Dear Referee

Thank you for your detailed comments. Please find below our answers.

> However, the analysis procedure has one critical point that has

> influence on all presented results. On the one hand the efficiencies

> for the track finding and muon identification are determined from data

> to be independent from the MC simulation, on the other hand a large

> correction of up to 30% is applied for correlations between the

> efficiencies of the two muons which is determined from the MC, without

> any possibility to cross check it in data. You claim that this is due

> to the bin size in which the efficiency is determined. Since no

> correlations between the efficiencies of the two muons are expected

> if they are determined as a function of the relevant variables, this

> means that the bin sizes for the muon efficiencies are too large and

> that you have a strong variation within some of the bins. Looking at

> reference [23] I see that the muon efficiencies are determined

> with a limited statistics sample in a binning that has only 1 to 3 bins

> in the threshold region (if these are not the relevant plots for this

> analysis, you should provide the relevant ones). On the other hand you

> have a very fine binning for the pt dependence of the J/psi cross

> section.

> I see several possibilities how to avoid this problem:

> - you could cut out the threshold region and start only at higher

> muon pt. Then you probably need to adjust your analysis region

> accordingly

> - you could use a smooth function, determined from a finer binning in

> the threshold region, for the correction

> - you could adjust the MC to describe the data, using a smooth

> function

> In any case the large correlation between the two muons is unphysical

> and should disappear.

We spent quite a lot of time on understanding the effect of the correlation and we believe it is very well under control.

First, to avoid any misunderstanding, the data sample that we have used in the data to compute the efficiencies with the Tag and Probe (T&P) method is larger with respect to those quoted in Reference 23. In fact the muon tagging efficiency has been measured using a sample of 3 pb-1 data, which includes the period of the 300 nb-1 used for the analysis, in which the detector behaved in a very stable way. We have verified that the reconstruction efficiencies did not vary over time. This is shown in Fig. 1, that show the data to data comparison for the two periods (in a smaller binning). Fig. 2 shows the data to MC comparison of the muon tagging efficiency for the 3 pb-1 period. As expected, there are differences in the values of the efficiency between the data and the simulation, but the trend as a function of pT and eta is very well reproduced. As for the trigger we have used the 300 nb-1 data only, since the trigger menu changed after that. Figure 3 shows the data to MC comparison for the trigger efficiency and - despite the larger error wrt the tagging efficiency - the trends are also very well reproduced.

