

An Improved Way to Propagate π^+ Production Uncertainties MiniBooNE Tech Note 257

S.J.Brice

October 7, 2008

1 The Original Way π^+ Production Uncertainties Were Propagated

For the oscillation result the uncertainties associated with π^+ production by 8 GeV protons were propagated through to an error matrix in reconstructed neutrino energy in the following way:-

1. Using the results from the HARP and E910 experiments [1, 2] the two dimensional angular and momentum distribution of π^+ produced by protons incident on Be is fit to the Sanford Wang functional form. The result is a set of 8 best fit Sanford Wang parameters and the 8x8 error matrix that goes with them [3]. The Sanford Wang function is

$$SW(p, \theta; \mathbf{c}) \equiv c_1 p^{c_2} \left(1 - \frac{p}{p_p - 1}\right) \exp\left(-c_3 \frac{p^{c_4}}{p_p^{c_5}} - c_6 \theta (p - c_7 p_p \cos^{c_8} \theta)\right) \quad (1)$$

2. In the MultisimMatrix analysis framework package the 8 best fit SW parameters and their error matrix are taken to describe an 8 dimensional Gaussian probability distribution for the SW parameters. This distribution is drawn from 1000 times to form 1000 SW parameter sets, each, presumably, allowed by the HARP/E910 data.
3. Define \mathbf{c}^{CV} to be the SW parameter set used for the Beam Monte Carlo simulation, and \mathbf{c}^α to be the 1000 alternative SW parameter sets drawn in the previous step. One can then use the ratio

$$\frac{SW(p, \theta; \mathbf{c}^\alpha)}{SW(p, \theta; \mathbf{c}^{CV})} \quad (2)$$

to reweight each event of a dataset 1000 times. In this way 1000 alternative datasets can be mocked up that span the range allowed by the Sanford Wang function and it's uncertainty.

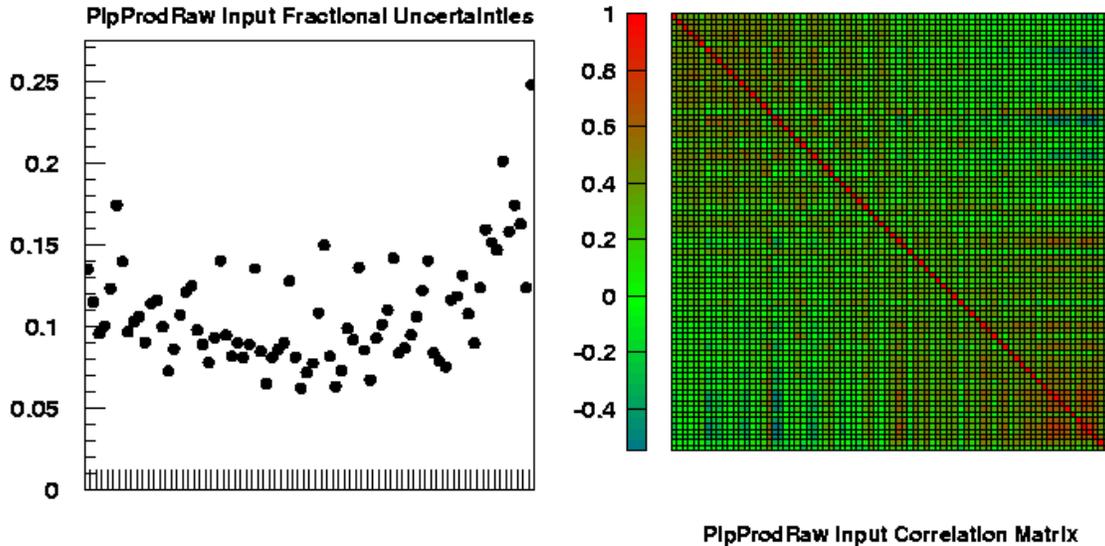


Figure 1: The HARP π^+ production error matrix. The fractional errors on the $p - \theta$ bins are shown on the left and the correlation matrix for those bins is shown on the right. For further details see Sec. 2.1

4. To form an error matrix for the EnuQE distribution of the dataset form the EnuQE histogram for each of the 1000 alternative datasets. These histograms are then combined to form the error matrix for the EnuQE distribution of the dataset. If the CV EnuQE histogram is labelled n_i^{CV} and the 1000 alternative histograms n_i^α (with i running over the bins) the expression to form the error matrix is

$$M_{ij} = \frac{1}{1000} \sum_{\alpha=1}^{1000} (n_i^\alpha - n_i^{CV}) (n_j^\alpha - n_j^{CV}) \quad (3)$$

2 The Original Method in Practice

2.1 The HARP Error Matrix

Fig. 1 shows a graphical representation of the HARP error matrix. The harp results are provided in 6 π^+ angle bins and 13 π^+ momentum bins. This is $6 \times 13 = 78$ numbers and a 78x78 error matrix to go with them. The left hand plot of Fig. 1 shows the 78 fractional errors on the harp bins. The first six bins are the uncertainties on the six angles of the lowest momentum, the next six bins are the errors in the six angles of the next lowest momentum bin, and so on. One can see that in the mid momentum bins that tend to create neutrinos that MiniBooNE sees, the errors on the harp results are roughly 7%. On the right hand side of Fig. 1 is a graphical representation of the 78x78 correlation matrix where the bin ordering is the same as the left hand side plot. One can see that the plot is mostly green, indicating errors that are largely uncorrelated, with some red creeping in at low and high π^+ momenta.

2.2 The Sanford-Wang Fit

The method for propagating the π^+ outlined in Sec. 1 works in principle, a perfectly valid approach, but it does rely on the Sanford Wang function fitting the HARP data very well. Unfortunately, the SW function is a tolerable, but not particularly good fit to the HARP data. This can be seen in Fig.11 of [3]. As a result of the fit quality the fit is adjusted to ensure a reasonable χ^2 . The details are given in [3], but the net effect is an 8x8 SW parameter error matrix where there is a $\sim 18\%$ error on c_1 , the first SW parameter. This translates into a $\sim 18\%$ normalization error, since c_1 just multiplies the whole SW function (see Eqn. 1). Therefore the $\sim 7\%$ uncorrelated errors of the HARP error matrix have been turned into $\sim 18\%$ fully, correlated normalization errors by the fitting procedure.

2.3 Propagating the SW fit to an Error Matrix in $E_{\nu QE}$

Following the last two steps of Sec. 1 the Sanford-Wang fit parameters error matrix of the previous section is propagated into an error matrix in reconstructed neutrino energy, $E_{\nu QE}$. Fig. 2 shows a graphical representation of this $E_{\nu QE}$ matrix using the same conventions as Fig. 1. In this figure there are 53 bins represented, the first 18 are bins of $E_{\nu QE}$ for the track based (TBA) cuts applied to a fully oscillated sample, the second 18 are bins of $E_{\nu QE}$ for the track based (TBA) electron cuts applied to a cocktail sample, and the final 17 bins are bins of $E_{\nu QE}$ for the track based ν_μ CCQE cuts applied to a cocktail sample. The middle sample is a mix of event types that are not subject to π^+ production errors (e.g. π^0 s) and those that are and so should be disregarded when looking to see the effect of π^+ production errors. The first sample (first 18 bins) are high purity ν_e CCQE events and the third sample is high purity ν_μ CCQE or CC π^+ events. These two samples show that the π^+ production errors have been turned into a $\sim 16\%$ almost, pure normalization error.

By using the Sanford-Wang fit the initially $\sim 7\%$, largely uncorrelated errors of the HARP error matrix have been turned into a $\sim 16\%$, almost completely normalization error matrix in $E_{\nu QE}$. This is not a faithful transmission of the information in the HARP error matrix.

3 Improving the Propagation of the π^+ Production Uncertainties

How can the SW fitting distortion of the π^+ errors be avoided? In the past parameterizations of π^+ production data were needed as the data was too sparse or not at the correct proton beam energy. This is no longer true with the high quality HARP data that was taken at exactly the MiniBooNE proton beam energy. One can therefore consider throwing away the step of parameterising the HARP data and replace it with a simple spline interpolation. With that in mind the HARP results can be propagated through to an error matrix in reconstructed neutrino energy in the following way:-

1. Draw from the 78x78 HARP error matrix 1000 times and add to the HARP

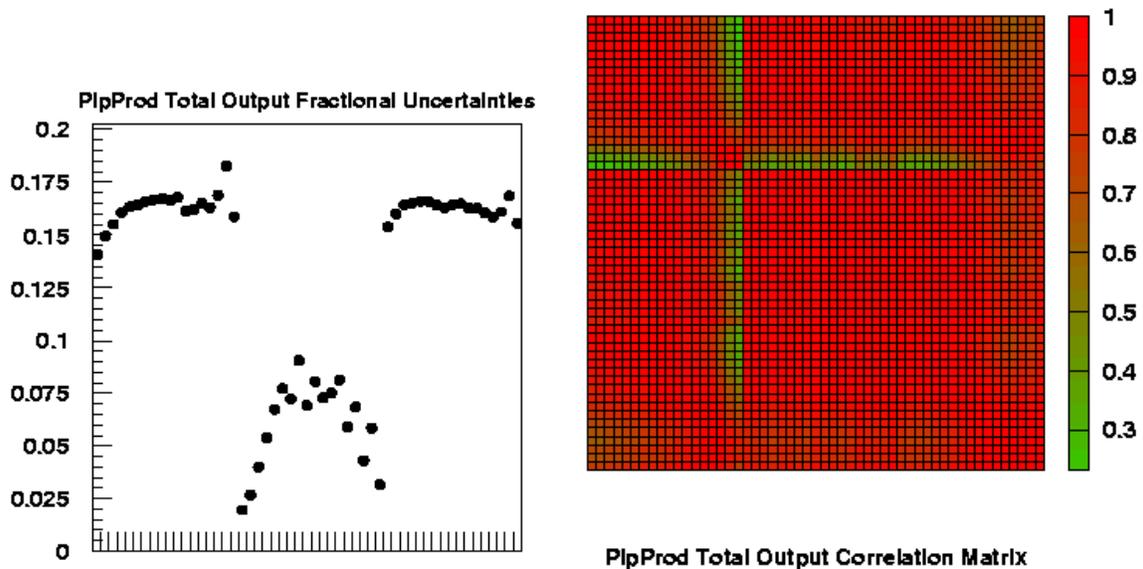


Figure 2: The π^+ production uncertainties propagated via a Sanford Wang fit into errors on the $E_{\nu Qe}$ distribution of fully oscillated events, nue candidate events, and numu CCQE candidate events. The left hand plot shows the fractional error in each $E_{\nu QE}$ bin and the right hand side shows the correlation matrix. See Sec. 2.3 for details.

