



Combining Results from Three Single Top Quark Cross Section Measurements using the BLUE Method

The DØ Collaboration
URL <http://www-d0.fnal.gov>
(Dated: March 17, 2007)

We apply the Best Linear Unbiased Estimate (BLUE) method to combine three measurements of the single top quark production cross section using decision trees (DT), matrix elements (ME) and Bayesian neural networks (BNN) on 0.9 fb^{-1} of DØ data. The resulting combined measurement is $\sigma(p\bar{p} \rightarrow tb + X, tqb + X) = 4.8 \pm 1.3 \text{ pb}$. The probability to measure this value of the cross section or higher in the absence of signal is 0.027%, corresponding to a 3.5 standard deviation significance.

Preliminary Results for Winter 2007 Conferences

I. INTRODUCTION

This note presents the combination of three measurements of the single top quark production cross section obtained using different multivariate techniques, based on 0.9 fb^{-1} of $D\bar{O}$ data [1]:

$$\begin{aligned}\sigma(p\bar{p} \rightarrow tb + X, tqb + X) &= 4.9_{-1.4}^{+1.4} \text{ pb (Decision trees)} \\ &= 4.6_{-1.5}^{+1.8} \text{ pb (Matrix elements)} \\ &= 5.0_{-1.9}^{+1.9} \text{ pb (Bayesian neural networks)}.\end{aligned}$$

It is now routine to have multiple analyses of the same dataset and to be faced with the associated problem of what to do with the multiple correlated results. One approach is to pick the result which *a priori* is the most precise, as determined by an ensemble study of pseudo-datasets [2]. The more collegial thing to do is to combine the results. In general, if the correlations between the different measurements are small, one expects to also improve the sensitivity from the combination.

II. BLE, BLUE, AND ALL THAT

The simplest way to combine results is to use a linear function

$$f(\hat{\sigma}, w) \equiv y = \sum_i w_i \hat{\sigma}_i, \quad (1)$$

of the measurements $\hat{\sigma}_i$, usually with the constraint $\sum_i w_i = 1$. The weights are determined by minimizing the mean square error (MSE)

$$\text{MSE} \equiv \text{E}((y - \sigma)^2), \quad (2)$$

$$= \text{Var}(y) + (\text{bias})^2 \quad (3)$$

where $\text{bias}(y) = \text{E}(y) - \sigma$ is the bias and $\text{Var}(y) = \text{E}(y^2) - \text{E}^2(y)$ is the variance, for a measurement y when the true value is σ . This method is called the *Best Linear Estimate* (BLE) method. If each individual measurement $\hat{\sigma}_i$ has a bias β_i then the ensemble average of y

$$\begin{aligned}\text{E}(y) &= \sum_i w_i (\beta_i + \sigma), \\ &= \sigma + \sum_i w_i \beta_i,\end{aligned} \quad (4)$$

will be biased absent a fortuitous cancellation. If, however, we have forced the measurements to be unbiased then the ensemble average of the combined measurement y will also be unbiased. Thus, we arrive at the *Best Linear Unbiased Estimate* (BLUE) method [3], in which the weights are obtained by minimizing the variance (since the bias is zero).

Note that in taking the ensemble average in Eq. (4) we have assumed the parameters w_i to be *constant* across the ensemble [4]. If the w_i depend on the estimates $\hat{\sigma}_i$ then the BLUE method will not, in general, yield a combined estimate that is unbiased, though of course the method can be corrected for bias after the fact.

For uncorrelated measurements $\hat{\sigma}_i$, the variance is given by

$$\text{Var}(y) = \sum_i w_i^2 \text{Var}(\hat{\sigma}_i) \quad (5)$$

and the minimum variance occurs when we set the weights to

$$w_i = \frac{1/\text{Var}(\hat{\sigma}_i)}{\sum_i 1/\text{Var}(\hat{\sigma}_i)}. \quad (6)$$

We note, again, that in the derivation of Eq. (5) the parameters w_i must be independent of the estimates $\hat{\sigma}_i$, though they may vary from one ensemble to another. It should also be noted that Eq. (6) holds true for *any* distribution with finite variance. In particular, it is not necessary that the distribution be Gaussian.