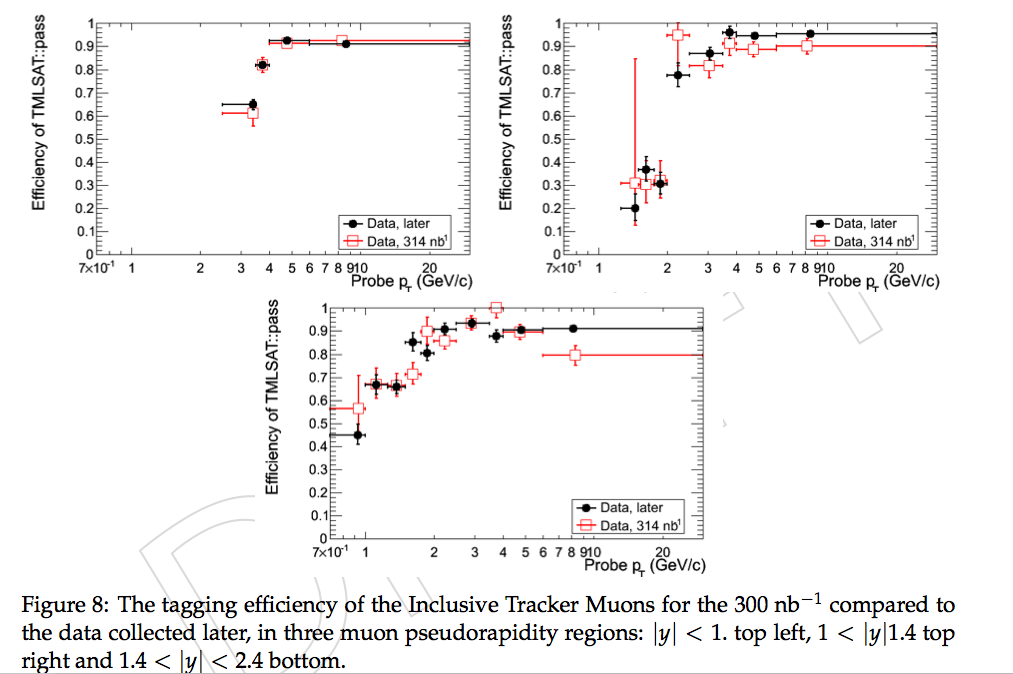


Figure . Tagging efficiency for the 300 nb-1 compared to the one collected later, in three muon pseudo-rapidity regions: |y|<1 (top left), 1<|y|<1.4 (top right) and 1.4<|y|<2.4 (bottom)

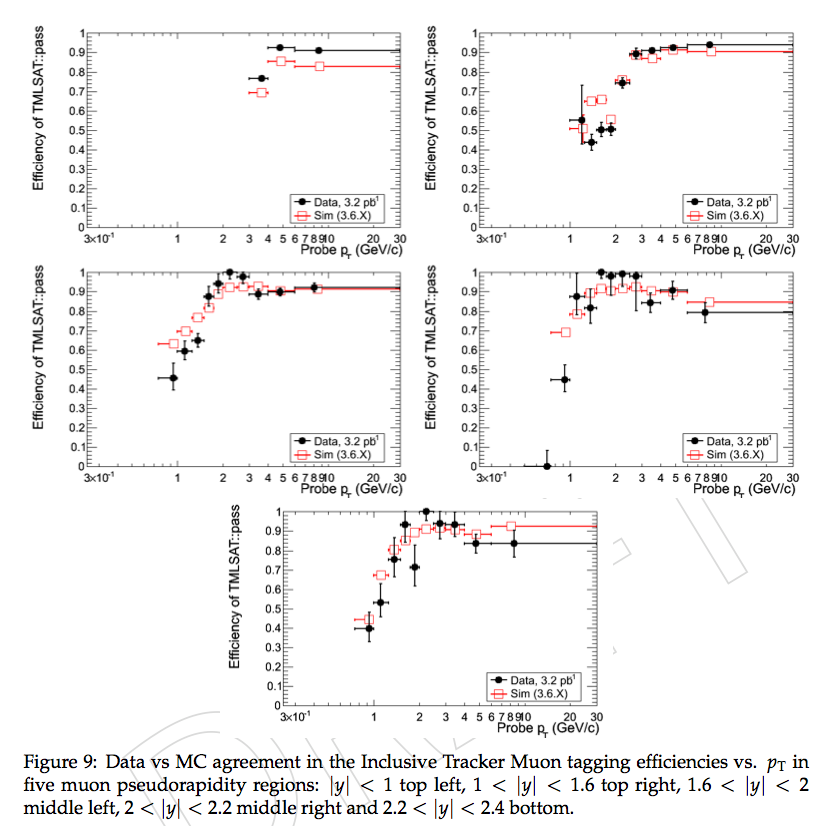


Figure 2. Data versus MC agreement in the muon tagging efficiencies vs. pT in five muon pseudorapidity regions: |y|<1 (top left), 1<|y|<1.6 (top right), 1.6<|y|<2 (middle left), 2<|y|<2.2 middle right and 2.2<|y|<2.4 (bottom).

