

---

# Utilizing Artificial Neural Networks in MATLAB to Achieve Parts-Per-Billion Mass Measurement Accuracy with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

D. Keith Williams Jr., Alexander L. Kovach, David C. Muddiman,  
and Kenneth W. Hanck

W. M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina, USA

Fourier transform ion cyclotron resonance mass spectrometry has the ability to realize exceptional mass measurement accuracy (MMA); MMA is one of the most significant attributes of mass spectrometric measurements as it affords extraordinary molecular specificity. However, due to space-charge effects, the achievable MMA significantly depends on the total number of ions trapped in the ICR cell for a particular measurement, as well as relative ion abundance of a given species. Artificial neural network calibration in conjunction with automatic gain control (AGC) is utilized in these experiments to formally account for the differences in total ion population in the ICR cell between the external calibration spectra and experimental spectra. In addition, artificial neural network calibration is used to account for both differences in total ion population in the ICR cell as well as relative ion abundance of a given species, which also affords mean MMA values at the parts-per-billion level. (*J Am Soc Mass Spectrom* 2009, 20, 1303–1310) © 2009 American Society for Mass Spectrometry

---

**F**ourier transform ion cyclotron resonance (FT-ICR) mass spectrometry has the ability to detect ions with unparalleled mass measurement accuracy, even when working with complex analytes such as proteins, metabolites, and other molecules found in biological, biochemical, and physical science fields [1–5]. The fundamental relationship between cyclotron frequency and  $m/z$  has been under investigation since the introduction of FT-ICR MS in 1974 by Comisarow and Marshall [1]. To reach the highest mass measurement accuracy (MMA) achievable for a given FT-ICR MS system, frequency shifts due to space-charge-effects must be accounted for using external or internal calibration strategies [6, 7]. While internal calibration [8–13] provides the best correction for space-charge effects, it often requires specialized hardware or software.

To account for frequency shifts, external calibration has been utilized to improve the MMA of measurements made by FT-ICR [14]. Amster and coworkers developed a calibration curve to account for the differences in ion populations between the calibration spectrum and subsequent spectra acquired for data analysis resulting in <10 parts-per-million MMA [15, 16]. Based

partly on both the work of Amster [15, 16] and Smith [17], Oberg and Muddiman reported a novel external calibration law which provided data with <5 ppm MMA [18]. Amster and coworkers have also developed stepwise-external calibration, in which calibration spectra are acquired at low trapping voltages that provide low ppm mass accuracy. This is then followed by collection of spectra at more customary trapping voltages where the major peaks, which appeared in the spectrum collected at low trapping voltages, were used as “internal calibrants” to calibrate the rest of the spectrum [19].

Hunt and coworkers have demonstrated the usefulness of combining external calibration with automatic gain control (AGC) where the number of ions in the ICR cell are controlled to fall within the external calibration range [20]. This approach allowed them to routinely achieve MMA of <2 ppm. This external calibration strategy has been implemented on commercially-available FT-ICR MS instruments equipped with AGC [20, 21]. Muddiman and coworkers were able to achieve high MMA using this approach [22]. The Smith and Gygi research groups commonly report MMA values of ~5 ppm when utilizing AGC for external calibration procedures [23–25]. Recently, Muddiman and coworkers utilized a combination of AGC with calibration laws that used multiple linear regression and were able to achieve an external

---

Address reprint requests to Dr. K. W. Hanck, Department of Chemistry, North Carolina State University, Raleigh, NC 27695, USA. E-mail: ken\_hanck@ncsu.edu

calibration with mass measurement accuracy in the parts-per-billion (ppb) range [26].

Artificial neural networks (ANNs) utilize back propagation techniques to establish the weights and biases needed to fit a target output using measured parameters as input data. Unlike multiple linear regression, one does not need to specify a mathematical functional relationship between the input and target data—the ANN is trained to produce a set of outputs that minimize the difference between the target data and the ANN output. The concept of an ANN has been applied in many different fields of chemistry, including various problems in spectroscopy such as the calibration of previously existing spectra as input functions in NMR spectra [27, 28], mass spectra [29–32], infrared spectra [33], and other forms [34]. They have also appeared in the study of chemical sensors applications [35, 36] and protein folding [37]. The concept of having a program that can generate solutions to an unknown via the input of a known is extremely useful [18]. Herein, data from a hybrid LTQ-FT-ICR MS have been applied to an artificial neural network system in an attempt to generate  $m/z$  values with high mass measurement accuracy by utilizing parameters to account for space charge effects within the ICR cell. These results are compared with that of previously published data utilizing multiple linear regression to provide a fit for this data [26].

## Experimental

### Materials

Poly(propylene glycol) with an average molecular weight of 1000 Da, ammonium acetate (>99%), and formic acid were purchased from Sigma (St. Louis, MO). HPLC grade acetonitrile and high-purity water was purchased from Burdick and Jackson (Muskegon, MI). 2-Propanol (HPLC grade) was purchased from Fischer Scientific (Fair Lawn, NJ). All materials were used as received.

