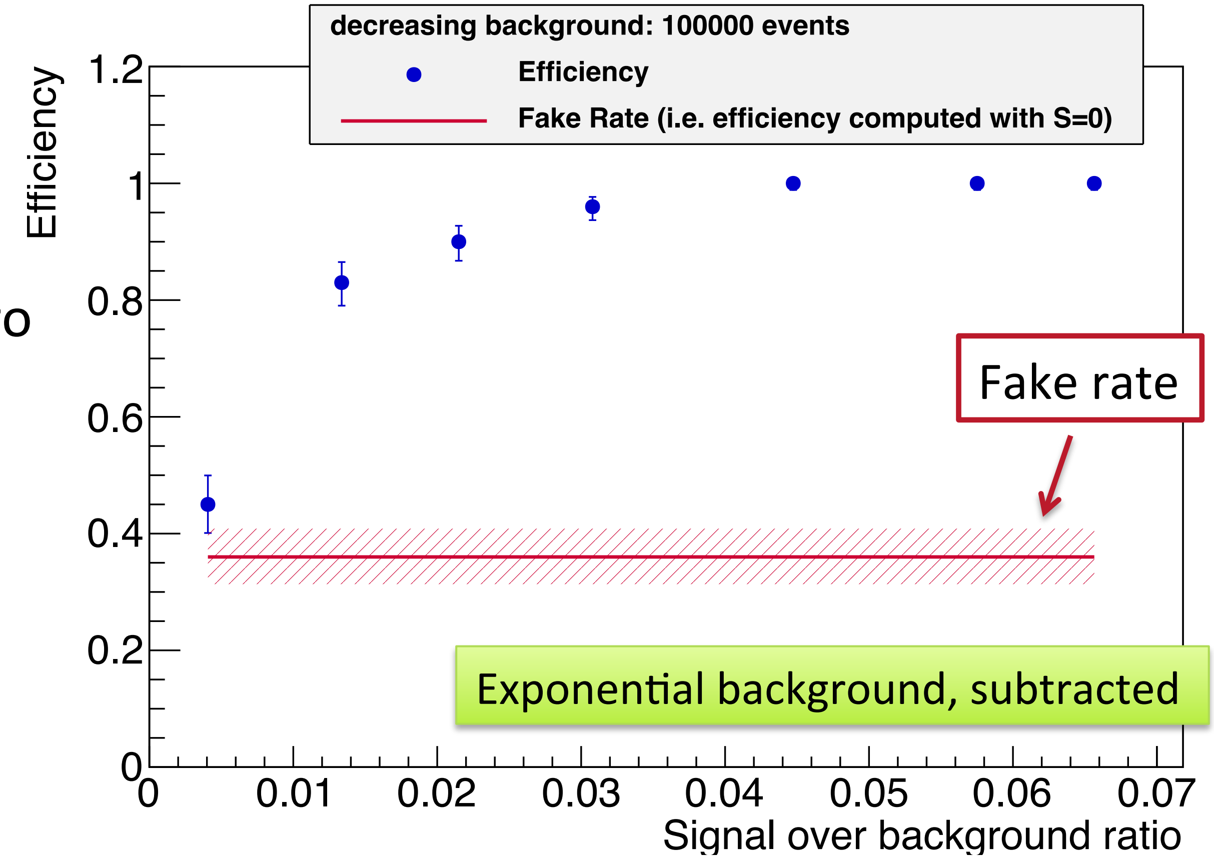
$$\mu_{flat} = 1.37 \pm 0.01$$
$$\mu_{exp} = 1.31 \pm 0.01$$

Efficiency

- Efficiency is computed injecting a gaussian signal (*N(μ,σ)*) of known mass on a simulated background.
  - Efficiency is the fraction of cases in which a *W(m,s)* peak is found in the mass window [μ-σ,μ+σ].

- Fake rate is computed by applying the same algorithm with zero signal events.

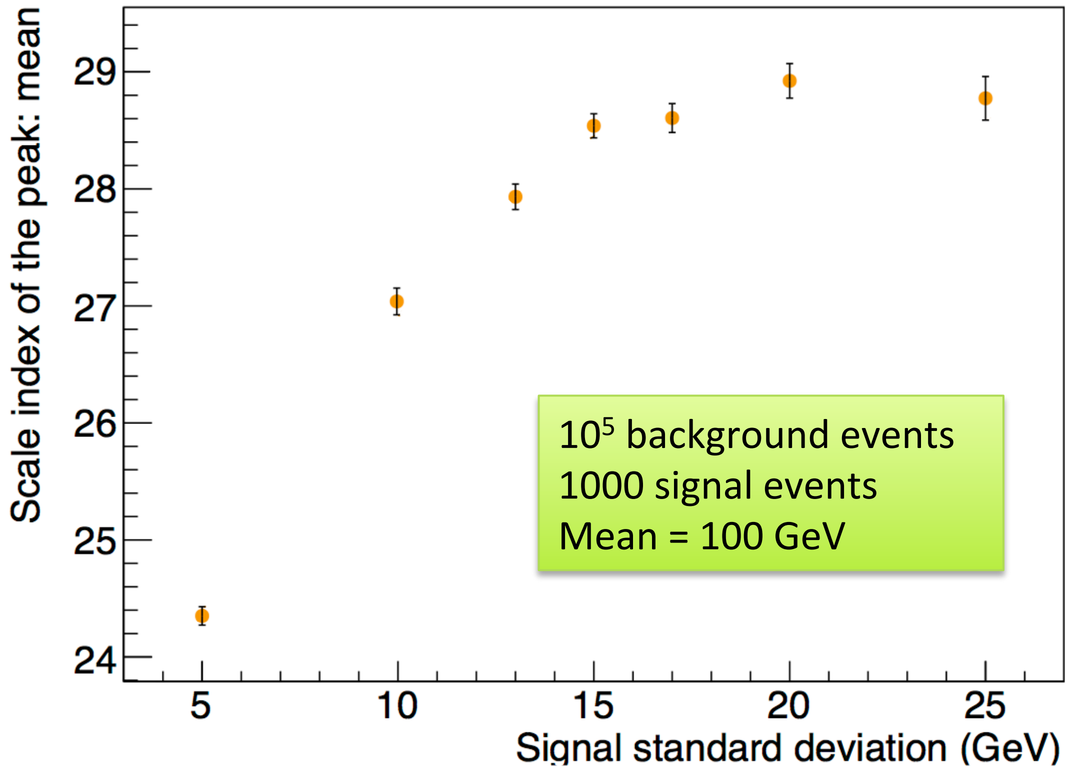
- No requests are made about the significance of the found peaks.



Estimation of the signal width

- From the theory, a wider signal is expected to peak at larger scales (i.e. at larger scale index j_s)
- This property has been investigated using toy MonteCarlos.

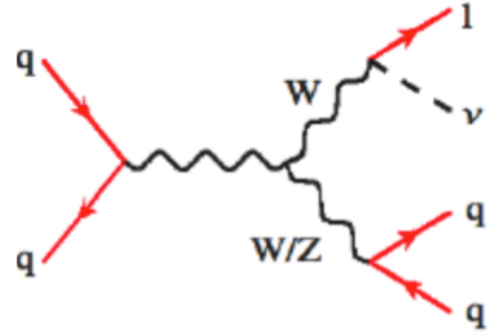
  - The scale index j_s of the peak has been measured varying the signal standard deviation.
  - The scale index of the peak tends to saturate for large signals.



Application to real data: available samples and event selection

- Invariant mass of jet pairs produced in association with a leptonically decaying *W*.
  - Data acquired by the ATLAS experiment in 2011: $\sqrt{s}=7$ TeV and L=4.702 fb^{-1}.
  - The Standard Model main contribute is the diboson production (*WW/WZ*).
    - Cross section from the ATLAS collaboration:

$$\sigma_{WW/WZ} = 72 \pm 9\ \text{(stat.)} \pm 15\ \text{(syst.)} \pm 13\ \text{(MC stat)}\ \ \text{pb}$$