In practice, measurements based on the same dataset will be correlated and we must replace Eq. (5) by

$$\text{Var}(y) = \sum_i \sum_j w_i w_j \text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j), \quad (7)$$

where $\text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j) \equiv \text{E}(\hat{\sigma}_i \hat{\sigma}_j) - \text{E}(\hat{\sigma}_i)\text{E}(\hat{\sigma}_j)$ are the matrix elements of the covariance matrix of the measurements. The minimization yields the result

$$w_i = \frac{\sum_j \text{Cov}^{-1}(\hat{\sigma}_i, \hat{\sigma}_j)}{\sum_i \sum_j \text{Cov}^{-1}(\hat{\sigma}_i, \hat{\sigma}_j)}, \quad (8)$$

where $\text{Cov}^{-1}(\hat{\sigma}_i, \hat{\sigma}_j)$ denotes the matrix elements of the *inverse* of the covariance matrix.

A. Ansatz for confidence intervals

The expression for the variance, Eq. (7), of the ensemble distribution of y , Eq. (1), suggests the following ansatz for a confidence interval. Compute

$$\delta_y = \sqrt{\sum_i \sum_j w_i w_j \rho_{i,j} \hat{\delta}_i \hat{\delta}_j}, \quad (9)$$

where the $\hat{\delta}_i$ are the uncertainties on $\hat{\sigma}_i$, and $\rho_{i,j} \equiv \text{Cov}(\hat{\sigma}_i, \hat{\sigma}_j) / \sqrt{\text{Var}(\hat{\sigma}_i)\text{Var}(\hat{\sigma}_j)}$ is the correlation matrix element. An approximate (symmetric) confidence interval is then given by $[y - \delta_y, y + \delta_y]$.

III. RESULTS USING THE BLUE METHOD

Since the bias is measured to be small in each of the three single top analyses, using decision trees, matrix elements, and Bayesian neural networks, we apply the BLUE method outlined above to combine the individual results. For this, we use corresponding measurements from the three analyses from the following two ensembles of pseudo-datasets:

- SM signal (2.9 pb) + background pseudo-datasets
- Background-only pseudo-datasets.

These ensembles of pseudo-datasets are generated from a pool of 1.6 million Monte Carlo events used in the modeling of the SM backgrounds and the single top quark signals. Each source of background or signal is fluctuated separately, according to the allowed variation on that source due to systematic and statistical uncertainties. The normalization to data imposed in the background model is also taken into account when fluctuating background sources that are anti-correlated by the normalization. The event weights (from trigger, object reconstruction, and b -jet identification efficiencies) are taken into account such that events with a higher weight will be more likely to be picked.

In order to take into account correlations between the different measurements (those arising from using the same set of observed events as well as those from the systematic uncertainties), we use results from the same ensemble-entry in the sums in Eqs. 1 and 9. There are about 1700 common entries between the three analyses in the SM signal+background pseudo-datasets, and about 63,000 entries in the background-only pseudo-datasets.

A. Weights, coverage probability, and combined measurement

We use the SM signal+background pseudo-datasets to determine the weights and to check the coverage probability of the confidence intervals. The cross section measurements from this ensemble are shown in Fig. 1 for the individual and combined analyses. The mean and square root of the variance obtained from these distributions are given in Table I. The weights w_i for each of the three analyses obtained using Eq. 8, are:

