Study of Uncertainties in NSTX Thomson Scattering Data

D. Johnson, R. Bell, and B. LeBlanc *PPPL*

Motivation

- Typical scattered light signals in the NSTX Thomson detectors are in the range 1,000-100,000 photoelectrons.
- In many cases, systematic calibration uncertainties predominate over photon statistics in determining the computed errors in the measured signals, which are used as weights for the least squares fits for T_e and n_e.
- The most significant contribution from the calibration is the quoted uncertainty in the the traceable calibration of the reference detector used to characterize the spectral response of the filter polychromators.
- A **statistical study of the residuals** between the measured and fitted signals for a large number of laser shots can illuminate systematic offsets, some of which are consistent with errors in the reference detector responsivity.
- This system from is currently being upgraded from 20 to 30 spatial channels, and a recalibration of the entire system is planned in the near future. This study provides insights into potential improvements in this process.
- This type of analysis has been applied previously to LIDAR Thomson data on JET. **"***A new method for estimating and identifying systematic error in Thomson scattering diagnostics***"**, H. Fajemirokun, C. Gowers, P. Nielsen, H. Saltzmann, and K. Hirsch. RSI, **61**, (1990)
- The above paper describes how this technique was able to discern a change in the window transmission, and there is potential that we could do the same on NSTX to monitor window conditions.

Strategy

A) Compare measured χ^2 **distribution** with theoretical distribution for 3 degrees of freedom:

$$P_k(x) = \frac{(x)^{3/2-1} e^{-x/2}}{\Gamma(3/2) 2^{3/2}} \quad (1) \qquad \text{for} \quad x_j = \chi_j^2 = \sum_i \frac{(y_{ij} - f_{ij})^2}{\sigma_{ij}^2} \quad (2)$$

which peaks at a value of $\chi^2=1$ for our case, assuming we do not use the Rayleigh channel in the fitting. Here y_{ij} is the measured number of photoelectrons, f_{ij} is fitted value, and σ_{ij} is the modeled uncertainty in spectral channel i of spatial channel j.

If, for a large number of samples, the measured distribution is different from (1) it could mean that one or more the following is untrue:

a. The assumed fitting function describes the physics
b. Calibrations are correct
c. Estimated errors have been properly assigned
d. Residual distributions are Gaussian

B) In order to obtain more information, it is useful to look at the **normalized residuals**:

$$z_{ij} = \frac{y_{ij} - f_{ij}}{\sigma_{ij}} \quad (3)$$

For a large sample size *(k=1, N)* of laser shots, r_{ij}^k should be distributed with a **Gaussian distribution**:

$$P(z_{ij}^k) = \frac{1}{s_{ij}\sqrt{2\pi}} e^{-(z_{ij}^k - \mu_{ij})^2/(2 s_{ij}^2)} \quad (4).$$

If a) through d) above are correct, then the distribution should be centered about $\mu_{ij} = 0$, with a width $s_{ij}=1$ or equivalently $FWHM = 2\sqrt{2\ln 2} = 2.35$. Nonzero values of μ_{ij} indicate a failure of one or more of a) through d), the most likely candidate being a systematic calibration error. $s_{ij} \neq 1$ indicates that errors have not been properly estimated.

C) If there are systematic offsets in the normalized residual distributions, and if one assumes these are due to systematic calibration errors of individual detector sensitivity, one can calculate **correction factors** c_{ij}

$$\sum_k \frac{y_{ij}^k c_{ij} - f_{ij}^k}{\sigma_{ij}^k} = 0 \qquad \text{or} \qquad c_{ij} = \sum_k \frac{f_{ij}^k}{\sigma_{ij}^k} \Big/ \sum_k \frac{y_{ij}^k}{\sigma_{ij}^k} \quad (5)$$

where the summation over k is for thousands of laser pulses.