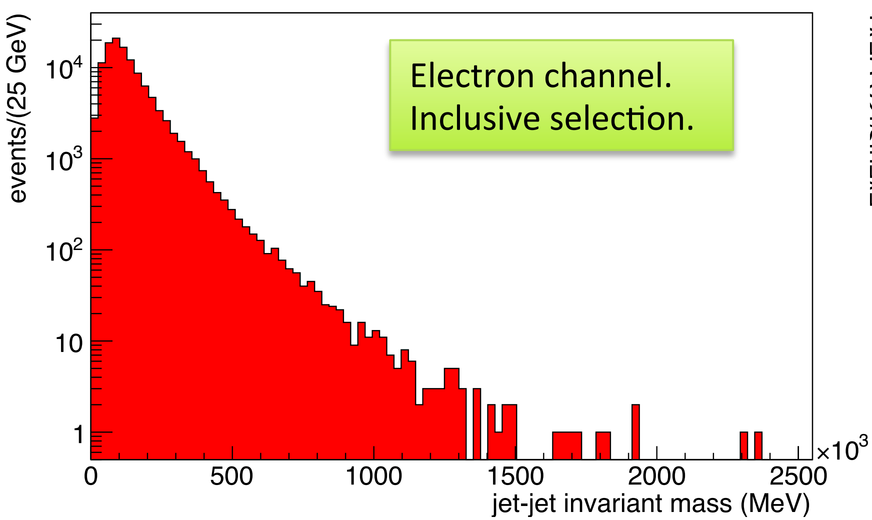


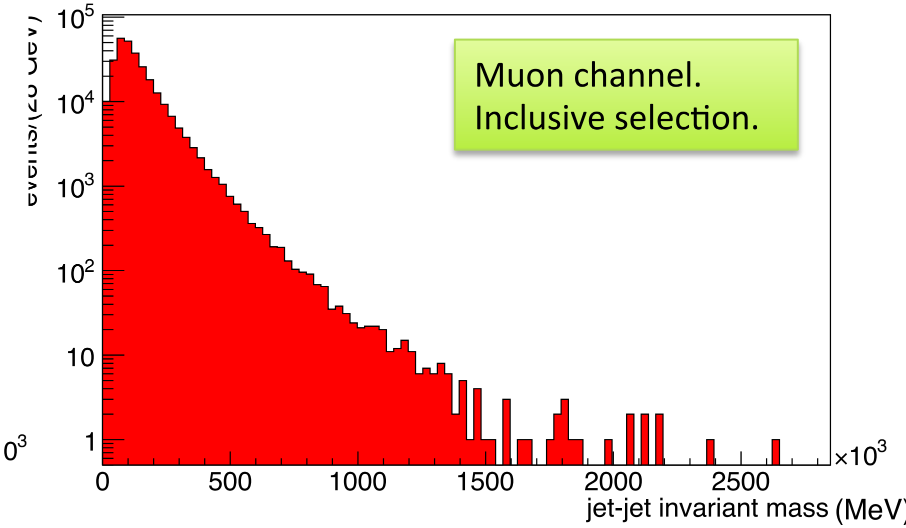
- Dijets events are selected by requiring a W ⟶ lν decay.
    - Events must contain one single charged lepton passing the object selection and large missing transverse energy (i.e. a neutrino)
    - Cut on lepton-neutrino transverse mass to select *W* events.
- A further selection is applied to jets to reduce background.

Application to real data: M_{jj} samples

- The analysis have been performed independently for two channels: electron channel and muon channel.
    - Depending if the selected lepron is an electron or a muon.

- For most of our analysis, the combined sample will be used.





- With a rough evaluation of the selection acceptance the number of expected diboson events is estimated to be of order 10^3 for both channels.

The fit problem

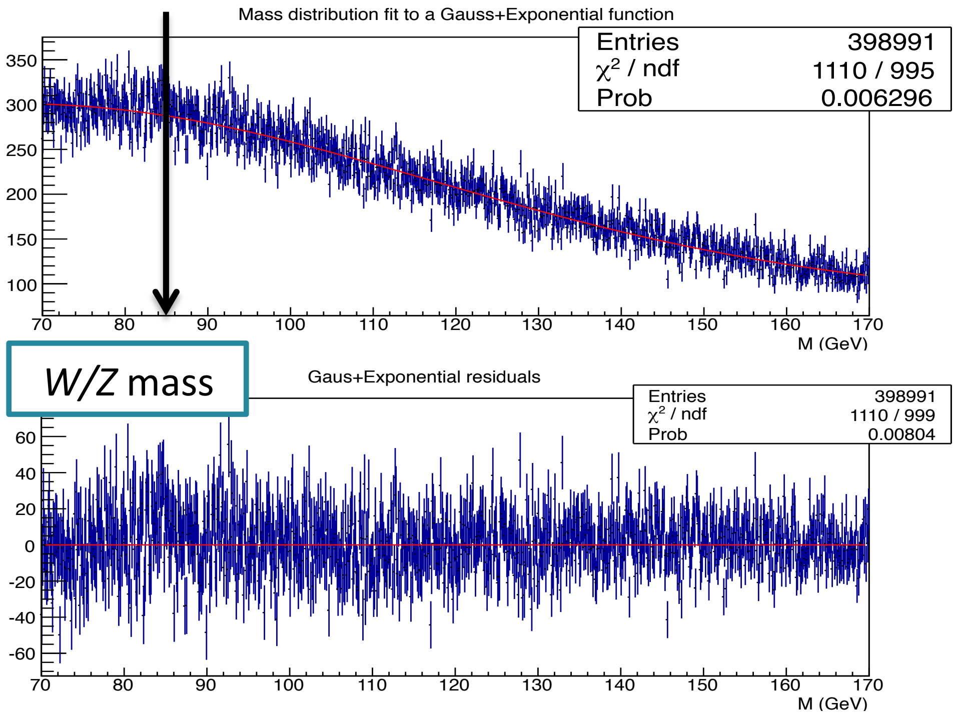
- In real data, the background shape is not a simple exponential as in toy models.
    - A combination of gaussian and exponential has been used as fit function.
- Test of the pure exponential approximation:
    - The mass range [130,330] GeV has been used as control region.
    - A simulated signal has been inserted at μ=200 GeV
    - <u>The signal has been detected at the right mass, but its width and number of events are underestimated.</u>

Inserted signal	Detected signal
$\mu = 200$ GeV	$m_{peak} = 197.9$ GeV
$\sigma = 15$GeV	$j_s = 25$
signal events $= 900$	$W_{peak} = 30.8$
	$p\text{-}value = 0.21$

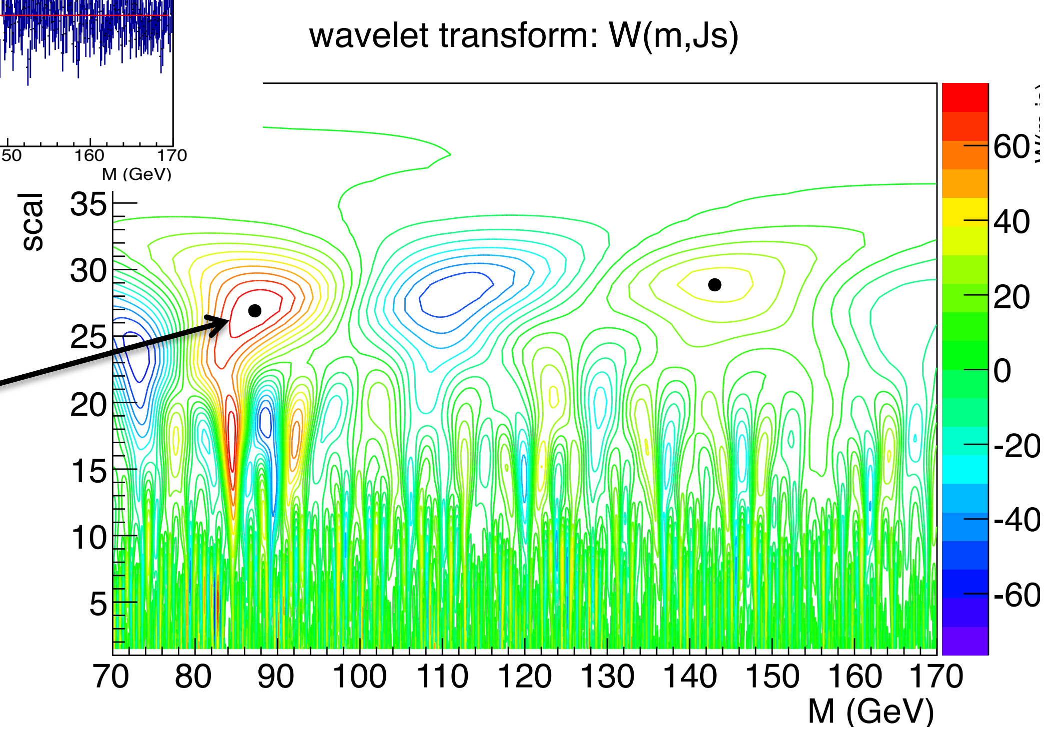
$\Rightarrow$ $\sigma \sim 6$ GeV, signal events ~ 460

- The method should be calibrated using real data or a more precise MC simulation.

Data analysis: combined (e+μ)



- Signal expected peak is very close to background peak: due to fit difficulties, this check of wavelet method is not much significative.



