Electron temperature profile reconstructions from multi-energy SXR measurements using neural networks

2013 Plasma Phys. Control. Fusion 55 095015

(http://iopscience.iop.org/0741-3335/55/9/095015)

Plasma Phys. Control. Fusion **55** (2013) 095015 (8pp) doi:10.1088/0741-3335/55/9/095015

Electron temperature profile reconstructions from multi-energy SXR measurements using neural networks

D J Clayton[1], K Tritz[1], D Stutman[1], R E Bell[2], A Diallo[2], B P LeBlanc[2] and M Podestà[2]

[1] Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA
[2] Princeton Plasma Physics Laboratory, Princeton, NJ 08543, USA

E-mail: ClaytoDJ@nv.doe.gov

Received 2 April 2013, in final form 14 June 2013
Published 7 August 2013
Online at stacks.iop.org/PPCF/55/095015

Abstract
Neural networks have been implemented to reconstruct electron temperature profiles from multi-energy soft-x-ray (ME-SXR) arrays and other plasma diagnostics with fast time resolution. On NSTX, electron temperature profiles are measured with a Thomson scattering diagnostic at 60 Hz, a speed limited by the repetition rate of the lasers. By training a neural network to match fast (>10 kHz) x-ray data with T_e profiles from Thomson scattering, the ME-SXR diagnostic can be used to produce T_e profiles with high time resolution. In particular, a new ME-SXR system will be used in conjunction with a new laser blow-off impurity injection system to measure cold pulse propagation in NSTX-U plasmas for direct, perturbative heat transport measurements. Synthetic ME-SXR data were used to optimize performance of the neural networks and study the impact of including data from various diagnostics in the networks. Initial tests on data from a previous-generation ME-SXR diagnostic on NSTX have proven successful.

(Some figures may appear in colour only in the online journal)

1. Introduction

In the National Spherical Torus eXperiment (NSTX) [1] and similar ST experiments, electron temperature and density are primarily measured via Thomson scattering [2]. While these diagnostics provide accurate, localized measurements of electron temperature T_e and electron density n_e, the time resolution is limited by the repetition rate of the powerful lasers used. While conventional tokamaks and other magnetic fusion devices may use ECE diagnostics for fast temperature measurements, the relatively low fields and high density of the ST result in a plasma frequency that far exceeds the electron cyclotron frequency and prohibits the use of these diagnostics. An alternative method to measure electron temperature with fast time response is to use soft-x-ray (SXR) emissivity measurements [3–8]. The limitations of using SXR emissivites for temperature measurements are due to the complex dependence of x-ray emissivity on T_e, n_e, and the concentration and charge state distribution of each impurity present in the plasma. Two-filter techniques, where the ratio of x-ray emissivity measurements from two x-ray diagnostics with different x-ray filters are compared to theoretical values, have been used to produce temperature estimates in some fusion devices, usually when the measured x-ray brightness is dominated by continuum emission or when impurity line emission is small and well-defined [5–8]. For NSTX, in which x-ray emissivity is dominated by line radiation, multi-energy SXR (ME-SXR) diagnostics, with three or more filtered x-ray arrays have been developed [9, 10]. The additional arrays place constraints on the impurity concentrations and improve temperature measurements [3, 4]. Previously, analytical methods have been used to extract T_e from ME-SXR data, involving atomic modeling of impurity line and continuum emission. While these models may be sufficient to simulate emission from carbon, the material used in NSTX plasma-facing components (PFCs), NSTX-Upgrade (NSTX-U) [11] plans to begin replacing the PFCs with high-Z materials such as molybdenum or tungsten. These high-Z impurities generate quasi-continuum emission that can be quite difficult to model. Additionally, impurity ions are generally assumed to be in

coronal equilibrium in these models, though in the presence of large impurity transport, this is no longer the case and impurity transport modeling should be used.