CV to create 1000 alternative HARP results each allowed by the HARP error matrix

2. Use splines to interpolate (and extrapolate) each of the 1000 alternate HARP results. One can define $SP(p, \theta; \mathbf{k}^\alpha)$ as the spline function in the π^+ momentum p and angle θ controlled by the spline parameters \mathbf{k}^α for multisim α . One can then use the ratio

$$\frac{SP(p, \theta; \mathbf{k}^\alpha)}{SW(p, \theta; \mathbf{c}^{CV})} \quad (4)$$

to reweight each event of a dataset 1000 times. Note that the CV Sanford-Wang function is in the denominator as this was the function used to generate the events. In this way 1000 alternative datasets can be mocked up that span the range allowed by the HARP result and it's uncertainty. Since the Sanford Wang function matches the HARP data fairly well one could also consider defining $SP(p, \theta; \mathbf{k}^{CV})$ to be the spline fit to the HARP CV and reweighting the events with the ratio

$$\frac{SP(p, \theta; \mathbf{k}^\alpha)}{SP(p, \theta; \mathbf{k}^{CV})} \quad (5)$$

Later sections will explore these two alternate weighting schemes.

3. To form an error matrix for the EnuQE distribution of the dataset form the EnuQE histogram for each of the 1000 alternative datasets. These histograms are then combined to form the error matrix for the EnuQE distribution of the dataset. If the CV EnuQE histogram is labelled n_i^{CV} and the 1000 alternative

histograms n_i^α (with i running over the bins) the expression to form the error matrix is

$$M_{ij} = \frac{1}{1000} \sum_{\alpha=1}^{1000} (n_i^\alpha - n_i^{CV}) (n_j^\alpha - n_j^{CV}) \quad (6)$$

4 The Improved Method in Practice

4.1 The Splines

As described in the previous section, the HARP 78x78 error matrix is drawn from 1000 times to produce 1000 alternate HARP results. A spline interpolation scheme is then used to interpolate (and extrapolate when needed) the fixed locations in p and θ of the HARP results. The spline fitting is to the 2D cross-section as a function of p and θ and is done by using the 1D CERNLIB spline routine DCSPLN. First 1D spline fits are done in θ at values of p close to the desired point, then a final 1D spline fit in p is done using the results of the θ splines.

Figures 3 and 5 show these splines overlaid with the HARP data and the SW fit. Fig. 3 has six panels, one for each of the HARP angle bins and each panel gives the π^+ momentum distribution for that angle bin. The red data points are the HARP results with their uncertainties, the blue curve is the SW fit and the spline multisims (actually only the first 40) are plotted in black. Fig. 5 is identical to Fig. 5 but shows the angular distributions in 13 panels of π^+ momentum. Figures 4 and 6 are identical to Figures 3 and 5, but a profile histogram of the spline curves is plotted instead of the splines themselves.

It is clear from Figures 4 and 6 that where there is HARP data the spread in the spline curves nicely matches the HARP error bars. Where there is no HARP data, particularly at π^+ momenta below 1 GeV the splines diverge from the SW fit. It should be noted that extremely few neutrinos detected by MiniBooNE come from π^+ with momenta below 1 GeV. This is shown in Figures 7 and 8 which show the parent meson momentum of the neutrinos passing nue and numu CCQE cuts respectively. The only place where the spline curves can get a little crazy and create unreasonable large weights for events is at large angles. For this reason events that have pion angles greater than the largest HARP pion angle bin are given a weight as if the pion angle was in the middle of the largest pion angle bin. This keeps the weights on these handful of events reasonable, and prevents them creating error matrices that erroneously explode at large pion angle.

4.2 Propagating the Spline Multisims to an Error Matrix in

$$E_{\nu QE}$$

As noted in Sec. 3, two different reweighting schemes have been used to turn the 1000 spline interpolated alternative HARP results into 1000 $E_{\nu QE}$ histograms from which an $E_{\nu QE}$ error matrix can be formed. The strictly correct scheme weights events with the ratio $\frac{SP(p,\theta;\mathbf{k}^\alpha)}{SW(p,\theta;\mathbf{c}^{CV})}$, whereas the approximate scheme uses $\frac{SP(p,\theta;\mathbf{k}^\alpha)}{SP(p,\theta;\mathbf{k}^{CV})}$, using the notation developed in Sec. 3. The approximate scheme makes the assumption that

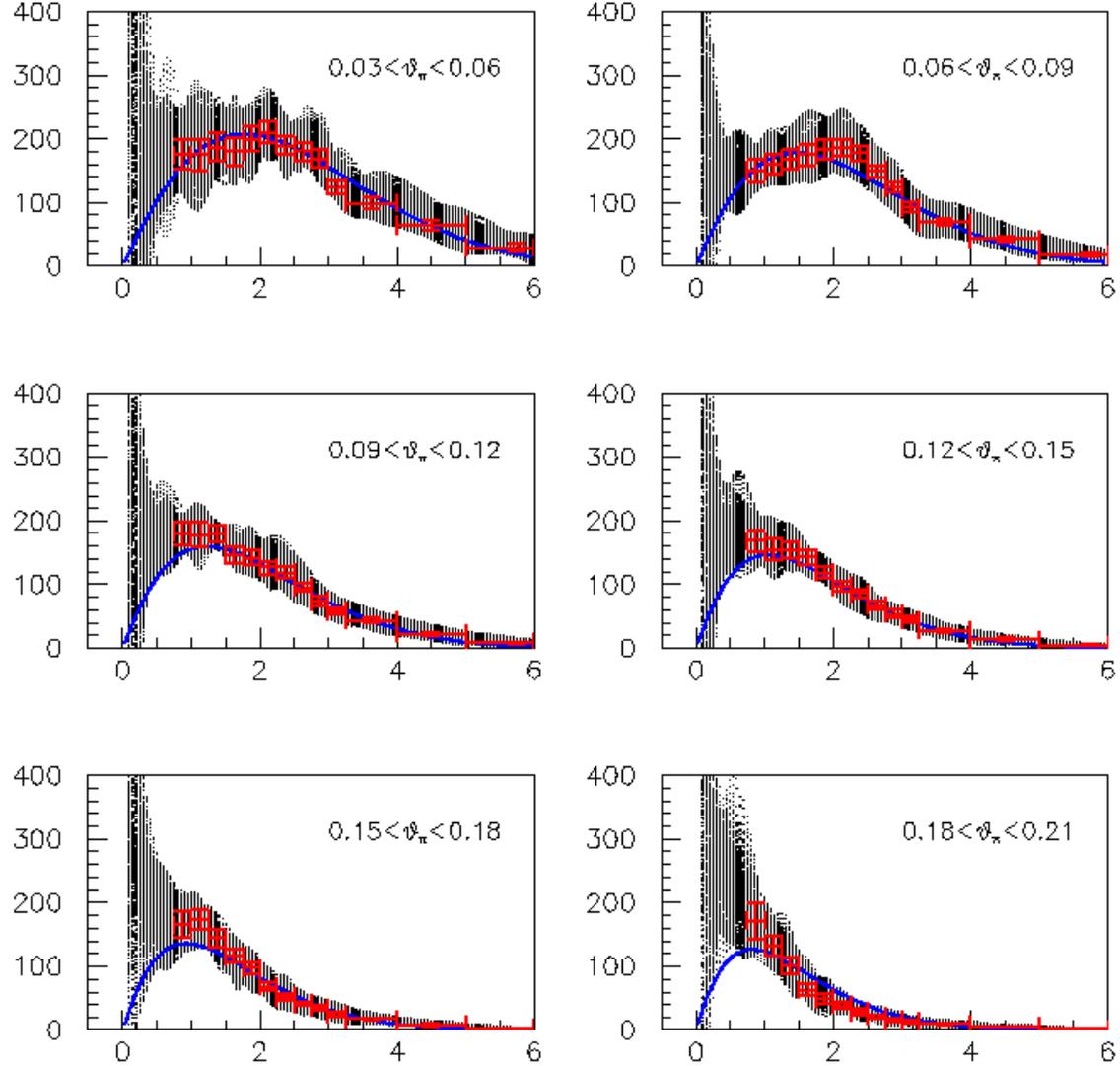


Figure 3: π^+ production cross-section as a function of π^+ momentum from 0 to 6 GeV/c. The six panels are for the six π^+ angular ranges indicated on the plots. The red points are the HARP results and uncertainties, the blue curve is the Sandford-Wang fit and the black points are the first 40 spline interpolated multisims.