- A peak is found at the expected mass.

$$m_{peak} = 87.2 \text{ GeV}$$
$$j_s = 27$$
$$W_{peak} = 76.7$$
$$p\text{-}value = 0.003$$

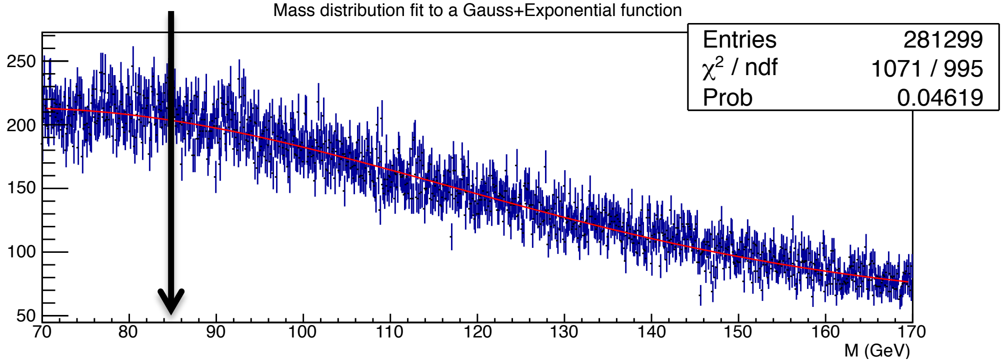
Data analysis: results

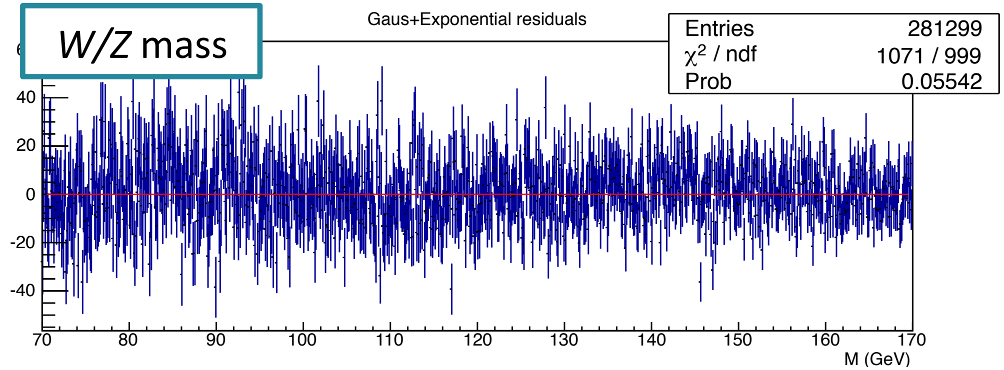
- A peak is found at the expected mass.
  - The detected number of events is about 1/5 of the expected one (a fewer fraction than obtained with the simulated signal).
  - The underestimation of the signal width is fewer than what observed in the calibration.

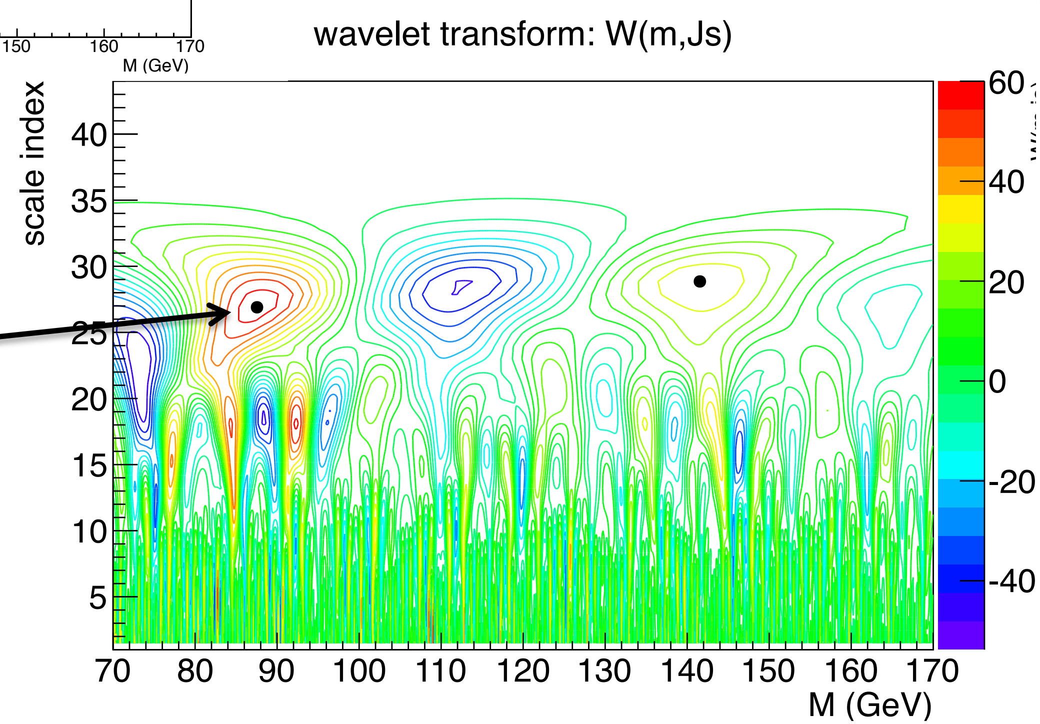
	Expected	Measured
Signal events	$\sim 5 \cdot 10^3$	$\sim 1.5 \cdot 10^3$
Signal standard deviation	~ 15 GeV	~ 10 GeV

- A peak is found in both electron and muon channel treated separately:
  - The masses are clearly compatible
  - The peaks have the same width.
  - The sum of the number of events detected in the two channels is compatible with the number of events measured in the combined case.

Data analysis: muon channel

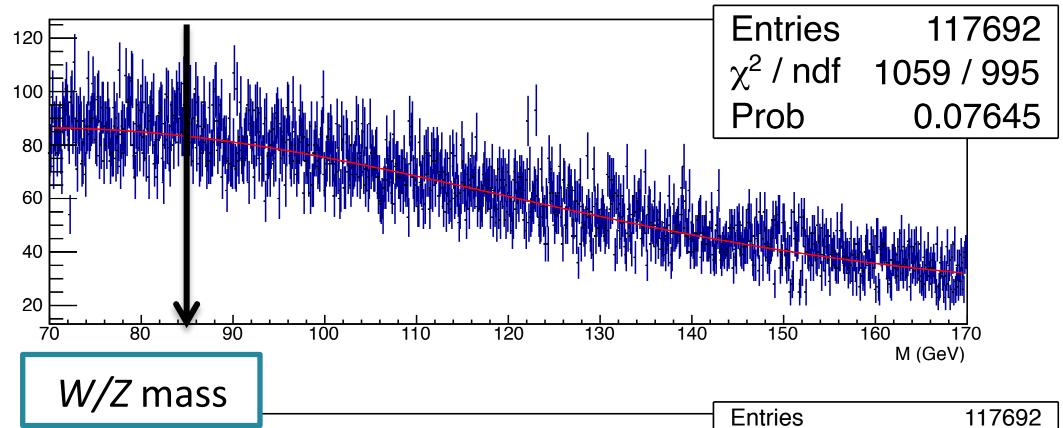


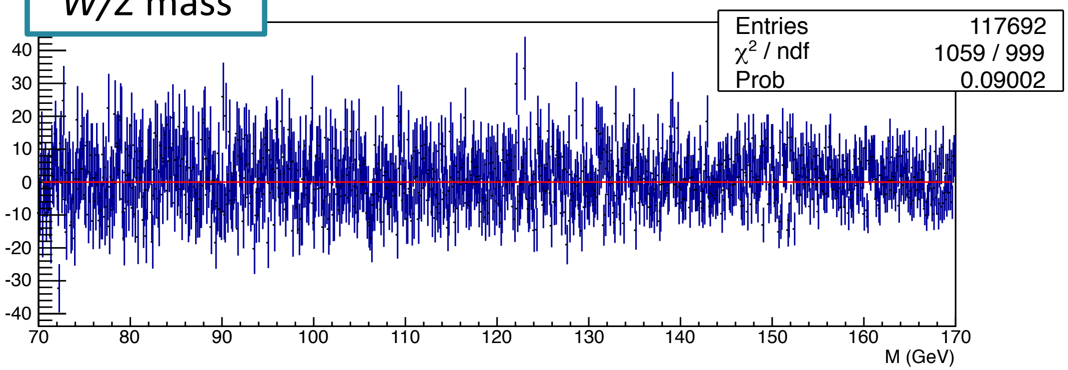




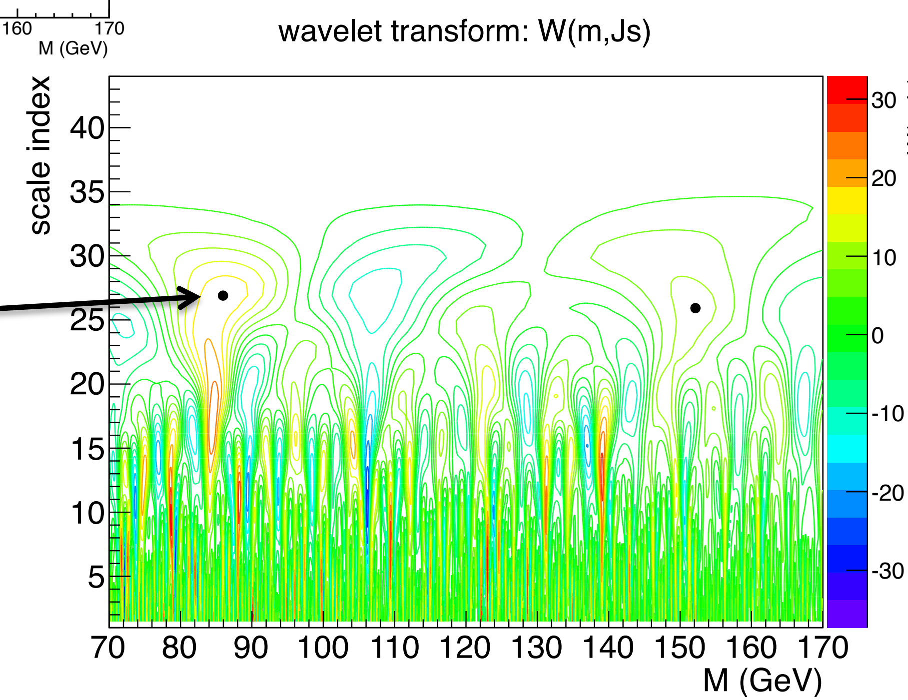
$$m_{peak} = 87.55 \text{ GeV}$$
$$j_s = 27$$
$$W_{peak} = 57.9$$
$$p\text{-}value = 0.01$$