An alternative approach to finding T_e profiles from ME-SXR data, without relying on complex atomic data or transport models, is to use neural networks. Neural networks have previously been employed in magnetic fusion research to solve a variety of problems, including fast magnetic equilibrium reconstruction (originally described in [12]), real-time disruption detection (originally in [13], see [14] for a recent summary of work in this area), fast tomographic reconstruction [15], and fast spectroscopic analysis [16–19]. They have been used to determine parametric dependencies of plasma quantities such as radiated power [20] and energy confinement time [21]. Neural networks have also been used for fast estimates of plasma parameters such as total radiated power [22] and deuterium density profiles [23]. A review of neural networks with some applications to magnetic fusion can be found in [24]. Similarly, a neural network can be trained by matching ME-SXR data with actual T_e and n_e profiles from Thomson scattering, then applied to x-ray data between Thomson measurements to provide increased time resolution. Additionally, once a network is trained with sufficient data, it can be applied to discharges without Thomson scattering, or it can be used to produce a real-time measurement of the T_e and n_e profiles. Such networks have been tested with synthetic x-ray data to study their behavior, and with real experimental data to demonstrate their capabilities. These results are described below, following a description of the ME-SXR diagnostics and the design of the neural networks.

2. Description of ME-SXR diagnostics

The basic concept of a ME-SXR diagnostic can be implemented in many ways. The results in section 5 were obtained with data from a scintillator-based diagnostic referred to as the optical SXR (OSXR) array [9]. This instrument consisted of three horizontal rows of 16 channels, each with overlapping mid-plane tangential views from the outboard scrape-off layer to just inside the magnetic axis. Each of the three rows had a different x-ray filter, with 10, 100 and 300 μm Be foils. The OSXR diagnostic had a time resolution of $\sim$10 kHz. An updated design, utilizing diode-based detectors in place of scintillators for better detector efficiency, was tested during the last NSTX run campaign [10]. One advantage of the diode-based system was that it utilized variable-gain preamplifiers, which could be remotely adjusted to account for changes in x-ray brightness in the plasma edge as well as accommodate a wide range of filters. This system, also with a tangential mid-plane view, consisted of five 20-channel diode arrays, with a view from $r/a \sim 0.6$ to outside the scrape-off layer. 0.3 μm Ti and 5, 15, and 50 μm Be foils were used with four of the arrays, while the fifth array had no filter, essentially operating as a bolometer. Depending on the preamplifier gains, this diagnostic had a time resolution $\sim$10–100 kHz. While the analysis described in this paper is being tested on data from these diagnostics, with the hope of gleaning new physics understanding from a large database collected during NSTX operations, it is primarily being developed for use with a new ME-SXR diagnostic, currently under construction, for NSTX-U. This diode-based diagnostic will be similar in design to the previous edge system, but will be installed inside the NSTX-U vacuum vessel and will consist of both core-viewing and edge-viewing arrays.

3. Description of the neural networks

The neural networks described here were implemented with the PyBrain 0.3.1 modular machine learning library for Python 2.7 [25]. The results that follow were obtained with simple feedforward networks; there were no feedback loops. These feedforward networks are multi-layer perceptrons with three layers: an input layer, a hidden layer and an output layer. The addition of a second hidden layer was briefly tested but the resulting networks required more time to train with no noticeable improvements in performance. The layers are fully connected, with all input nodes connected to all hidden nodes, and all hidden nodes connected to all output nodes. No alternative connection schemes were tested. Figure 1 shows the basic design of the networks. Diagnostic data fed into each input node are scaled linearly from 0 to 1, normalized to the maximum input value for that node in the training dataset. The inputs are propagated to the hidden nodes through each connection, multiplied by a weight associated with that connection. A sigmoid activation function, in the form of the logistic function $f(x) = 1/(1+e^{-x})$, is applied to the summed inputs of each hidden node. The outputs of the hidden nodes are then propagated to the output nodes, again multiplied by weights associated with each connection. The output nodes simply sum all of their inputs. In the training datasets, the outputs of the network are also normalized to the maximum values of the dataset, scaled from 0 to 1.

The Rprop- (resilient backpropagation without weight-backtracking) learning algorithm performed best amongst the supervised training algorithms available in PyBrain and was used to train the networks [26]. The algorithm optimizes the connection weights to minimize the mean square error between the network output and training data. The datasets used to train the networks were collections of data from different instances in time in a set of NSTX discharges, with an input node for each radial chord of each diagnostic. There are typically 16 or 20 input nodes per ME-SXR array. The T_e output nodes were given the same radial resolution as an ME-SXR array (again, typically 16 or 20 nodes). Unless otherwise stated, the number of hidden nodes was set to 40, for reasons discussed in the next section. The networks were trained for 1000 epochs, with 25% of the data used as a validation dataset to prevent over-fitting. Each network takes minutes to tens of minutes to train on a single desktop CPU.