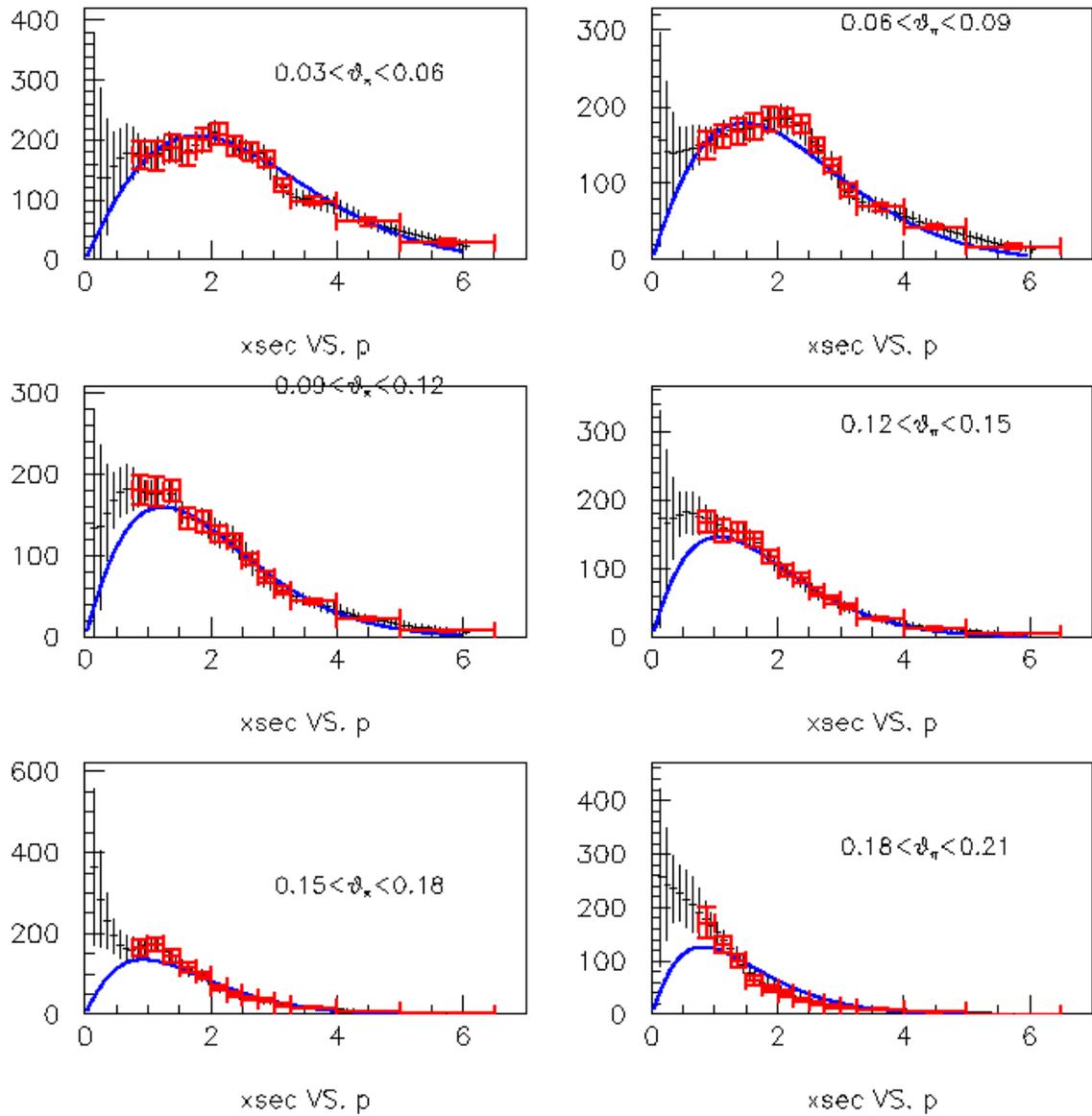


Figure 4: Identical to Fig. 3, but the 40 spline multisims have been replaced by a profile histogram of all 1000 spline multisims

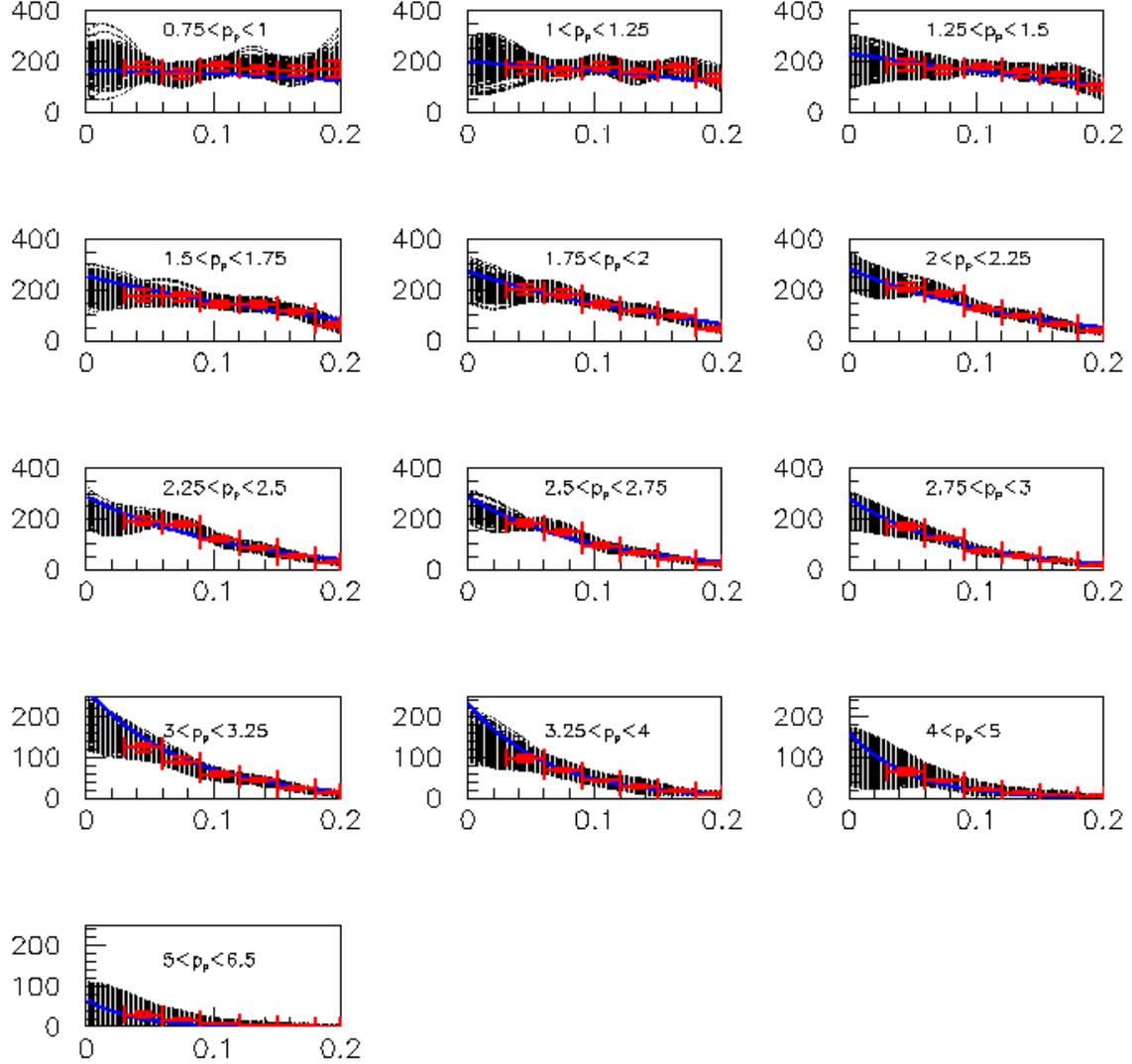


Figure 5: π^+ production cross-section as a function of π^+ angle from 0 to 0.25 rad. The thirteen panels are for the thirteen π^+ momentum ranges indicated on the plots. The red points are the HARP results and uncertainties, the blue curve is the Sandford-Wang fit and the black points are the first 40 spline interpolated multisims.