Data analysis: electron channel





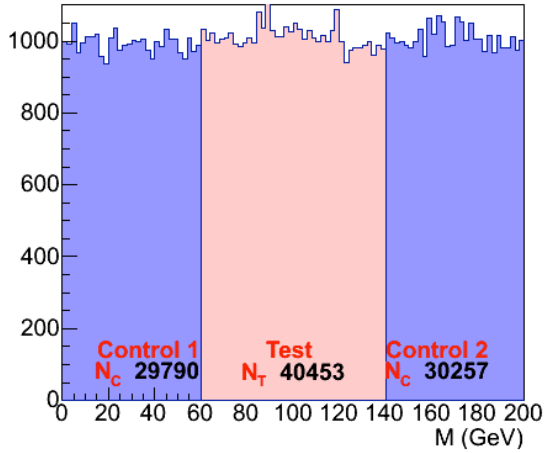
$$
\begin{aligned}
m_{peak} &= 85.95 \text{ GeV} \\
j_s &= 27 \\
W_{peak} &= 17.9 \\
p\text{-}value &= 0.3
\end{aligned}
$$



Comparison with other methods

- Simple hypothesis test based on the likelihood ratio.
  - Consider a variable *x* of PDF $f(x; \vartheta_1 \ldots \vartheta_n)$. Given a measurement x_0 of *x*, the likelihood (1) is a function of the parameters ϑ_i

$$L(x_i|\theta_1 \ldots \theta_n) = f(x_i;\theta_1 \ldots \theta_n) \qquad (1)$$



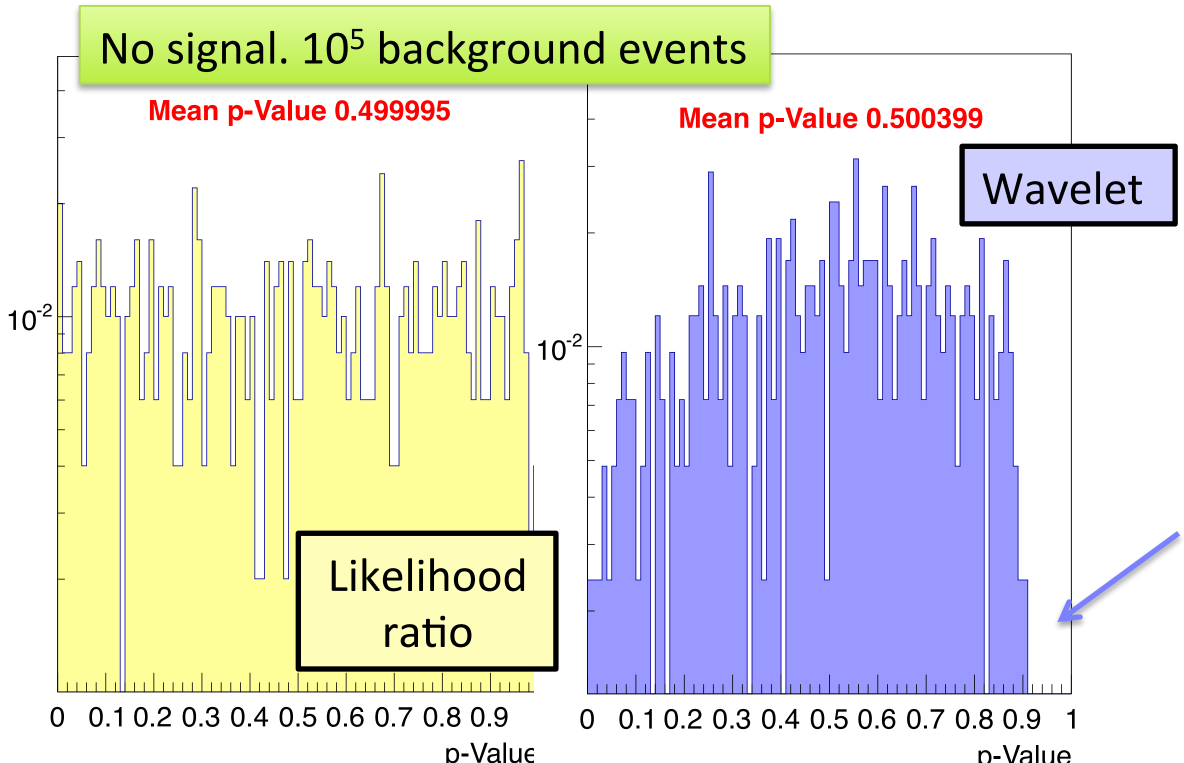
- The mass spectrum is divided in a test region, where the signal is expected, and a control region.
  - The number of events in the control region is used to estimate the background.
- The test statistic is the likelihood ratio Λ .

$$\Lambda = \frac{L(N_T, N_C|S=0)}{L(N_T, N_C|S\neq 0)} \qquad -2\log\Lambda \sim \chi^2(1)$$

- For this test, the expected mass of the signal must be known. More general cases are being considered at present.

Direct comparison with wavelet analysis

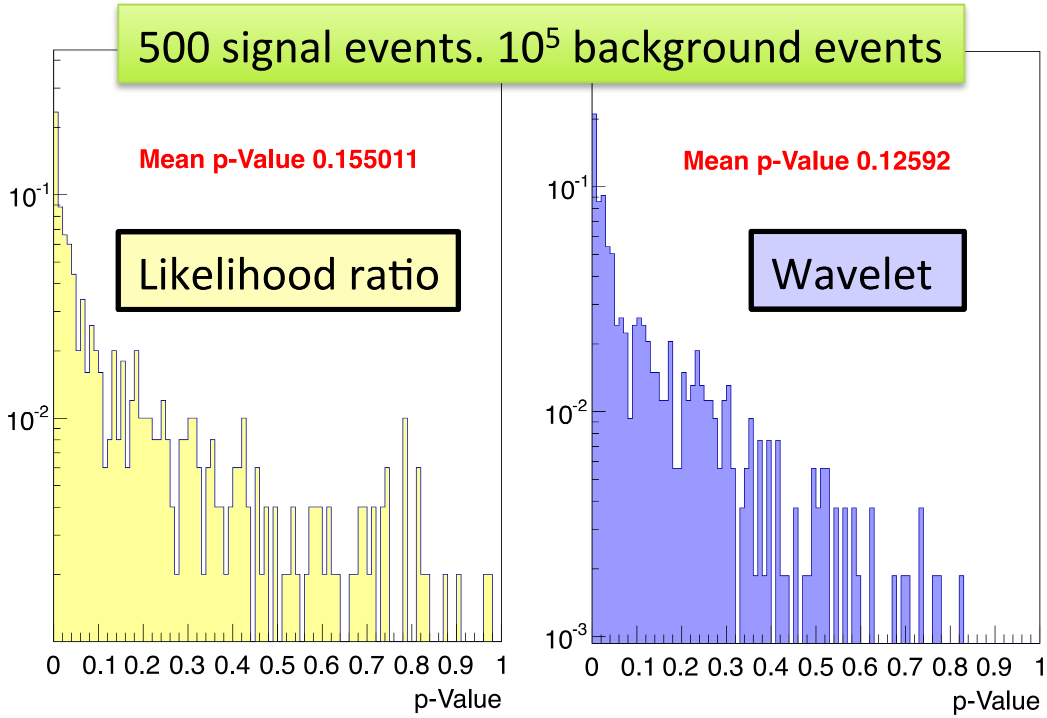
- Generate pseudo-experiments given B and S, for each one compute the *p-value* and plot it.
  - If the null hypothesis is true, the *p-value* is uniformly distributed between 0 and 1.
  - To reproduce the same conditions in the two test, the wavelet peaks are required to have a mass difference smaller than a standard deviation from the inserted signal.