4. Characterization of the networks using synthetic x-ray data

4.1. Generating synthetic data

The neural networks were first tested with synthetic x-ray data to better understand their performance. Synthetic datasets

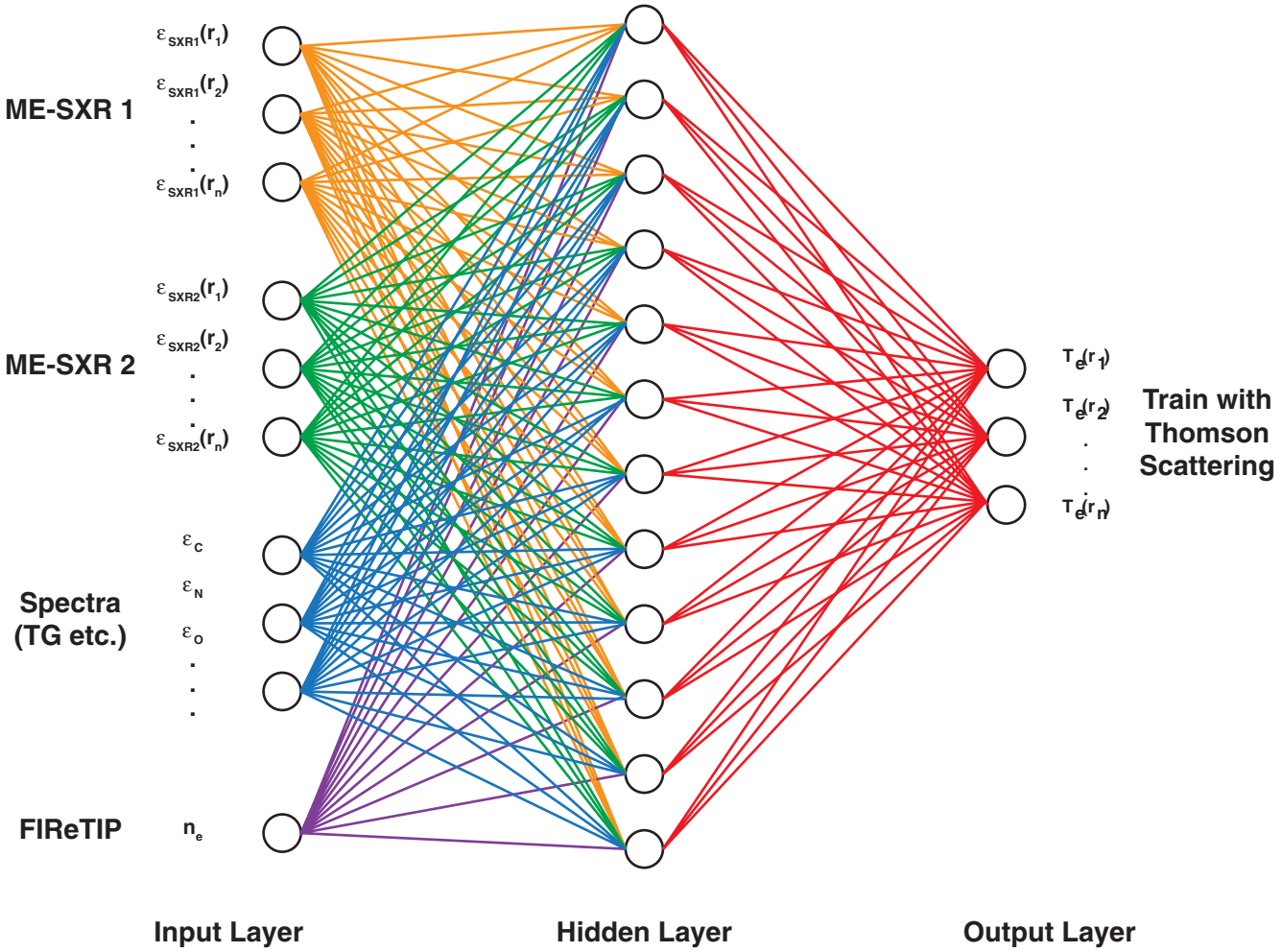

Figure 1. Representation of a fully connected three-layer feedforward neural network. Raw data is scaled linearly from 0 to 1 and fed into the input nodes, with one node for each radial position for each diagnostic; results from the output nodes are similarly scaled.