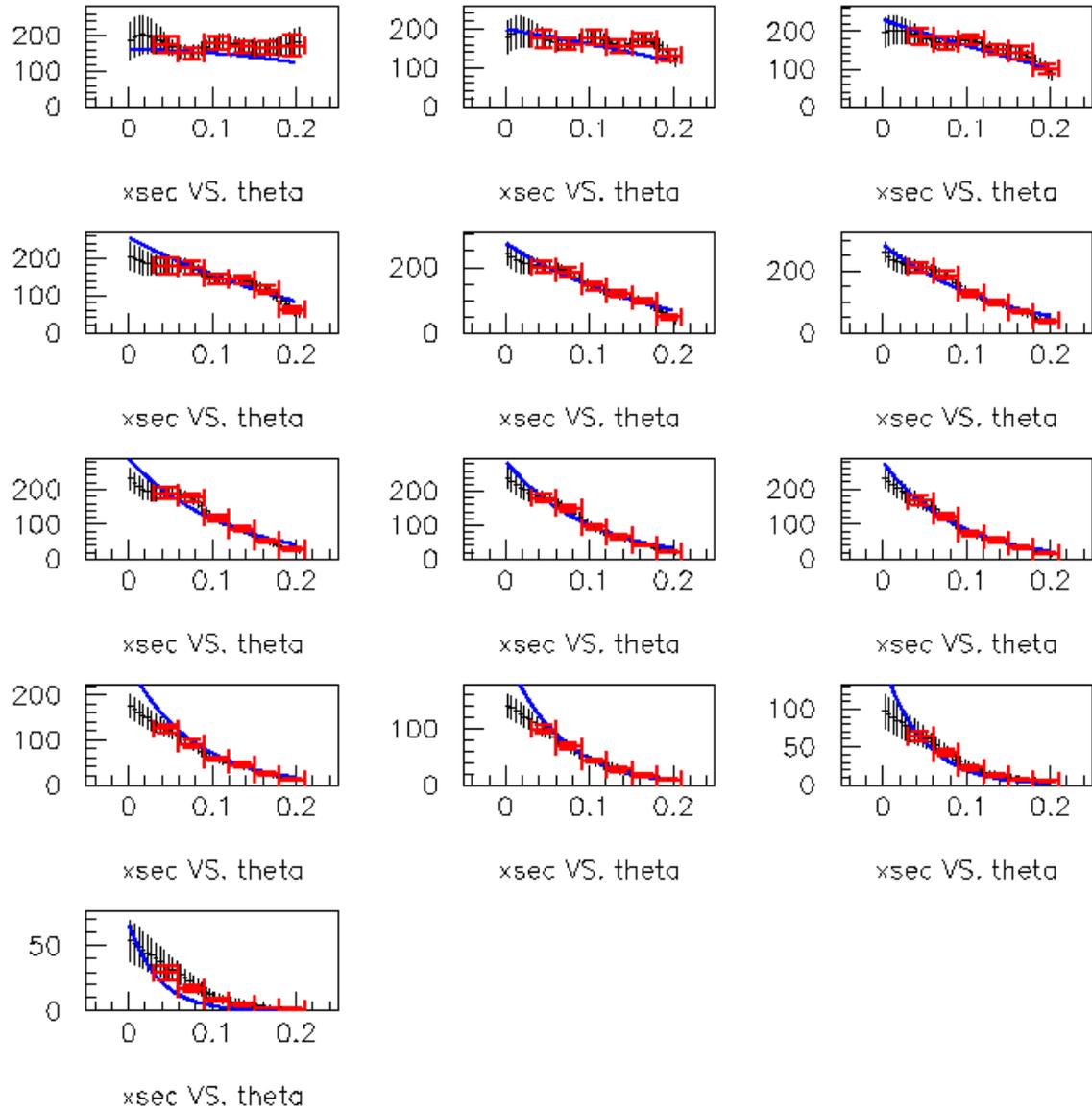


Figure 6: Identical to Fig. 5, but the 40 spline multisims have been replaced by a profile histogram of all 1000 spline multisims

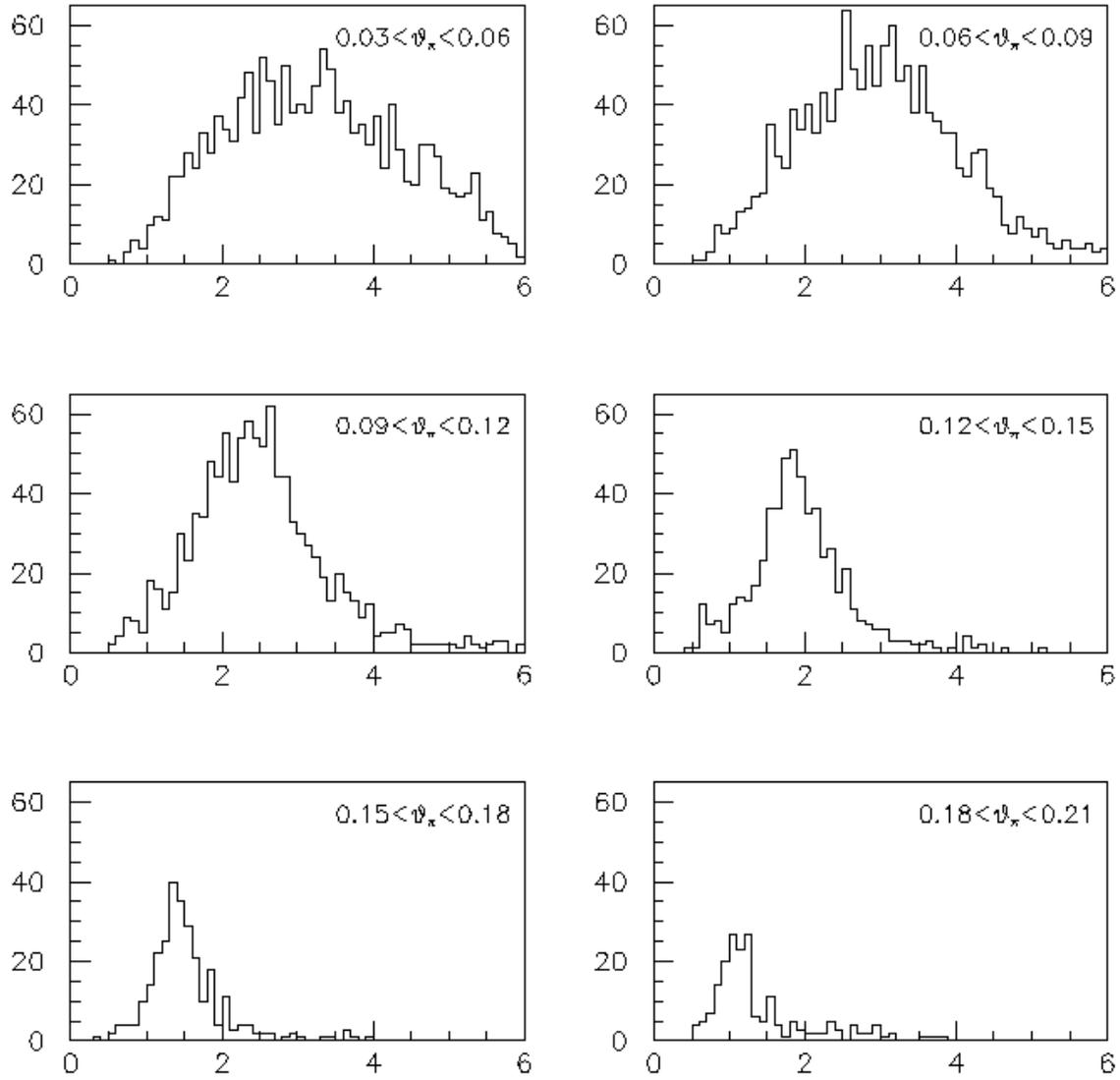


Figure 7: The momentum distribution of the parent meson of neutrinos passing the TBA nue cuts. The six panels are for the six meson angle ranges indicated. The normalization is arbitrary