- $w_{\text{DT}} = 0.401$
- $w_{\text{ME}} = 0.452$
- $w_{\text{BNN}} = 0.146$

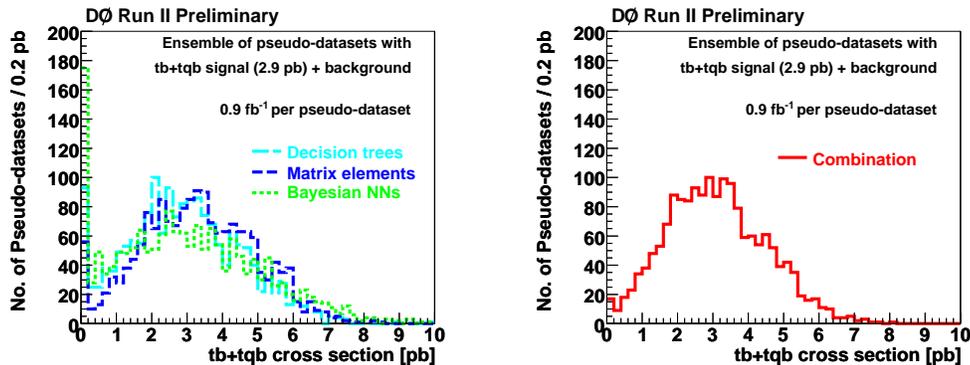


FIG. 1: Distributions of the measured cross sections from the individual analyses (left), and the combined analysis (right), using SM signal+background ensembles.

TABLE I: Mean and square root of variance from the SM signal (2.9 pb) + background ensembles for the different analyses.

Analysis	Mean	$\sqrt{\text{Var}}$	$\sigma/\Delta\sigma$
	σ [pb]	$\Delta\sigma$ [pb]	
Decision trees (DT)	2.9	1.62	1.8
Matrix elements (ME)	3.3	1.60	2.1
Bayesian neural networks (BNN)	3.0	2.05	1.4
Combined	3.1	1.40	2.2

To check the coverage probability of the confidence intervals as discussed in Sec. II A, we determine the correlation matrix using the same ensemble. This is found to be:

$$\rho = \begin{pmatrix} DT & ME & BNN \\ 1 & 0.57 & 0.51 & DT \\ 0.57 & 1 & 0.45 & ME \\ 0.51 & 0.45 & 1 & BNN \end{pmatrix}.$$

The resulting coverage probability of $[y - \delta_y, y + \delta_y]$ is 0.64, which is close to 68% of a one-standard-deviation confidence level interval. We may therefore use this definition of δ_y to compute the uncertainty on the combined measurement from real data.

The combined result and its uncertainty for the single top quark cross section measurement is, thus, obtained to be:

$$\sigma(p\bar{p} \rightarrow tb + X, tqb + X) = 4.8 \pm 1.3 \text{ pb (DT + ME + BNN combined)},$$

using the measurements listed in Sec. I. Fig. 2 summarizes the measurements from the individual analyses as well as the combination.

B. Significance

We use the background-only ensemble to determine the expected and observed significance of the combined cross section. Here too, results from the individual analyses are combined using Eq. 1. Distributions of the results from all the analyses are shown in Fig. 3.

The expected p -value (and the associated significance in Gaussian-like standard deviations) is obtained by counting how many background-only pseudo-datasets result in a measured cross section above the expected SM value of 2.9 pb. These are shown in Table II for the different analyses.

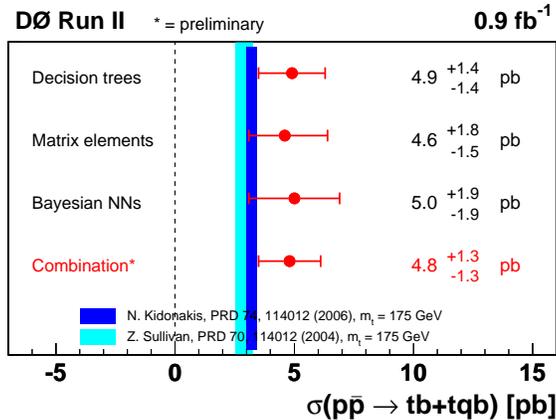


FIG. 2: The single top cross section measurements using real data, from the individual analyses and the combination.

