Likelihood analysis of 2008 MEG data with a Bayesian approach

Gianluca Cavoto, Giancarlo Piredda, Cecilia Voena, Francesco Renga
Università di Roma "La Sapienza" and INFN Roma

September 25, 2009

Abstract

We study the 2008 data sample with a maximum likelihood analysis, estimating the confidence level intervals on the number of signal events and on the branching ratio of the $\mu^+ \to e^+\gamma$ decay using a Bayesian approach. We find BR($\mu^+ \to e^+\gamma$)$< 2.8 \times 10^{-11}$ @90 % C.L. where systematic effects have not been included. We cross-check our result computing the upper limit using a frequentistic technique with a Feldman-Cousins ordering and we find consistent results.

Contents

1 Introduction

We describe here the likelihood analysis of the 2008 data sample. The discriminating variables used in the analysis are the photon and positron energy (E_γ, E_{e^+}) , the relative photon-positron time and angle ($T_{e^+\gamma}$, $\theta_{e^+\gamma}$). The likelihood function for signal and background is built from a combination of studies on data and on full simulation. The result on the $\mu^+ \to e^+\gamma$ branching ratio is obtained with a Bayesian analysis. A frequentistic interpretation with a Feldman Cousins ordering is also used as a cross-check. Section 2 describes the selection and the data sample used and section 3 describes the likelihood function. Section 5 quotes the results of a likelihood fit to the data (the likelihood is minimized through the RooFit interface). Section 5 and 6 report the upper limit on the branching fraction from the Bayesian and frequentistic analysis, respectively.

2 Data sample and selection

We use the 2008 data with the *Standard* selection reported in [1] , with the exception of the cut on the track T0. We apply:

- track quality cuts:
  - number of hits in the DCH >=7
  - χ_2 of the fitted track, normalized to the number of degrees of freedom, <=12
  - uncertainty on the positron energy <= 0.7 MeV
  - uncertainty on the polar angle <=0.6 deg
  - uncertainty on the azimuthal angle <=1.5 deg
  - |dR (=difference of the track projection at the TC and TC measurement in the radial coordinate)+2.2cm| <3 cm
  - |dZ (=difference of the track projection at the TC and TC measurement in the longitudinal coordinate)+0.1cm| <6 cm
  - number of chambers >3
  - chamber span>4
  - number of multi-hit chambers > 1
  - $TCIter$ variable >0.
- requirements on beam and target spots: Target ellipse y axis < 2.8 cm, target ellipse z axis <8.24 cm, beam ellipse y axis <2.8 cm, beam ellipse z axis <7.5 cm, beam y displacement <-0.26 cm, beam z displacement <-1.5 cm. The best rank ghost is chosen after all the above cuts are applied.
- XEC cuts: |u|<71 cm, |v|<25 cm, pile-up requirements (npeakid>1 or xectimefit.timechisq0<3), cosmic rejection (ratio of the charge of inner over outer PMTs >0.3).

- additional pile-up cut, described in [2].

We define as analysis region, where we perform the likelihood fit, with the following cuts in the four discriminating variables: $46 < E_\gamma < 60$ MeV, $50 < E_{e^+} < 56$ MeV ,$-1 < T_{e^+\gamma} < 1$ ns, $|\theta_{e^+\gamma}| > 174.68$ deg. We also consider the left and right sideband regions in the $T_{e^+\gamma}$ variable with $-3.5 < T_{e^+\gamma} < -1.5$ ns and $1.5 < T_{e^+\gamma} < 3.5$ ns, respectively. We use the algorithm $XECEnergyPileUpEliminated$ for the photon energy measurement and the $TimeFit$ algorithm for the photon time measurement. To cross-check the result, we use also alternative algorithms for both the photon energy and time, $TRGEnergy$ and $TimeWeightedAverage$.