were generated using real electron temperature and density profiles from Thomson scattering and carbon density profiles from charge exchange recombination spectroscopy (CHERS) measurements [27]. Additional impurities were assumed to have the same radial profile as carbon, with an oxygen density of 0.2 n_C, nitrogen density of 0.1 n_C, and iron density of 0.001 n_C, values consistent with spectroscopic measurements of typical NSTX discharges [28, 29]. The charge state distributions of all impurities were assumed to be in coronal equilibrium. The CHIANTI atomic database [30, 31] was used to calculate line and continuum emission for each impurity. Synthetic ME-SXR data were generated by accounting for the spectral response of each array and line-integrating the emission along each line of sight. Gaussian noise was added to each signal to mimic real data. The final dataset was obtained from a set of thirteen discharges that included a scan of both plasma current ($I_p = 0.8$–1.1 MA) and toroidal field ($B_t = 0.4$–0.55 T). Only temperature profiles with $T_e > 200$ eV were included; each discharge provided about 50 Thomson profile measurements, generating a total of 658 training sets.

Simulated data from other diagnostics were generated to add additional constraints to the neural networks. To better constrain the impurity concentrations in the plasma, synthetic spectra from the NSTX UV transmission grating based imaging spectrometer (TGIS) were generated [32]. The TGIS diagnostic has a tangential view of the plasma mid-plane with radial resolution from the magnetic axis to the outboard scrape-off layer. The synthetic TGIS data simulated line-integrated UV brightness, spectrally resolved into 20 bins from 30 to 700 Å and spatially resolved into 20 radial bins. These data were generating using the same methods described for the ME-SXR data. To account for the slow time response of the TGIS diagnostic (400 ms in NSTX will be upgraded to ~10 ms in NSTX-U), an additional, uniform, Gaussian offset was applied across all channels to simulate an overall drift in impurity concentrations in time. As an additional test, electron density was constrained by including line-integrated electron density in the network, as would be measured by the NSTX far infrared tangential interferometer/polarimeter (FIReTIP) [33].

4.2. Results with synthetic data

The first goal accomplished with the synthetic data was to optimize the structure of the neural networks. After settling on the basic design of the network (feedforward, three layers,

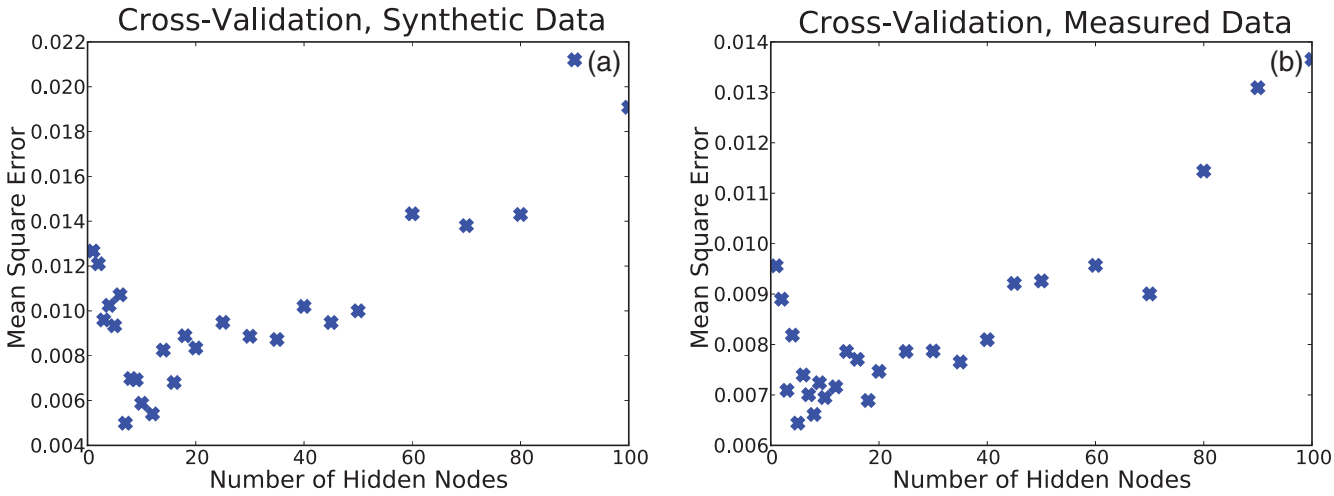

Figure 2. Mean square error from cross-validation of neural networks for varying numbers of hidden nodes using (*a*) the synthetic dataset described in this section and (*b*) the measured dataset described in the next section.