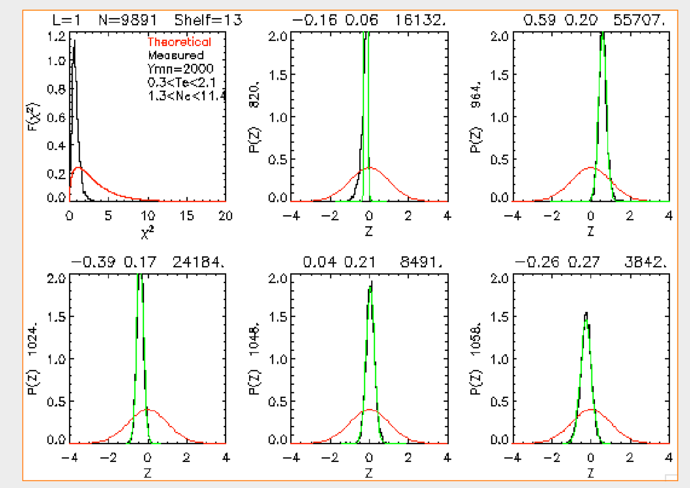
- Thus knowing the observed signals, the fitted values and the estimated errors permits calculation of correction factors for individual detectors.
- Multiplying the observed signal in each detector by the appropriate correction factor and refitting the data allows the process to be iterated to convergence.

Baseline Analysis - Case 1

χ^2 Distribution: The distribution of χ^2 for one of the positions is shown in the upper left panel at the left. The experimental distribution in black has a very different shape from the theoretical curve in red, with a higher peak at lower values. This can be understood by looking at the normalized residual distributions.

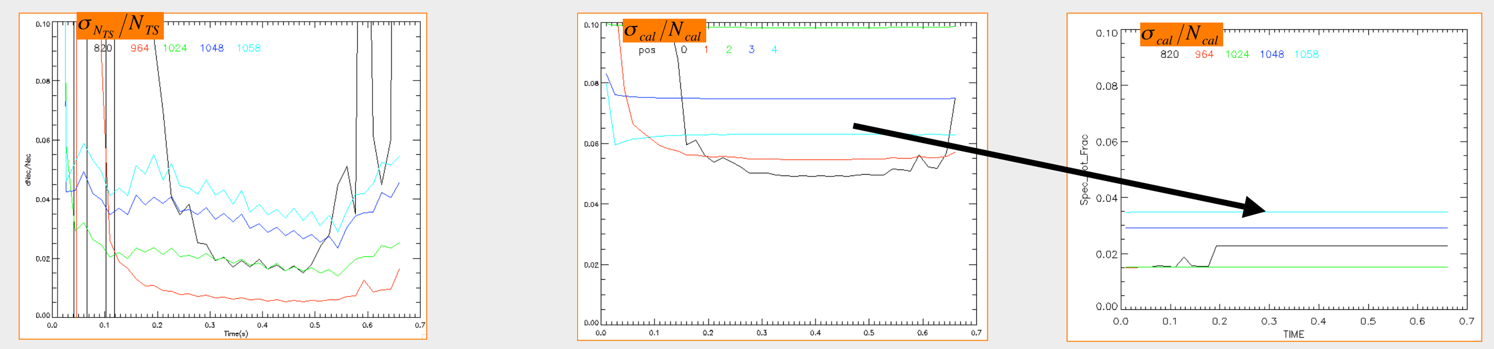
Normalized Residual Distributions: Distributions of the normalized residuals *z* are shown in the other panels at the right, one for each of the five wavelength channels for this position. The red curves are what one would expect if there were no systematic calibration errors, and if the σ^k are being modeled correctly. Note that *z=1* corresponds to a residual equal to one standard deviation. Both experimental and theoretical curves have been normalized to have an integral equal to one. The green curves are Gaussian fits to the observed distributions. Above each P(z) curve are three numbers. The first is the **fitted offset**, the second is the **fitted width**, and the third is the **mean number of photoelectrons** for the data in the distribution.

Note that the experimental distributions in black are much **narrower than expected**, indicating that the **modeled** σ^k **are too large**, consistent with the experimental χ^2 peak being too low.