3 Likelihood analysis

We implement the likelihood analysis in a RooFit-based stand-alone package. *Minuit* minimization algorithms are used. The extended likelihood function $\mathcal{L}$ is written as:

$$\mathcal{L}(\mathrm{N_{sig}}, \mathrm{N_{RD}}, \mathrm{N_{BG}}) = \frac{\mathrm{N^{N_{obs}}exp^{-N}}}{\mathrm{N_{obs}!}} \prod_{\mathrm{i=1}}^{\mathrm{N_{obs}}} [\frac{\mathrm{N_{sig}}}{\mathrm{N}}\mathrm{S} + \frac{\mathrm{N_{RD}}}{\mathrm{N}}\mathrm{RD} + \frac{\mathrm{N_B}}{\mathrm{N}}\mathrm{B}], \tag{1}$$

where $\mathrm{N_S}$ is the number of signal events, $\mathrm{N_{RD}}$ is the number of radiative decay events, $\mathrm{N_B}$ is the number of accidental background events, S, RD, B are the probability density functions (pdfs) for the three components, respectively, $\mathrm{N{=}N_S{+}N_B{+}N_{RD}}$ and $\mathrm{N_{obs}}$ is the number of observed events in the analysis region. For each component the pdf is written as the product of the pdfs of the individual analysis variables, assuming that the correlations are negligible. The pdfs are described below.

1. Signal

   - E_γ pdf: response function as extracted from the CEX runs reported in [3] for $XECEnergyPilupEliminated$. The $TRGEnergy$ pdf is extracted from a combination of data E_γ sideband and of MC simulation by fitting the endpoint of the background photon energy distribution.
   - E_{e^+} pdf: sum of three Gaussian functions as measured from Michel events[4].
   - $T_{e^+\gamma}$ pdf: single Gaussian function with σ=147.7 ps as measured from the radiative decay peak outside the E_γ signal region.
   - $\theta_{e^+\gamma}$ pdf: it is built with a toy Monte Carlo (MC) technique using the polar and azimuthal angle resolution from a study of full simulated MC signal events taking into account the various correlations. The polar angle resolution for the positron has been enlarged by summing (in quadrature) 15 mrad to take into account differences from data [5].

2. Radiative decay

- joint E_γ, E_{e^+}, $\theta_{e^+\gamma}$ pdf: it is obtained from the Kuno-Okada theoretical spectrum[6] weighted with the acceptance and smeared with the experimental resolutions. The acceptance for the positron is taken from the analysis of the Michel spectrum reported in[4] while the trigger acceptance for the photon is parametrized with the same function but with different parameters (the function is $\frac{(1+erf(x))}{2}$ where $erf(x)$ is an error function; the parameters used for the photon are μ =39.2 MeV and σ = 2.3 MeV). The same resolutions as for the signal are used.
- $T_{e^+\gamma}$ pdf: same as the signal.

3. Accidental background

   - $\theta_{e^+\gamma}$ and $T_{e^+\gamma}$ pdfs: they are a 2^{nd}-order polynomial and a constant function respectively, fitted on data sidebands.
   - E_{e^+} pdf: it is the theoretical Michel spectrum multiplied by the acceptance and convoluted with the resolution, fitted in the sideband region.
   - E_γ pdf: it is an empirical function fitted in the sidebands. The function is a polynomial before a fitted threshold and an exponential above (which describes the residual pile-up).

Figure 1 shows the pdf for signal and accidental background events from a toy MC simulation. In order to verify the fitting procedure we perform toy MC studies. We generate sets of experiments with a given number of signal and background events using our likelihood function and we fit them. Of particular interest is the pull distribution of a variable computed as the difference of the fitted and the generated value divided for the error that the fit returns. If no bias on the fitted variable is introduced and if the errors are correctly estimated the pull distribution must be a Gaussian, with $\mu = 0$ and $\sigma = 1$. We consider two cases:

- $\mathrm{N_{sig}}$(gen)=32, $\mathrm{N_{RD}}$(gen)=32, $\mathrm{N_B}$(gen)=970. Figure 2 shows the pull distribution for the number of signal and of radiative decay events. Even if we do not expect such a large number of signals events this test is useful to debug the fitting code, in particular to check the likelihood normalization.

- $\mathrm{N_{sig}}$(gen)=0, $\mathrm{N_{RD}}$(gen)=32, $\mathrm{N_B}$(gen)=970. Figure 3 shows the pull distribution for the number of signal and of radiative decay events.

As it can be seen, the pull distribution for the number of signal and radiative decay events has no significant bias and correct sigma in the case of sizeable signal, while in the case of null signal the pull distribution for the number of signal events become non-Gaussian, as expected. This means that the fit result and error for $\mathrm{N_{sig}}$ do not have a correct statistical meaning in our case, since a small number of signal events is expected. We then use the full likelihood function as explained in the next sections. It should be noted that we do not apply any constraint to the fitted parameters and thus the minimum of the likelihood function can be in correspondence of a negative number of signal events. This is not a problem in the Bayesian interpretation of the likelihood where the a-priori pdf is zero outside the physical region.

