Subject: MisID vs Impurity
Posted by Klaus Götzen on Wed, 21 Nov 2012 14:48:34 GMT
View Forum Message <> Reply to Message

Dear all,

since my mic died during the EVO meeting yesterday, I'll try to explain what I wanted to say concerning misID and impurity.

The issue I wanted to point out is, that the impurity is a quantity which is not independent of the fluxes, whereas the misID is. The misID is the fraction of false positive identified particles of a certain species. E.g. the pion-misID of a kaon selector is

misID(pi|K) = #selected pi / #total pi

Obviously this quantity is flux independent, since the flux would go into nominator and denominator. Please note, that there is not information about the selected number of kaons in this quantity, although it is a property of the kaon selector. It's just the probability for another particle species to be accepted by the kaon selector.

On the other hand the impurity is defined as

impurity(pi|K) = 1 - purity = 1 - #sel. K / (#sel. pi + #sel. K) = #sel. pi / (#sel K + #sel pi)

First of all, these two things are different quantities. The other issue is, that a change in relative fluxes would change the impurity (as well as the purity of course), since (with F being the relative pi/K flux factor change in the upper equation)

impurity = F·#sel pi / (#sel. K + F·#sel pi) != #sel. pi / (#sel K + #sel pi)

I think the quality measure of a particular selection algorithm should not depend on the current physical environment (like relative fluxes).

Of course I'm open for discussions in that respect.

Cheers,
Klaus

Subject: Re: MisID vs Impurity
Posted by Malgorzata Gumberidze on Thu, 22 Nov 2012 09:17:47 GMT
View Forum Message <> Reply to Message

Hi,

Unfortunatly i have missed full discussion after i have presented my slides ... In principle what i do i think is
correct, but i simple call it differently. Reading entry of
Klaus i realized, that what i call impurities, he is calling
mis-identification.

Just to be sure I recall what i do. I run simulation for of the
electrons and pions for example and then what i do:

X is some value of the cut on the PID for given particle to be
in this case electron.

electron efficiency:

ele_eff = PID_{e} > X / e_all

and than to study what i call impurity (but probably should be called mis-identification):

pi_imp = PID_{e} > X / pi_all

PID_{e} - i call here probablity of being an electron for given particle. In this case for pion.

greetings,
gosia

Subject: Re: MisID vs Impurity
Posted by Klaus Götzen on Thu, 22 Nov 2012 15:48:21 GMT
View Forum Message <> Reply to Message

Dear Gosia,

thanks for pointing that out! Then your impurity is indeed the same I refer to as 'mis-ID'. And thank you for studying the PID performance in such detail, showing possible bugs in the mechanism!

Why sometimes the mis-ID levels get higher when adding more information I nevertheless don't understand.

I think something is going wrong with the computation and normalization of the PID values.

Best regards,
Klaus

Subject: Re: MisID vs Impurity
Posted by StefanoSpataro on Thu, 22 Nov 2012 16:21:31 GMT

Hi Klaus,
I believe that if you add more detectors, the meaning of the cut in the value we call "probability" has a different meaning, then cutting P>90% for EMC is different from P>90% in EMC+STT. If you compare the plots, once you add more detectors your efficiency with the same cut increases -> this means that you have more signal candidates in your selection -> your selection is looser, the misId increases.
In order to crosscheck really the numbers, one should select two algorithms cutting a value which provides the same efficiency. In such case one can compare purity and understand what is good and what is not good. The ROC curves could be helpful,maybe for the moment just integrated in separated momentum ranges.

Subject: Re: MisID vs Impurity
Posted by donghee on Fri, 23 Nov 2012 16:29:10 GMT

Dear Gosia,

First of all, thank you for your great effort.
You showed to me the good guide line for PID study.

But as usual, I have still few question to better understand your definition of impurity.
pi_imp = PID_{e} > X / pi_all
I can understand about the nominator, that is reconstructed true pion, which is tested MC true PID matching.

