

Table of Contents

| | |
|--|----------|
| Analysis Review Twiki for SUS-21-003: Search for 4-body decays of stop in 1 lepton final states using a multivariate approach in Run II Legacy Dataset..... | 1 |
| After full-status report and comments on the AN..... | 2 |
| Comments on AN_v2..... | 2 |
| Comments on AN_v3..... | 10 |

Analysis Review Twiki for SUS-21-003: Search for 4-body decays of stop in 1 lepton final states using a multivariate approach in Run II Legacy Dataset

Cadi entry: [SUS-21-003](#)

Key

- Questions and answers start by **Q** and **A:** respectively.
- In cases where a request and/or question has been replied to, a  is used to indicate that it has been answered.
- In cases where a request and/or question has not yet been replied to, a  is used to indicate that the requested is still pending.
- If there is a reason the requested change has not been made, an  is used along with a following explanation.
- DB stands for replies from Diogo Bastos
- PB stands for replies from Pedrame Bargassa

Expand all Collapse all

After full-status report and comments on the AN

Comments on AN_v2

▣ Show... ▣ Hide

Q1: L106 and appendix A2: I am surprised to see that you observe identical p_{miss} with and without 2017 corrections. It is true that you are not selecting jets in the noise regions, but unless you explicitly veto events with jets in those regions the p_{miss} should still be affected. How do you apply the recipe on plot 26?

A DB: I applied the recipe as it's suggested here: MissingETUncertaintyPrescription loading a producer. Then compared the MET distribution without any selection, and at preselection level. It was at preselection level we saw no difference in the distribution. Not seeing any difference is the criterion to assess if the analysis is sensitive to this problem or not and it happens not to be.

Q2: Table 1: with respect to the previous version of the AN, Flag_eeBadScFilter disappeared from the table. Are you applying it on data?

A DB: I'm not applying Flag_eeBadScFilter because it's stated as "not suggested" in MissingETOptionalFiltersRun2

Q3: L 117 and others. So I am surprised you have to use the loose cut-based working point. Probably this is because it is not optimized. Have you looked at MVA ID (using for example the WP used by SOS) to check whether you could gain considerably. The cut-based ID is not really optimized for low p_{T} leptons.

A DB: We looked into using MVA ID and compared with cutBasedID efficiencies. It's true that MVA ID is more efficient at low p_{T} electrons but in order to keep consistency with the previous analysis and later combine with the results, we decided to keep with the cutBased ID.

Q4: L123: so muon ID scale factors for $p_{\text{T}} < 20$ GeV are provided by the POG?

A DB: Yes, you can find them for 2017 and 2018.

Q5: L188: please, provide more detail on the W+jets reweighting, maybe in a dedicated appendix.

A DB: I agree. Now included in version 3 of the AN, section B.1.

Q6: L140: do you have the lepton SFs for IP+Iso requirements?

A DB: Since we have a good agreement between Data/MC in the lepton related variables using only the ID SFs (that don't depend on IP or ISO) for both years and in both final states, given the time and manpower constraints, we moved forward without deriving the IP+Iso SFs. Nonetheless, we include an uncertainty of 1% (a conservative estimation of the sys. unc. taken from 2016 analysis) in the end to account for this.

Q7: L 213: why do you preselect the MET this high? Does it not give you low acceptance? I think many analyses use it at lower MET

A PB: The values of the cut on MET, $p_{\text{T}}(\text{jet1})$ (and also HT as far as I remember) have all been determined with a FOM maximization, precisely to be systematic and not choose cuts 'by eye'. What I did was to pick up a signal point of rather low ΔM , here (300,270), vary the cut on MET (and other variables) and see which cut was yielding the highest FOM. This procedure was done when (1) changing the cut at preselection, (2) then training a new BDT, (3) observing the highest FOM. The procedure was a bit horrendous, but it was/is to ensure that it yields to the best result in-situ, i.e. in the real and full conditions of the search. In brief, the cut isn't too high, its value having gone through the maximization of a number which accounts for both the stat &

sys uncertainties.

It is not giving a low acceptance, you can get an idea of this on the 4th plot of Fig 6. Each analysis is different, and again, in our case, we took a full picture of the analysis by maximizing the FOM.

Q8: L226: so this analysis is not disjoint with other soft analyses, like SOS, right? Is there really a gain of keeping the lepton veto at higher pt. Analyses like the stop-1l saw that the dilepton backgrounds are hard to kill in the 1l analysis and therefore they use a tight lepton veto (at 5 GeV), so I wonder whether you can explain why your case is different.