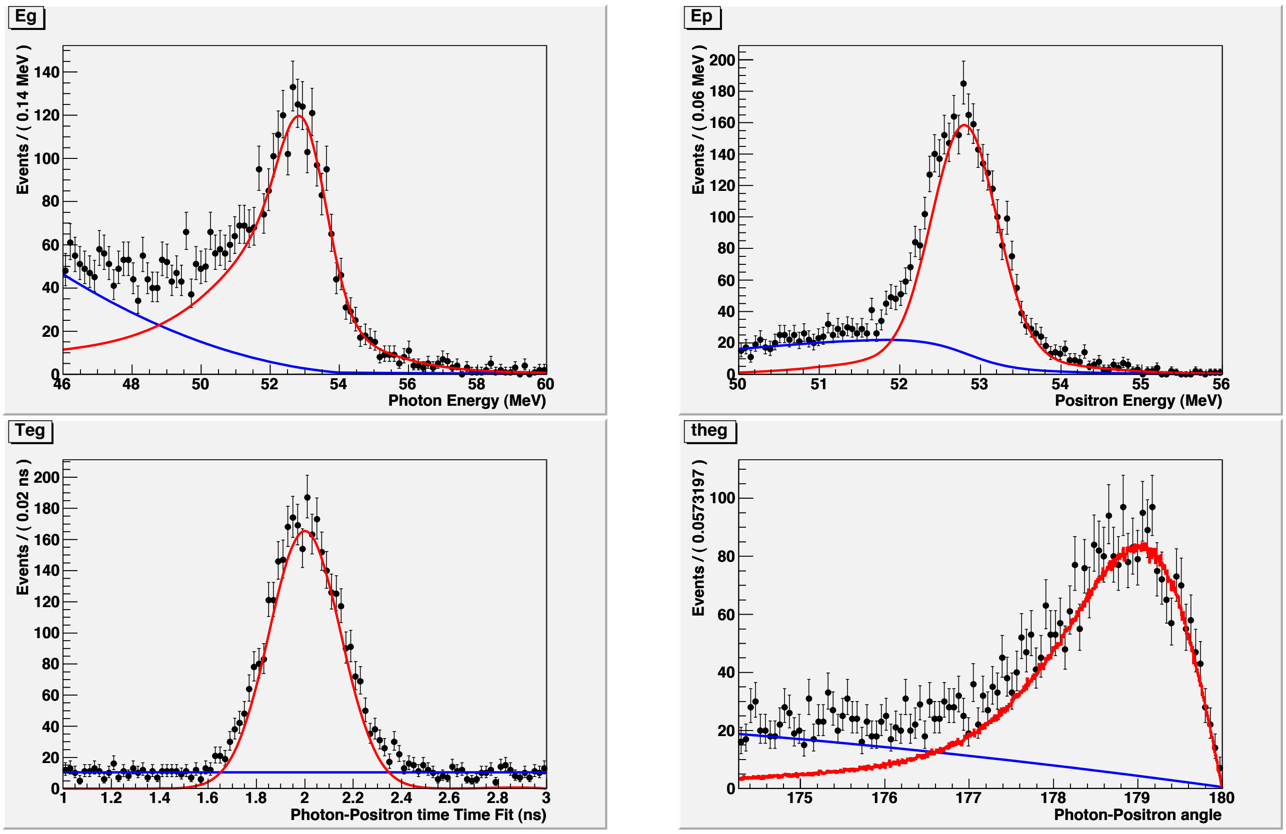

Figure 1: *pdfs for signal and accidental background events for the various discriminating variables. The red curve shows the pdfs for signal, the blue curve shows the pdf for accidental background. The dots are the sum of the two (from a toy MC generation with an arbitrary normalization).*