A modified version of an electrospray ionization (ESI) source developed previously in this laboratory was coupled to a hybrid ThermoFisher Scientific (San Jose, CA) LTQ-FT Ultra MS equipped with an Oxford Instruments actively shielded 7 T superconducting magnet (Concord, MA). All spectra were acquired with a resolving power at 400  $m/z$  set to 100,000<sub>fwhm</sub>, along with AGC settings ranging from  $5.0 \times 10^5$  to  $3.0 \times 10^6$ . Samples were introduced by direct infusion using a 100  $\mu\text{L}$  gastight syringe (Hamilton, Las Vegas, NV) and the syringe pump on the LTQ-FT Ultra at a flow rate of 0.5  $\mu\text{L}/\text{min}$ . The ESI emitter tips used were 360  $\mu\text{m}$  o.d., 50  $\mu\text{m}$  i.d., and tapered to  $30 \pm 1.0 \mu\text{m}$  i.d. (New Objective, Woburn, MA) and held at a constant potential of 2200 V for all experiments. Electrospray solutions were comprised of 70:30 2-propanol/water with 0.5 mM ammonium acetate ( $\text{NH}_4\text{OAc}$ ).

Calibration of the instrument was completed utilizing a user-defined list of eleven monoisotopic  $m/z$

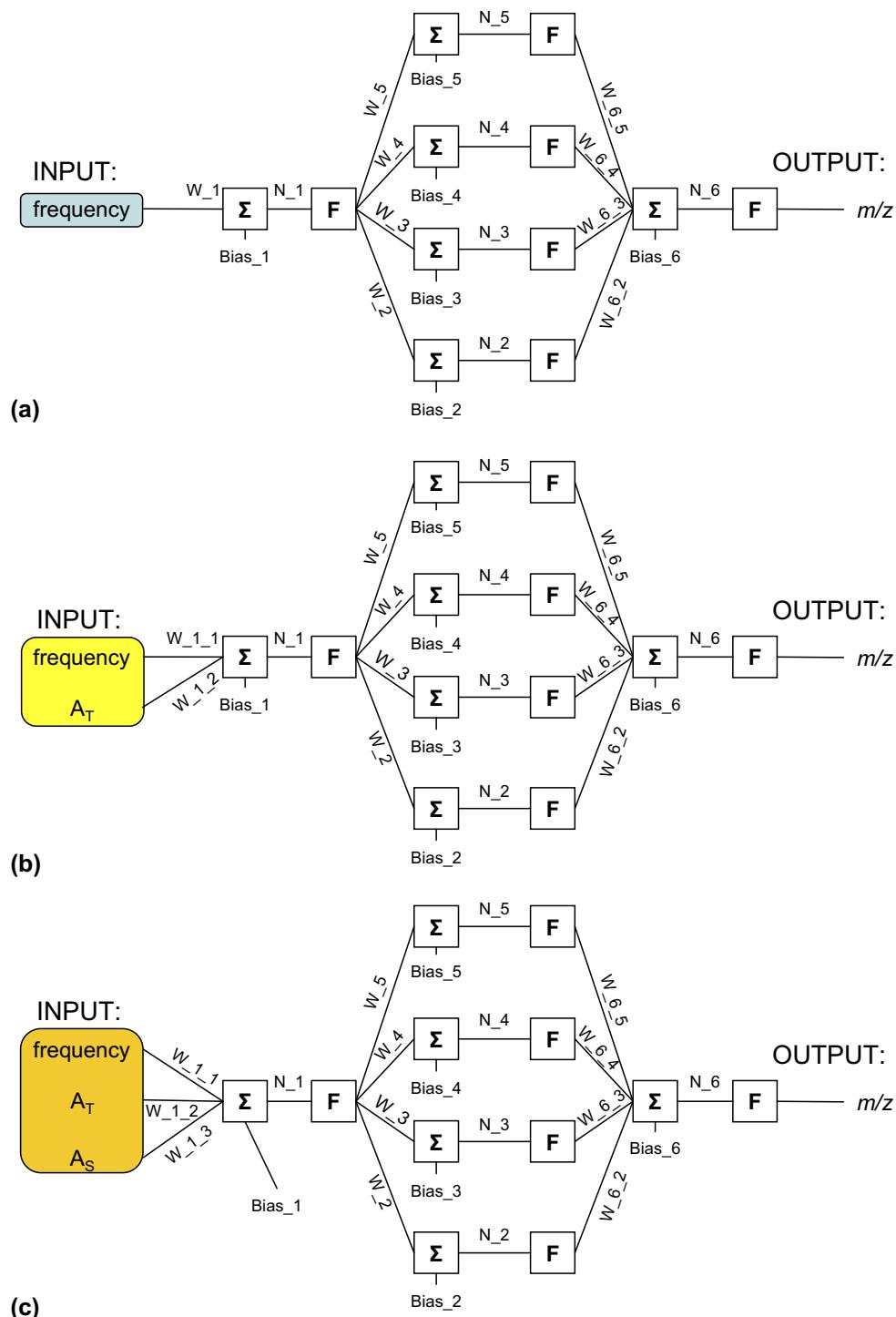
values for ammonium-adducted PPG-1000 oligomers, which ranged from  $m/z$  732 to 1312. This calibration was conducted using the manufacturer's protocol, which generates coefficients for five different AGC target values of which the largest three were used:  $5.0 \times 10^5$ ,  $1.0 \times 10^6$ , and  $3.0 \times 10^6$ .