A PB: I don't know what you mean by disjoint with other analysis :). What I can say is that it is obviously a different analysis. Regarding the ttbar(dilepton):

* For the selection aspect: I did not make specific attempts to kill the dilepton channel, since the Wjets is the main challenging background, this from pre-selection until the final selection level for most of DMs. Now since the ttbar sample with which we train the BDT is an inclusive sample, whatever fraction of ttbar(dilepton) survives is taken into account in the training, and diminished (not necessarily killed) with the application of the cut on BDT's output. Tables e.g. 5 to 8 report the MC expected yield of ttbar, including the dilepton contribution. Now one might say that fraction of ttbar(dilepton) isn't the same in MC and Data. For the selection aspect of things, the answer is that by not training the BDT specifically versus ttbar(dilepton), we pay a certain price in residual background, fine... Overall, the good Data-MC agreement of the BDT output gives us good confidence in the fact that the MC ttbar, including the dilepton component, is

* For the background prediction aspect: the equation (1) is also inclusive: the term $N^{\{CR\}}(\text{Data})$ includes events purified in ttbar (after the $N_b(\text{tight}) > 0$ cut) for both 1-lepton and whatever fraction of 2-lepton survives the pre-selection; similarly for the $N_{\{\text{prompt}\}}$ terms derived from MC and used in a ratio.

Q9: L241: could you please explain the figure of merit used, it is mentioned a few times but never explained

A PB: May I refer you to the section 4.2.2 of the supporting AN for the 2018 publication [iii] where the full formula and reference are given. Very shortly: this is a much more advanced and complete version of the already quite complete $S / \sqrt{S + B + \text{sys}^2(B)}$ figure. Diogo will add it back to the AN.

[iii] https://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/035

NB: Fig. 10 gives a good illustration of where the FOM is used, i.e. for finding the most discriminating set of input variables to the BDT. Now since this aspect of the work isn't done anymore in the present AN, we didn't think that giving the FOM was very necessary, particularly that Diogo is citing the AN in [iii].

Q10: L270: why is only the highest b-jet discriminant used and not the second one? For the second one I would also expect some discriminating power, but maybe it is already fully in the $N(\text{loose } b)$?

A PB: This happens to have been covered in my studies leading to the publication: I did tests of BDT including this variable, and its inclusion didn't improve the performance of the BDT, so wanting to have the 'best' performance with a most reduced set of input variables, I simply decided to keep it out.

Q11: Figure 6: why are these plots split between W+jets and ttbar plots. Since the training is done to the sum, it would be nice to see that here as well.

A PB: Simply because a given variable is sometimes discriminating versus one background only. Take the example of the charge of the lepton: by definition of pp collisions, ttbar production is symmetric in this variable, while the W+jets isn't. So displaying the ttbar distribution wouldn't bring anything, being identical to the signal.

Q12: Something which would be nice to see would be to see some of the variable distributions after applying the BDT cut (maybe for 2 mass points), just to see what kind of events we select in the end with the BDT.

A DB: That's a good idea, added an appendix in AN_v3, section C.5.

Q13: I was also wondering whether you ever consider a parametrized BDT (similar to the parametrized NN that SUS-19-012 uses), since it might help with the statistical power of your sample.

A PB: Parametrized ML tools bring an added value where the number of unknown parameters (here masses) is too large for having its picture captured by an already existing quantity. Allow me to explain. In SUS-19-012, there are (at least ?) 3 unknown parameters: the mass of the LSP, NSLP, and of the e.g. chargino. There isn't an experimental parameter which allows to easily capture these 3 degrees of freedom. Hence indeed the point of using a parametrized ML tool, here taken to be NN. In our case, (1) things are simpler, and (2) we capture the picture of the unknown parameters with DM. (1) We have only 2 unknown parameters: $m(\text{stop}_1)$ and $m(\text{LSP})$. (2) We show that the main parameters governing the variance of the kinematics across signal points, thus the selection of the signal, is ΔM (see Fig. 9), and we precisely train one BDT per one DM; we thus have 8 different selections for optimally adapting the selection to 8 type of different signals. In short words, we are somehow doing the parametric search, but by ourselves, while being guided by the (as far as we know only) parameter which governs the kinematic variance.

Please note that the point (2) is valid for a number of other searches where the number of unknown parameters is 2, at most 3 (in which case we consider slices along the 3rd parameter). I applied this in SUS-14-015 (in this case, the 2-body decays $\text{stop} \rightarrow t \text{ LSP}$ perfectly fall under what I say; for the decays involving an intermediate chargino, it might probably been better to use something like in SUS-19-012, but parametric ML wasn't very known at that time).