I'm wondering about the PID_{e} > X.
Is it "selected true pion after doing MC PID match and requiring PID probability"
or
"any kind of tracks just passed given probability condition"?
I am not clear for this which one have to be applied to see whatever impurity or misID.

concerning single and more tracks
You have selected only one reconstructed particle, which has a closest momentum value to MC one.
We can have usually more than one track after reconstruction even from one event.
That means, all values of impurity shown in your categories are too ideal in some point. Could you tell me the number, how much % of such event can we have from tracking?

Thank you for your teaching.
Donghee

Subject: Re: MisID vs Impurity

Posted by donghee on Fri, 23 Nov 2012 16:59:54 GMT
View Forum Message <> Reply to Message

Dear all,

I have a idea about the definition of impurity and misID.
before getting the answer from Gosia

Gorsia showed us...
pi_imp = PID_{e} > X / pi_all
PID_{e} - it calls here probablity of being an electron for given particle. In this case for pion.

If we define PID_{e} > X as a
"selected true pion after doing MC PID match and requiring PID probability", then it refer to impurity.
or
"any kind of tracks passed required PID probability", then this quantity should be misID.

Above one is impurity as 1-purity and tell us how much % of wrong type particles are contributed in given PID selection.
Below quantity can have a meaning, how much % of a particle type can participate to other list of track candidate according given particle type.

Could we think two definition with this way?
Have a nice weekend,

Donghee

Subject: Re: MisID vs Impurity
Posted by Malgorzata Gumberidze on Sat, 24 Nov 2012 06:04:39 GMT
View Forum Message <> Reply to Message

Hello

I'm wondering about the PID_{e} > X.
Is it "selected true pion after doing MC PID match and requiring PID probability"

Yes this is exactly what am I doing. I ask, that the particles is really pion by cross-checking MC id and than i apply to the primary PION cut on PID to be an electron.

You have selected only one reconstructed particle, which has a closest momentum value to MC one.
We can have usually more than one track after reconstruction even from one event.
That means, all values of impurity shown in your categories are too ideal in some point.

Could you tell me the number, how much % of such event can we have from tracking?

Of course we have more than one particle per event, but ideally you will try later in the experimental analysis also select one particle per event, assuming that you want to analyze one particle per event. If you do not do such selection, than in case of efficiency you will get more than 1.

For example in our case pbarp->e+e- in the old framework we were doing all combination of +- pair in the event, and than we were selecting to have only one
per event. The best one, looking to CHi2 from kinematical fit for example.

So I THINK that in this case it is fine what i do.

gosia

Subject: Re: MisID vs Impurity
Posted by Malgorzata Gumberidze on Sat, 24 Nov 2012 20:29:36 GMT
View Forum Message <> Reply to Message

hello

please find here plot which is showing multiplicity of reconstructed tracks for one true MC electron track :

and here are some numbers (number of primary electrons=972500):

mult ==0 : 83628
mult ==1 : 863344
mult ==2 : 24060
mult ==3 : 1116
mult ==4 : 239
mult ==5 : 64