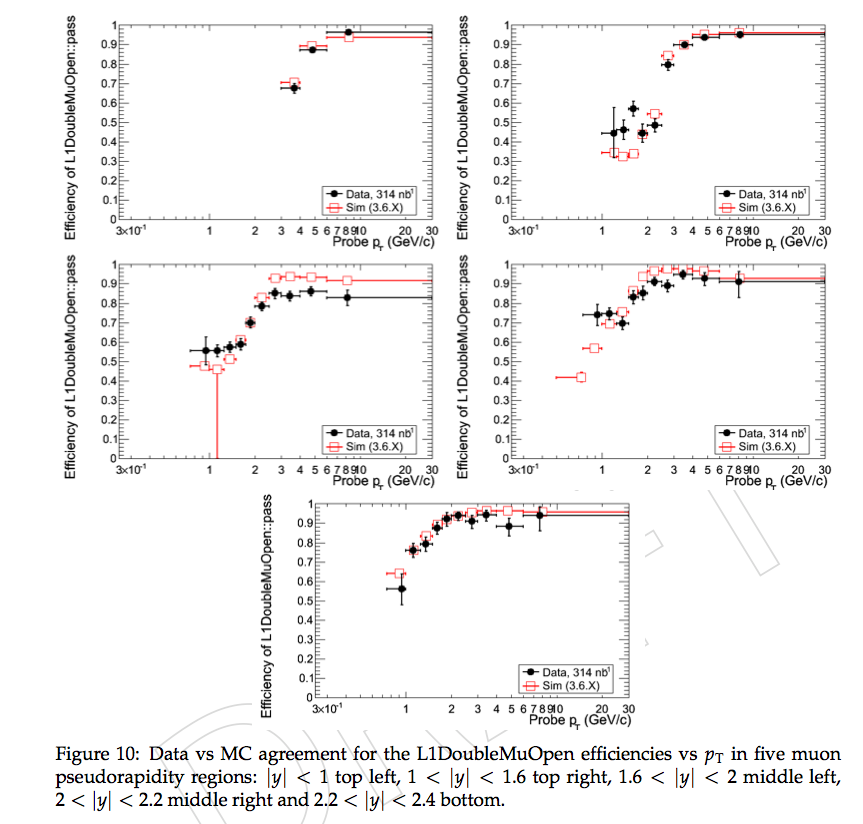


Figure 3. Data versus MC agreement for the muon trigger efficiency vs pT in five muons pseudorapidity regions (the same as Fig. 2).

Since the behavior of the efficiency turn on curves in the different five regions of the muon pseudorapidity are very well reproduced by the MC simulation, we believe that taking the correction from the MC is justified.

Also notice that the variation of the efficiency as a function of pt is a steep function for low pT, given that most of the muons are in the turn-on trigger and tagging efficiency. Unfortunately the integrated statistics does not allow to use both very small eta and pt bins of the muons (we in fact use 10 bins in pT and 5 in eta), without spoiling completely the measurement.

In fact the structure of the muon efficiencies is more complex than this picture. In fact Fig. 5 shows the variation of the trigger and tagging efficiency as a function of eta for different muon pt bins - using a high statistics (2 Million Jpsi) MC sample. For instance, the efficiencies in the bin between 3 and 6 GeV vary only slightly (less than 10%) as a function of eta. On the contrary, the bins at lower pt vary a lot (between 30 and 60%, roughly, in the bin between 1 and 1.6 in eta). Notice that the muons pt-eta bins which populate mostly the Jpsi bins with low rho value (or close to zero) are those for which the efficiencies are not varying very much as a function of eta, while those which have a large value of rho, have the largest

sample of muons coming from the regions where the variation as a function of eta is the largest.