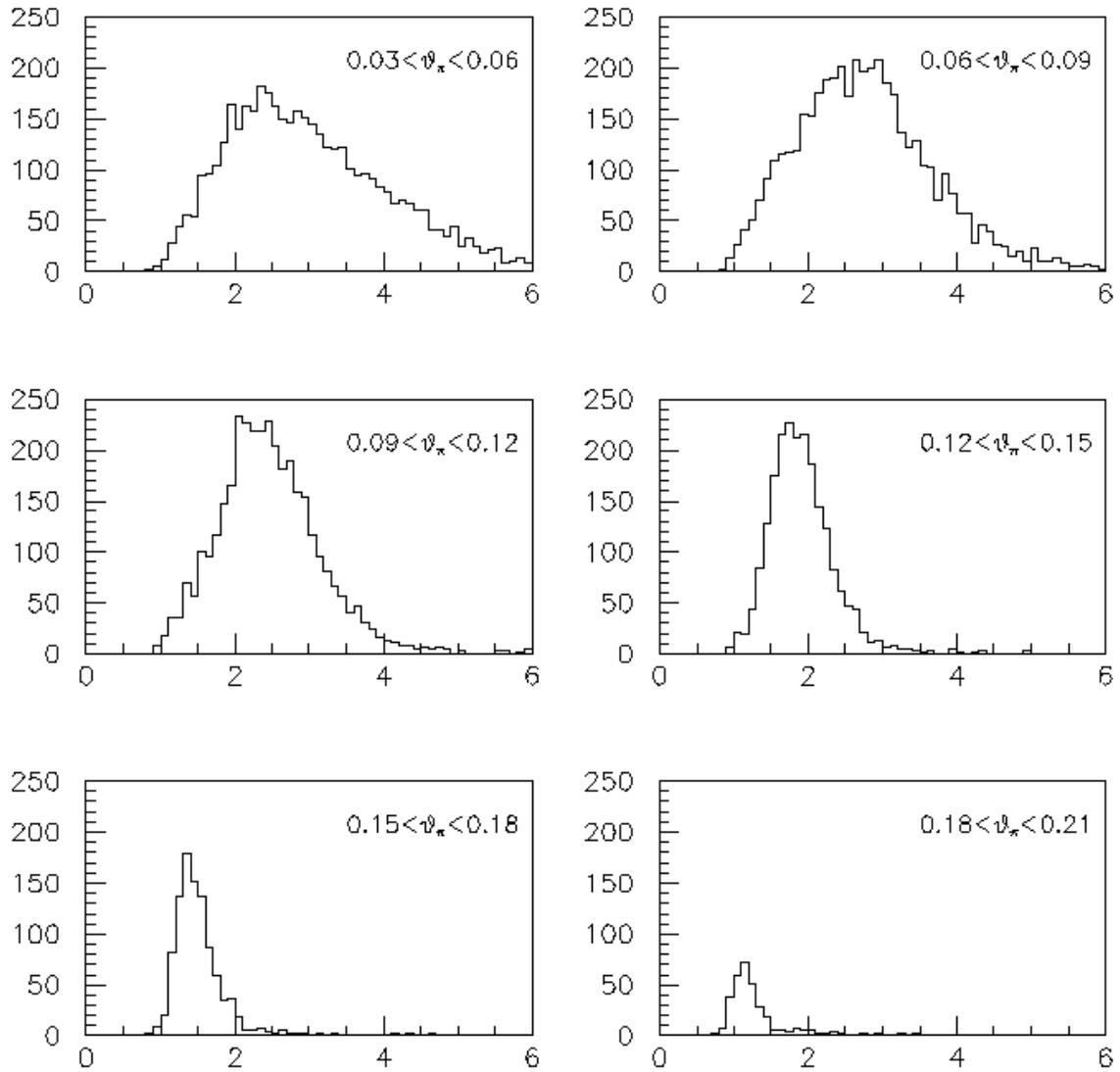


Figure 8: The momentum distribution of the parent meson of neutrinos passing the Likelihood based numu CCQE cuts. The six panels are for the six meson angle ranges indicated. The normalization is arbitrary.

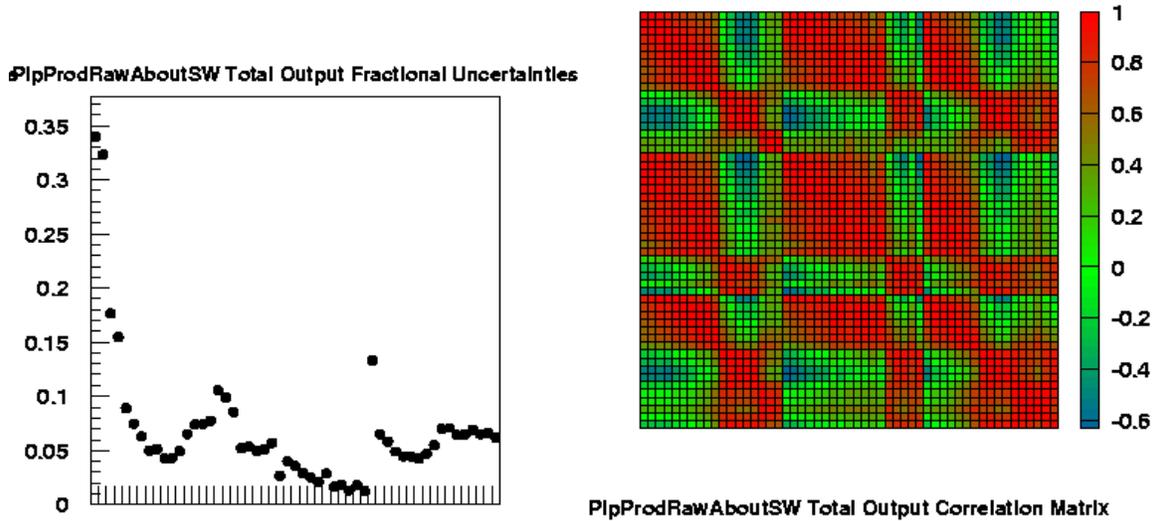


Figure 9: The π^+ production uncertainties propagated via spline interpolation and the strictly correct reweighting scheme into errors on the $E_{\nu Qe}$ distribution of fully oscillated events, nue candidate events, and numu CCQE candidate events. The left hand plot shows the fractional error in each $E_{\nu QE}$ bin and the right hand side shows the correlation matrix. See Sec. 4.2 for details.

the SW curve is very close to the HARP data points. It exists largely as an historical accident:- the π^+ propagation uncertainties for the low energy electron candidate excess paper uses the approximate scheme. It will be shown in a later section that the two schemes yield almost exactly the same constrained uncertainties.

Fig. 9 shows the fractional errors and correlation matrix in $E_{\nu QE}$ that results when the exact scheme is used and Fig. 10 shows the same for the approximate scheme. The binning and datasets used in these figures are identical to earlier figures of this type in this document.

The correct and approximate weighting schemes have $E_{\nu QE}$ errors that are very similar except at the very lowest couple of energy bins where the correct scheme has noticeable larger errors. The correct scheme also has less correlated errors, as evidenced by the amount of green in the two correlation matrix plots on the right hand side of Figures 9 and 10.

It is clear that both correct or approximate weighting schemes produce final $E_{\nu QE}$ error matrices that are less correlated and have much smaller errors than the matrix produced by propagating SW errors. They seem to do a much more accurate job of propagating the HARP error matrix.

4.3 The Effect on Low E Constrained Errors

Table 1 shows the effect of the different ways of propagating the π^+ production uncertainties on the nue candidates. For the three standard energy bins and for the TBA se;ection with the full 6.462×10^{20} POT the unconstrained and constrained number of predicted background events is show for the full set of uncertainties and the π^+ production uncertainties propagated either via a Sanford Wang fit, via the

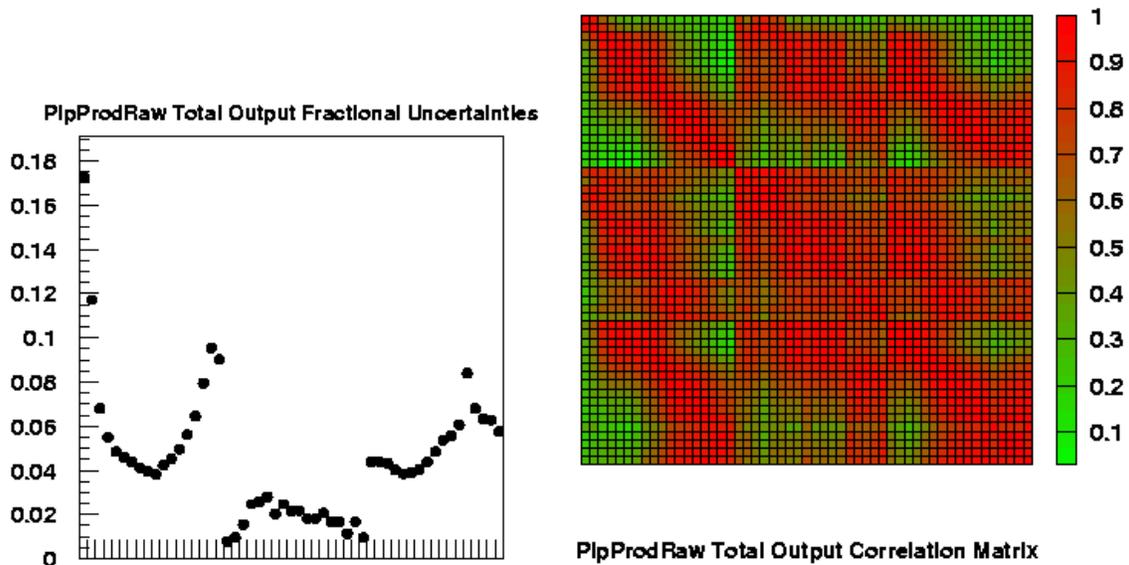


Figure 10: The π^+ production uncertainties propagated via spline interpolation and the approximate reweighting scheme into errors on the $E_{\nu Qe}$ distribution of fully oscillated events, nue candidate events, and numu CCQE candidate events. The left hand plot shows the fractional error in each $E_{\nu QE}$ bin and the right hand side shows the correlation matrix. See Sec. 4.2 for details.

correct Spline approach, or via the approximate spline approach described in the previous section. Table 2 shows exactly the same thing, but for the TBA+Dirtcut event selection. In all cases the method of propagating the π^+ errors doesn't make much difference to the final answer. Whilst analyses should, in general, use the correct spline method it is a tolerable approximation for the low E analysis to use the approximate spline method.

A π^- Splines

Results on π^- have also been produced by HARP and can be propagated into error matrices in exactly the same way as the π^+ results. Figs 11 to 16 are the π^- versions of Figs 3 to 8. It is clear that, as expected, the π^- splines have similar deviations from their CV as the π^+ splines do. Finally, Fig. 17 shows the π^- production $E_{\nu QE}$ error matrix. Unsurprisingly it is very similar to Fig. 9, the π^+ version.

References

- [1] D. Schmitz, PhD Thesis, 2008
- [2] The E910 paper
- [3] MiniBooNE Technote 204, "Determination of the π^+ Production Cross Section from External Data", M. Tzanov, 9/20/06, revised 11/16/06

Energy Bin	Data	Bkgd Pred.	SW Fit	Spline	Approx. Spline
200-300 MeV	427	unconstr.	380.5±48.3	380.5±54.5	380.5±47.6
		constr.	384.4±44.5	388.9±45.9	386.0±44.3
		signif.	0.96 σ	0.83 σ	0.93 σ
300-475 MeV	428	unconstr.	330.6±42.9	330.6±43.4	330.6±40.6
		constr.	328.4±31.9	330.2±32.1	330.0±31.8
		signif.	3.12 σ	3.05 σ	3.08 σ
475-1250 MeV	431	unconstr.	418.9±70.3	418.9±64.4	418.9±63.4
		constr.	410.7±39.2	412.0±37.8	412.7±37.6
		signif.	0.52 σ	0.50 σ	0.49 σ

Table 1: The effect of different ways of propagating π^+ production uncertainties into the predicted number of nue candidates passing TBA cuts and the uncertainty in that number. Numbers are given both the case where the numu constraint is applied and where it is not. The significance of the excesses is also given.