The frequencies were obtained utilizing diagnostic mode in the instrument software, which allowed for calibration utilizing artificial neural networks. The AGC level was first set to  $5.0 \times 10^5$  and five calibration spectra were recorded, along with their respective cyclotron frequencies. These frequencies provided 55 points (11 data points, 5 spectra) which were used to train the neural network. This was followed by the collection of five validation spectra including their cyclotron frequencies, to determine the achievable MMA through the use of neural networks. Both training and validation datasets were the result of analysis of spectra of ammonium-adducted PPG-1000 oligomers, as these experiments were for proof-of-principle. In an effort to reduce systematic error with respect to AGC levels, this procedure was then repeated for AGC target levels of  $3.0 \times 10^6$  and  $1.0 \times 10^6$  in that order. Values for total ion population ( $A_T$ ) and relative ion population ( $A_S$ ) were calculated as described previously [26].

### Artificial Neural Networks in MATLAB

The basic computational unit of a neural network is the neuron. Our network consisted of six neurons arranged in three layers. The initial layer contained one neuron with a logsig output function. The middle layer contained four neurons each with a logsig output function. The outer layer contained a single neuron with a linear output. Each neuron computes the value of  $N$  via the equation:  $N = W_i * P_i + B_i$  and applies  $N$  to the input of its output function.  $W_i$  is a weight applied to the data by the neuron and  $B_i$  is a bias. The input data (matrix  $P$ ) contains one row for each mass spectrum and one column for each experimentally measured independent variable upon which the calibration is based. Our experiments typically utilized 55 rows and 1, 2, or 3 columns (i.e., frequency,  $A_T$  and  $A_S$ ). A simplified diagram of the neural networks used in this work is shown in Figure 1. The maximum number of weights/biases available to our ANN was 9W/6B, 10W/6B, or 11W/6B depending on the number of independent variables used.

For training of the neural network, the input layer used the calibration dataset, which was obtained through a set of known compounds (PPG-1000 oligomers) that were analyzed by FT-ICR MS. This calibration data allowed the neural network to calculate an output that was compared with the target ( $m/z$  data) and through several iterations determined the weights and biases, which provided outputs closest to those provided in the calibration dataset. Once the desired output is obtained, the same weights and biases are applied to the validation data, which then results in  $m/z$

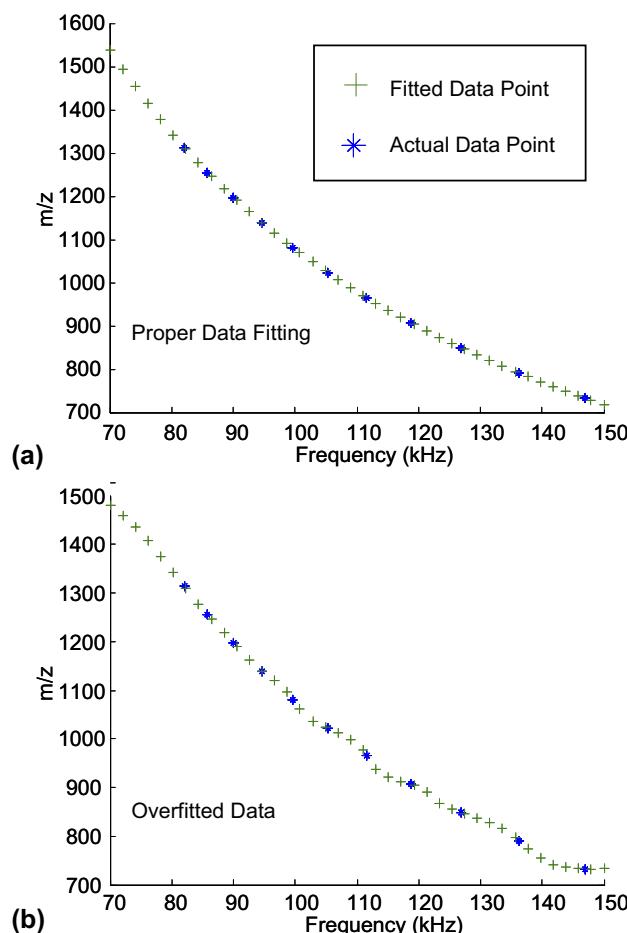


**Figure 1.** (a) Neural network map utilizing frequency data with a single neuron, (b) neural network map utilizing frequency along with total ion population ( $A_T$ ), (c) neural network map utilizing frequency, total ion population ( $A_T$ ), and relative ion abundance of a given species ( $A_S$ ). The first F is the logsig function and last F is the purelin function, as stated in text.

values upon which to base the mass measurement accuracy for the validation data.