Q14: L296: what is the motivation in not including Z_{nu} in the training. I understand multijets are impossible due to the high weights, but this sample should work. In principle I would be tempted to include as many background sources as possible, so the BDT can just optimize. But since you mention Z_{nu} is not negligible after the BDT, I would for sure be tempted to use it.

A PB: I did tests with and without including Z_{nu} in the training, this versus the same signal. In terms of performance, the training including Z_{nu} somehow under-performed when compared to the training without Z_{nu} , so this latter wasn't included in the final training. On the explanation side: this might be due to the fact that the 12 input variables, while being pretty good as discriminating signal from $t\bar{t}$ and/or W jets, aren't very decisive for Z_{nu} . On the to-do-for-future side: one might do either a multi-class training, or a specific training for Z_{nu} , but again, given that our (so far) best variables aren't very good at distinguishing signal from Z_{nu} , I am rather skeptical, but will do it once we have the time and opportunity.

This is why one of the the only ways to decrease this background is simply to require a tighter lepton ID, which Diogo did.

Q15: Table 5-12: this is connected to a question asked at the full status: the signal and background yields are very different in 2017 than in 2018 due to the different operating points (2018 a factor 3-4 higher in a few bins). But somehow, we would expect it to be quite similar. Do you expect there to be a real physics reason or do you expect this to be due to statistical fluctuations in the samples. In case it is the second, is there a way to be less dependent on this and how much difference do we have if we move to the same WP?

A DB: There are generally two effects: 1) different trainings per DM within the same year, and 2) for the same DM samples statistical fluctuations between the years.

1) Within the same year from one dm to another, it's a different and independent analysis. The best way to see this is by looking at the BDT shapes in the overtraining check as it was presented in the Full Status (section C.1 of AN_v3). This is the reason why within the same year one should not expect the optimal cut to happen

at the same value, thus not expect the yields of the background (a fortiori signal) not to be the same.

2) Within the same DM but between the years we are affected by the statistical fluctuations in the samples. New section C.7 develops on this. To put it shortly, the cut setting is affected by rather large statistical fluctuations due to the smaller size of the WJets samples in 2017 and 2018. As a result of this, we relax the BDT cut to be in a region where we have enough statistics (for the Data driven background prediction). Since the size of the samples between the two years are different, this results into different cuts and therefore different predicted yields.

The only way not to be hit by this problem is to have statistically significant samples which we don't have. If we change the WP in one year for it to resemble to the cut value of the other year, the yields would still be different because the statistics of the samples are different.

Q16: L324: is this split of datasets in test- and train-samples only used for this check or also in deriving the final training used in the analysis?

A DB: The overtraining check is performed after the samples are split into training and testing. The trained BDT is used afterwards in the analysis and all the plots and tables are on the test dataset.

Q17: L329: does this paragraph really belong here?

A DB: You're right. I think it's better suited for the beginning of the chapter when the BDT method is introduced. Changed it.

Q18: L362: not clear to me: the BDT is re-trained on the VRS, or you use the same training from the AR?

A DB: We used the same training as that is the training we want to validate.

Q19: L379: is it really the highest for the signal with the highest cross-section? Since the acceptance might change a lot too. Could you show the full image? Also, is this used in the final fit to correct?

A DB: The highest XS is for Stop Mass of 250 GeV, it's not 300 GeV, you're correct. I've update that and included the plots. According to tests done in the 2018 publication, signal contamination < 5% are negligible.

Q20: About the automated determination of cut point: is this done with the testing sample? And what is used for the SF calculations in the background method? Also this one? The reason that I am asking is that for low stats you risk biasing yourself in that case, since you will optimize the BDT cut taking into account the statistical fluctuation and then the same fluctuation will make the SR/CR ratio in MC small. I think the possible issue is slightly stronger here than usual due to the automated method.

A PB: The determination of the cut is all but automated. As explained by Diogo in his talks, the cut value found by the UL minimization is:

* checked by the efficiency curves of signal and background: if statistical fluctuations are observed there, the cut is made softer as to not expose the final selection to the lack of statistics,

* checked iteratively, and the cut chosen as to minimize the systematics.

Q21: Table 13: there is a big jump on the BDT cut for $\Delta M=70$ in 2017. Can you please provide more detail on the origin of this behaviour?

A DB: This goes into my reply for the question about Table 5-12.