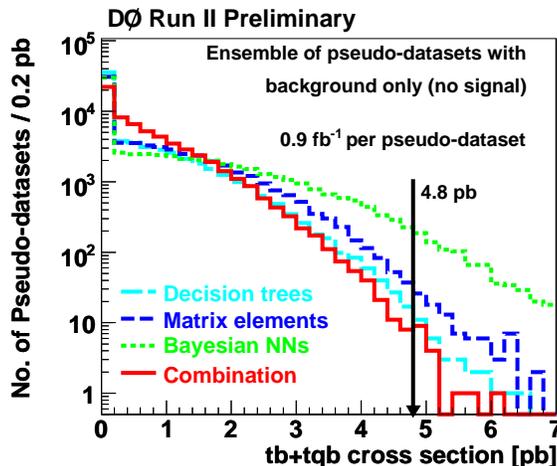


FIG. 3: Distributions of the measured cross sections from the different analyses, using background-only ensemble. The arrow shows the combined cross section measurement (4.8 pb) using real data.

TABLE II: The expected p -values and significances for the individual and the combined analyses, using the SM value of 2.9 pb for signal cross section as the reference point in Fig. 3.

Analysis	Expected p -value	Expected significance [std. dev.]
Decision trees (DT)	0.0177	2.1
Matrix elements (ME)	0.0358	1.8
Bayesian neural networks (BNN)	0.0992	1.3
Combined	0.0137	2.2

The observed p -value is similarly calculated by counting how many background-only pseudo-datasets result in a measured cross section above the observed value of 4.8 pb. The result is 0.027% or 3.5 standard deviations. The observed cross sections, p -values, and significances from all the analyses are summarized in Table III.

Finally, using the SM signal+background pseudo-datasets, we obtain the compatibility with the SM expectation by counting how many pseudo-datasets result in a cross section with the observed value or higher for each of the

TABLE III: The measured cross sections, p -values, and significances for the individual and combined analyses, the latter two obtained using the background-only ensemble.

Analysis	Measured cross section [pb]	p -value	Significance [std. dev.]
Decision trees (DT)	4.9	0.00040	3.4
Matrix elements (ME)	4.6	0.00201	2.9
Bayesian neural networks (BNN)	5.0	0.01157	2.3
Combined	4.8	0.00027	3.5

analyses. The probabilities for the different analyses are 11% for the DT, 22% for the ME, 17% for the BNN, and 12% for the combined analyses.

IV. CONCLUSIONS

To conclude, the measured single top quark production cross section after combining results from the DT, ME and BNN analyses, is 4.8 ± 1.3 pb with a significance of 3.5 standard deviations.

-
- [1] V. M. Abazov *et al.* [D0 Collaboration], “Evidence for production of single top quarks and first direct measurement of $|V_{tb}|$,” submitted to Phys. Rev. Lett., arXiv:hep-ex/0612052.
 - [2] In principle, using ensembles of pseudo-datasets, one could study different protocols for picking a single result *a posteriori*, such as picking, each time, the most precise one. However, this “Maxwell demon” approach was considered too radical when it was tried in Run I. One presumes it would still be considered radical now.
 - [3] L. Lyons, D. Gibaut, and P. Clifford, Nucl. Instrum. Methods **A 270**, 110 (1988); R. J. Barlow, *Statistics: A Guide To The Use Of Statistical Methods In The Physical Sciences*, The Manchester Physics Series, John Wiley and Sons, New York (1989); G. Cowan, *Statistical Data Analysis*, Oxford (1998).
 - [4] In the BLUE method, statistical and systematic uncertainties can be treated separately. This however is not necessary if one uses ensembles of pseudo-datasets appropriately randomized for statistical and systematic uncertainties.