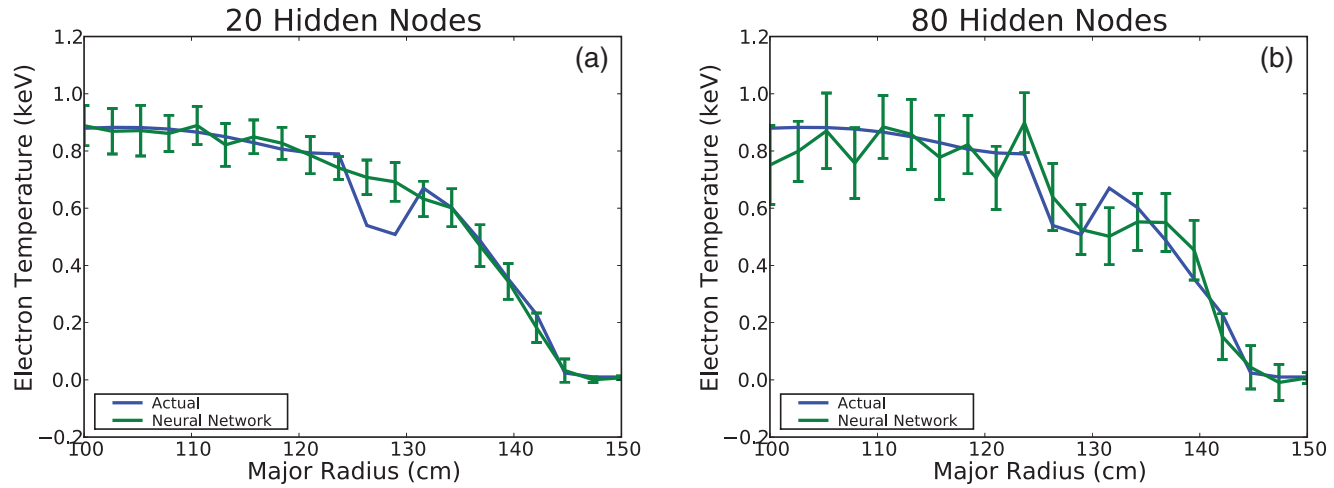

Figure 3. For tests with synthetic data, neural network outputs are compared to actual T_e profiles from Thomson scattering. The error bars represent the rms difference between the two, averaged over the test dataset. While the rms values decrease with the number of nodes in the hidden layer, spatial features may be missed. Here, an artificial dip in the T_e profile is introduced to test networks with (*a*) 20 nodes and (*b*) 80 nodes.

fully connected), the optimal number of hidden nodes needed to be determined. This was accomplished by using five-fold cross-validation, in which the dataset from the thirteen discharges was randomly divided into five subsets, then each of the five subsets were used to validate networks trained with the remaining 4/5 of the data. The validation metric used was the mean square error averaged over the five subsets, which was calculated for varying numbers of hidden nodes and is plotted in figure 2(*a*). While networks with few hidden nodes provide the closest fits, interesting radial features may be washed out in the resulting smooth temperature profile. Figure 3 shows that when the synthetic temperature was depressed within a small radial band, a neural network with 20 hidden nodes did not recognize this radial feature. By comparison, when 80 hidden nodes were used, the total rms error increased, but the network provided better radial detail, capturing the localized dip in temperature. To balance the overall accuracy of the network with radial resolution, 40 hidden nodes were used in all subsequent examples. Cross-validation was also applied to the measured x-ray dataset described in the next section, with similar results shown in figure 2(*b*).

