

Correcting Saturation in Mass Spectrometry Data using Principal Components Analysis

S. A. DeJong, J. Hochrein, L. Miller, and M. H. Van Benthem
Sandia National Laboratories, Albuquerque NM 87185; sdejong@sandia.gov

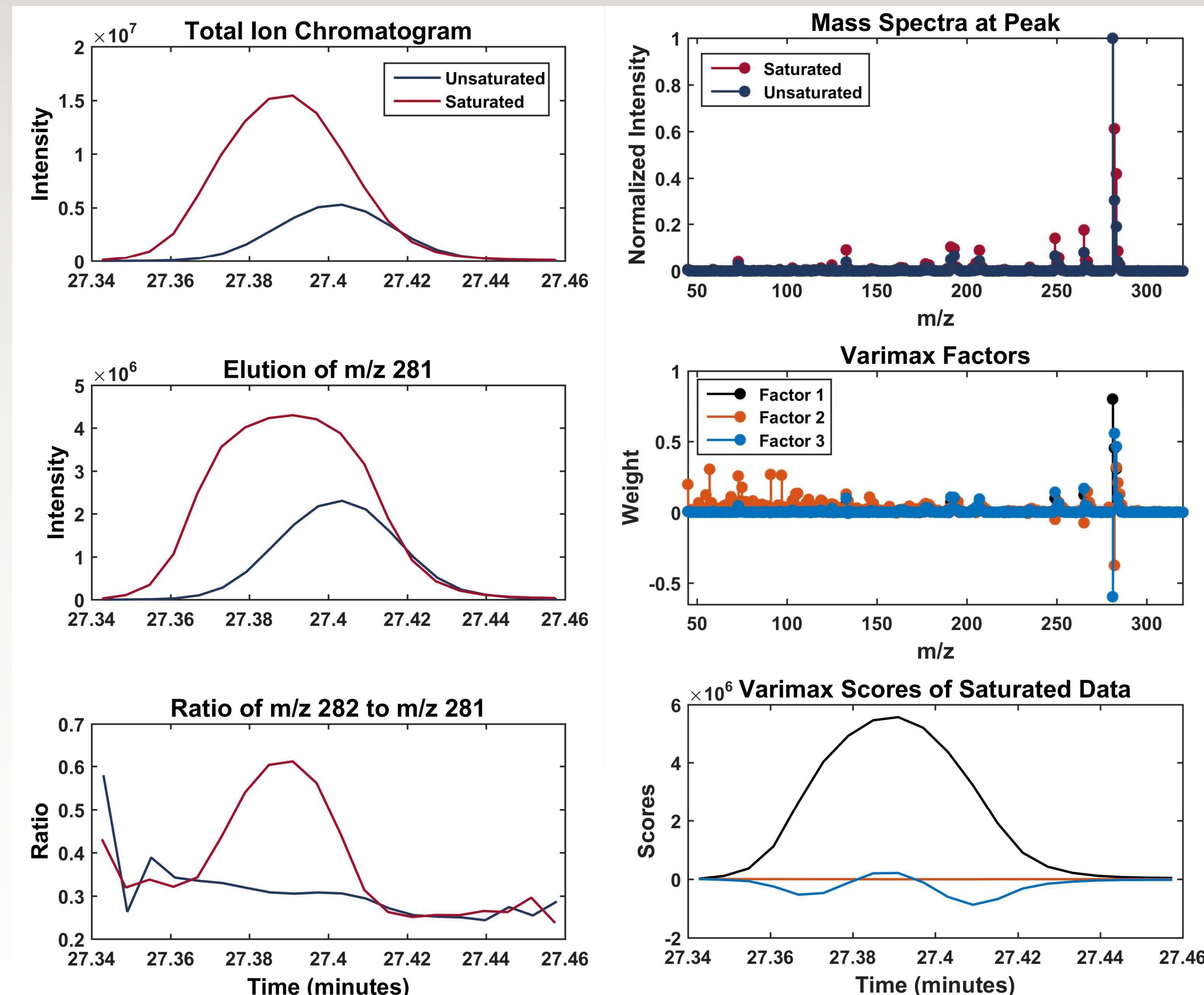
ABSTRACT

Some gas chromatography mass spectrometry data contain distortions that arise from saturation of the instrument's detector. The common features observed include a broadened, rounded chromatographic peak shape accompanied by changing isotope ratios in the mass spectra across the elution profile. These spectral artifacts complicate factor analysis of the data, requiring the incorporation of more factors to account for the distortions and impeding quantitative analysis of the data. We have developed a method to correct the saturated intensities using principal components analysis of the unsaturated regions of the elution profile to predict new values across the saturated region. The efficacy of this correction is demonstrated by examining the isotope ratios across the elution profile before and after correction, along with performing factor analysis. In the simplest case we present here – a single eluting compound demonstrating saturation – three factors were required to model the data. After the saturation correction, the model was reduced to two factors – one for the background and one for the eluting compound.