Regardless of the number of inputs or outputs, the number of neurons can be changed, which may change the quality of the fit provided by the artificial neural network. There is usually an optimal number of neu-

rons that would allow for the best fit within the system and provide for acceptable computational time; in this case it was determined to be six neurons [38]. This allowed for a high quality, smooth fit of the line through the known points as seen in Figure 2a. Too few neurons (one or two) and the ANN did not converge



**Figure 2.** (a) Graph consisting of an appropriate number of neurons to provide a proper fit, note that the line is a smooth curved line. (b) Graph consisting of too many neurons which resulted in an over fitting of the data.

well or took an extensive amount of time trying to find what would be the best solution. The counter example was taking too many neurons ( $\geq 8$ ), which resulted in the neural network over-fitting the line, which leads to inaccurate prediction of the  $m/z$  values corresponding to particular frequency points, as is shown in Figure 2b. For the experiments described in this paper, all runs and data available in this publication were completed utilizing six neurons. Statistically, the residuals were interpreted as errors and as such would be expected to follow a normal distribution. Several steps were taken to examine the residuals and evaluate whether they were normally distributed.

The data for the Neural Networks program was collected and applied using MATLAB 7.1.4 R14 SP3 for Windows XP or MATLAB 7.2.0.294 (R2006A) for Redhat Enterprise Linux 5. Input data were from AGC target values of  $5.0 \times 10^5$ ,  $1.0 \times 10^6$ , and  $3.0 \times 10^6$ . The trainbr neural network function from the MATLAB Neural Network Toolbox was utilized. Trainbr is a network training function that updates the weight and bias values according to the Levenberg-Marquardt optimization algorithm [38]. It minimizes a combination of

squared errors and then determines the correct combination so as to produce a network that generalizes well. This process is called Bayesian regularization. The output also contained an estimate of the effective number of parameters needed for the fit.

The input and target data were normalized to fall between +1 and -1 using the MATLAB routines map minmax before initiation of the neural network training, but upon output these values are converted back to their proper numerical values. In addition, the training performance target was set to zero and was terminated when validation performance decreased more than two times since the last decrease in performance. Initial neural network weights and biases were assigned using random numbers generated by MATLAB. Typically, 30 training runs were made per dataset. This allowed a wide range of initial weights and biases to be explored. Our criteria for the network that provided the best fit was one, which produced the lowest mean square of residuals, termed MSE, while producing residuals that passed at least four out of five normalcy tests.

Data were analyzed through variations of a script which utilized MATLAB to train a neural network from one dataset, and exploit the trained neural network to produce  $m/z$  values from another dataset, referred to as the "test" dataset. These programs were based on equations previously published by Muddiman [18, 26], which utilized the relative ion abundance of a given species ( $A_S$ ) and the total ion population ( $A_T$ ) of a FT-ICR MS dataset and their effect on mass measurement accuracy. The first program utilized frequency alone and did not incorporate relative ion abundance of a given species nor the total ion population. The second program incorporated the cyclotron frequency and total ion population of a given species ( $A_T$ ), but not the relative ion abundance, and the third program incorporated the cyclotron frequency, total ion population, and the relative ion abundance of a given species.

Key data that was determined via the MATLAB Neural Network program was saved in a data matrix, which included values for the calibration dataset including MMA, median, standard deviation, and other values that would allow for interpretation of the neural network training. The MATLAB program also exported values for the test dataset, which are the values of particular interest since these are the points that the trained neural network actually predicts.

## Results and Discussion

Before applying our ANN to FT-ICR data, the effectiveness of the ANN was evaluated by applying the network to the Filip dataset available from NIST [39]. The initial 82 points were divided into 54 points for training and 28 for validation. The results are summarized in Table 1. The residuals produced by both MLR and ANN were normally distributed and in excellent agreement with the NIST certified values. The MSE values produced by ANN during training were lower than that

**Table 1.** Comparison of MLR and ANN fits to NIST data

	MSE	Residuals	Number parameter	Function
MLR calibration	1.14E-05	Normal	11	10th order polynomial
MLR validation	1.07E-05	Normal		10th order polynomial
ANN training	1.00E-05	Normal	9.7	None needed
ANN validation	2.31E-05	Normal		None needed

produced by MLR; the MSE values produced by ANN during validation were about twice as large as those obtained by MLR. The effective number of parameters utilized by the ANN for the Filip dataset was 9.7 compared with the 11 used by MLR. The effective number of parameters utilized by the ANN were 9.7 (frequency alone), 10.6 (frequency and  $A_T$ ), and 12.3 (frequency,  $A_T$  and  $A_S$ ). The fact that the effective number of parameters increases by about one parameter each time another independent variable is added is consistent with the importance of  $A_T$  and  $A_S$  in the calibration of FT-ICR experiments. The importance of both  $A_T$  and  $A_S$  has been previously reported [18].