Q22: L426: The signal contamination is not very small, going up to 7%. This should for sure be taken into

account in the final estimate. It is not clear it is.

A DB: Thank you for pointing this out Luca! This was a table and results from a while ago when we were still using the BDT cut of the 2016 analysis... We reviewed the cut as the BDT shapes for 2017/2018 are also different. I updated to what we are currently using and included the table with the yields to support it. Now, with the correct cuts applied the signal contamination is ~3%. Tests have been done in 2016 where signal contamination of 5% have zero effect on the minimization of UL XS.

Q23: L438: this paragraph would read better if the tables were introduced and referenced following a precise order.

A DB: Here, I prefer just to reference the example of only one DM in the main body of the text and complete the picture in an appendix, in section "Test of the prediction of WJets and TTbar for different DM", because the total number of tables is 56. And it's organize by year, then VR, then wjets and then per dm and within the same VR ttbar and the dm's. Since we have 2 years, 2 VRs and 14 SR in each VR (8 for VR2 and 6 for VR3).

Q24: Table 15 and following: I am not sure I follow the math. Let's take Table 15. The data yield in the CR is lower than the expected MC yield from W+jets. So data/MC is < 1 but then the data-driven estimate is considerably larger than the MC yield for W+jets. Maybe I am misunderstanding the labels, but could you explain for Table 15 exactly how you get to the numbers. Then we can see how to present it.

A DB: I don't know if I made a typo the first time I was adding these tables to the AN. Meanwhile I've updated them, now in version 3 of the AN, with the correct trigger efficiency, the event filter that was missing and without ttbar ISR reweigh as you pointed in the previous set of comments. Since I had to repeat all the measurements of the systematics and the validation of the MVA, I took some time to review my code and now it outputs the table in tex format so that I just have to copy/paste it to the AN, in order to reduce similar mistakes. I think it was not you who didn't follow the math but rather me who left there a typo in one of the numbers. So, if it's ok with you, let's follow now table 36 in AN_v3 with the updated values.

By replacing the variable in eq 1 with the number in table 36 we have the following:

$$370.2/102445 \times (109256 - 9302) = 361.2 = N^{\{SR\}}_{DD}$$

$$\text{To } N^{\text{SR}}_{DD} \text{ we add } N^{\text{SR}}(\text{Other}) = 646.8 = N^{\text{SR}}(\text{Predicted})$$

$$\text{We do the difference between } N^{\text{SR}}(\text{Data}) \text{ and } N^{\text{SR}}(\text{Predicted}) = 768 - 646.8 = 121,2$$

In this case, this difference is higher than $\sigma(N^{\text{SR}}_{DD})$ and the relative systematic uncertainty is $\text{Diff}/N^{\text{SR}}_{DD} = 121,2/361,2 \times 100 = 33,5 \% \rightarrow$ Table 44, dm=30 in VR2

Q25: Also, how much of the 'other' that is deducted is the other major background source? I am just wondering whether it makes to couple the estimates by using both CRs to get a more correct estimate.

A DB: Whatever is the amplitude of the cross contamination within the CR of the VRs, it has to be taken into account and this is what we do. I don't think we can couple VR2 and VR3 as they are by definition orthogonal to each other.

Q26: Table 37: the fakes are taken from the data-driven method, right?

A DB: Correct, $N^{\text{CR}}_{DD\text{fake}}$ is the data driven prediction of the fake leptons in the CR.

Q27: Just to be sure: ttbar fakes are included from the fake lepton estimate and not the MC?

A DB: Yes.

Q28: ttbar 2l will be more pronounced in SR than in CR due to the variables involved, which one of your uncertainties covers this effect?

A PB: Again, ttbar(dilepton) is included in the data-driven prediction of the ttbar background, see eq. (1). Now differences between CR and SR are taken into account via 2 different methods explained in section 5.1.1, this, for any background. Very shortly: it is method (2) (see eq. 6) which looks at the same ratio (Data/MC) both in CR and SR: if there is a difference of this ratio from CR to SR, it is taken into account as sys.

Q29: L451: I guess in the SR [...] in the corresponding VR means in the SR of the VR both for observed and expected events.

A DB: You guess correctly.

Q30: L 518 and following: do you have some plots in MC to show how much the flavor dependence changes and whether this could cover the difference between fake estimation region and the application region? In past analyses asking for 1 b to enrich the b-flavored background also added a deltaphi requirement between the b and the fake lepton to make it even more likely that the fake lepton came from the second b.