Next, the effect of different diagnostics on the performance of the neural networks was tested. To begin, the importance of the ME-SXR diagnostic was studied by testing varying numbers of filtered SXR arrays. As seen in figure 4(*a*), for a 900 eV test case not included in the original training dataset, a single array was sufficient to recreate the temperature profile, with additional arrays providing little added benefit. The same held true when the electron density was artificially increased in the synthetic test data; in fact, the network with only one array produced the lowest rms error. However, when the electron temperature is increased, as in figure 4(*b*), it is revealed that in fact the network with one array is unresponsive to changes in the data, and that two or more arrays are needed to reconstruct the temperature. This is expected, since data from one array should be insufficient to distinguish changes in temperature from density. Figure 4(*c*) shows that the networks with 2–3 arrays perform reasonably well when the impurity

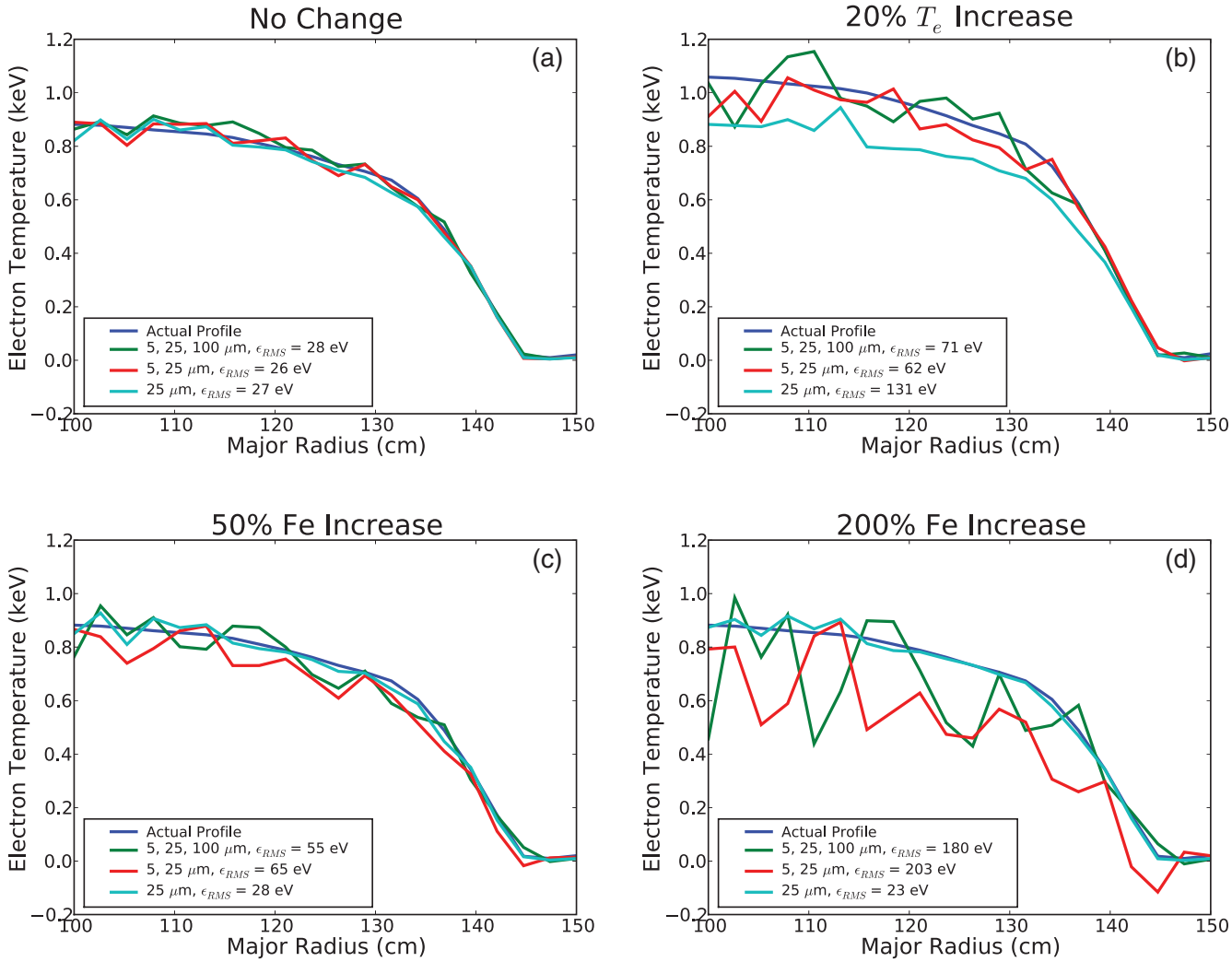

Figure 4. (*a*) A synthetic test case is generated to compare results from networks with 1, 2 and 3 ME-SXR arrays. This same test case is then retried with (*b*) a 20% increase in T_e, (*c*) a 50% increase in n_{Fe}, and (*d*) a 200% increase in n_{Fe}. Note in each case, the network with only one array was unresponsive to changes in the data. The rms errors quoted are averaged over the 20 radial nodes.

concentrations used to generate the test data are varied within reasonable bounds. However, when there is a large increase in impurity density, such as might be the case when a flake of material from the plasma-facing components enters the plasma, the networks perform poorly, shown in figure 4(*d*). This should not be surprising, as this is a scenario not included in the training dataset.