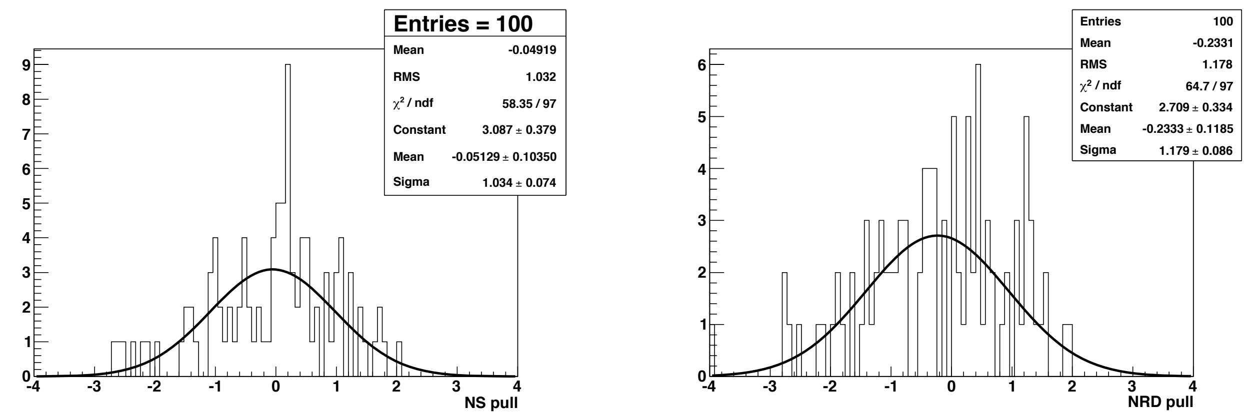

Figure 2: *Left plot: pull distribution for* $N_{\rm sig}$. *Right plot: pull distribution for* $N_{\rm RD}$. *In this toy MC study* $N_{\rm sig}$*(gen)=32,* $N_{\rm RD}$*(gen)=32,* $N_{\rm B}$*(gen)=970.*

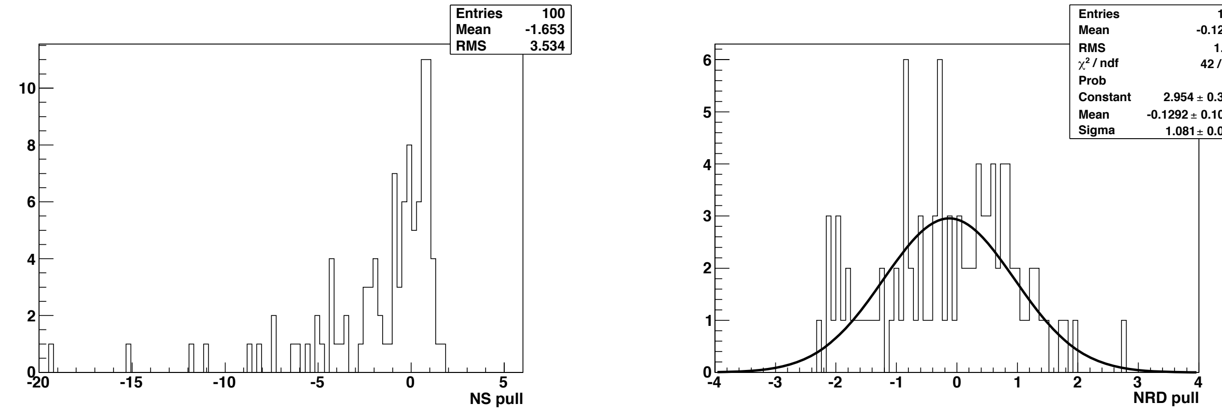

Figure 3: *Left plot: pull distribution for* $N_{\rm sig}$*. Right plot: pull distribution for* $N_{\rm RD}$*. In this toy MC study* $N_{\rm sig}$*(gen)=0,* $N_{\rm RD}$*(gen)=32,* $N_{\rm B}$*(gen)=970.*

4 Fit results on data

With the selection described in section 2 we select 1007 events in the analysis region, 1004 events in the left sideband and 1059 events in the right sideband region. The result of the maximum likelihood unbinned fit in the analysis region is:

$$N_{\rm sig} = 5 \pm 4, N_{\rm RD} = 32 \pm 16, N_{\rm B} = 969 \pm 34, \tag{2}$$

where the HESSE errors (i.e. calculated by inverting the full-second derivative matrix) are quoted. Figure 4 shows the data in the analysis region with the pdf projection superimposed. As a cross-check, we performed the fit to the $T_{e^+\gamma}$ variable only and we find a consistent result:

$$N_{\rm sig} + N_{\rm RD} = 40 \pm 30, N_{\rm B} = 940 \pm 40. \tag{3}$$

4.1 Fit results in sideband regions

On the left $T_{e^+\gamma}$ sideband we find:

$$N_{\rm sig} = -33 \pm 2, N_{\rm RD} = -11 \pm 2, N_{\rm B} = 1048 \pm 33, \tag{4}$$

while on the right data sideband we find:

$$N_{\rm sig} = -9 \pm 1, N_{\rm RD} = -14 \pm 2, N_{\rm B} = 1082 \pm 33. \tag{5}$$