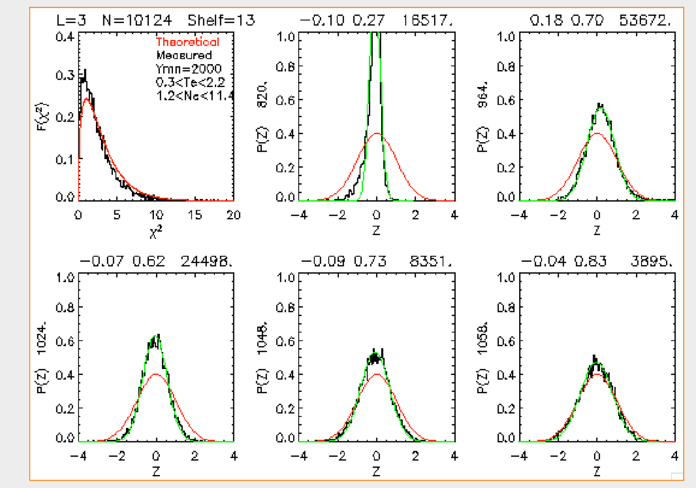
Correcting the uncertainty model: Motivated by the above discrepancies, we reexamined our model for the uncertainties σ. We found two partially offsetting problems. First, there was an error in coding the uncertainty due to the various calibration contributions. Second, the error contribution from the reference detector responsivity was incorrectly treated as a statistical measurement error rather than as a systematic error.

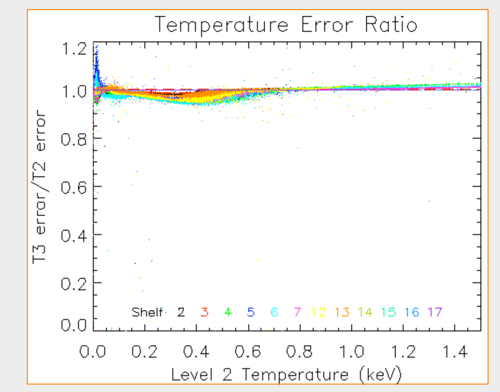
Impact of the improved analysis: The panel at left below shows the fractional error due to scattered photoelectrons, $\sigma_{N_{TS}}/N_{TS}$, as a function of time for the 5 spectral channels at a central position. The central panel shows the fractional error from the old calculation of the combined calibration components σ_{cal}/N_{cal} from all but the first row of the table at the lower left. The right panel shows the corrected analysis for σ_{cal}/N_{cal}, showing a significant decrease in this contribution. In the corrected analysis, the error due to photon statistics dominates the total σ.

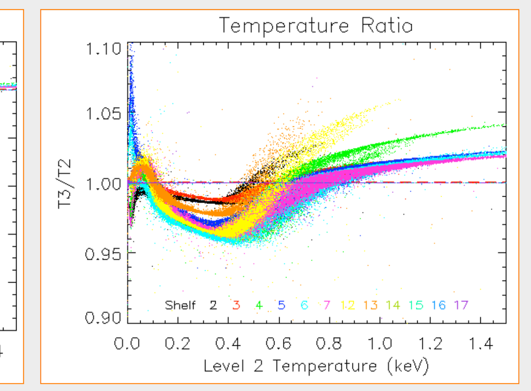


Analysis Applying Statistical Correction Factors - Case 3

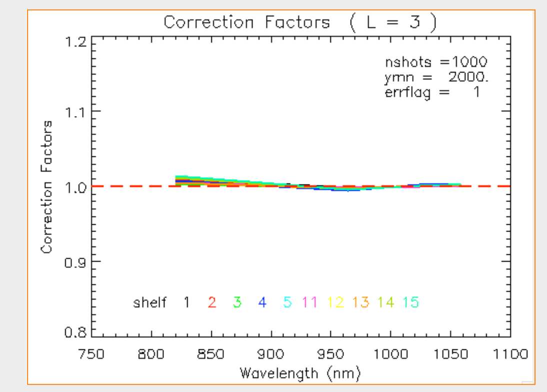
- The correction factors computed in the case 2 analysis were used to scale the signals prior to a new iteration of fitting the data.
- With the correction, the offsets are largely removed, and the χ^2 distribution and the normalized residual distributions P(z) are much closer to the theoretical predictions.
- There is still a slight discrepancy in the fitted widths of the distributions (middle of the top 3 numbers should be 1).
- There is a trend for the fitted widths to be narrower for lower wavelength channels.
- Small offsets persist (left of top 3 numbers should be 0), which would probably be eliminated with one more iteration.