Several tests were applied to assess whether the residuals were normally distributed. These were the Lilliefors test for residuals [40], the Shapiro-Wilk parametric hypothesis test of composite normality [41], the D'Agostino-Pearson's L2 test for assessing normality of data using skewness of kurtosis [42], the Jarque-Bera test [43], and the Anderson-Darling test for assessing normality of sample data [44]. The results from these tests, histograms, and normal probability plots demonstrated that a fit using artificial neural networks provided a better fit of data than the MLR method previously published. This is because in all but one case the data provided by the neural network passed more of these normalcy tests than the previously produced MLR results. Also of importance is the mean square of error values (MSE) calculated for the results of each run of the ANN; as this value approaches zero it indicates a better fit.

Figure 3 demonstrates one comparison between the previous scheme utilizing MLR [26] and the results achieved by the ANN method to perform a fit from the calibration dataset. This calibration data were also utilized to generate the coefficients for the MLR equations and to train the ANNs. Figure 3 contains data acquired at an AGC level of  $5.0 \times 10^5$  and utilized only frequency. Figure 3a illustrates the residuals generated by the fitting of the calibration data by MLR. For this example, the points on the normal probability plot deviate from the line that represents where theoretical residuals representing a normal distribution would be located. In addition, the box plot identifies several outlier points (+) that are produced by utilizing the MLR methodology on this particular data. In contrast, Figure 3b shows that utilizing ANN to fit the same data produces residuals with a much more normal distribution. Many more data points of the normal probability plot fall along the line that represents a normal distri-

bution of residuals, and there are no outliers present in the boxplot. In addition, the residuals from ANN fits pass at least four of the five normality tests, which is an improvement over the residuals produced by MLR (failed all five).

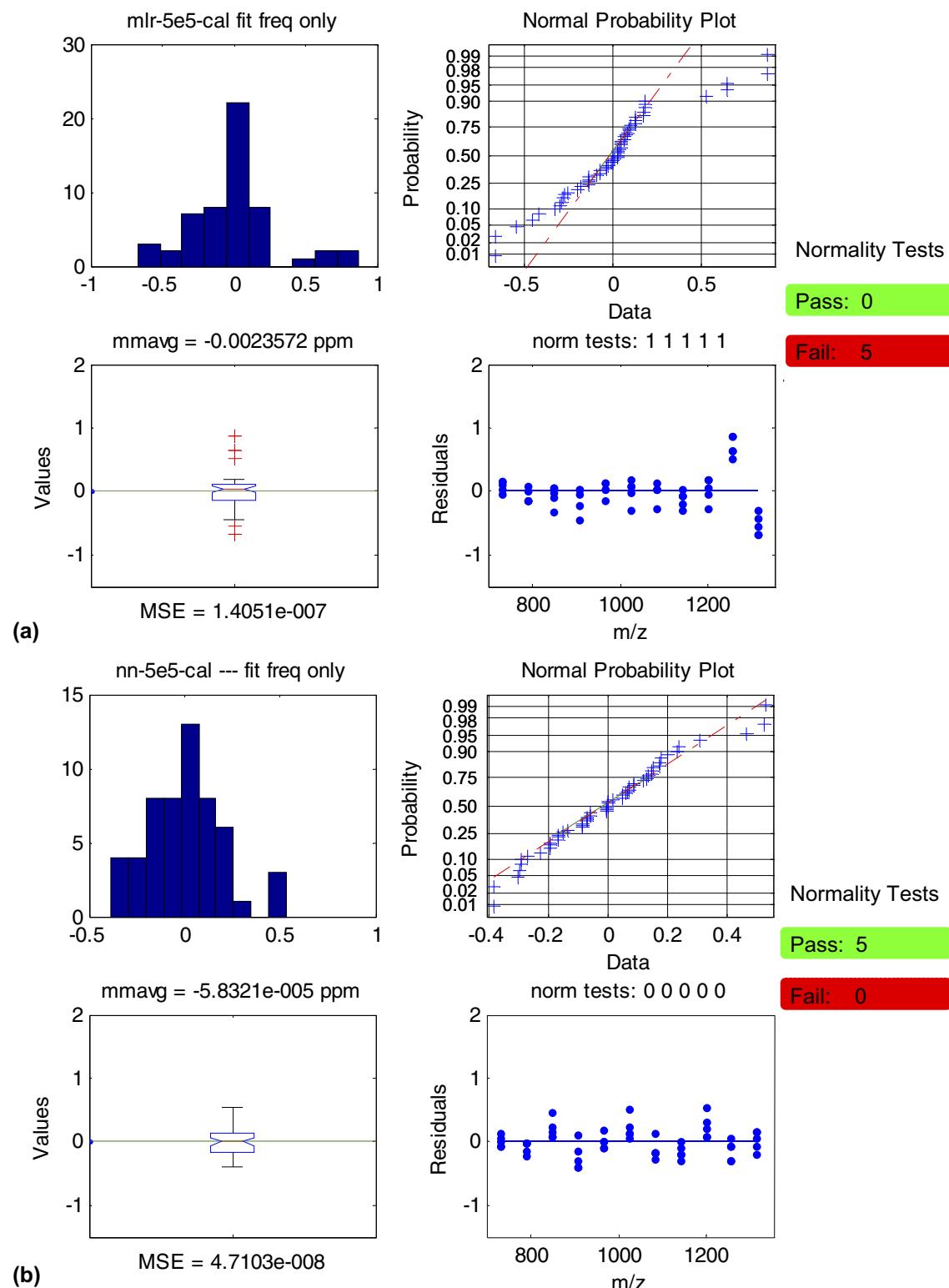
Included in Figure 3 are the average mass measurement accuracy (MMA Avg) and the mean square of error (MSE). If systematic error is eliminated, MMA average should approach zero and the MSE should also approach zero. The MLR method was able to achieve 24 ppb for MMA average versus -0.05 ppb for the fit provided by the ANN, for this dataset. In addition, the MSE value for the ANN fit is also smaller than that provided by the MLR fit of this data. Figure 3 demonstrates only one example of the many datasets that were analyzed in this study. The data for all additional experiments can be found in the Supplementary Material, which can be found in the electronic version of this article.