4.2 Fit results using TRGEnergy and TimeWeightedAverage

As explained in section 2 we use also alternative algorithms for the XEC energy ans time. We find a consistent result:

$$N_{\rm sig} = 7 \pm 5, N_{\rm RD} = 27 \pm 10, N_{\rm B} = 1070 \pm 40, \tag{6}$$

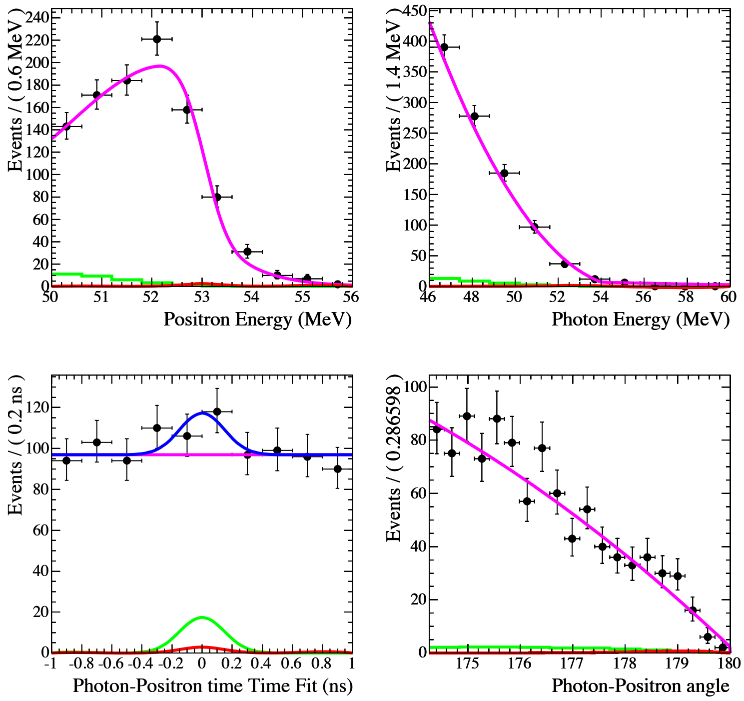

Figure 4: *Fit result in the analysis region. The dots are the data in the various analysis variables. Pdf projections are superimposed. The red curve is the signal pdf, the green curve is the radiative decay pdf and the violet curve is the accidental background component.*

where the higher number of selected events is probably due to the fact that the TRGEnergy algorithm has an higher pile-up contamination with respect to XECEnergyPilepEliminated (which uses the DRS digitizer).

5 Bayesian analysis

We use a Bayesian interpretation of the likelihood to derive the a-posteriori probability density function of the unknown $\mu^+ \rightarrow e^+\gamma$ branching ratio (BR). The Bayes theorem allows to update the a-priori probability of the BR, P(BR), producing an a-posteriori pdf:

$$P(BR|data) = P(data|BR) \times P(BR) = \mathcal{L}(\mathrm{data}|\mathrm{N_{sig}} = \mathrm{k} \times \mathrm{BR}) \times \mathrm{P(k)} \times \mathrm{P(BR)}, \qquad (7)$$

where $\mathcal{L}$ is the likelihood function, k is a normalization factor and P(k) its pdf. The symbol $''|''$ indicates the conditional probability. The procedure is the following:

1. we extract a value for the branching ratio BR_{test} between 0 and a maximum value (chosen to be 0.5×10^{-10}) according to a uniform distribution (this is the a-priori pdf). It should be noted that the choice of the maximum value of the BR has no impact on the result, since it has been chosen so that the likelihood value is practically zero beyond. The choice of a flat prior is somehow arbitrary.

2. we extract a value of the normalization factor according to a Gaussian distribution with μ=4.7×10^{11} and a σ of 10%.