Energy Bin	Data	Bkgd Pred.	SW Fit	Spline	Approx. Spline
200-300 MeV	232	unconstr.	181.1±30.0	181.1±32.7	181.1±29.5
		constr.	185.5±26.1	188.5±26.6	186.8±26.0
		signif.	1.78 σ	1.64 σ	1.74 σ
300-475 MeV	312	unconstr.	228.8±33.0	228.8±33.0	228.8±31.3
		constr.	226.8±24.6	229.0±24.8	228.3±24.5
		signif.	3.46 σ	3.35 σ	3.42 σ
475-1250 MeV	408	unconstr.	391.4±67.0	391.4±61.5	391.4±60.6
		constr.	384.1±37.4	385.6±36.0	385.9±35.7
		signif.	0.64 σ	0.62 σ	0.62 σ

Table 2: The effect of different ways of propagating π^+ production uncertainties into the predicted number of nue candidates passing TBA+dirtcut cuts and the uncertainty in that number. Numbers are given both the case where the numu constraint is applied and where it is not. The significance of the excesses is also given.

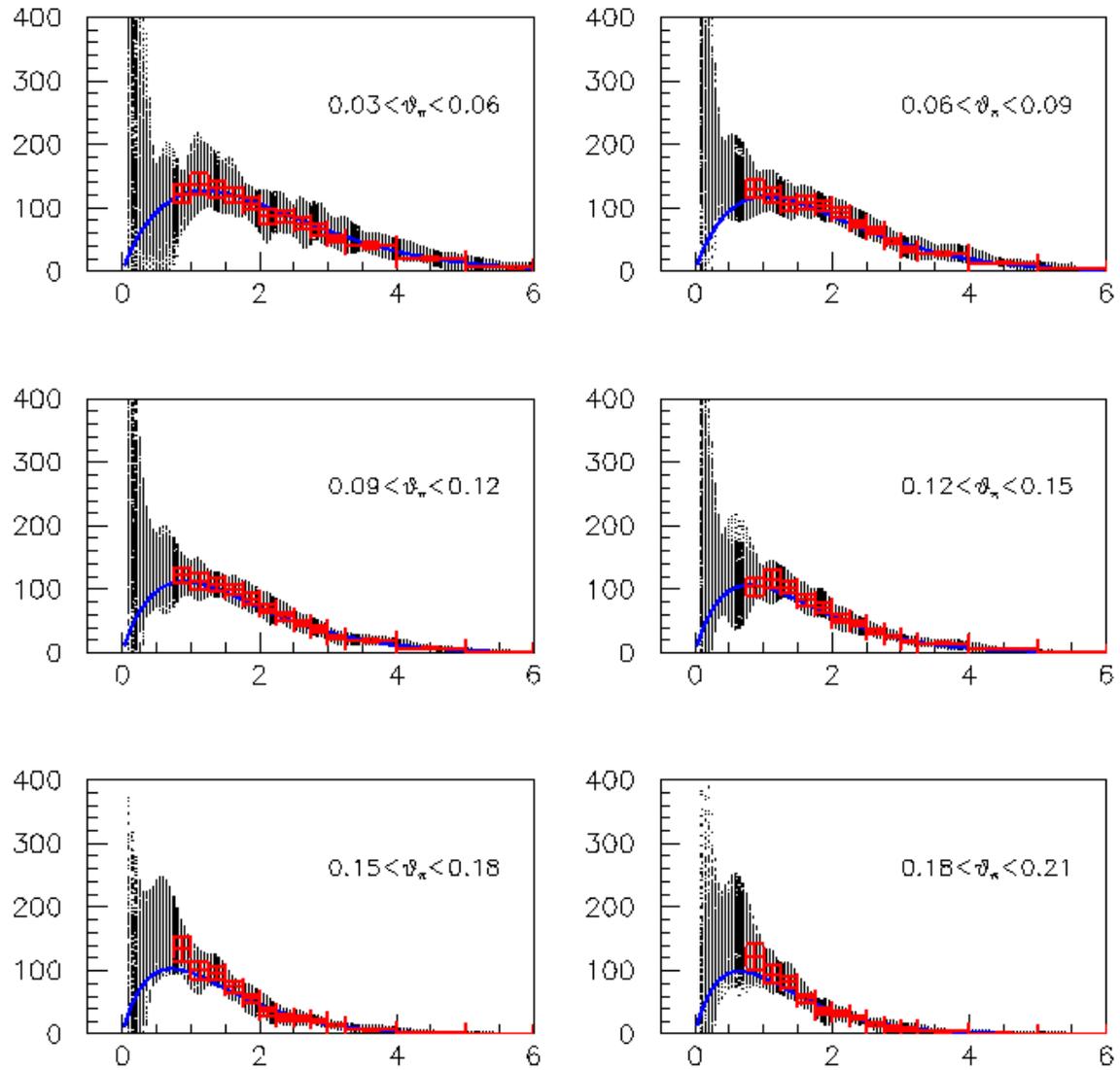


Figure 11: The π^- version of Fig. 3. The π^- production cross-section as a function of π^- momentum from 0 to 6 GeV/c. The six panels are for the six π^- angular ranges indicated on the plots. The red points are the HARP results and uncertainties, the blue curve is the Sandford-Wang fit and the black points are the first 40 spline interpolated multisims.

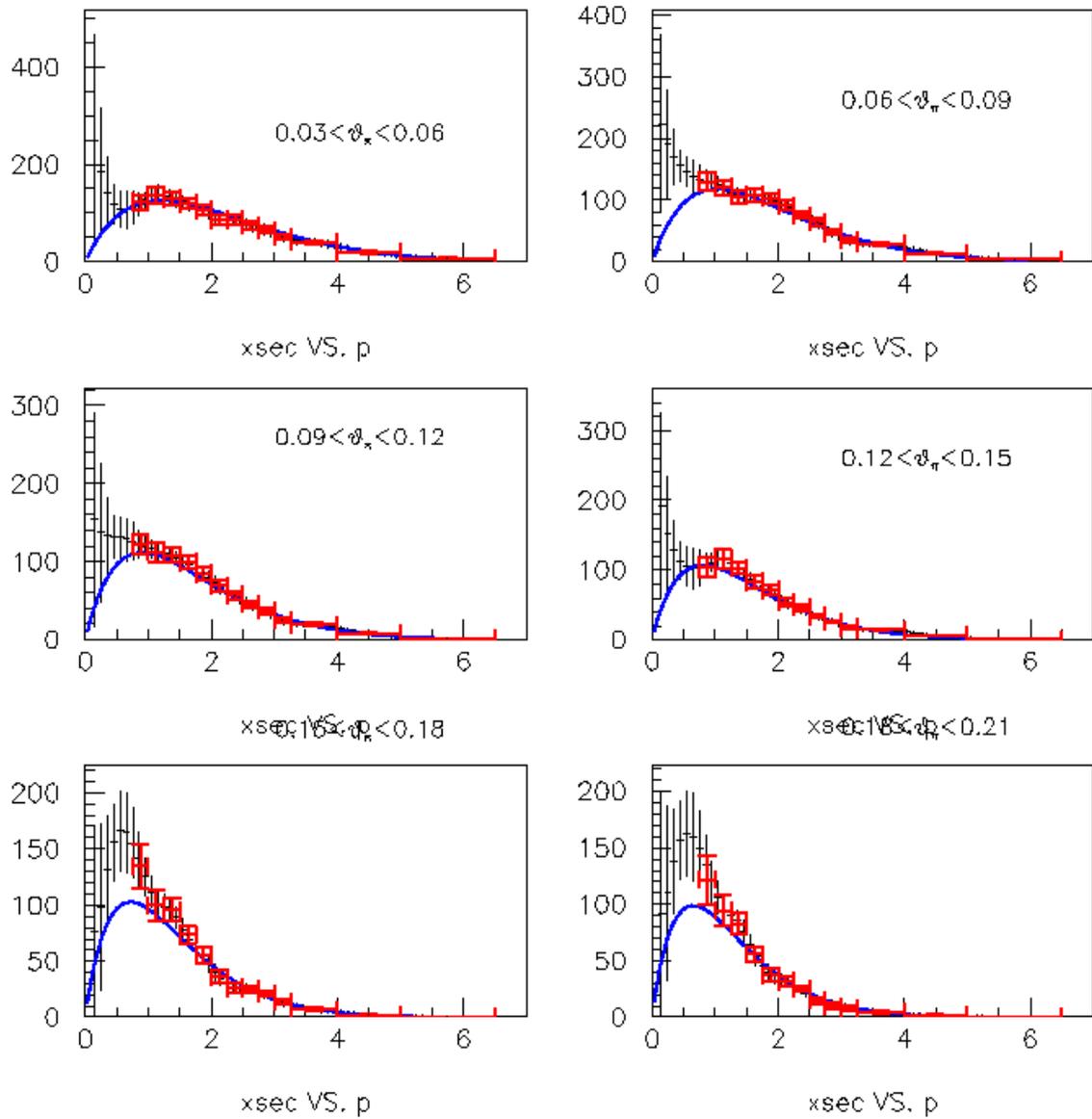


Figure 12: The π^- version of Fig. 4. Identical to Fig. 11, but the 40 spline multisims have been replaced by a profile histogram of all 1000 spline multisims