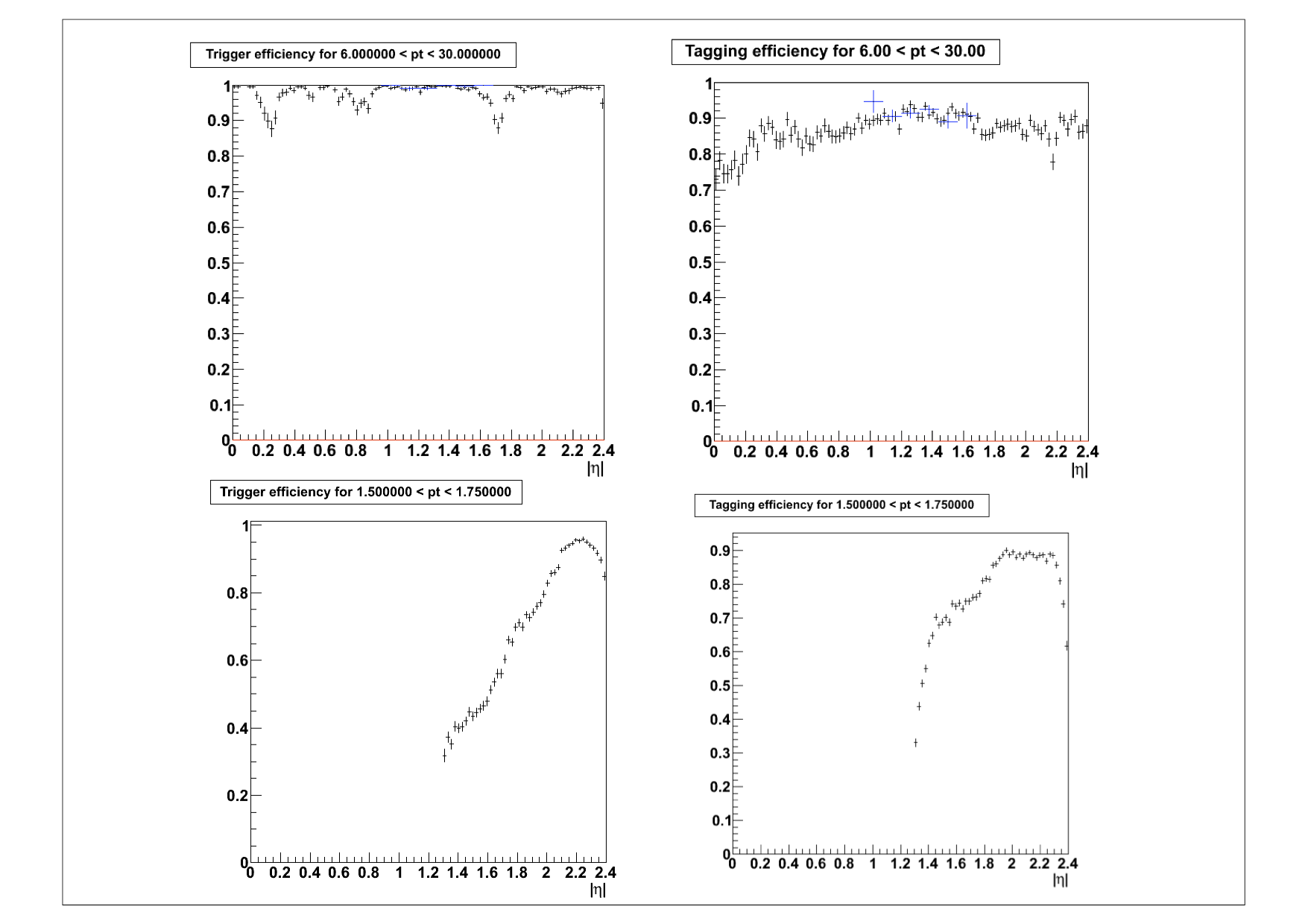


Figure 4. Muon trigger (left) and tagging (right) efficiencies as a function of pseudorapidity, for two bins in muon pT: 6<pT<30 GeV (top) and 1.5<pT<1.75 GeV (bottom). The upper bins contribute to Jpsi with rho~0 while the smaller pT to Jpsi with rho~0.3.

This effect is the responsible of the "correlation". As said before, having more statistics in data , we could have afforded to model it more precisely.

The rho factor arises from the fact that the bins chosen to describe the muon tagging and trigger efficiencies are too wide to properly account for their variation within the bin itself, as you suggest. This is shown in Fig. 4, which

shows 1+rho in Monte Carlo for the bin sizes as in the paper, and very fine bin choice:

MUON ETA BINS = 0., .1, .2, .3, .4, .5, .6, .7, .8, .9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4

MUON PT BINS = .1, .2, .3, .4, .5, .6, .7, .8, .9, 1., 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.2, 3.4, 3.6, 3.8, 4., 4.25, 4.5, 4.75, 5., 6., 30.