A DB: The plots where we show the flavor dependence of the tight-to-loose ratio are in figure 20 (AN_v2) / 21 (AN_v3) although we didn't add any extra requirement besides the enrichment in b-flavored background.

Q31: L569: it is not clear to me from the text how the up/down variation of the weights are computed.

A DB: I've explained this part better in the appendix B.1 of AN_v3. The idea is that we get the weights, which have a statistical uncertainty associated to them, and then we see how much varying each weight up and down within this uncertainty affects the total prediction of Wjets and TTbar (because wjets enters in the ttbar prediction in the Other).

Q32: L583: Again, what about lepton isolation?

A DB: See reply to L140.

Q33: L614: it would be more clear if you put the tables in the order they are introduced in the text.

A DB: I completely agree. I've moved the tables in AN_v3.

Q34: Table 42 and following: can we see the inputs? It would be interesting to see whether it comes from the precision of the test or from the discrepancies in general. We might have too many uncertainties here and we might have double-counting effects.

A DB: The inputs are on appendix "Test of the prediction of WJets and TTbar for different DM" and follow the math of the reply to table 15

Q35: L627 maybe worth stating explicitly here that the following studies address the modeling of BDT output.

A DB: Done!

Q36: L 630-632: How can we know whether this would cover us or maybe overcovers us? Is there a good reason to think that the effect in the VR directly translates?

A DB: The method (2) is, in itself, a cross-check of the method (1) for the assessment of the systematic uncertainties. Method (1) is measuring the difference between observation and prediction of a given

background in the SR of the VR where the prediction is a function of the CR of the VR. Method (2) directly compares the ratio R of the CR to the one in SR, both of them in the VR. As you can see, the two methods use the same blocks and can be used as cross-check of one another. On a historical note: for the publication, we were originally using only method (1), and were suggested to use method (2) as cross-check. So one should see the 2 methods as a whole, i.e. as cross-checking each other, and this is the reason why we take the maximum of the two per DM ensuring that we cover the systematic uncertainty of the DD method.

As to know whether the effect in the VR directly translates, allow me to explain how and why the VRs are constructed first. They are built in kinematically orthogonal regions to the Analysis Region as to decrease the presence of the signal, and to allow to gauge the BDT output across its entire range, i.e. up to SR while the signal is negligible; this is useful to assess the systematics. In order to render the VR as comparable as possible to the AR, (1) we take them as close as possible to the AR, (2) when possible (6 out of 8 cases) we consider two of them, please see Fig. 14. Taking VRs "surrounding" the AR and as close as possible to it is the best because the most complete exercise we can do to gauge the effects happening in the SR of the VR while having to be blind to this latter.

Q37: L 635 formula. I was a bit confused by the formula. What is $\Delta_{\text{sta}}(\text{DD})$.

A DB: It's the statistical uncertainty of the Data-driven prediction of the background in the SR of VR.

Q38: L637: same as for L614.

A DB: Fixed.

Q39: Table 46: I am a bit confused. I would think that the percentage effect was roughly $\Delta_{\text{prime}}^{\text{DD}}/D$. Why is this not the case and is it $\Delta_{\text{prime}}^{\text{DD}}/N^{\text{SR}}_{\text{DD}}$? I thought the percental bias on D would directly affect you.

A DB: No problem. For both methods (1) & (2), we first assess the absolute systematic uncertainty, then divide it by the DD prediction to get a percentage because this is how systematics are implemented in datacards. By absolute sys, we mean the sys uncertainty for a given yield of background. Let us take the example of method (1). The absolute sys is defined through equations 3 and 4. Then, in eq. 5, it's simply divided by the data-driven prediction of the background in the SR of the VR ($N^{\text{SR}}_{\text{DD}}(X)$). It's the same for method (2) across equations 6 & 7, which are then normalized to the same $N^{\text{SR}}_{\text{DD}}(X)$ in the VR.

Q40: Eq. 7: can you comment about the different choice for the first member of the Max function with respect to Eq. 4? Also, is there a difference between Δ_D and $\Delta_{\text{stat}}(\text{DD})$?

A DB: In eq. 4, we have to take into account the full precision of the method. On the left side, we measure how well the method predicts the observation in data. On the right side we include the precision of such a prediction, $\Delta_S(\text{DD})$. This precision includes the statistical uncertainty in Data, the uncertainty on rare MC processes due to 50% variation of the cross-section, and 20 % uncertainty for the cross-contamination process. We need to take into account such precision as Data includes X process, rare MC processes and cross-contamination process.