and here is a plot for negative pions

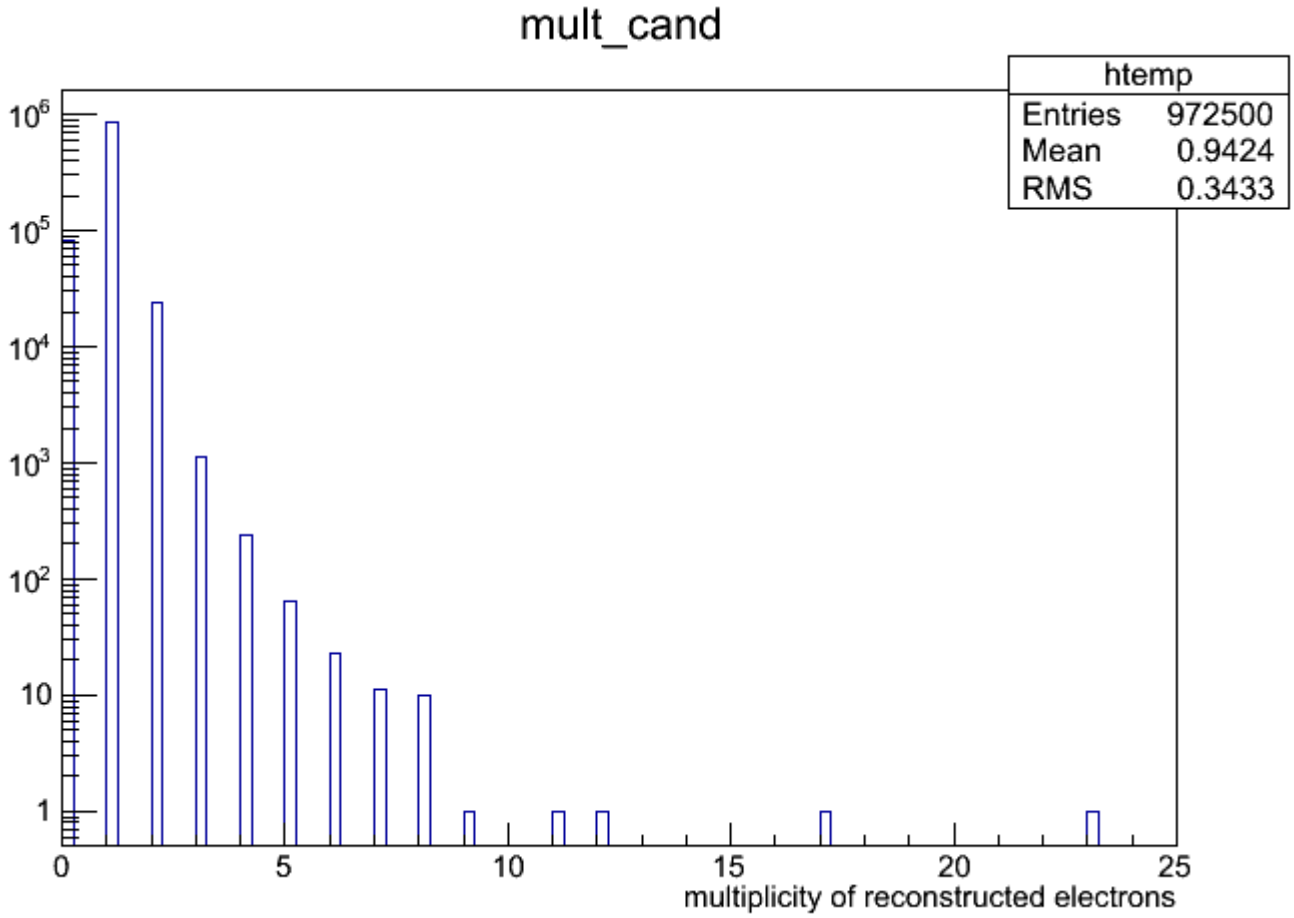
and numbers (number of primary pions = 986500):

mult ==0 : 110975
mult ==1 : 854022
mult ==2 : 20824
mult ==3 : 550
mult ==4 : 90
mult ==5 : 29
mult ==6 : 5