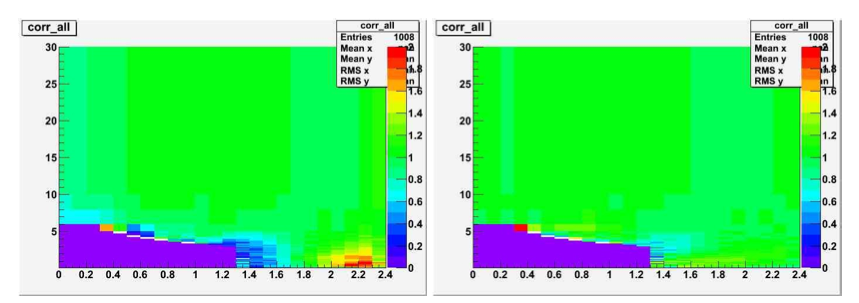


Figure 5. 1+rho factor for two different sizes of the muon pt-pseudorapidity bins (see text): those used for the paper (left) and very fine binning (right)

As it could be seen, the second - fine - binning completely decorrelates the two efficiencies.

The Table below shows that only very few bins in the J/Psi pt-y plane have a rho substantially different from zero. The typical error from the MC statistics (~2%) is such that only in a few cases rho is substantially different from zero.

|y|(JPsi) \* Pt Jpsi(GeV) \* 1+rho

0-1.2 \* 6.5-8 \* 1.09

\* 8-10 \* 1.09

\* 10-12 \* 1.15

\* 12-30 \* 1.08

1.2-1.6 \* 2-3.5 \* 0.81

\* 3.5-4.5 \* 0.83

\* 4.5-5.5 \* 0.94

\* 5.5-6.5 \* 1.023

\* 6.5-8 \* 1.06

\* 8-10 \* 1.07

\* 10-30 \* 1.09

1.6-2.4 \* 0-0.5 \* 1.120

\* 0.5-0.75 \* 1.29

\* 0.75-1 \* 1.30

\* 1-1.25 \* 1.25

\* 1.25-1.5 \* 1.27

\* 1.5-1.75 \* 1.26

\* 1.75-2 \* 1.19

\* 2-2.25 \* 1.14

\* 2.25-2.5 \* 1.12

\* 2.5-2.75 \* 1.10

\* 2.75-3 \* 1.08

\* 3-3.25 \* 1.07

\* 3.25-3.5 \* 1.06

\* 3.5-4 \* 1.04

\* 4-4.5 \* 1.03

\* 4.5-5.5 \* 1.03

\* 5.5-6.5 \* 1.03

\* 6.5-8 \* 1.04

\* 8-10 \* 1.03

\* 10-30 \* 1.02

Given the premises, it is unavoidable to have used the muon pt-eta binning to model the behavior of the efficiency and use a correction factor for the “large” bin sizes from the simulation. Having very fine grained bins is not affordable with the statistics used in the article. (in the example it is 1000 bins against 50)

The other possibility, which is to use a smoothing function by taking the high-statistics MC sample as an ansatz, would be totally equivalent to use the MC to predict the rho.

Finally, to model the effect of the generator we have in fact accounted for the variation of the spectrum using different theory predictions. Given that the three models have different behaviours in the low pt end and since we always took the largest variation of the three we believe that our estimate is a conservative one, even in the problematic bins.

> In addition I have a number of smaller comments:

>

> page 3, first paragraph: please add a short explanation of what the two

> different muon reconstruction algorithms are so the reader doesn't have to

> look this up in ref [23]. Especially since this reference is not a

> publication.

Ok, we will add that.

>

> figure 1: since the fit parameters are not quoted it is hard to judge,

> but my impression is that the radiative tail increases from the central

> to the forward region. If so, what is the reason for this?