IDENTIFYING SATURATED VALUES

- Broad mass elution profiles
- Rounded peaks
- Changing Isotope Ratios
- Higher Rank

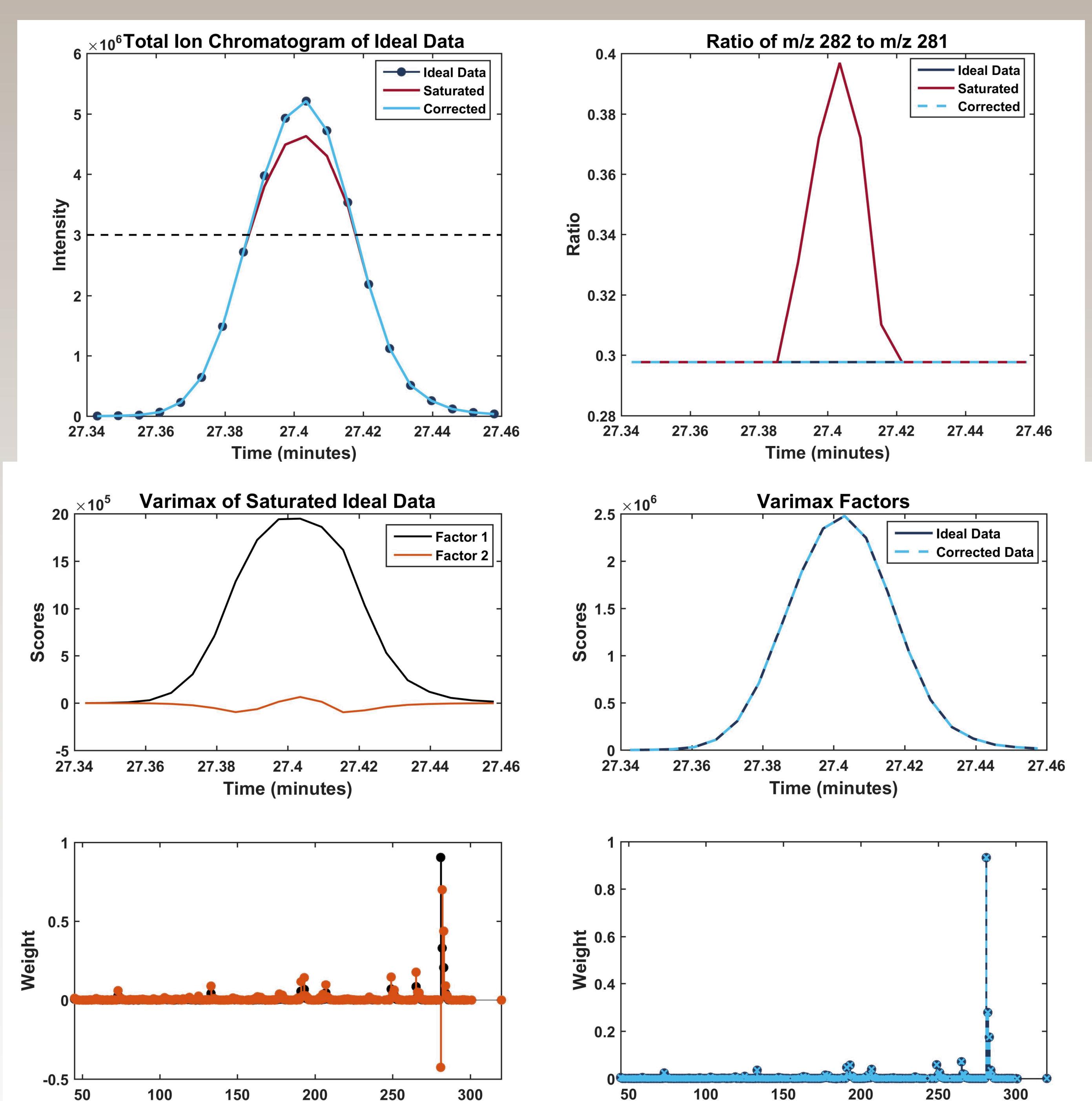
These saturation effects are shown in the figures below.¹