In addition, data from the validation dataset was utilized to test how well the ANN performed versus the MLR calibration method. The data from the validation dataset also illustrated that artificial neural network provided a better fit for the data. However, the MMA average provided by the ANN did not achieve the same levels as that provided by MLR, with one exception. One possible explanation is the fact that the residuals from the ANN data are normally distributed while those from the MLR are not. The MMA average for every ANN fit did remain sub-ppm for the validation dataset, ranging from 602 ppb to 40 ppb. This and the rest of the data are illustrated in the Supplementary Data section of this manuscript.

All MMA values achieved through the use of artificial neural network calibration and the application of  $A_T$  and  $A_S$  parameters to account for global and local space-charge effects improved those provided strictly from the instrument [26].

## Conclusions

The methods reported herein continue to reinforce the importance of accounting for global and local space-charge effects in FT-ICR to achieve the best possible MMA values. It was demonstrated that mean MMA values in the ppb range can be achieved across a range of AGC target levels utilizing artificial neural networks to calibrate data from a hybrid LTQ-FT-ICR MS. Taking into account the automatic gain control settings, it is important to note that the calibration provides the most



**Figure 3.** (a) For an AGC target value of  $5.0 \times 10^5$ , fit using multiple linear regressions, a histogram of the residuals, normal probability plot, MMA box plot (outliers marked by +), and residual scatter plot are shown. These plots illustrate how well the MLR model utilizing frequency alone fit the data. The results of normalcy tests are also illustrated (0 indicates a pass and a value of 1 indicates fail). (b) The same model using frequency alone is fit using artificial neural networks. This fit provided a more normal distribution of residuals, as demonstrated by the increased number of normalcy tests passed, and the decrease in outliers present in the box plot.

improvement when used at the highest possible population of ions, as expected. Also of importance is the fact that calibration with ANN did provide a better overall fit for these data (as evidenced from the MSE of the residuals), even though the average MMA was not as good as that provided for by the MLR methodology [26]. When the methods described herein are executed properly, we are able to obtain average MMAs between 600 and 40 ppb.

## Supplemental Information

A summary of all data analyzed utilizing ANN and MLR is contained in the Supplementary Information. This summary consists of tabulated data as well as figures in a similar format to Figure 3 for every dataset analyzed.

## Acknowledgments

The authors gratefully acknowledge financial support received from the National Cancer Institute, National Institutes of Health (R33 CA105295), the W. M. Keck Foundation, and North Carolina State University. In addition, D.K.W. thanks Merck and Co. and the ACS Division of Analytical Chemistry for their support.

## Appendix A Supplementary Material

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jasms.2009.02.030](https://doi.org/10.1016/j.jasms.2009.02.030).