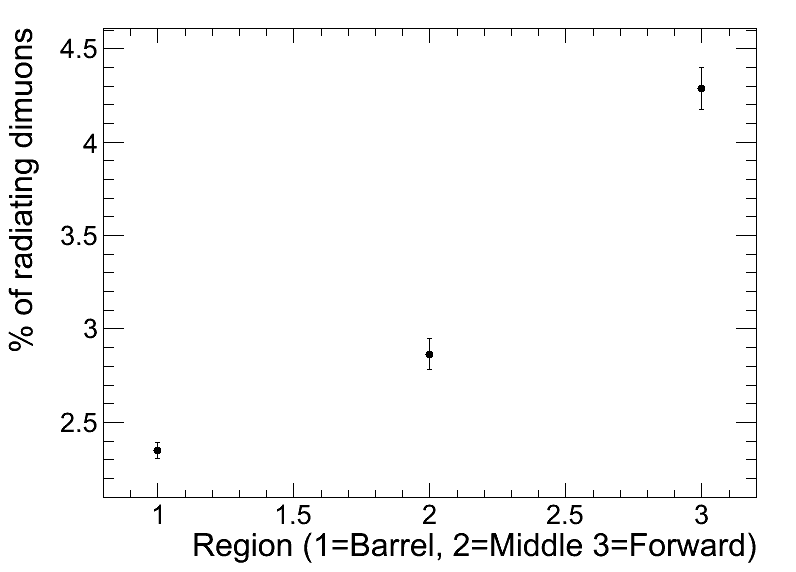
Your impression is right as you can see in Fig. 6, where the percentage of events in the low mass tail of the J/psi peak is plotted for the different rapidity regions considered in the analysis. The effect however is small, at the percent level. 

Figure 6. Fraction of radiative tail in the data, in the three rapidity regions considered in the paper.

The same behavior is also visible in the generated mass of the J/psi taking into account FSR radiation, as you can see in Fig. 7, where

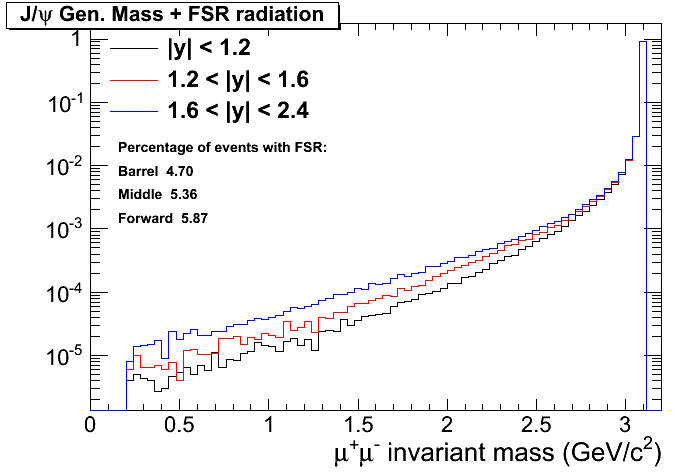


Figure 7. Fraction of radiative tail in the Monte Carlo simulation without the detector effects, for different rapidity regions. (See text for explanation)

in the forward region the FSR radiation is enanched. This is due to the harder momentum spectrum of muons in the forward region (see Fig. 8): given that the photons from FSR take on average the same muon momentum fraction in the three rapidity regions, in the forward part they have, on average, an higher energy.

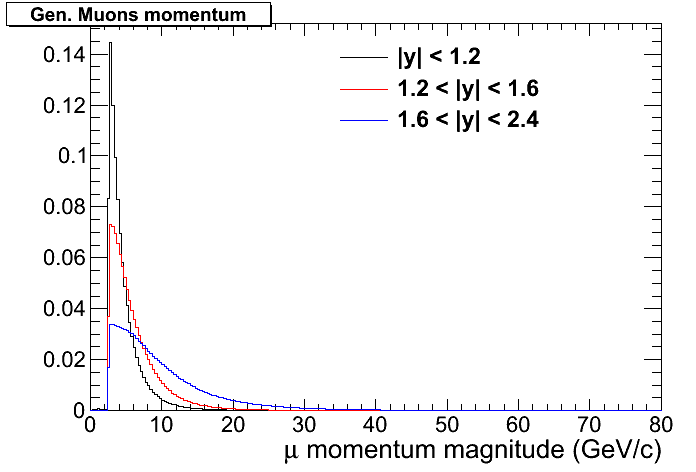


Figure 8. Generated muon momentum (before acceptance cuts) in the three Jpsi rapidity regions.