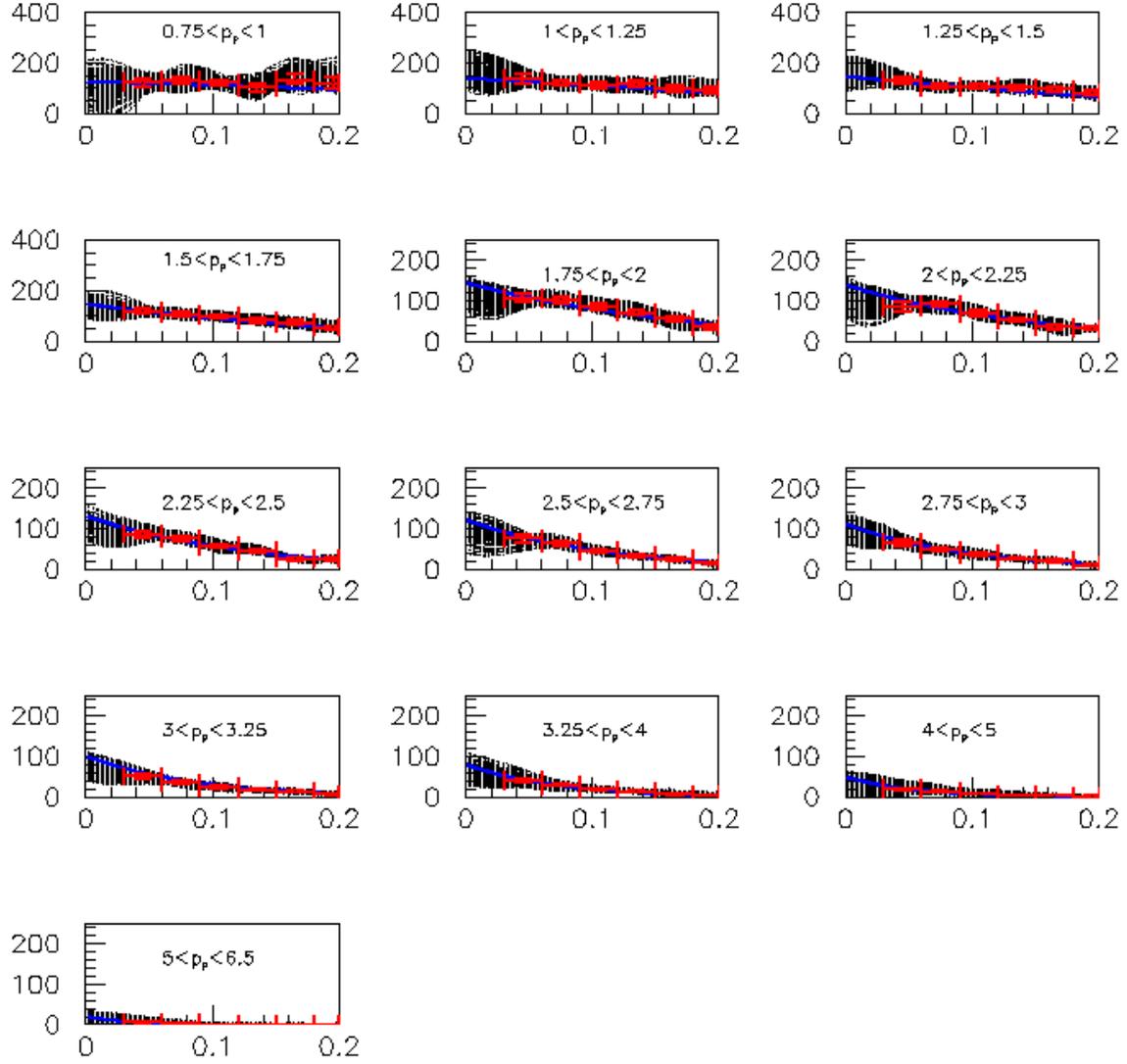


Figure 13: The π^- version of Fig. 5. The π^- production cross-section as a function of π^- angle from 0 to 0.25 rad. The thirteen panels are for the thirteen π^- momentum ranges indicated on the plots. The red points are the HARP results and uncertainties, the blue curve is the Sandford-Wang fit and the black points are the first 40 spline interpolated multisims.

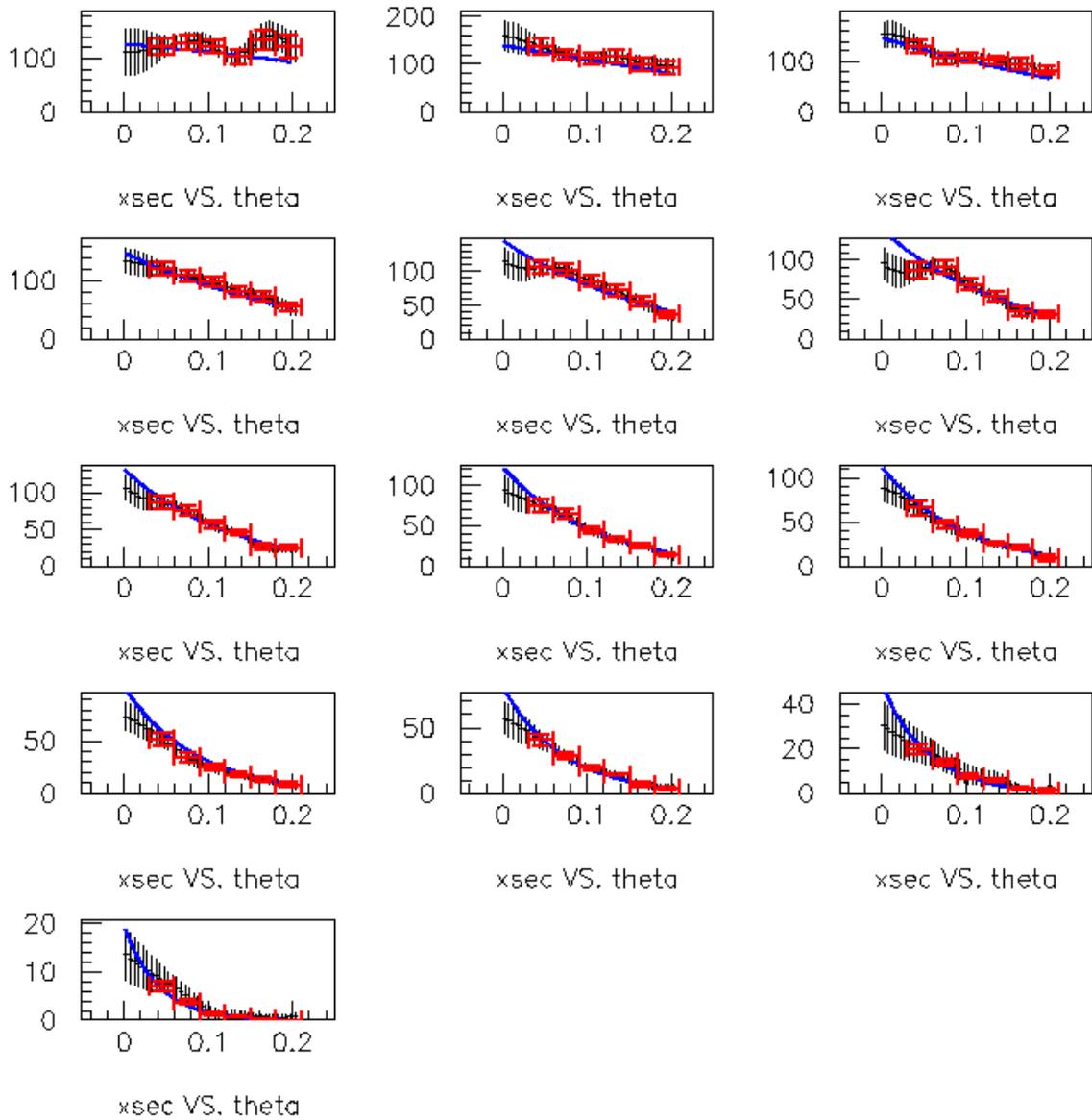


Figure 14: The π^- version of Fig. 6. Identical to Fig. 13, but the 40 spline multisims have been replaced by a profile histogram of all 1000 spline multisims

