# Supplemental Material B
# Statistical Comparison of 3-term and 4-term fits

While it is clear from Fig. 6 and Fig. 7 in the paper (reproduced in Fig. 1 below) that the 4-term fit provides a superior fit to the 3-term fit according to all fit metrics: sum of squares error (SSE), root mean square error (RMSE), R-square, and adjust R-square, one could question whether the improved fit quality achieved with the addition of the extra term is required by the PIP distribution. This need is illustrated most effectively through visualization of the residual, also reproduced in Fig. 1. It is clear that trends remain in the residual with a 3-term fit, which implies that functional form remains in the distribution which the fit was unable to characterize. In comparison, the residual is largely trend-free for the 4-term fit.

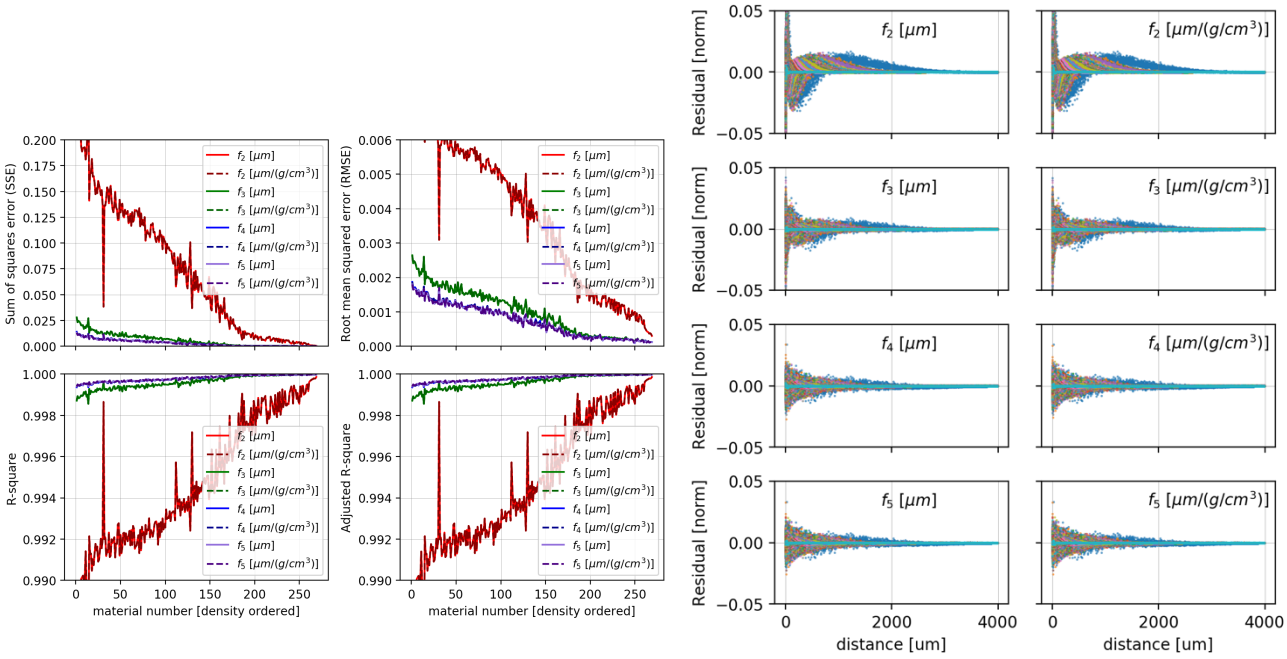


Fig. 1 The goodness of fit metrics and residuals for the fit algorithm conducted in density order without weights. This is the superior fit published with this publication. This is a reproduction of Fig. 6 and Fig. 7 from the paper.

If the residuals were normally distributed, this would imply that the model developed properly accounts for the functional form of the PIP. To test this for the 3-term and 4-term fits, the fraction of annihilations occurring at greater depth than each data point is calculated and a dataset is taken for all materials at percentages scaling from 95% - 40%. These datasets were tested statistically for normality. The Shapiro-

Wilk test is employed, which quantifies through the p-value the likelihood that a distribution could have been drawn from a Gaussian distribution. The results of these tests for the 3-term and 4-term fits are given in Table 1. It is observed (at an alpha value of 0.05) that one should fail to reject the null hypothesis that the data could have been drawn from a Gaussian distribution for the fractional range $> 50\%$ for the 4-term fit. This is not found to be the case for the 3-term fit. These results make intuitive sense because the additional term in the 4-term fits contributes to short length-scales and increases the fidelity of the model in this region, as shown in Fig. 8 in the manuscript. As such, the 4-term fit function captures more of the data functional form and the residuals take on a more normal distribution.

Table 1. Results of application of the Shapiro-Wilk test to material residuals at distances such that 95% - 40% of counts lie above the given distance. It is observed, at an alpha value of 0.05, that for the 4-term fit for fractional values of greater than 50%, one should fail to reject the null hypothesis that the data came from a normally distributed population. This is not found to be the case for the 3-term fit.

| Count Fraction Exceeding Value | Fail to Reject Null Hypothesis of Shapiro-Wilk Test | |
|---|---|---|
| | 3-term Fit | 4-term Fit |
| 0.95 | TRUE | TRUE |
| 0.9 | FALSE | TRUE |
| 0.85 | FALSE | TRUE |
| 0.8 | TRUE | TRUE |
| 0.75 | TRUE | TRUE |
| 0.7 | FALSE | TRUE |
| 0.65 | TRUE | TRUE |
| 0.6 | TRUE | TRUE |
| 0.5 | FALSE | FALSE |
| 0.4 | TRUE | TRUE |