In eq. 7, D is a difference between ratios that gives us the percentage of the mismatch of the BDT shape in the CR vs SR. This D has a statistical uncertainty in itself that we have to take into account. We can only trust D within its statistical uncertainty, so in the left side we do $D^2 - \sigma_D^2$. To illustrate this, let us consider cases where σ_D^2 is bigger than D^2 , which means that within the statistical uncertainty the ratio in the CR is compatible with the one in SR, therefore the uncertainty should be 0%. Which (as we've been discussing for the Raw Closure of the fakes) is not the best way to assess such systematic uncertainties. So, we do the maximization between the estimator $D^2 - \sigma_D^2$ and $\sigma_{N^{\text{SR}}_{\text{DD}}(X)}^2$ in the VR. Which is saying we can only trust the left member of eq. 7 with a certainty as low as the statistical uncertainty of running the DD method in the SR of VR. On the right side we don't include the precision in the same way

as of method 1 as we are not concern with the precise prediction of the total background in the SR. This is why compared to method 1, we take only into account the statistical uncertainty of the prediction method.

Q41: L651: I am not sure why this paragraph is here. Is the transfer function the on appearing on Eq. 1? Are you just repeating that W+jets background is also affected by the uncertainties on the reweighting or there is something more I am missing?

A DB: No, it's redundant here, I've removed it.

Q42: Table 50. The details about the closure systematic are not described in the main text. Can we move some of the information from appendix B.5 here?

A DB: We prefer to leave the this technical part of the analysis in order no to clog the main body in the appendix.

Q43: Eq 11: what is the reason for using a different formula than Eq. 7? It seems that the second member of the Max should be the uncertainty, not just 0.

A DB: You're right. I spent the last week looking to the DD systematics, mainly, the fakes and updated the equation to what you suggested. For the new estimate, please see eq. 10 in AN_v3. I recalculated the systematics due to raw closure and the limits I present in AN_v3 already include those.

Q44: Fig. 22: which options were used to run this? It seems like the nuisances were not allowed to be constrained, since the pull is always perfect. I would expect some uncertainties to be slightly affected. I also wondered how the systematic uncertainties on the fake rate are split. Here it seems there is just 1 nuisance for this? Is that correct?

A DB: To get the impact of the nuisance parameters I followed exactly the suggestions in the SUSPAGPreapprovalChecks that point to the combine tutorial [\[1\]](#). From the SUSPAG twiki page: "All the pulls should be centered at 0 and have width 1" which is what we observe. Regarding the impacts, they are not expected to be necessarily symmetric, and they are not. Although, the asymmetry is reasonable as one can see.

I didn't split the systematics of the fake rate, neither of the other DD methods. In principle, I'd like to separate them per source (in the case of the fakes: WJets correction, eTL, closure and statistical uncertainty) and combine the systematic uncertainties between the years while the statistical part would not be combined. Unfortunately this was not done in 2016, meaning all the uncertainties are quadratically added. In order to combine my results with the ones from 2016, I had to follow the same procedure.

Q45: Table 105: do we actually understand what is the reason for the non-closure? As we already noted before, it seems to be that the DD method almost always underestimates the MC. Is it due to the flavor dependence? Or do you have another hypothesis? Wouldn't it be more appropriate to derive a correction factor rather than setting an uncertainty?

A PB: What is the most important about this test is to see whether the method closes or not, irrespective of the reason; in the latter case, we take into account a systematic uncertainty to account for how much we are (systematically) off by predicting a known quantity from a given method. As such, we see the attribution of a sys uncertainty more appropriate.

Now we know that the $N_{\text{Tight}}^{\text{SR}}(\text{np})$ number suffers from lack of statistics given the samples we have to evaluate it: these are non-prompt leptons qualifying tight ID (for our analysis), this, in the SR; two reasons for this number to be small. On top of this, we have tightened the definition of our tight leptons compared to the 2016 search [i]: this is the reason why the uncertainty is larger for $N_{\text{Tight}}^{\text{SR}}(\text{np})$, and why the method closes less well than in [i].

It is not true that the DD always underestimates the MC, please see the cases DM=60, 70, 80 on Tab. 104. Hence our point that the disagreement, apart from being accounted for, is a reflection of lack of stats rather than anything something systematically wrong.

The flavour dependent sys is already accounted for as reported in Diogo's talks, see Figures 20 & 21 + Tab. 39 (and corresponding text: 1515-523), and Tables 50, 51 (column: "Lepton universality").

Q46: L691: please, describe the fit in Section 6.