## References

- Comisarow, M. B.; Marshall, A. G. Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974**, *25*(2), 282–283.
- Comisarow, M. B.; Marshall, A. G. Frequency-Sweep Fourier-Transform Ion-Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974**, *26*(4), 489–490.
- Gorshkov, M. V.; Guan, S. H.; Marshall, A. G. Masses of Stable Neon Isotopes Determined at Parts-Per-Billion Precision by Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry. *Int. J. Mass Spectrom. Ion Processes* **1993**, *128*(1/2), 47–60.
- Rodgers, R. P.; White, F. M.; Hendrickson, C. L.; Marshall, A. G.; Andersen, K. V. Resolution, Elemental Composition, and Simultaneous Monitoring by Fourier Transform Ion Cyclotron Resonance Mass Spectrometry of Organosulfur Species Before and After Diesel Fuel Processing. *Anal. Chem.* **1998**, *70*(22), 4743–4750.
- He, F.; Hendrickson, C. L.; Marshall, A. G. Baseline Mass Resolution of Peptide Isobars: A Record for Molecular Mass Resolution. *Anal. Chem.* **2001**, *73*(3), 647–650.
- Jeffries, J. B.; Barlow, S. E.; Dunn, G. H. Theory of Space-Charge Shift of Ion-Cyclotron Resonance Frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*(1/2), 169–187.
- Franc, T. J.; Sherman, M. G.; Hunter, R. L.; et al. Experimental-Determination of the Effects of Space-Charge on Ion-Cyclotron Resonance Frequencies. *Int. J. Mass Spectrom. Ion Processes* **1983**, *54*(1/2), 189–199.
- Kaiser, N. K.; Anderson, G. A.; Bruce, J. E. Improved Mass Accuracy for Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(4), 463–470.
- Hanns, J. C.; Muddiman, D. C. A Dual Electrospray Ionization Source Combined with Hexapole Accumulation to Achieve High Mass Accuracy of Biopolymers in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*(10), 876–883.
- Flora, J. W.; Hanns, J. C.; Muddiman, D. C. High-Mass Accuracy of Product Ions Produced by SORI-CID Using a Dual Electrospray Ionization Source Coupled with FTICR Mass Spectrometry. *Anal. Chem.* **2001**, *73*(6), 1247–1251.
- Nepomuceno, A. I.; Mason, C. J.; Flora, J. W.; Hanns, J. C.; Muddiman, D. C. Detection of Genetic Variants of Transthyretin by Liquid Chromatography-Dual Electrospray Ionization Fourier-Transform Ion-Cyclotron-Resonance Mass Spectrometry. *Clin. Chem.* **2004**, *50*(9), 1535–1543.
- Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. Obtaining More Accurate Fourier-Transform Ion-Cyclotron-Resonance Mass Measurements Without Internal Standards Using Multiply Charged Ions. *J. Am. Soc. Mass Spectrom.* **2000**, *11*(5), 416–421.
- Williams, K.; Hawkridge, A. M.; Muddiman, D. C. Sub-Parts-Per-Million Mass Measurement Accuracy of Intact Proteins and Product Ions Achieved Using a Dual Electrospray Ionization Quadrupole Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2007**, *18*(1), 1–7.
- Zhang, L. K.; Rempel, D.; Pramanik, B. N.; Gross, M. L. Accurate Mass Measurements by Fourier Transform Mass Spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, (2), 286–309.
- Easterling, M. L.; Mize, T. H.; Amster, I. J. Routine Part-Per-Million Mass Accuracy for High-Mass Ions: Space-Charge Effects in MALDI FT-ICR. *Anal. Chem.* **1999**, *71*(3), 624–632.
- Taylor, P. K.; Amster, I. J. Space Charge Effects on Mass Accuracy for Multiply Charged Ions in ESI-FTICR. *Int. J. Mass Spectrom.* **2003**, *222*(1/3), 351–361.
- Masselot, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. Mass Measurement Errors Caused by "Local" Frequency Perturbations in FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*(1), 99–106.
- Muddiman, D. C.; Oberg, A. L. Statistical Evaluation of Internal and External Mass Calibration Laws Utilized in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Anal. Chem.* **2005**, *77*(8), 2406–2414.
- Wong, R. L.; Amster, I. J. Sub Part-Per-Million Mass Accuracy by Using Stepwise-External Calibration in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17*(12), 1681–1691.
- Syka, J. E. P.; Marto, J. A.; Bai, D. L.; Horning, S.; Senko, M. W.; Schwartz, J. C.; Ueberheide, B.; Garcia, B.; Busby, S.; Muratore, T.; Shabanowitz, J.; Hunt, D. F. Novel Linear Quadrupole Ion Trap/FT Mass Spectrometer: Performance Characterization and Use in the Comparative Analysis of Histone H3 Post-Translational Modifications. *J. Proteome Res.* **2004**, *3*(3), 621–626.
- Peterman, S. M.; Mulholland, J. J. A Novel Approach for Identification and Characterization of Glycoproteins Using a Hybrid Linear Ion Trap/FT-ICR Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2006**, *17*(2), 168–179.
- Johnson, K. L.; Ovsyannikova, I. G.; Madden, B. J.; Poland, G. A.; Muddiman, D. C. Accurate Mass Precursor Ion Data and Tandem Mass Spectrometry Identify a Class I Human Leukocyte Antigen A\*0201-Presented Peptide Originating from Vaccinia virus. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(11), 1812–1817.
- Dieguez-Acuna, F. J.; Gerber, S. A.; Kodama, S.; Elias, J. E.; Beausoleil, S. A.; Faustman, D.; Gygi, S. P. Characterization of Mouse Spleen Cells by Subtractive Proteomics. *Mol. Cell. Proteom.* **2005**, *4*(10), 1459–1470.
- Belov, M. E.; Zhang, R.; Strittmatter, E. F.; Prior, D. C.; Tang, K.; Smith, R. D. Automated Gain Control and Internal Calibration with External Ion Accumulation Capillary Liquid Chromatography-Electrospray Ionization-Fourier Transform Ion Cyclotron Resonance. *Anal. Chem.* **2003**, *75*(16), 4195–4205.
- Page, J. S.; Bogdanov, B.; Vilkov, A. N.; Prior, D. C.; Buschbach, M. A.; Tang, K.; Smith, R. D. Automatic Gain Control in Mass Spectrometry Using a Jet Disrupter Electrode in an Electrodynamic Ion Funnel. *J. Am. Soc. Mass Spectrom.* **2005**, *16*(2), 244–253.
- Williams, D. K.; Muddiman, D. C. Parts-Per-Billion Mass Measurement Accuracy Achieved Through the Combination of Multiple Linear Regression and Automatic Gain Control in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *Anal. Chem.* **2007**, *79*(13), 5058–5063.
- Radomski, J. P.; Vanhalbeek, H.; Meyer, B. Neural-Network-Based Recognition of Oligosaccharide H-1-NMR Specter. *Nat. Struct. Biol.* **1994**, *1*(4), 217–218.
- Hare, B. J.; Prestegard, J. H. Application of Neural Networks to Automated Assignment of NMR-Spectra of Proteins. *J. Biomol. NMR* **1994**, *4*(1), 35–46.
- Gasteiger, J.; Li, X.; Simon, V.; Novic, M.; Zupan, J. Neural Nets for Mass and Vibrational Spectra. *J. Mol. Struct.* **1993**, *292*, 141–159.
- Werther, W.; Lohninger, H.; Varmuza, K.; Stael, F. Classification of Mass-Spectra—a Comparison of Yes/No Classification Methods for the Recognition of Simple Structural Properties. *Chemom. Intell. Lab. Syst.* **1994**, *22*(1), 63–76.
- Goodacre, R.; Karim, A.; Kaderbhai, M. A.; Kell, D. B. Rapid and Quantitative Analysis of Recombinant Protein Expression Using Pyrolysis Mass Spectrometry and Artificial Neural Networks—Application to Mammalian Cytochrome B(5) in *Escherichia coli*. *J. Biotechnol.* **1994**, *34*(2), 185–193.
- Goodacre, R.; Neal, M. J.; Kell, D. B. Rapid and Quantitative-Analysis of the Pyrolysis Mass Spectra of Complex Binary and Tertiary Mixtures Using Multivariate Calibration and Artificial Neural Networks. *Anal. Chem.* **1994**, *66*(7), 1070–1085.
- Alam, M. K.; Stanton, S. L.; Hebnar, G. A. Near-Infrared Spectroscopy and Neural Networks for Resin Identification. *Spectroscopy* **1994**, *9*(2), 30–38.
- Svozil, D.; Kvasnicka, V.; Pospichal, J. Introduction to Multi-Layer Feed-Forward Neural Networks. *Chemom. Intell. Lab. Syst.* **1997**, *39*(1), 43–62.