3. we compute the number of signal events $N_{\mathrm{sig,test}}$ as the product of the two numbers above.

4. we generate a number of radiative decay events and a number of accidental background events according to a uniform distribution in a range of [0,+5σ] and $\pm5\sigma$, respectively, of the value fitted on data (see eq. 4) using the HESSE error as σ.

5. we compute the weight to be assigned to the corresponding branching ratio as:

$$w = exp{-}(-\mathrm{log}\mathcal{L} + \mathrm{minlog}\mathcal{L}), \qquad (8)$$

   where the experimental likelihood described in section 3 is computed using the values of $N_{\mathrm{sig,test}}$, N_{RD}, N_{acc} obtained as described in point 3 and 4.

6. the procedure above is repeated 100 times and the a-posteriori distribution of the branching ratio is obtained by weighting each BR_{test} for the corresponding w.

In order to obtain the a-posteriori bi-dimensional distribution of N_{sig}, N_{RD} a similar procedure is applied but in this case $N_{\mathrm{sig,test}}$ is generated with a uniform distribution between 0 and 30. Figure 5 shows the bi-dimensional a-posteriori 68% and 90% regions for N_{sig} and N_{RD}, while figure 6 shows the a-posteriori distributions for N_{sig} (obtained by projecting the

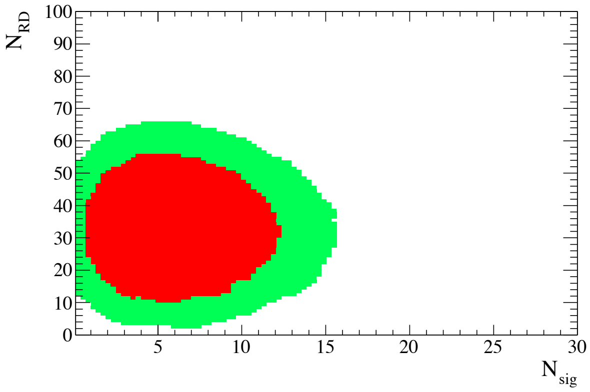

Figure 5: *A-posteriori distribution of $N_{\rm sig}$, $N_{\rm RD}$ obtained with the Bayesian technique described in the text. The red area is the region at 68%C.L. level, while the green area is at 90% C.L.*

bi-dimensional pdf) and for BR($\mu^+ \to e^+\gamma$). We can set the following upper limits @90% C.L. [1]:

$$N_{\rm sig} < 12.75 \text{ @90\%}C.L., \; BR(\mu^+ \to e^+\gamma) < 2.8 \times 10^{-11} \text{ @90\%}C.L. \qquad (9)$$

5.1 Results in sideband regions

With the same procedure, we obtain

$$N_{\rm sig} < 2.8 \text{ @90\%}C.L., \; BR(\mu^+ \to e^+\gamma) < 5.5 \times 10^{-12} \text{ @90\%}C.L \qquad (10)$$

on the left sideband and

$$N_{\rm sig} < 2.8 \text{ @90\%}C.L., \; BR(\mu^+ \to e^+\gamma) < 6 \times 10^{-12} \text{ @90\%}C.L \qquad (11)$$

on the right sideband.

5.2 Expected upper limit from a toy MC study

In order to estimate the expected upper limit on our data if no signal is present, we perform a toy MC study with 100 experiments with $N_{\rm sig,gen}$=0, $N_{\rm RD,gen}$=32, $N_{\rm B,gen}$=970 (as from the data fit results). We find the distribution of the upper limits in figure 7. As it can be seen, the upper limit we find on data is on the tail of the distribution meaning that we have been unlucky or that the experiment starts to see some signal. On the basis of this toy MC study the probability of having a limit worse than what we got is $\sim 2\%$.

[1]In the case of a Bayesian analysis, the term “probability level” would be more correct than “confidence level”.