A DB: I've included in AN_v3 the details of the Asymptotic Limits.

Comments on AN_v3

Show... Hide

Q47: Follow up to Q1: does this mean you did see a difference without any selection? If so, it is even stranger that you see identical distributions after the preselection, which cuts on MET.

A DB: I'm sorry to have induced you in error, this were some tests I did quite a while ago, while still working with miniAOD samples, I made the comparison at Preselection level. In any case, right now I'm using nanoAOD samples and using the METFixEE2017 variable that completely excludes jets with $p_T < 50$ GeV in the [2.650, 3.139] region from the calculation of pMET. This is the only recipe I'm aware that treats this issue.

I'll take this plots out and just leave the part the refers to the application of the recipe in nanoAOD to avoid confusion.

Q48: Follow up to Q2: the not suggested in the table refers to MC. For data, the Flag_eeBadScFilter is marked with

A DB: Ok, I was not sure about this. I've added this filter to Data only as suggested by the twiki.

Q49: Follow up to Q3: in principle it is not a problem to have a different ID in different years. How large/small was the gain seen with the other ID?

A DB: From the previous links I sent, the improvement in efficiency is ~5% at $10 < \text{electron } p_T < 20$ GeV. For lower p_T electrons, the efficiency is not calculated. It's hard to say without doing further studies. We prefer to keep with the same electron criteria as this was optimized by the cut-based analysis for the 2016 analysis and because of time constraints.

Q50: Follow up to Q4: ok, then just add this information on the AN, please.

A DB: Done!

Q51: Follow up to Q6: it is not clear to me that a 1% uncertainty is sufficient to account for the missing efficiency corrections. How was this estimated?

A DB: This estimate is taken from the 2016 analysis, where the full estimate of the systematic uncertainty was estimated. The value of 1% was a conservative one, as the total uncertainty due to lepton ISO was always < 1%.

Q52: Follow up to Q21: I understand that the jump is due to one WJets event from a low-pt bin (very high cross section) accidentally making it into the SR. That's life. Just wondering if there is any recommendation from StatComm about how to deal with these cases...

A DB: We will contact StatComm as soon as possible.

Q53: Follow up to Q25: not sure we are talking about the same thing. The idea was to couple ttbar and WJets CRs, so that these two backgrounds can be estimated simultaneously, rather than having to compute one of them by subtracting the uncorrected contribution from the other.

A DB: Whether in the CR of AR, or in the CR of VR, we take into account the cross contamination. Since the CRs for WJets and TTbar are orthogonal to each other (see lines: 460-464) we don't see a straightforward way to predict both processes simultaneously and in a Data driven way.

As you can see in table 16, the TTbar contamination in a WJets CR is $\leq 5\%$ and the one of WJets in TTbar CR is $< 9\%$. Even though rather small, we do take into account the cross contamination in the systematics where we attribute a specific uncertainty of 20 % to this effect (651-653).

Q54: Follow up to Q45: while your comment that the method does not always underpredict may be true for 2017, the method does seem to always underestimate the true yield for 2018, and dramatically so for high dM. This would seem to point to a systematic effect, e.g. some potential dependencies of the fake rate that are not accounted for, and it would be good to have some idea of what might be causing this. What MC samples are used for the closure test? As discussed on Friday, it would be nice to see if using the LO W+jets samples you get a better closing.

A DB: Working on this!

Q55: In general the large number of tables (one for each VR/CR/prediction/dM) make it very hard to digest the numbers and inputs to the background predictions. Please consider combining some of them to condense the information so that it is more easily readable (some suggestions below).

A DB: Thank you for the suggestions, I'm looking of ways to compress the tables.

Q56: L240-2: Just to make sure do you correct the signal yields for the filter (in)efficiency?

A DB: I take into account the values of 2016.

Actually, for 2017 and 2018 signal samples, I have a technical question. I want to recompute these efficiencies but I'm having some troubles because the info needed is not available in nanoAOD. I'm looking into Carlos script [script](#), to run it for our signal I need to change line 10. My problem is that I think I have to configure this to run through AAA in condor. The latter part is where I'm having trouble. Marco, do you think you can give me some pointers on how to do so?

Q57: L321-3: Since you use the DeepCSV discriminant shape in the BDT, do you use the shape-based reweighting for the b-tag corrections? From L617-26 it wasn't clear that this was the case.

A DB: Yes, I clarified it now.

Q58: L337-9: Can you provide the exact list of variables used for each dM after this optimization? Or is the list provided in l357-8 valid for all dM?