35. Kavuri, S. N.; Venkatasubramanian, V. Using Fuzzy Clustering with Ellipsoidal Units in Neural Networks for Robust Fault Classification. *Comput. Chem. Eng.* **1993**, *17*(8), 765–784.
36. Hoskins, J. C.; Himmelblau, D. M. Artificial Neural Network Models of Knowledge Representation in Chemical Engineering. *Comput. Chem. Eng.* **1988**, *12*(9/10), 881–890.
37. Qian, N.; Sejnowski, T. J. Predicting the Secondary Structure of Globular Proteins Using Neural Network Models. *J. Mol. Biol.* **1988**, *202*(4), 865–884.
38. Hagan, M. T.; Menhaj, M. B. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Trans Neural Networks* **1994**, *5*(6), 989–993.
39. Filippelli,, A. Statistical Reference Datasets, Online DG. *NIST*, <http://www.itl.nist.gov/div898/strd/l1s/data/Filip.shtml>.
40. Lilliefors HW. On Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *J. Am. Stat. Assoc.* **1967**, *62*(318), 399–402.
41. Shapiro, S. S.; Wilk, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611.
42. D'Agostino, R. B.; Pearson, E. S. Tests for Departure from Normality. Empirical Results for the Distributions of  $b_2$  and  $\sqrt{b_1}$ . *Biometrika* **1973**, *60*, 613–622.
43. Jarque, C. M.; Bera, A. K. Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals. *Econ. Lett.* **1980**, *6*(3), 255–259.
44. Anderson, T. W.; Darling, D. A. Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *Ann. Math. Stat.* **1952**, *23*(2), 193–212.