One way to improve the neural network performance, particularly in cases such as large impurity injections described above, is to add data from additional diagnostics into the network, particularly spectroscopic data. When synthetic TGIS data were included in the network, performance improved in virtually all cases, and drastically in the case of the large impurity injection. These results are summarized in figures 5(*a*)–(*c*). In each case, adding TGIS data to the network decreased the rms error by as much as 50%. Other NSTX-U spectrometers, while not providing spatial resolution, might still contribute additional constraints to the networks, and have the added advantage that they could be used in real time. Additional diagnostics that could improve network performance include any diagnostic signals relating to the electron density. One such diagnostic is the FIR interferometer (FIReTIP), with a single-chord, line-integrated density measurement that could also be used in real time. Adding synthetic FIReTIP data improves performance by up to 50%, as shown in figure 5(*d*). For best performance, as many relevant diagnostics should be included in the neural networks as is reasonable.

Finally, in addition to using these neural networks to find T_e profiles from ME-SXR data, one might consider using these networks to calculate electron density profiles. This was attempted with the same synthetic data, and while the neural network found a reasonable match to the density profile for the test case used in figure 4, it was unresponsive to changes in the simulated n_e profile, just as the network with only one ME-SXR array was unresponsive to changes in T_e. This makes some sense, since with the exception of the one-channel FIReTIP data, all other data scales as $n_e n_I$, where the electron and impurity densities cannot be deconvolved. In the future, additional data that depends on n_e alone, such as beam emission spectroscopy (BES) data, may be added to the neural networks to find fast n_e profiles in addition to T_e profiles.