SATURATION CORRECTION

1. Removed saturated time points from D to form $u\tilde{D}$. Poisson scale as $u\tilde{D} = uH uD uK$, where uH and uK contain the inverse of variance estimates in the temporal and spectral domains, respectively.²
2. Perform SVD to estimate spectral factors $u\tilde{P}$ from $u\tilde{D}$
3. Remove masses containing saturated points from D and $u\tilde{P}$ to form D_u and $u\tilde{P}_{-s}$
4. Estimate scores on $u\tilde{P}_{-s}$: $T = (D_u uK_{-s}) / u\tilde{P}_{-s}$
5. Estimate corrected values for saturated points: $\hat{D} = T u\tilde{P}^T / uK$

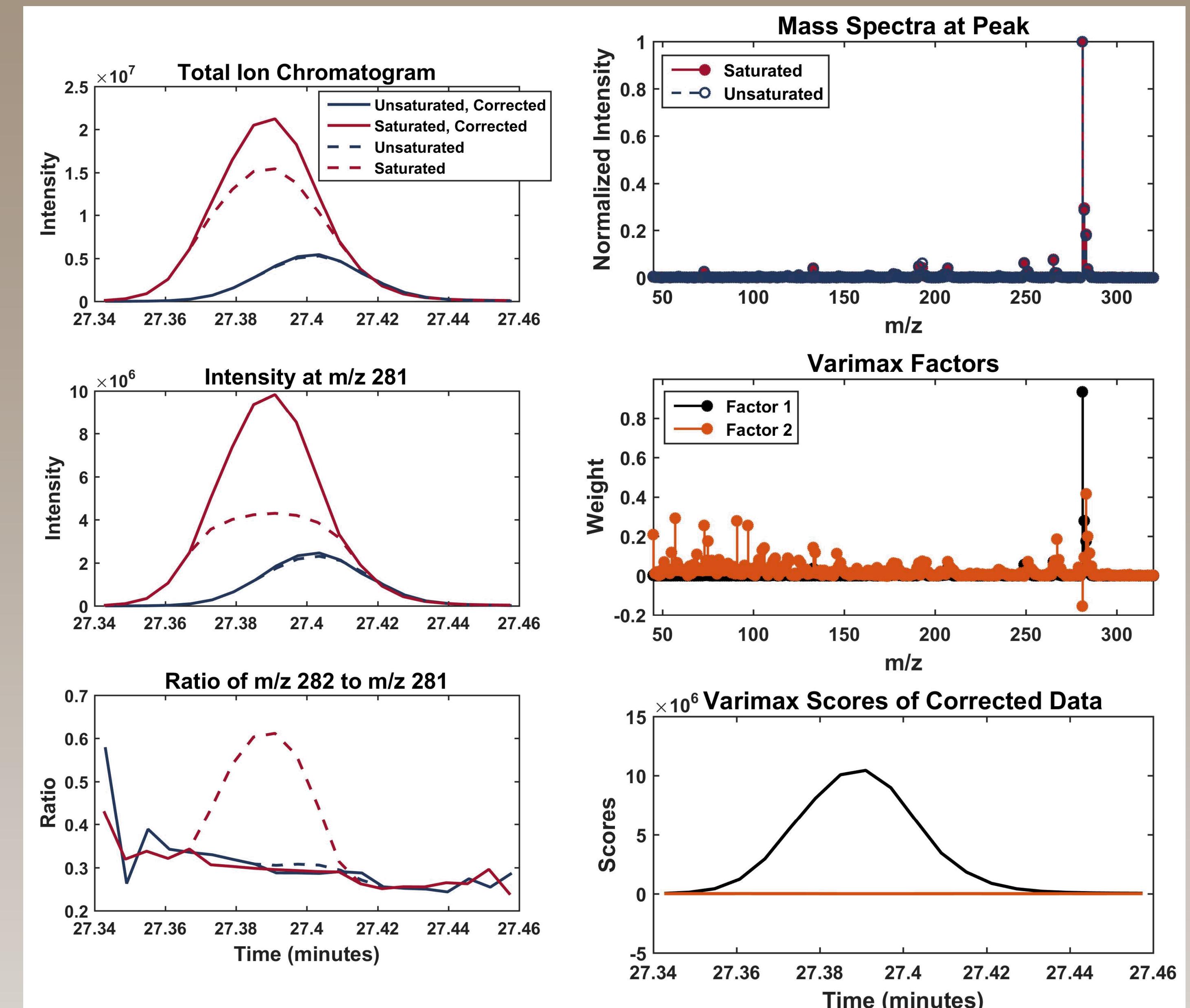
TEST ON IDEAL DATA



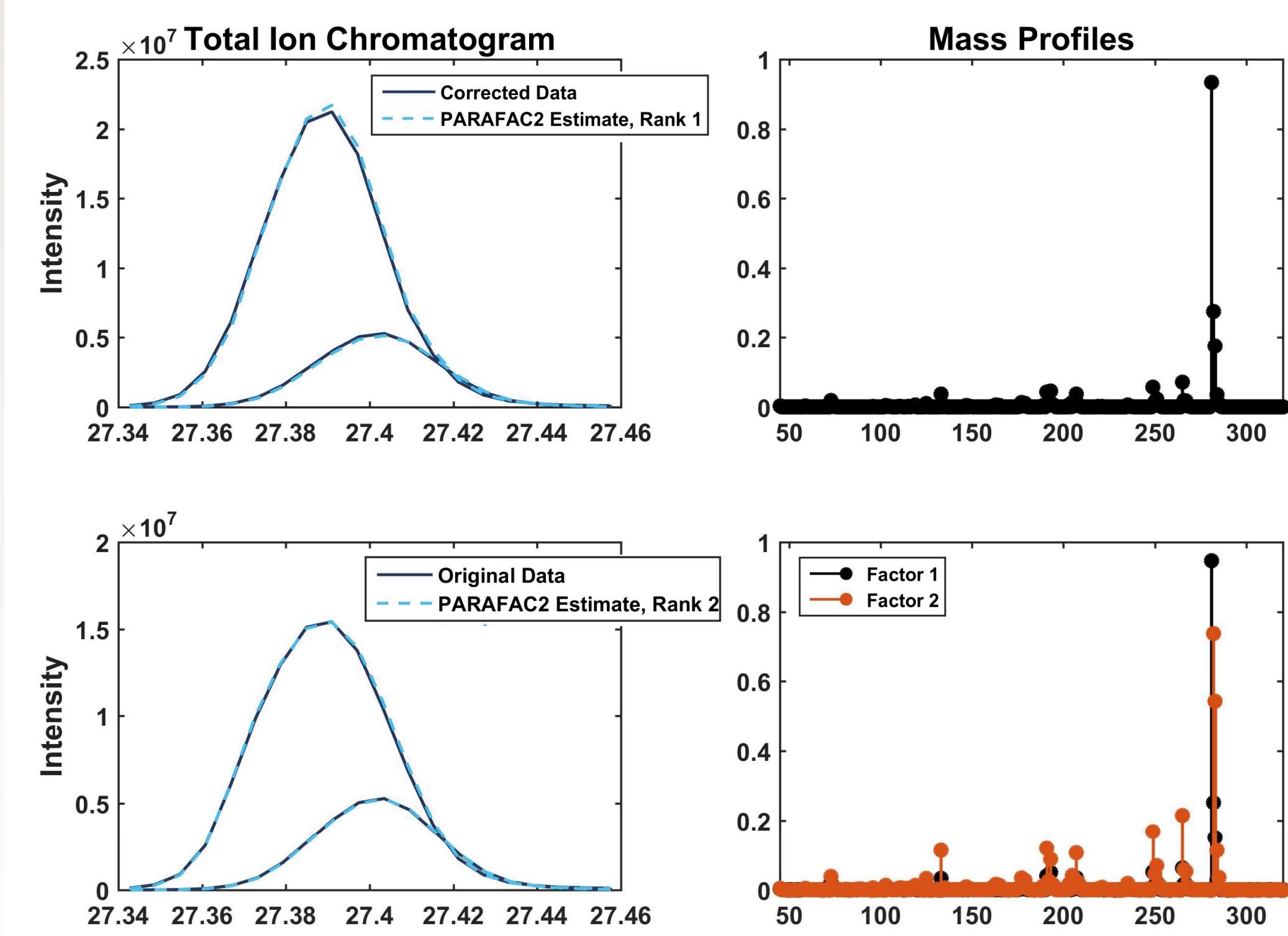
We created an ideal data set using the first principal component of the unsaturated data shown to the left. We created saturated points above a threshold, and corrected those points as described above.

Saturation increased the ratio between the isotope peak at m/z 282 and the peak at m/z 281. The model rank also increased. After the correction, the model rank and isotope ratios corresponded to the unsaturated data.

TEST ON REAL DATA



The corrected, saturated data are now explained by 2 factors (1 background, 1 analyte). The isotope ratios are also nearly flat across the peak. The performance of subsequent multivariate calibration can be improved by this correction. Below are the results of PARAFAC2 models fitted across 10 data sets, corrected and uncorrected. The model fitted to the corrected data required 1 factor while the model fitted to the uncorrected data required 2 factors.



REFERENCES

1. P. Ausloos, et al. "The Critical Evaluation of a Comprehensive Mass Spectral Library". *J. Am. Soc. Mass Spectrom.* 1999; **10**: 287:299
2. M. R. Keenan and P. G. Kotula. "Accounting for Poisson noise in the multivariate analysis of ToF-SIMS spectrum images". *Surf. Interface Anal.* 2004; **36**: 203:212.