A DB: We use the same set of discriminant variables for all dM. I've clarified it in the text now.

Q59: Please combine Tables 14-15.

A DB: Good point. Done.

Q60: Section 5.1: Do you have data/MC plots specifically in the CRs somewhere? If so, I missed them.

A DB: I have them but didn't upload them to the AN. I'm adding an extra appendix for those.

Q61: L495-6: Can you please comment on the cases where the data and prediction do not agree well in the VRs? It is very hard to parse through the 48(!) tables in Appendix C.4 to judge this – please find a way of compressing this information into fewer tables, and possibly making some plots to show this information better.

A DB: Whenever the prediction and Data observation disagree in the VR, we take it as an estimator of the systematics of the method, as we should. These systematics are covered in section 6.1.

Q62: Tables 17-32: again, please find a way of combining some of the tables, e.g. you can show the different dMs for each year in the same table.

A DB: I'll do my best.

Q63: L584: What is the motivation for the 50%?

A DB: As far as I understand, this amplitude was suggested by the ARC / SUSY conveners at the time of the publication of 2018.

Now, we've observed in a number of non-SUSY and SUSY analysis that the amplitude of this systematics is 30%. Can we proceed with 30% ?

Q64: Please combine Tables 36-39, 40-43.

A DB: I'm looking on how to improve this.

Q65: L649: Do you use the data-driven prediction for W+jets/ttbar when deriving the uncertainty for the other process? Or just the MC-based prediction?

A DB: When deriving the systematics (eq.3) for one given process X (X=WJets or TTbar), the other contributions are taken from MC. This is clear for rare backgrounds (as there's no other way to estimate them), but also for the fake lepton background, and the cross-contaminating process Y (Y=TTbar or WJets) because when determining one process X from we determine the cross contamination of Y from MC.

Q66: Tables 44-45: It would be useful to add the actual $S_{\{DD\}}$, $\Delta_{\{SDD\}}$ values here too, so we can see if the uncertainty is limited by the non-closure or the statistical uncertainty of the test in each case.

A DB: I agree, I'm updating this for the, soon to be released, version 4.

Q67: L674: What do you mean by the control and signal regions of the VR? Would be good to state that explicitly here (it is all a bit confusing with the different CR/VR/SR definitions).

A DB: CR of the VR: low BDT region (enriched in WJets or TTbar process) of VR. For example, VR2: $200 < MET < 280$ GeV and $LepPt < 30$ for lower dM (see figure 14)

SR of the VR: high BDT region of exactly the same

I'm clarifying this in the updated version of the AN.

Q68: L674-6: Can you please expand a bit on the statement that normalization differences between the two regions should account for a potential shape bias in the background prediction? This was raised at the full status, but I did not fully get the explanation.

A DB: I'm going to add plots in the VRs for the computation of the DD sys as you suggested. That's the best way to see the difference in shape between the CR and SR of the VR and how our methods cover that.

Please note that it's not a normalization, but a ratio. First, we measure the Data/MC ratio in the CR, then in the SR. Let's take an example, table 48, $dM=30$ GeV. In the CR, we see that Data/MC is close to 1, meaning we have an overall good agreement between Data and MC. The same is not observed in the SR, the ratio is 1.30, MC underestimates Data by 30%. This is not an abrupt transition of Data/MC from 0.98 to 1.3 and what it shows up is a trend in the SR, basically, a shape disagreement between CR and SR of the VR. I think after adding these plots, this will become more clear.

So, if there is a problem in the shape of the BDT (for example MC not reproducing the Data, and this Data/MC difference changing vs BDT output value), we can "catch it" by following this difference in the ratio from CR to SR. A change in this ratio vs the BDT output is an estimator of systematic uncertainties due to the shape of the BDT.

Q69: L720-2: Please list explicitly which FastSim corrections/uncertainties are included.

A DB: Added in the newer version.

Q70: Table 55: Why are cross section uncertainties applied to W+jets and ttbar if you use the data-driven predictions for these? Also, only electron ID uncertainties are listed, what about muon ID?

A DB: We have to propagate the uncertainty of the cross-section of rare process in the prediction of WJets and TTbar (eq.1).

It's a typo, it should be "Lepton". Thank you for noticing!

-- DiogoBASTOS - 2021-06-20

This topic: Sandbox > SUS21003Review

Topic revision: r1 - 2021-06-20 - DiogoBASTOS



Copyright &© 2008-2021 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

or Ideas, requests, problems regarding TWiki? use Discourse or Send feedback