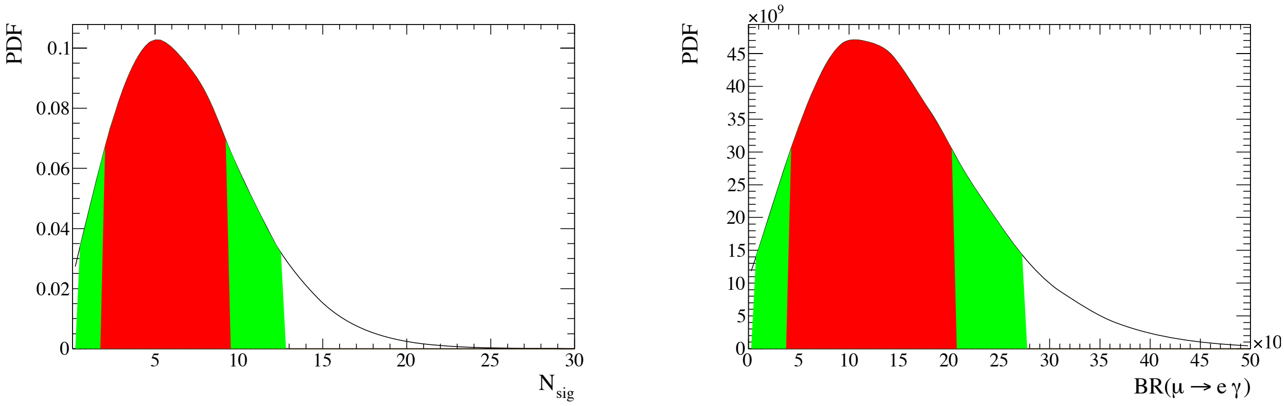

Figure 6: *A-posteriori distribution of* $N_{\rm sig}$ *(left) and BR(*$\mu^+ \to e^+\gamma$*) (right) obtained with the Bayesian technique described in the text. The red area is the region at 68%C.L. level, while the green area is at 90% C.L.*

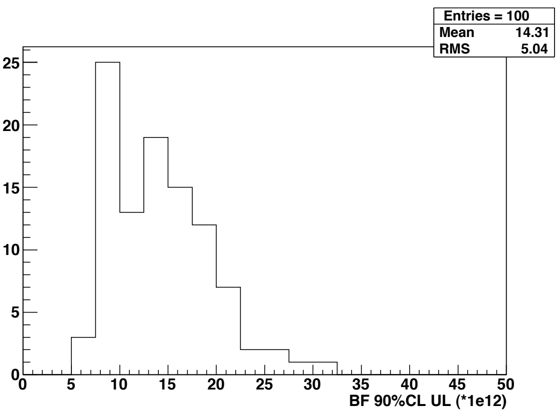

Figure 7: *Distribution of Upper Limits obtained in a toy MC simulation with* $N_{\rm sig,gen}$*=0,* $N_{\rm RD,gen}$*=32,* $N_{\rm B,gen}$*=970.*

6 Cross-check with Feldman Cousins approach

In order to cross-check our result we re-compute the upper limit using a frequentistic method with a Feldman Cousins ordering [7]:

- we perform a loop on the possible number of signal events ($N_{\rm sig,test}$) from 0 to 20 in steps of 1 event.
- for each $N_{\rm sig,test}$, We perform a loop over the possible number of radiative decay events ($N_{\rm RD,test}$) from 0 to 75 in steps of 5 events.
- for each pair ($N_{\rm sig,test}$,$N_{\rm RD,test}$) we compute the ratio R(data):

$$R(data) = \frac{\mathcal{L}(\rm N_{sig,test}, N_{RD,test}, N_{B,fit})}{\mathcal{L}(\rm max)}, \tag{12}$$

  where $\mathcal{L}$(max) is the maximum value of the likelihood from the fit to the data. Note that here and in what follows a lower limit at 0 to the number of fitted signal events is applied in performing the fit (as opposed to what we do in the Bayesian method). $N_{\rm B,fit}$ is the value from the fit to the data.
- also, for each pair ($N_{\rm sig,test}$,$N_{\rm RD,test}$) we generate several toy MC experiments where the number of signal, radiative decay and accidental background events are extracted according to a Poisson distribution with μ= $N_{\rm sig,test}$, $N_{\rm RD,test}$, $N_{\rm B,fit}$ respectively.
- for each toy experiment the ratio R(MC) is computed as:

$$R(MC) = \frac{\mathcal{L}_{\mathcal{MC}}(\rm N_{sig,test}, N_{RD,test}, N_{B,fit})}{\mathcal{L}_{\mathcal{MC}}(\rm max)}. \tag{13}$$