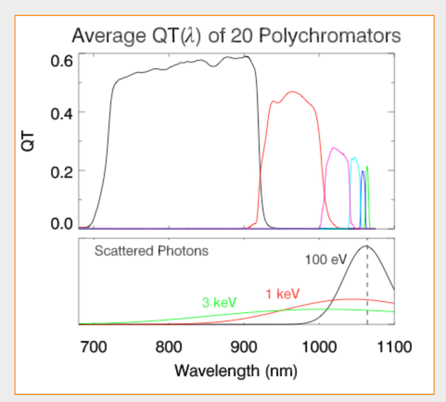
- Compared to case 2, the T_e's resulting from the case 3 analysis are a few % lower below 0.7 keV and a few % higher above this value.
- There is only a slight decrease in δT_e at low T_e.

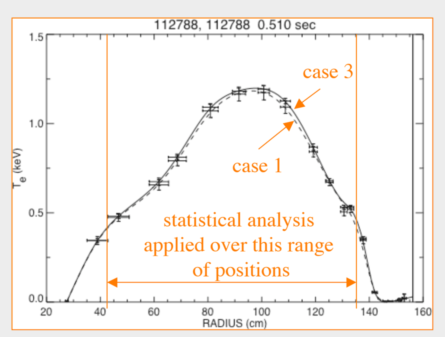


- After one iteration, correction factors have decreased an order of magnitude, roughly consistent with results from Fajemirokun, *et al.*
- Expect factors to be within a part in a thousand of 1 after 1-2 more iterations.
- Each iteration takes 2-3 hours of computer time.

NSTX Thomson Scattering System

- Operational for several years, the NSTX Thomson system currently has 20 spatial channels, each with a 6 wavelength filter polychromator designed by General Atomics.
- The filter passbands are shown at right, convolved with the quantum efficiency of the APD detectors.
- Excluding the channel at the laser wavelength used for Rayleigh scattering, this statistical study made use of signals from the 5 other spectral channels, used to fit the scattered spectrum.
- We have used a set of ~ 1000 recent NSTX discharges each with ~ 40 laser shots. We have constrained the data used in this study such that there are at least 2000 photoelectrons in the two short wavelength polychromator channels.





- Because we were interested in discerning systematic effects over a broad spectral range, we confined the study to the central set of positions, which see a wide range in temperatures.
- The position range is shown at left, which also shows the change in the temperature profile which occurred between the baseline case 1 analysis and the final case 3 analysis.

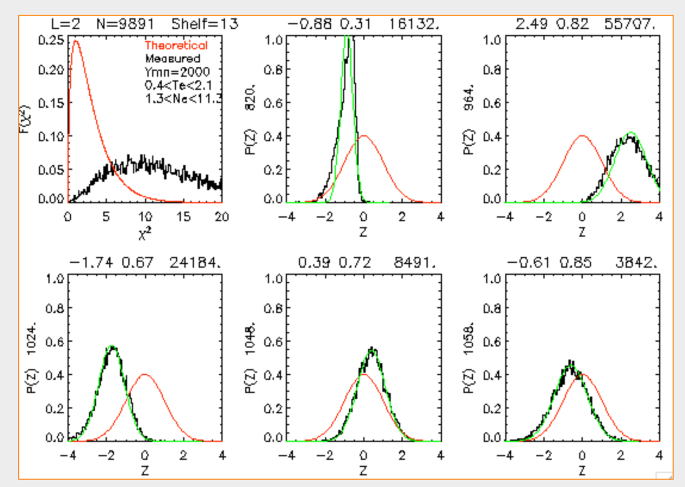
Uncertainty Estimates

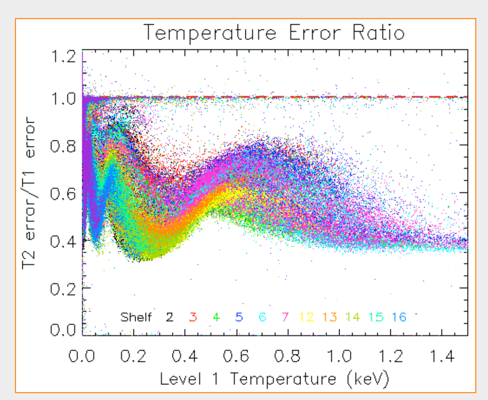
- Sources of uncertainty in the Thomson signals are shown at the right.
- The first row shows the "statistical" errors due to photoelectron statistics in the scattered and plasma light.
- The other rows list the uncertainties resulting from various calibrations.
- The terms in red always dominate the modeled uncertainty