> section 4: it is not really clear to me if a bin-by-bin correction or

> a matrix unfolding is used. If it is a bin-by-bin correction, some

> information about the bin purity and a comparison of the MC used for the

> correction with the data would be interesting. On page 5 it is stated

> that the spectra of other theoretical predictions are used to determine

> the systematic uncertainty, but from figure 6 it is obvious that none of

> them describes the pt distribution in the forward direction of the data.

> This means that the systematics might be under- or overestimated.

See above. Given that the three models have different behaviours in the low pt end and since we always took the largest variation of the three we believe that our estimate is a conservative one, even in the problematic bins.

> page 5, systematic uncertainty due to b-fraction: why don't you use your

> own measurement (section 6) and it's uncertainty?

Thanks for spotting this out. In fact this is what is being done, but the clarity of the text could be improved. The 20% is an average of the b-fraction uncertainties (an average is taken to simplify things since this systematic uncertainty is not dominant).

> page 6, systematic uncertainty on rho: concerning the reweighting to the

> different theoretical predictions, the same applies as above.

See above.

> table 1: what is the uncertainty quoted for the average acceptance

> times efficiency?

The error is the total error (stat + syst) - where the systematics of course does not include those from the fit function and the luminosity. The error on the yield is only statistical.

> table 3: how are the <pt> values determined?

Here it is the mean pt of events in an invariant mass region of +/-

100 MeV/c^2 around the J/psi peak value, after subtracting the background contribution, estimated by the sidebands.

>

> section 6.1: it is not clear to me what are the free parameters of the

> fit. Obviously the fractions f\_Sig and f\_B, but also the f\_i of the

> background? and why are the lambda\_i fixed in the separate fit although

> you use the full mass spectrum and the signal and background shapes in

> the fit (especially since, in the forward bin, the signal extends into

> your sidebands)?

Yes, also the f\_i of the background are floated. Fixing lambda\_i ensures fit convergence in some bins where the fit is more critical (low pT, with poorer resolution and higher background), so we adopted this as a general fit strategy.

The lambda\_i values are varied by +/-1sigma (statistical uncertainty in the sideband-only fit) as a separate test, and the change in the b fraction is found to be negligible: this part is not detailed in the paper.

> page 12, background shape: are all three 'long-lived' components needed?

> CDF states that the sources for the negative and the symmetric

> contribution are unknown, so a bit more information about them would

> be helpful.

We found a much better fit quality with three long-lived components than with two. Assessing the source of the three components exactly is not easy because it would need a huge amount of MC events and, also, we found fake muons at low pT not to be reproduced very well in the simulation.

Naively, background is made of are random particle pairs that usually do not come from a single decay: one of the two tracks (the more "precise" in term of number of hits, chi2 etc.) "drives" the secondary vertex position. So you may have different cases:

- in most cases, the driving track is a real muon from a b, c (or even K, pi) decay and this gives the positive component

- in some other cases, the two tracks have the same weight in the vertex determination, so the displaced position can end up on either side of the primary vertex

- the choice of the negative component is only made based on an improved description of the negative part of the background lifetime

> figure 4: why does the total fit change curvature at large positive

> l\_Jpsi?

As explained in the text, the non-prompt component is given by a

MC template, extracted from a finite-size samples. Oscillations in the few MC entries at very large l\_Jpsi cause the curvature change. One can infer the relevance for the fit result is small judging from the size of the "B-lifetime model" systematic uncertainty, which instead makes use of smooth analytical functions with no curvature changes.

> table 4: how is the average pt determined?

See above.

> and what is the rms supposed to tell the reader?

The RMS is intended to give the second moment of the distribuition of the pt. If you really insist we could remove it.

> page 15, section 7: please add a short statement on the model

> implemented in CASCADE (colour singlet, kt-unintegrated PDF)

Thanks for alerting us, this description is indeed missing and will be put.

> page 15, section 7: I agree with you that for the CSM with higher order

> corrections not all contributions are available, but LO NRQCD

> predictions including feed-down from chi\_c exist (see for example

> hep-ph/0003142). In fact, you probably use the colour-octet matrix

> elements in Pythia. So why do you not show the calculations?

This matrix elements that are reported in this publication (Table 13) and the references are indeed what PYTHIA is based on. But this is not the same as what Pierre Artoisenet was calculating for us, he said he could only make predictions for direct production and based his calculations on MadOnia.