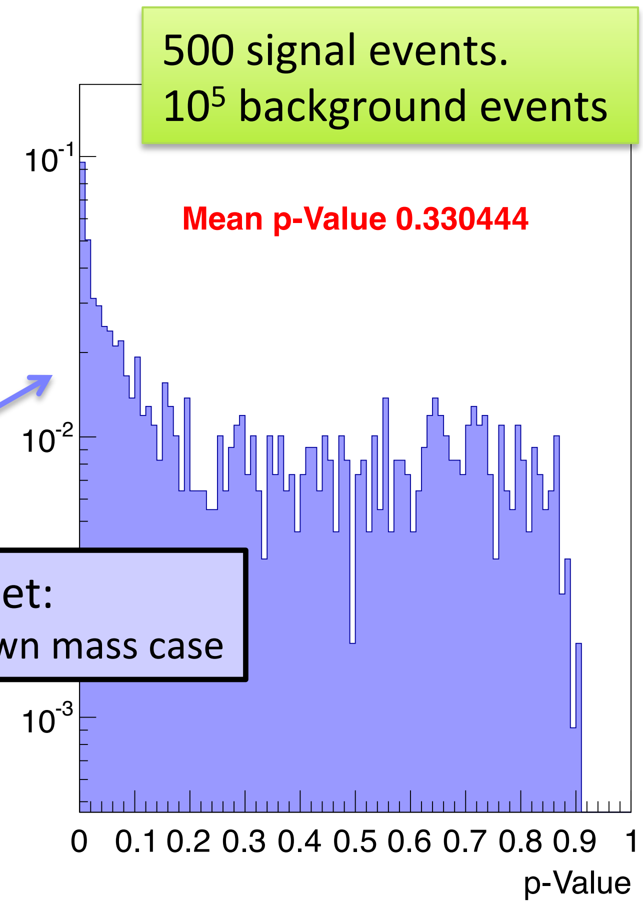
Wavelet *p-value* is only approximately flat: the gaussian approximation for W/σ is not perfect.

The distribution does not reach 1 because we are considering local maxima, not single bins.

Direct comparison with wavelet: signal



Mean *p-values* are comparable in the two cases.



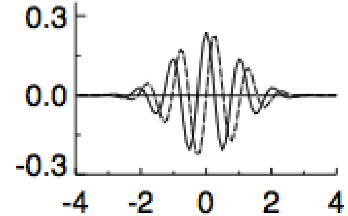
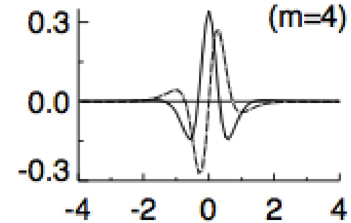
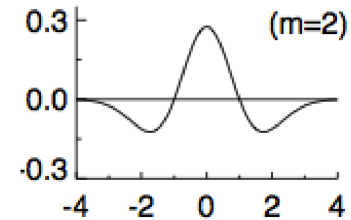
Eliminating the requirement on the wavelet peak's mass, the *p-value* still peaks at zero.

Extending other tests to the case of unknown mass requires much more complications.

Conclusions

- Wavelet analysis provided very promising results with toy MonteCarlo simulations:
  - It is highly sensitive in the detection of small signals over large background
  - It's response is linear in the number of signal events
- When considering more realistic background distributions, the method appears less performant:
  - Further calibration studies should be done in more realistic conditions.
- Further test on known resonances should be done, avoiding patological background conditions.
- Comparative studies with standard research methods should be developed.

BACKUP

Name	$\psi_0(\eta)$	$\hat{\psi}_0(s\omega)$	$\psi_0(\eta)$ (graphic)
Morlet (ω_0 = frequency)	$\pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}$	$\pi^{-1/4} H(\omega) e^{-(s\omega-\omega_0)^2/2}$	
Paul (m = order)	$\frac{2^m i^m m!}{\sqrt{\pi(2m)!}}(1-i\eta)^{-(m+1)}$	$\frac{2^m}{\sqrt{m(2m-1)!}} H(\omega)(s\omega)^m e^{-s\omega}$	
DOG (m = derivative)	$\frac{(-1)^{m+1}}{\sqrt{\Gamma\left(m+\frac{1}{2}\right)}} \frac{d^m}{d\eta^m}\left(e^{-\eta^2/2}\right)$	$\frac{-i^m}{\sqrt{\Gamma\left(m+\frac{1}{2}\right)}}(s\omega)^m e^{-(s\omega)^2/2}$	

$H(\omega)$ = Heaviside step function, $H(\omega) = 1$ if $\omega > 0$, $H(\omega) = 0$ otherwise.
DOG = derivative of a Gaussian; $m = 2$ is the Marr or Mexican hat wavelet.

Three wavelet mother functions and their Fourier transform. Constant factors for ψ_0 and $\hat{\psi}_0$ are for normalisation. The plots on the right give the real part (solid) and imaginery part (dashed) for the wavelets as functions of the parameter η.

Reference:

C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.

Details on wavelet transform calculation

$$W(m,s) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\psi}^*(s\omega_k) e^{i\omega_k n\delta m}$$

- It is considerably faster to compute the wavelet transform in Fourier space.
  - The discrete Fourier transform of x_n is: $\hat{x}_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N}$
  - $\hat{\psi}(s\omega_k)$ is the Fourier transform of a (continuous) function ψ(*m*/*s*).

$$\omega_k = \begin{cases} \frac{2\pi k}{N\delta m} & \text{if } k \leqslant \frac{N}{2} \\ -\frac{2\pi k}{N\delta m} & \text{if } k > \frac{N}{2} \end{cases}$$

- *W(m,s)*, as a continuous function of *s*, can be approximated by computing the wavelet transform for a set of scales.
  - s_0 is the smallest resolvable scale: $s_0 = \delta m$
  - *δj* sets the smallest wavelet resolution: *δj = 0.25*
  - *J* sets the value of the largest scale: *J = 44*

$$s_j = s_0 2^{j\delta j}, \qquad j = 0, 1, \ldots, J$$

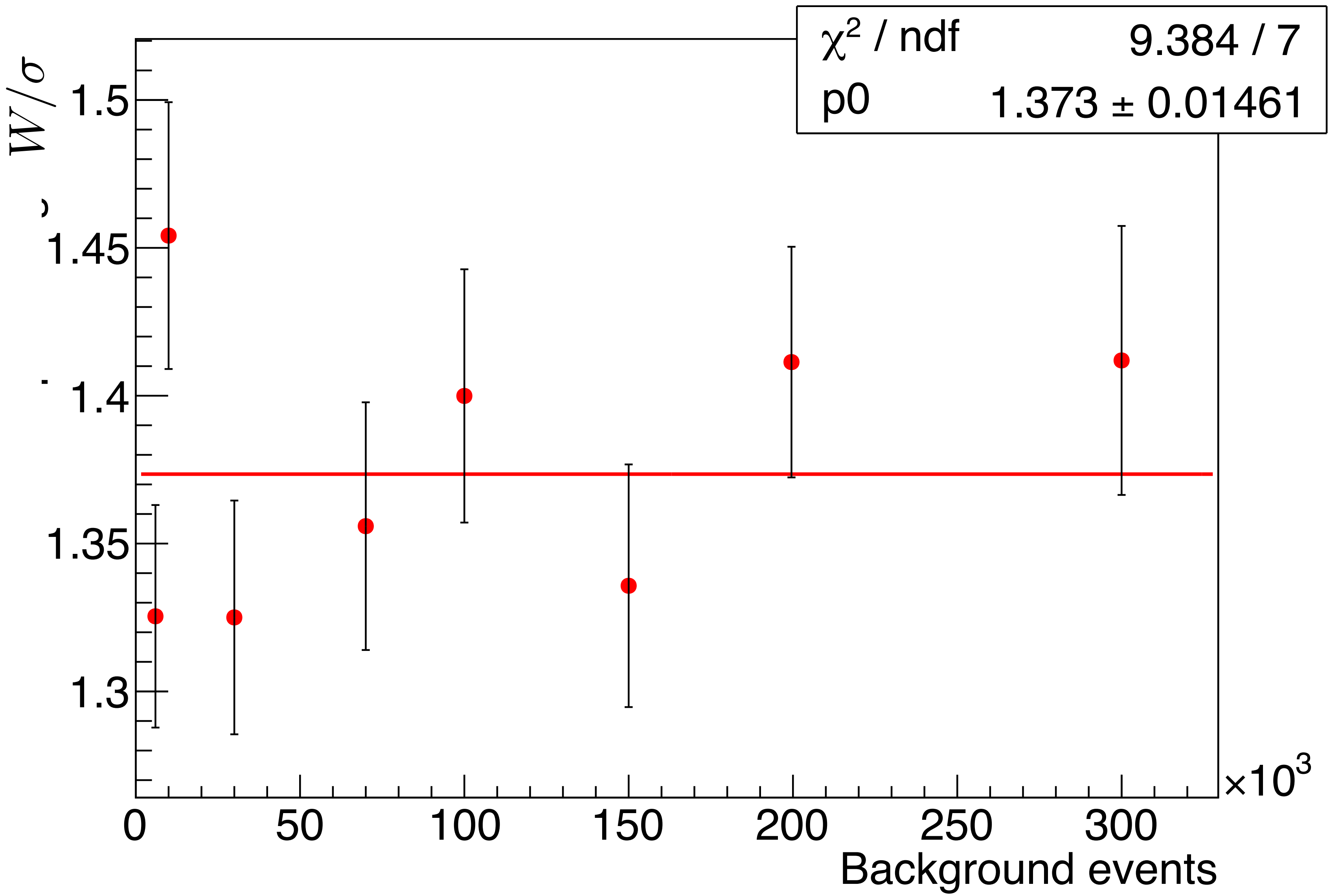
- Normalization: *W(m,s)* at different scales must be directly compared, therefore it is necessary that they all have the same normalization.
  - The normalization is fixed for the Fourier transform of the *mother* wavelet function: it is normalized to have unit energy.
  - The wavelet *daughter* are normalized in the same way adding a normalization constant to their Fourier transform.