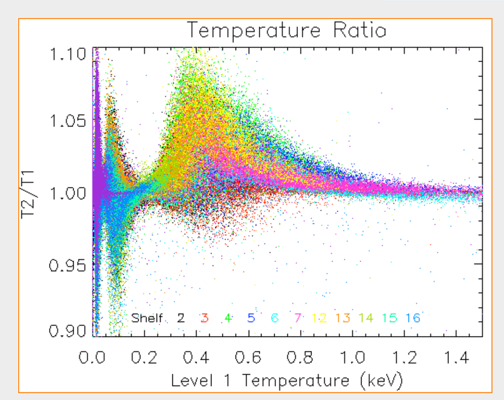
fractional uncertainty	description	range of values	comments
$\frac{\sigma^2_{N_{TS}^{ij}}}{(N_{TS}^{ij})^2}$	estimate of the photoelectron noise normally dominatied by the scattered photoelectrons N_{TS} and/or the plasma signal N_{pl}	.005 – 0.10	$\sigma^2_{N_{TS}^{ij}} = (N_o^{ij})^2 + [F(M)(N_{TS} + N_{Pl})]^{ij}$ readout noise, excess noise, scat pe, plasma pe
$\frac{\sigma_{v_{fs}}}{v_{fs}}, \frac{\sigma_{V_{fs}}}{V_{fs}}$	errors in the fast and slow outputs in the calibration that relates fast and slow sensitivities	.001 - .003	statistical measurement error, can reduce with further averaging
$\frac{\sigma_{(M_x/M_{sp})}}{(M_x/M_{sp})}$	error in the ratio of APD gains used in the TS measurement and the spectral calibration	not yet measured	not currently measured
$\frac{\sigma_{V_{sp}}}{V_{sp}(\lambda)}$	errors in the slow outputs during the spectral calibration	.002 - ,004	statistical measurement error, can reduce with further averaging, finer wavelength intervals
$\frac{\sigma_{V_{ref}}}{V_{ref}(\lambda)}$	errors in the reference detector signals during the spectral calibration	.002 - .005	statistical measurement error, can reduce with further averaging, finer wavelength intervals
$\frac{\sigma_{S_{ref}}}{S_{ref}(\lambda)}$	quoted uncertainty in the responsivity of the reference detector	.015 - .03 (.03-.06)	systematic error, cannot be reduced by more averaging or finer wavelength intervals (range in parentheses indicates discrepancy between two calibrated reference detectors)
$\frac{\sigma_{T_c^j}}{T_c^j(\lambda)}$	errors in the determination of the spectral transmission of the collection optics and fiber bundles	.0005 - .0015	statistical measurement error, can reduce with further averaging, finer wavelength intervals

Analysis with Improved Uncertainty Model - Case 2

- The Thomson data from the same 1000 plasma shots were re-analyzed using the improved uncertainty model.
- The figure to the right shows the resulting distributions. Here we have used $\sigma = \sigma_{N_{TS}}$ in the calculation of the χ^2 distribution and the normalized residual distributions P(z).
- For most channels, the width of P(z) is much closer to the theoretical width, however, there are significant offsets in the centers of these distributions from *z=0*.
- These offsets cause the large discrepancy between the observed and theoretical χ^2 distributions.