See:

http://www.physics.ohio-state.edu/~partois/DirectJpsi\_LHC.pdf

So, if we want to display predictions as given in hep-ph/0003142, perhaps we should have contacted different authors, but we think people like Pierre Artoisenet or Jean-Philippe are the most active these past years, so were the right ones to ask.

> page 16, discussion of figure 7: the statement on the good agreement of

> the calculations with the data seems too positive for Pythia in the

> forward direction.

Yes, we realize that the comment might give too emphasis to Pythia, while we wanted to stress more the agreement with the CASCADE and FONLL, which indeed are the calculations.

> figures 6 and 7: why do you use Pythia to calculate the abscissa

> although it does not describe the data in some regions? showing data

> and predictions in the same bins would avoid this problem.

We have checked that using the other curves instead of Pythia give very similar results, and visually there is no difference among the three.

> figure 6: showing the different contributions in Pythia (feed-down,

> colour singlet, colour octet) might enhance the physics message.

The implementation of direct vs indirect production in PYTHIA doesn't make so much sense, we know it doesn't agree with CDF data. So why display something we do not believe in....

In fact, CASCADE and CEM are more solid predictions, but since we are not dealing with measuring direct vs indirect prediction, displaying just the sum should be ok.

> figures: please use the same abbreviation BR as in the text

OK.

> references: many references are CMS notes. They are available also to

> external people, but they are not journal publications. So if the

> information is really important for the analysis you should include it

> in the paper so a referee has a chance to comment on them.

> There is a preponderance of results in the list of references, namely 14, 15, 23, 25 and 29 that have not been considered worthy of a proper publication by CMS and hence have not been peer reviewed. Yet several analysis aspects for this report are directly drawn from the notes with little further explanation. Reference [23] is based on much smaller statistics than this paper and yet drives the systematic uncertainty. The analysis introduces formally a correlation rho, which seems to be driven by a binning artifact rather than adequate analysis of the full data. This is one example where it would be much preferable to see the relevant assessment of the data carried out for and in the publication.

>

> In general references to notes should be avoided. Please review the references and modify the text or list according to the following priority list:

> 1) If at all possible such references should be replaced by references in a refereed journal or should be avoided altogether.

> 2) If such a reference does not exist and the reference is vital for the paper consider publishing the reference first.

> 3) If you do not want to publish the note in a refereed journal first explain the core context of the measurement in this publication and mention the quantitative result, which should be made plausible. The unpublished note could be referenced as supportive material for the interested reader and should be indicated as such. The paper must however be understandable and trustworthy without the note.

>

> Case 3 is certainly discouraged.

>

> Further observations on the reference list are indicated below and should be addressed.

>

> [1] provide a link

> [2] add doi:10.1142/S0217751X06033180

> [9] misspelling in title: heavy flavor -> heavy-flavour

> [10] misspelling in title: heavy flavor -> heavy-flavour

> [11] misspelling in title: heavy flavor -> heavy flavour

> [14] CMS PAS EWK-10-004 is an unpublished and unrefereed note. It does not have an author nor an address. There is only an email address

> [15] CMS-PAS-TRK-10-005 (2010) is an unpublished and unrefereed note. It does not have an author nor an address. There is only an email address

> [17] misspelling in title: perspectives at LHC -> perspectives at LHC(b). Please provide a link to the cds server. This is an unpublished and unrefereed note.

> [21] misquote in titlte: G4 -> GEANT4

> [23] CMS-PAS-MUO-10-002 (2010) is an unpublished and unrefereed note. It does not have an author nor an address. There is only an email address

> [25] CMS-PAS-TRK-10-003 (2010) is an unpublished and unrefereed note. It does not have an author nor an address. There is only an email address

> [28] misquote in title: in B decay -> in B decays

> [29] CMS-PAS-TRK-10-00 (2010) does not exist

> [34] add doi:10.1016/S0010-4655(01)00438-6

> [36] misquote in title: flavors -> flavours

> [45] wrong doi: use doi:10.1103/PhysRevD.53.6203

PhysCoord taking care of that.