$$\int_{-\infty}^{+\infty} |\hat{\psi}_0(s\omega)|^2 d\omega = 1$$

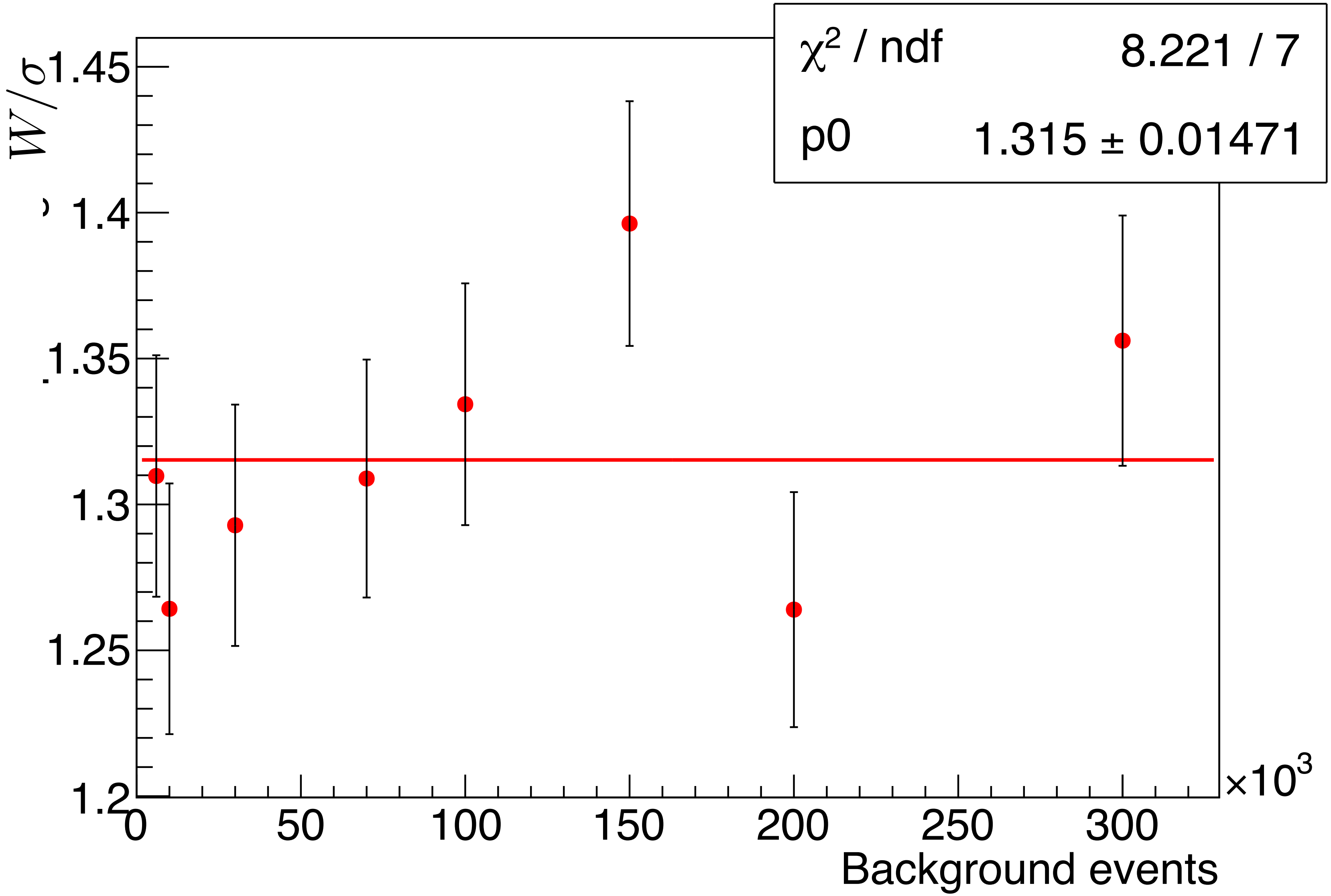
$$\hat{\psi}(s\omega_k) = \left(\frac{2\pi s}{\delta m}\right)^{1/2} \hat{\psi}_0(s\omega_k)$$

- Fourier transform is computed padding with zeroes the end of the mass range: this influence *W(m,s)* in the region close to the edges.
  - The *Cone of Influence* (COI) is the region in *m* × *s* plane where edge effects are important. Discontinuities at the edges decrease exponentially: at each scale, COI is defined by the 'characteristic length' of this decrease.