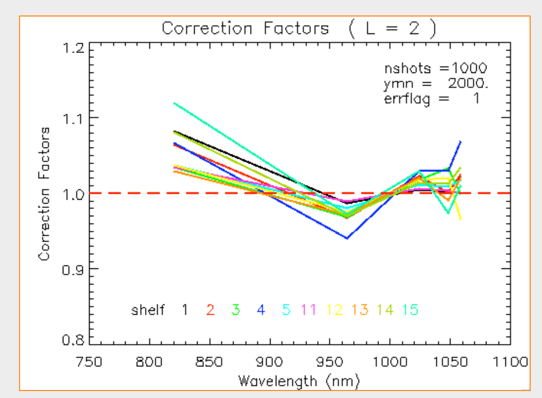






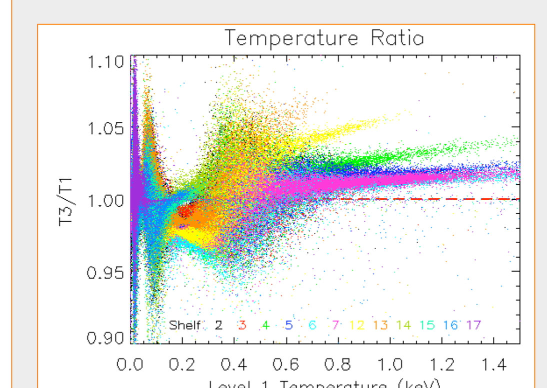
- The improved uncertainty model had a significant effect on the fitted values of δT_e.
- As shown in the figure at far left. these values decrease by more than a factor of 2 at high T_e.
- The effect on T_e is smaller, but significant, particularly at moderate T_e. where increases of 5-8% are seen.
- However, there was no significant change at high T_e.

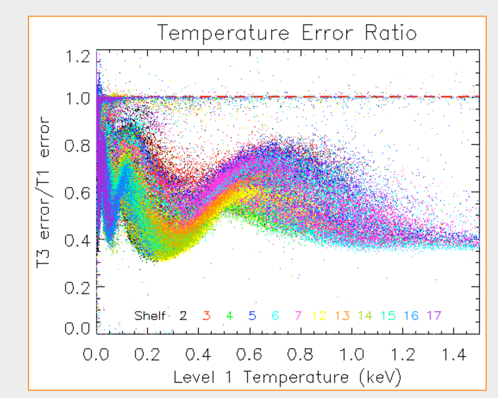
- The newly fitted data was used to calculate the correction factors as defined in equation (5) in the upper left grey panel.
- The results are shown at right where factors are plotted vs the central wavelength, overlaying the results from the central spatial channels.
- These correction factors are quite robust for different shot range selection, and different ways of filtering the data.
- For the different spatial positions, there is a systematic wavelength dependence apparent in the correction factors, which is consistent with a systematic error of order 5-10 % in the spectral calibration.

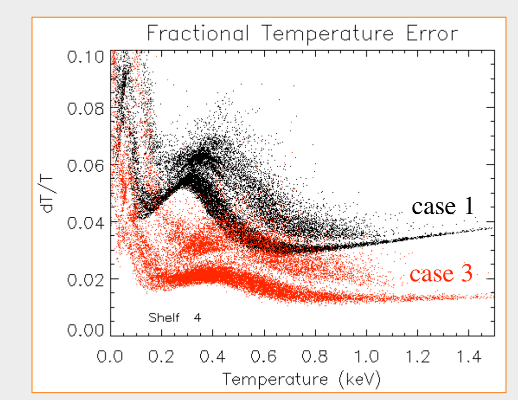


Conclusions

- This study has resulted in a new analysis for the Thomson data, with the following significant impacts:
  - We have identified errors in our previous model for uncertainties used as weights in the least square fits for T_e and n_e.
  - Using the improved model results in a roughly x2 reduction in the fitted uncertainties δT_e and δn_e.
  - Applying the statistically determined correction factors results in a few % change in the fitted T_e and n_e.
  - The combined results of the improved uncertainty model and application of the correction factors is shown in the left two panels below. The reduction in the fractional error $\delta T_e / T_e$ is shown for a central position in the plot at right below.
  - The correction factors have a common spectral dependence across many sample positions in the plasma, consistent with a systematic error in the spectral calibration of the system.
  - Further work is needed to validate this technique for the NSTX data and to understand its full potential.







Future Work

- This is a 'work in progress', and there are some clear directions to proceed further.
- One more iteration will likely bring the fractional correction factor change to < .001, and this should be tried.
- It should be possible to test the sensitivity of this technique to various spectral 'shapes' of systematic error.
- A database of 100-200 discharges are needed to discern correction factors of ~ 1%. It would be interesting to analyse data from periods where window coatings were known to cause a relative change in the window transmission by > 5% across the spectrum.
- If the data is filtered to permit laser shots resulting in signal levels lower than 1000 pe in the widest channels, distinctly non-Gaussian P(z) distributions result. This needs to be understood.
- The trend toward narrower P(z) distributions at lower wavelength needs to be investigated. Perhaps in indicates an error in the treatment of the plasma light component of the uncertainty.
- We have obtained a new calibrated reference detector, and it should be used, along with the statistical tools developed in this study, to investigate whether we can better understand the true source of the systematic errors.