The confidence level contour, for example at 90% C.L. is built calculating the fraction of experiments, for a given pair $N_{\rm sig,test}$,$N_{\rm RD,test}$, for which R(MC) is below R(data): if the fraction of experiment for which R(MC)<R(DATA) is at least 10% of the cases, then the sample point belongs to the contour. Figure 8 shows the 68% and 90% confidence level contours in $N_{\rm sig}$, $N_{\rm RD}$. Figure 9 shows the result of the procedure above where no scan over the number of radiative decay events is performed; the number of $N_{\rm RD,test}$ is fixed to 35 since it can be seen from figure 8 that the most conservative limit on the number of signal events is obtained in this case. There is not a standard way to project a bi-dimensional distribution into a one-dimensional one in the frequentistic approach: the way we chose is somehow arbitrary and the 90% coverage is not guaranteed. The black curve represents the data, the red curve represents the average from a set of toy MC experiments and the yellow region shows the 68% region for these experiments. The dashed curve represents the limit of the 90% level contour from the toy MC, i.e. the R-value for which the fraction of experiments with R greater than this value is 90%. We can set the limit:

$$N_{\rm sig} < 17 \ @90\% C.L., \tag{14}$$

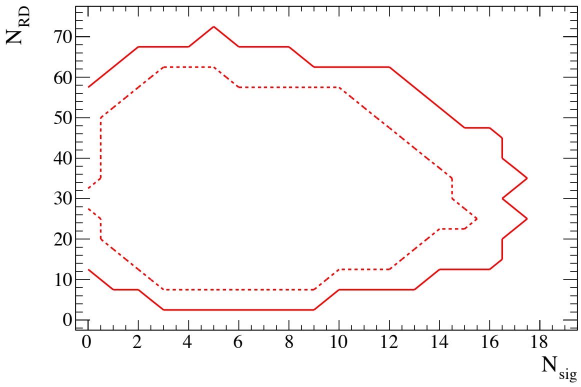

Figure 8: *68% and 90% confidence level contours in N_{sig}, N_{RD} obtained with the Feldman Cousins apprach described in the text.*

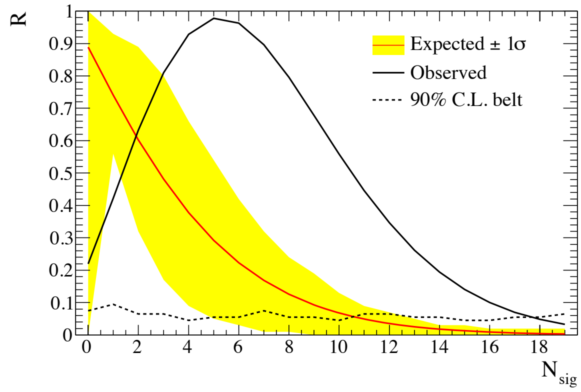

Figure 9: *R ratio ($=\frac{\mathcal{L}(\mathrm{N_{sig}})}{\mathcal{L}(\mathrm{max})}$) as a function of the number of signal events for N_{RD} = 35 and N_B=970. The black curve represents the data, the red curve represent the average from a set of toy MC experiments and the yellow region shows the 68% region for these experiments. The dashed curve represents the limit of the 90% level contour from the toy MC.*

7 Conclusions

We have analyzed the MEG 2008 data with an unbinned maximum likelihood technique and we interpreted the result both in a Bayesian and in a frequentistic (a-la Feldman Cousins) framework. We find:

$$N_{\mathrm{sig}} < 12.75 \text{ @}90\% C.L., \; BR(\mu^+ \to e^+\gamma) < 2.8 \times 10^{-11} \text{ @}90\% C.L, \tag{15}$$

with a Bayesian approach and:

$$N_{\mathrm{sig}} < 17 \text{ @}90\% C.L., \tag{16}$$

with a Feldman Cousins approach.

References

[1] https://midas.psi.ch/elogs/Physics+Analysis/205.

[2] https://midas.psi.ch/elogs/Physics+Analysis/184.

[3] https://midas.psi.ch/elogs/Physics+Analysis/160.

[4] Technical Note TN054.

[5] http://meg.web.psi.ch/restricted/talks/pawg/Meeting090729/Betta_week_meet_29_07_09.pdf.

[6] Y. Kuno and Y. Okada, Rev. Mod. Phys. **73** (2001) 151.

[7] G.J. Feldman and R.D. Cousins, Phys. Rev D **57** (1998) 3873.