W/σ mean: flat background

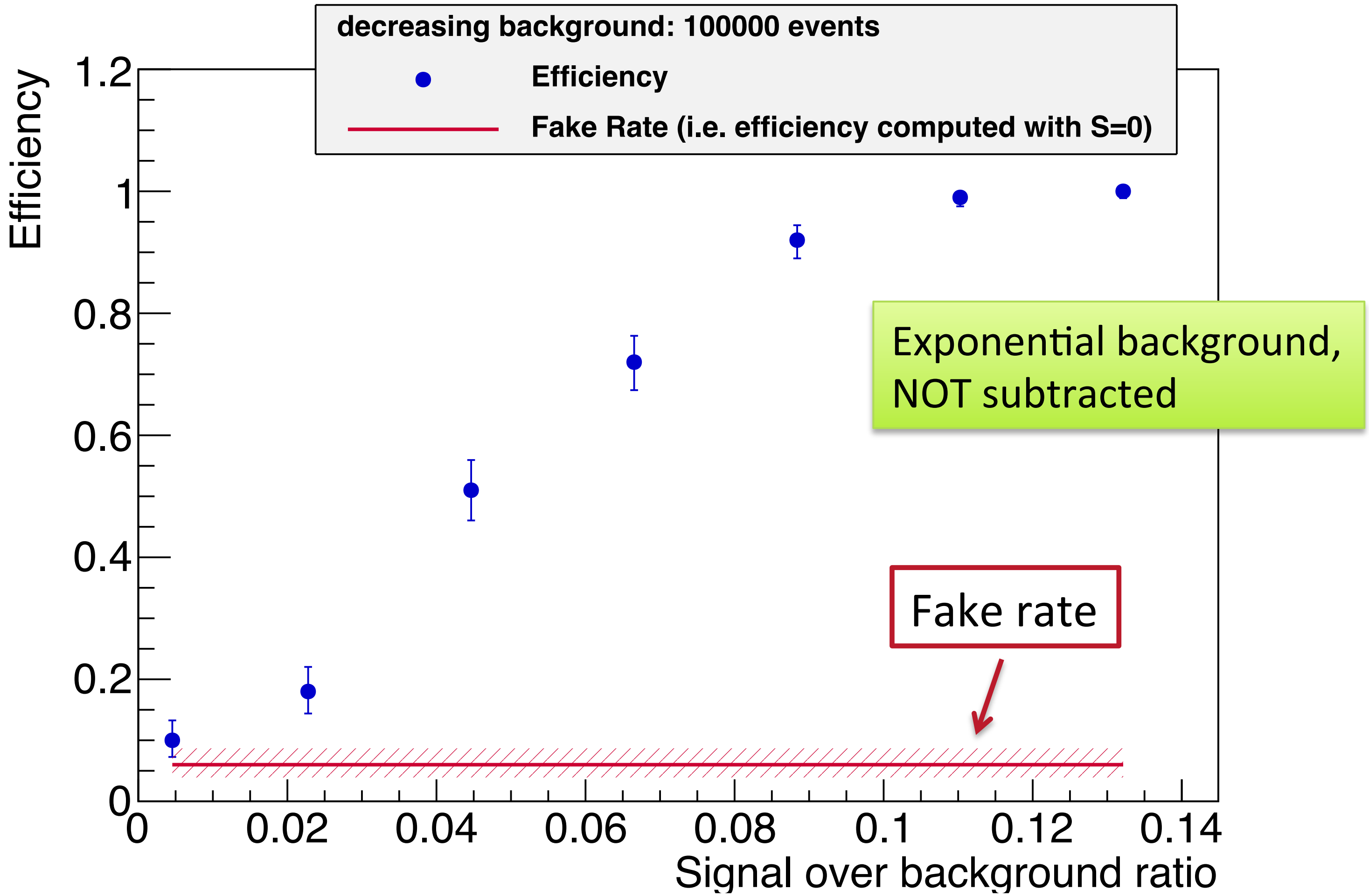


W/σ mean: exponential background

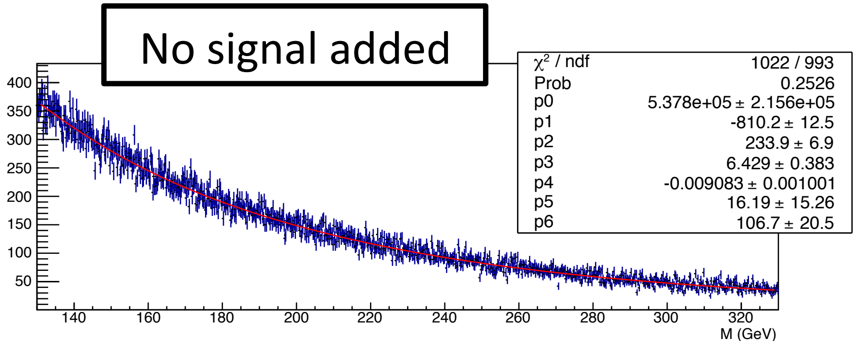


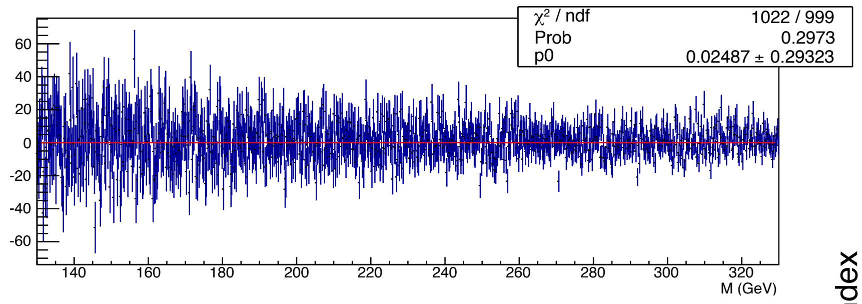
Efficiency without background subtraction

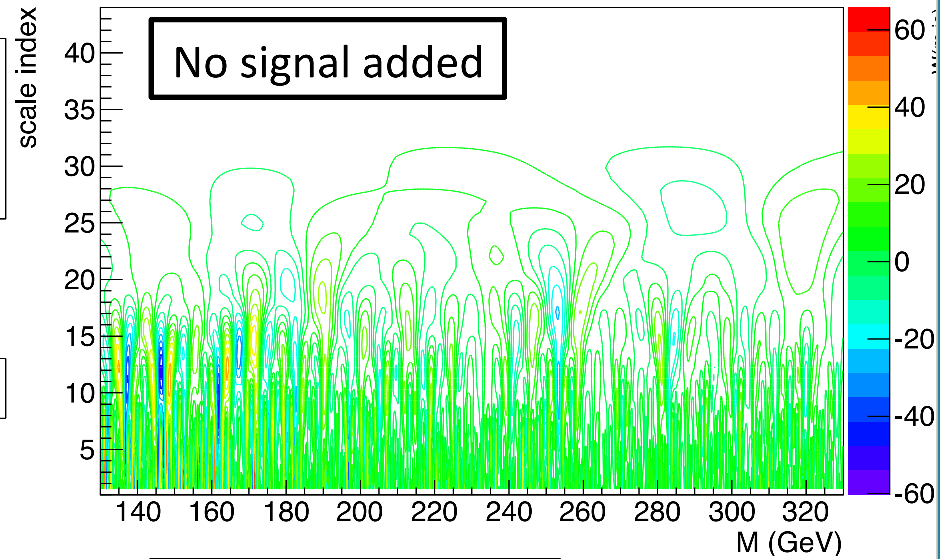
- Efficiency is the fraction of cases in which a *W(m,s)* peak is found in the mass window [μ-σ,μ+σ].

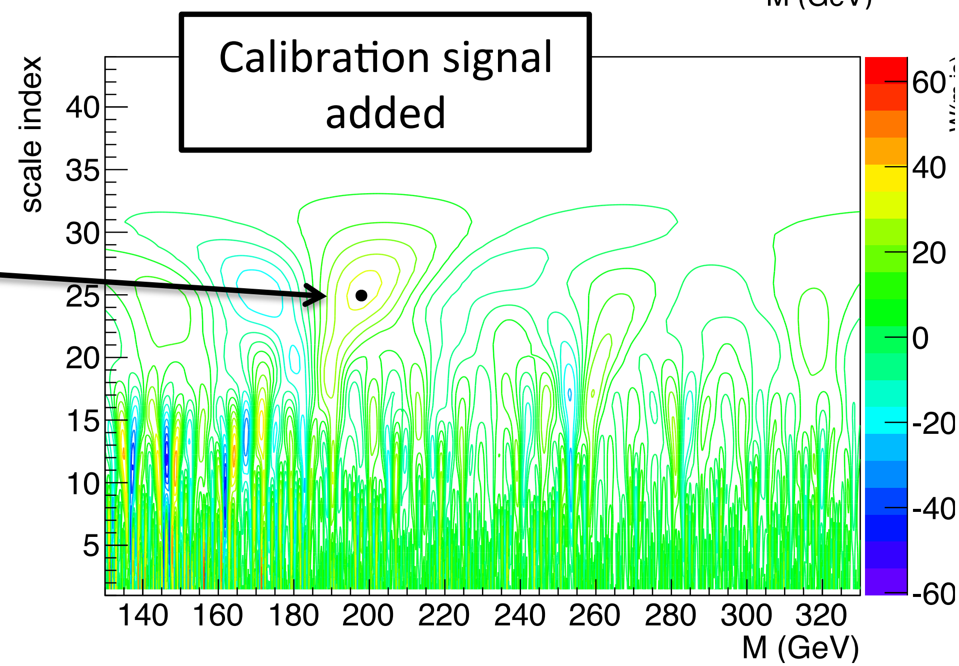


Calibration control region: [130,330] GeV





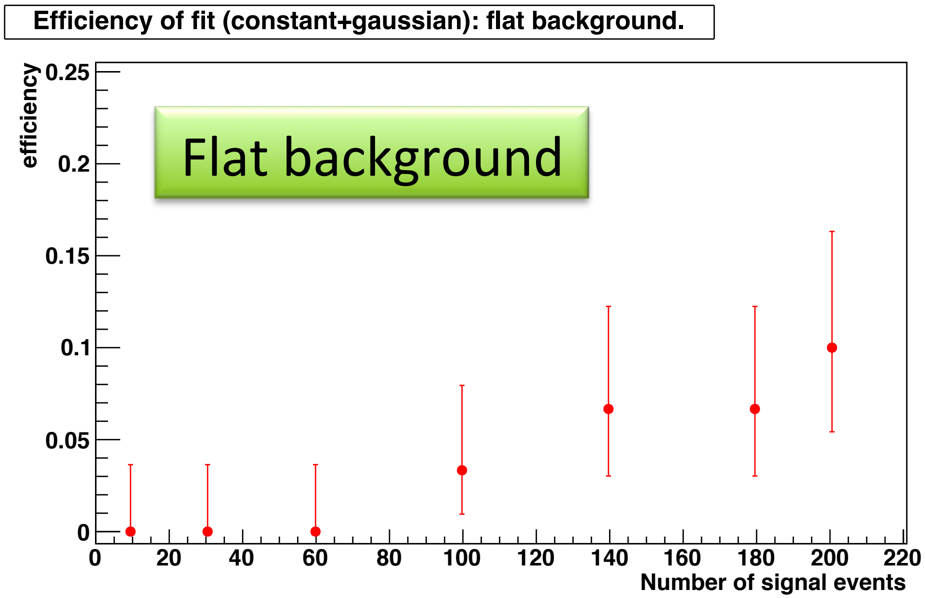




Calibration signal:

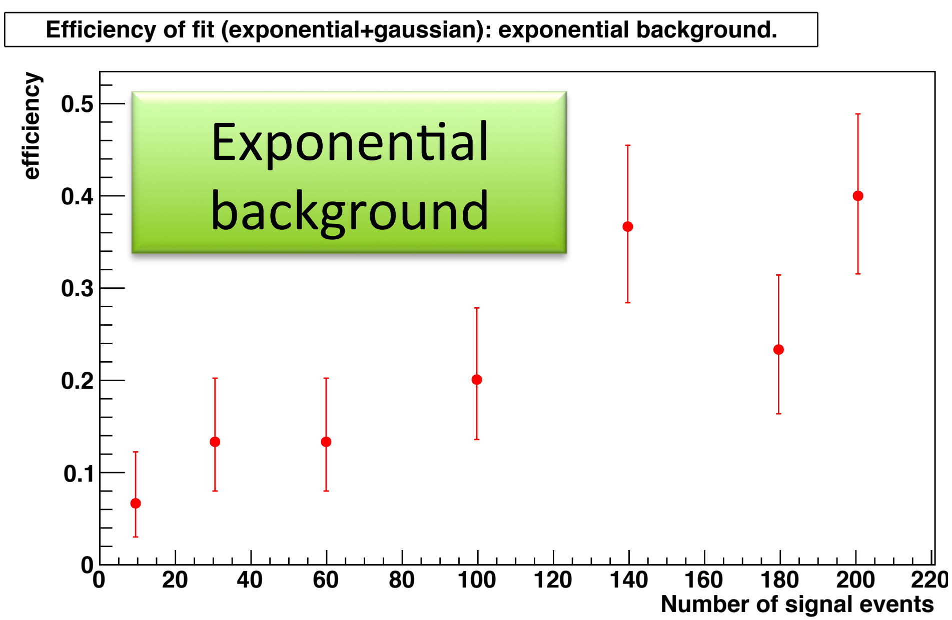
$$\mu = 200 \text{ GeV}$$
$$\sigma = 15\text{GeV}$$
$$\text{signal events} = 900$$

NOTE: only the interesting maximum is marked by a black dot.