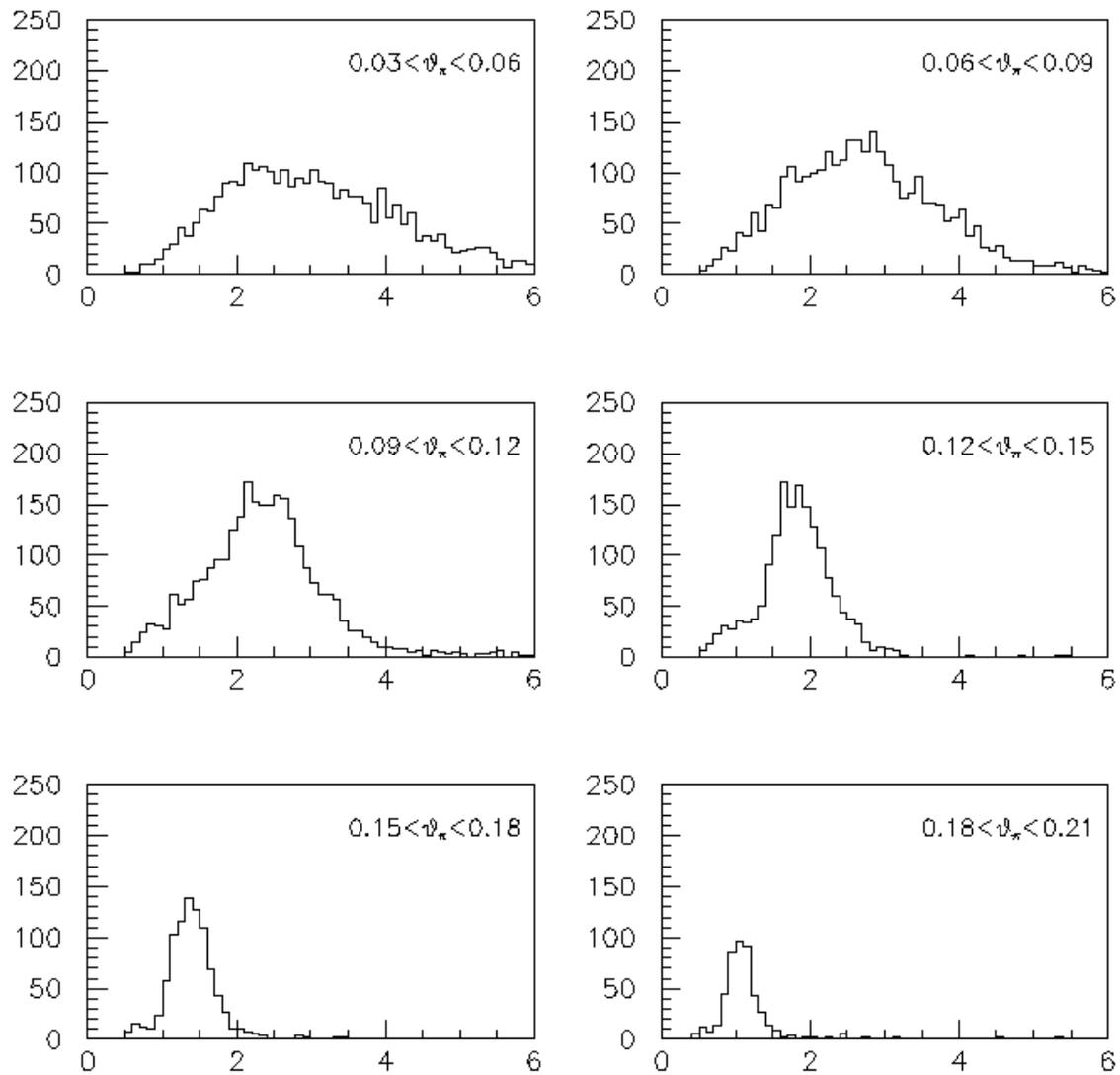


Figure 15: The momentum distribution of the parent meson of neutrinos passing the TBA nue cuts in anti-neutrino mode. The six panels are for the six meson angle ranges indicated. The normalization is arbitrary

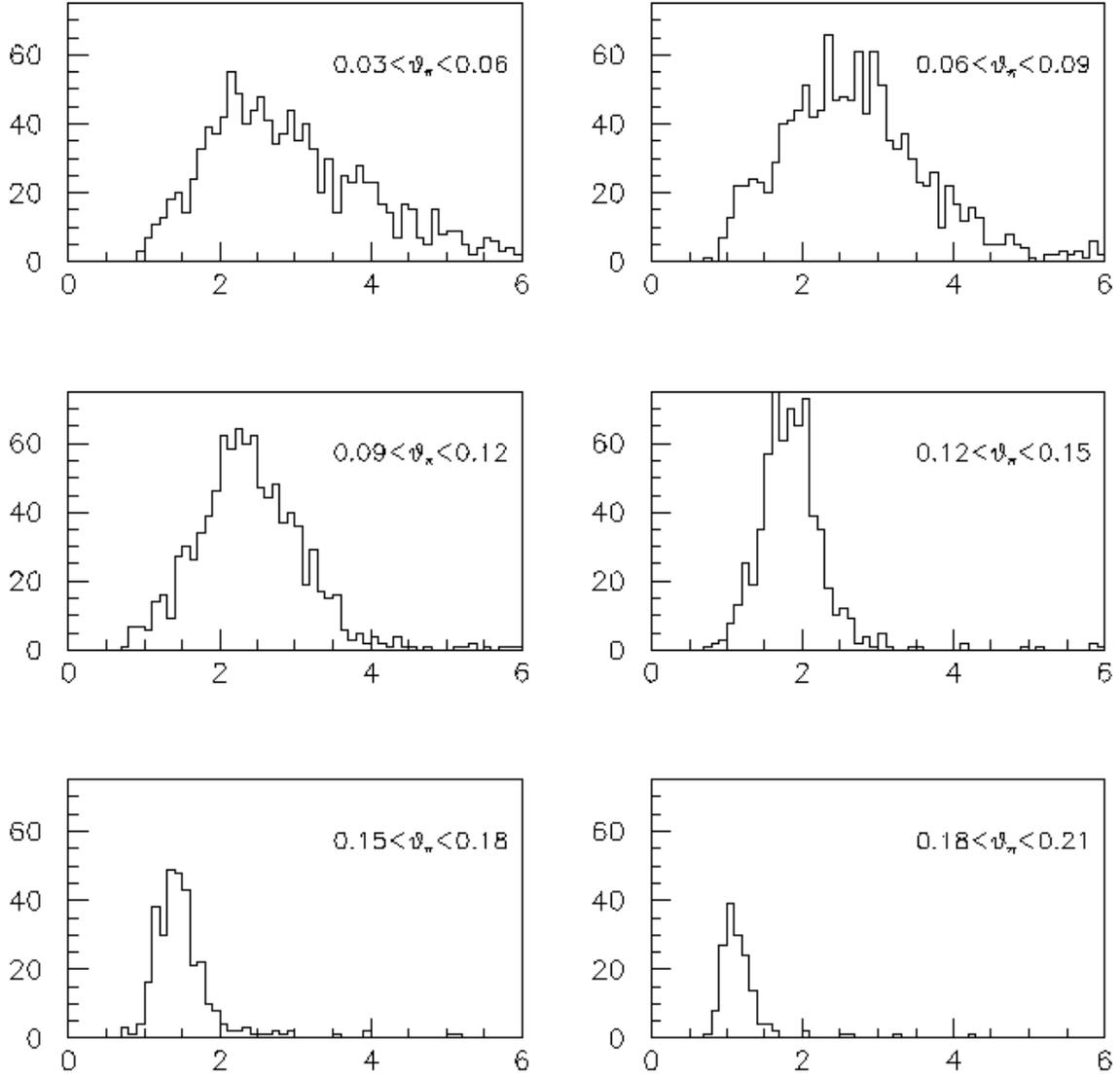


Figure 16: The momentum distribution of the parent meson of neutrinos passing the Likelihood based numu CCQE cuts in anti-neutrino mode. The six panels are for the six meson angle ranges indicated. The normalization is arbitrary.

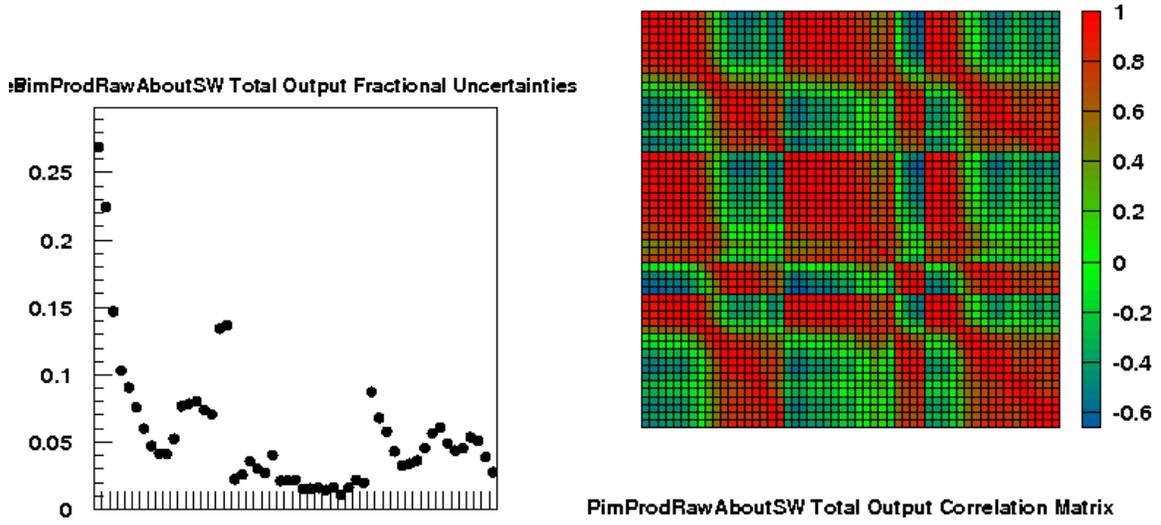


Figure 17: This is the π^- version of Fig. 9. The π^- production uncertainties propagated via spline interpolation and the strictly correct reweighting scheme into errors on the $E_{\nu Qe}$ distribution of fully oscillated nuebar events, nuebar candidate events, and numubar CCQE candidate events. The left hand plot shows the fractional error in each $E_{\nu QE}$ bin and the right hand side shows the correlation matrix. See Sec. 4.2 for details.