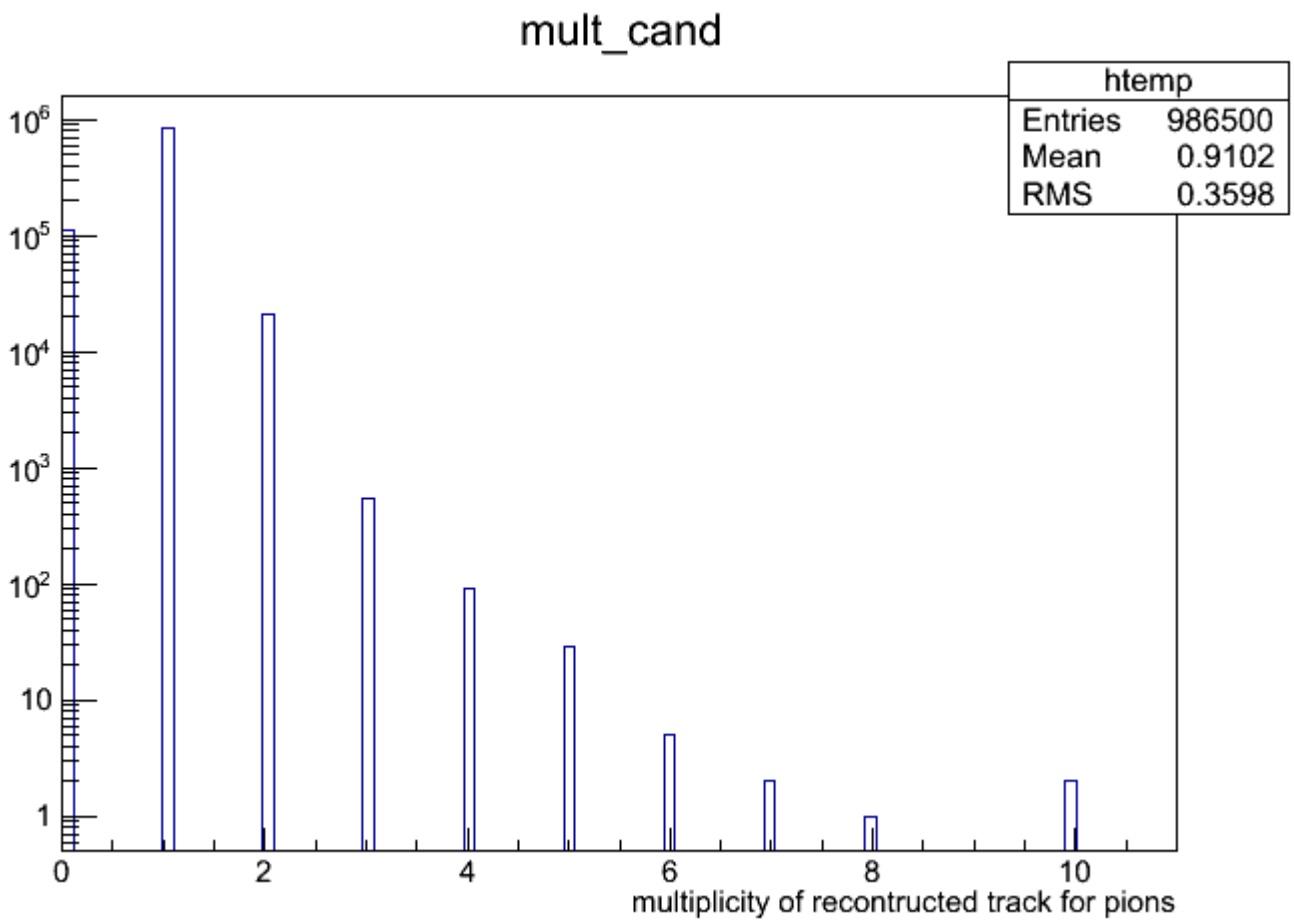
3% of the events has more than 1 particles reconstructed per MC track in both cases for electrons and pions. So in this way i would say that selection of one good track per MC true one does not introduce big bias.

File Attachments

1) mult_ele.gif, downloaded 1168 times



2) mult_pions.gif, downloaded 1064 times



Subject: Re: MisID vs Impurity
Posted by Malgorzata Gumberidze on Sat, 24 Nov 2012 20:37:10 GMT
View Forum Message <> Reply to Message

few more explanations:

But as usual, I have still few question to better understand your definition of impurity.
pi_imp = PID_{e} > X / pi_all
I can understand about the nominator, that is reconstructed true pion, which is tested MC true PID matching.

I'm wondering about the PID_{e} > X.
Is it "selected true pion after doing MC PID match and requiring PID probability"
or
"any kind of tracks just passed given probability condition"?
I am not clear for this which one have to be applied to see whatever impurity or misID.

in both case i select MC true pions.
pi_all - are all TRUE MC pions without any condition on PID
PID_{e} - are TRUE MC pions with condition on PID of being electron

i hope that soon we will come to the common point with definitions

---

Subject: Re: MisID vs Impurity
Posted by Ronald Kunne on Sun, 25 Nov 2012 09:07:56 GMT
View Forum Message <> Reply to Message

A physicist in Kuala Lumpur
said that her sample was rather impure:
"There's no denyin'
I mis-ID'd that pion."
Klaus Goetzen is right, for sure!