Efficiency in identifying a gaussian signal over a flat background, by fitting the data with a gaussian function superimposed to a constant term.

Flat background: 6000 events.
Signal: 100 events, μ=100 GeV, σ=15 GeV.



The signal width have been fixed to 15 GeV in the fit.

Half width of the acceptance interval:

15 GeV = signal width

The "Bump Hunter"

- Standard ATLAS tool to extent the hypothesis test to the case of unknown mass.

  1. The invariant mass spectrum is divided in regions of varying center and width.
  2. For each region, a *p-value* is computed, given the expected background events in each region (obtained with data fitting or MC studies).
  3. The smallest *p-value* is chosen: this is the test statistics *X*.
  4. A *global p-value* is computed using the PDF of *X.*

- The tool will return the most significant interval and the corresponding global *p-value.*

  - Ref:
    - ATLAS internal web resources.
    - http://arxiv.org/abs/1101.0390

SELECTION APPLIED TO DATA: OBJECT SELECTION

Objects passing the selection are defined as *good* objects.

MUON SELECTION.

- *Combined muons* are used.
- Trigger: EF_mu18_MG, EF_mu18_MG_medium. p_T>25 GeV is required to restrict to the trigger efficiency plateau.
- Track quality cuts.
- $|\eta| < 2.4$
- Impact parameter: $|d_0/\sqrt{\sigma(d_0)}| < 3$ and z_0<1 mm.
- Isolation.

  Track: $\Sigma(p_T^{track})/p_T < 0.15$ in a cone of radius R=0.3

  Calorimeter: $\Sigma(E_T^{corr})/p_T < 0.14$ in a cone of radius R=0.3

ELECTRON SELECTION.

- Candidates satisfying the *tight++* identification criteria.
- Trigger: EF_e20_medium, EF_e22_medium, EF_e22vh_medium1. p_T>25 GeV is required to restrict to the trigger efficiency plateau.
- $|\eta| < 2.47$, excluding $1.37 < |\eta| < 1.52$.
- Impact parameter: $|d_0/\sqrt{\sigma(d_0)}| < 10$ and z_0<1 mm.
- Isolation.

  Track: $\Sigma(p_T^{track})/p_T < 0.14$ in a cone of R=0.3

  Calorimeter: $\Sigma(E_T^{corr})/p_T < 0.13$ in a cone of R=0.3

JET SELECTION.

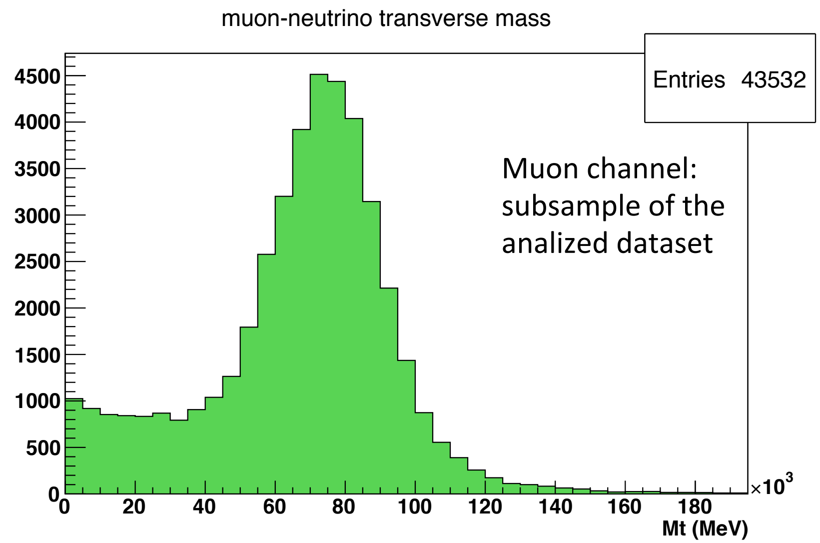
- Jets reconstructed with *Anti-kt* algorithm, passing *looser* quality criteria.
- $p_T > 25$ GeV
- $|\eta| < 2.8$
- Jet Vertex Fraction > 0.75 to reject jets from pile-up interactions.
- $\Delta R(j,l) > 0.5$, *l* is the selected lepton. This to remove overlap between jets and energy deposits due to leptons.

Event selection

Dijets events are triggered by requiring a $W \longrightarrow l\nu$ decay.

- Events are firstly pre-selected applying cuts on event quality:
  - Stable beam conditions, absence of large noise bursts or data integrity errors in the LAr, no jets of p_T>20 GeV pointing to the Lar non-sensitive area (*Lar hole*).
  - A reconstructed primary vertex with at least three associated tracks of p_T>0.5 GeV

- Events with one charged lepton passing the object selection.
  - Events are discarded if a second lepton passes the object selection.
  - *Trigger-matching*: a check to verify that the selected lepton is the one that fired the trigger in the event.



- Events containing also a neutrino: E_t^{miss} >25 GeV
  - Cleaning cuts are applied to the jets before E_T^{miss} cut to avoid non-physical E_T^{miss} due to jet reconstruction errors.

- Cut on the lepton-neutrino transverse mass: M_T > 40 GeV

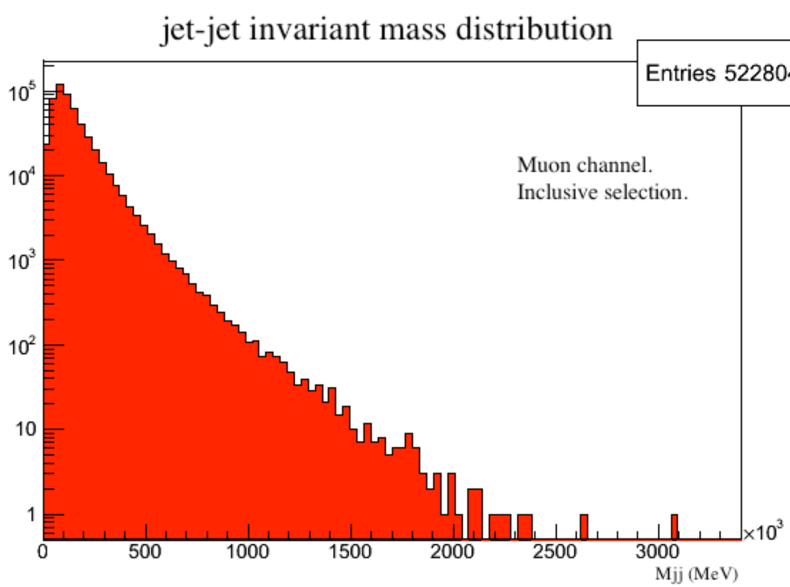
Event selection

<u>Once $W \longrightarrow l\nu$ events are selected, further cuts are applied to jets.</u>

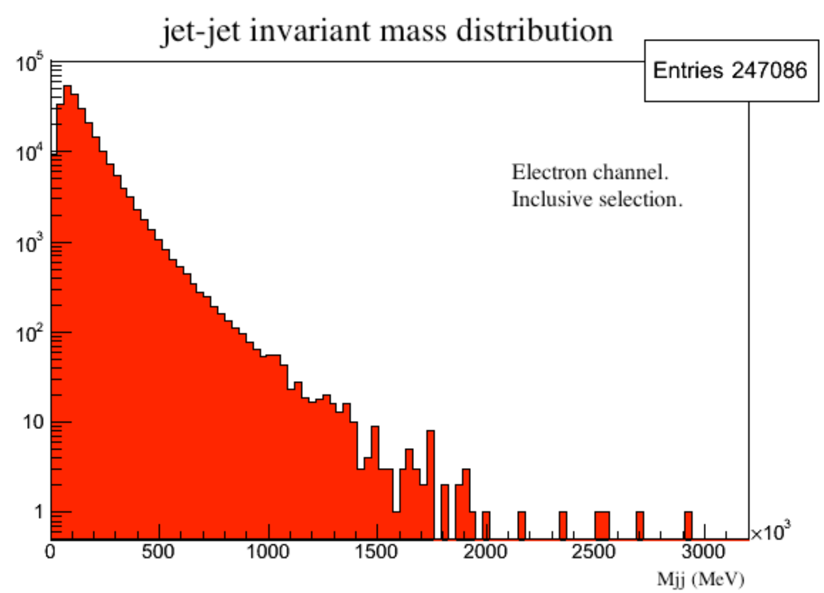
➞ with respect to the selection used in Standard Model diboson measurement, fewer cuts are applied to apply wavelet analysis at a more inclusive level.

- At least two jets passing the object selection
- $\Delta\phi(E_t^{miss}, j_1) > 0.8$. Where j_1 is the jet of highest p_T
- The dijet invariant mass is built using the two selected jets of highest p_T

Jet-Jet invariant mass (logarithmic scale), obtained with $L_{int} = 4702\ \mathrm{pb}^{-1}$.



(a)



(b)