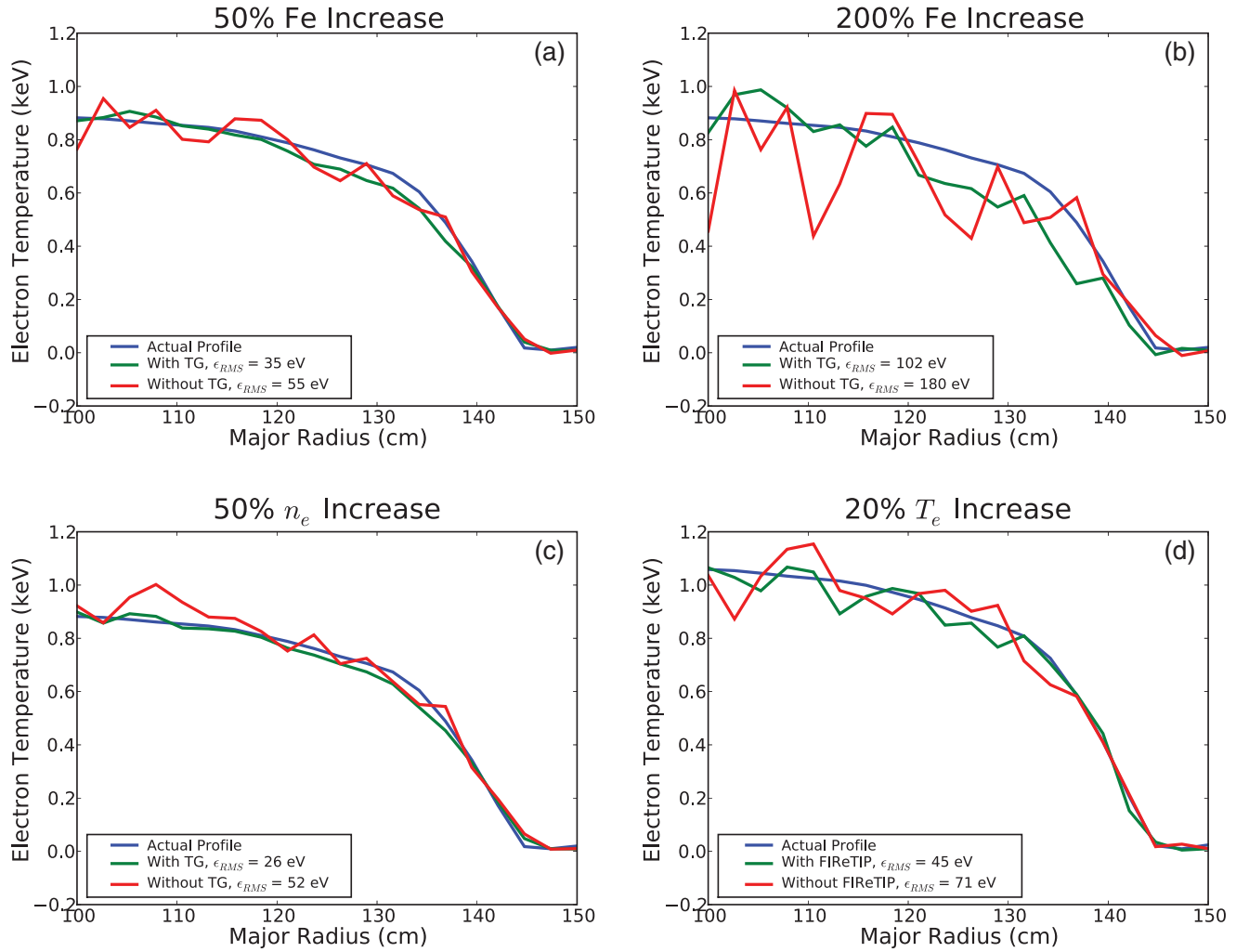

Figure 5. Using the test case from figure 4, results from networks with and without TGIS data (with data from three ME-SXR arrays in each case) are compared for (*a*) a 50% increase in n_{Fe}, (*b*) a 200% increase in n_{Fe} and (*c*) a 50% increase in n_{e}. (*d*) A comparison with and without FIReTIP data for a 20% increase in T_{e}. The rms errors quoted are averaged over the 20 radial nodes.

5. Initial tests with experimental data

To confirm that neural networks could perform with real data as well as synthetic data, this method was applied to data from the previous-generation OSXR diagnostic. Data were used from one run day of NSTX operation; day-to-day variability in the diagnostic configuration prevented the inclusion of data from additional run days. The dataset contains 30 discharges with a total of 1831 input patterns, roughly half of which included gas puffs of neon, a bright emitter in the SXR range. Neon concentration was negligible in the remainder of the discharges. No other diagnostics were included in this test, thus the network had 48 input nodes (16 channels $\times 3$ arrays), 40 hidden nodes, and 16 output nodes. Once the network was trained with this dataset, an additional discharge from this run day was tested.

The results of these initial tests are illustrated in figures 6 and 7, and appear quite promising. Figure 6 is a contour plot of temperature as a function of time and major radius obtained from the neural network and shows significant variation in the temperature profile on time scales faster

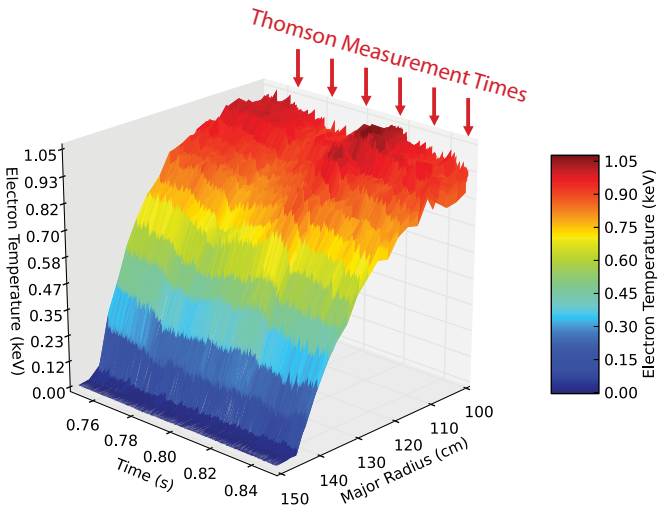

Figure 6. Time-evolving T_{e} profile found with a neural network applied to optical ME-SXR data. The red arrows indicate when the Thomson scattering diagnostic makes a measurement.

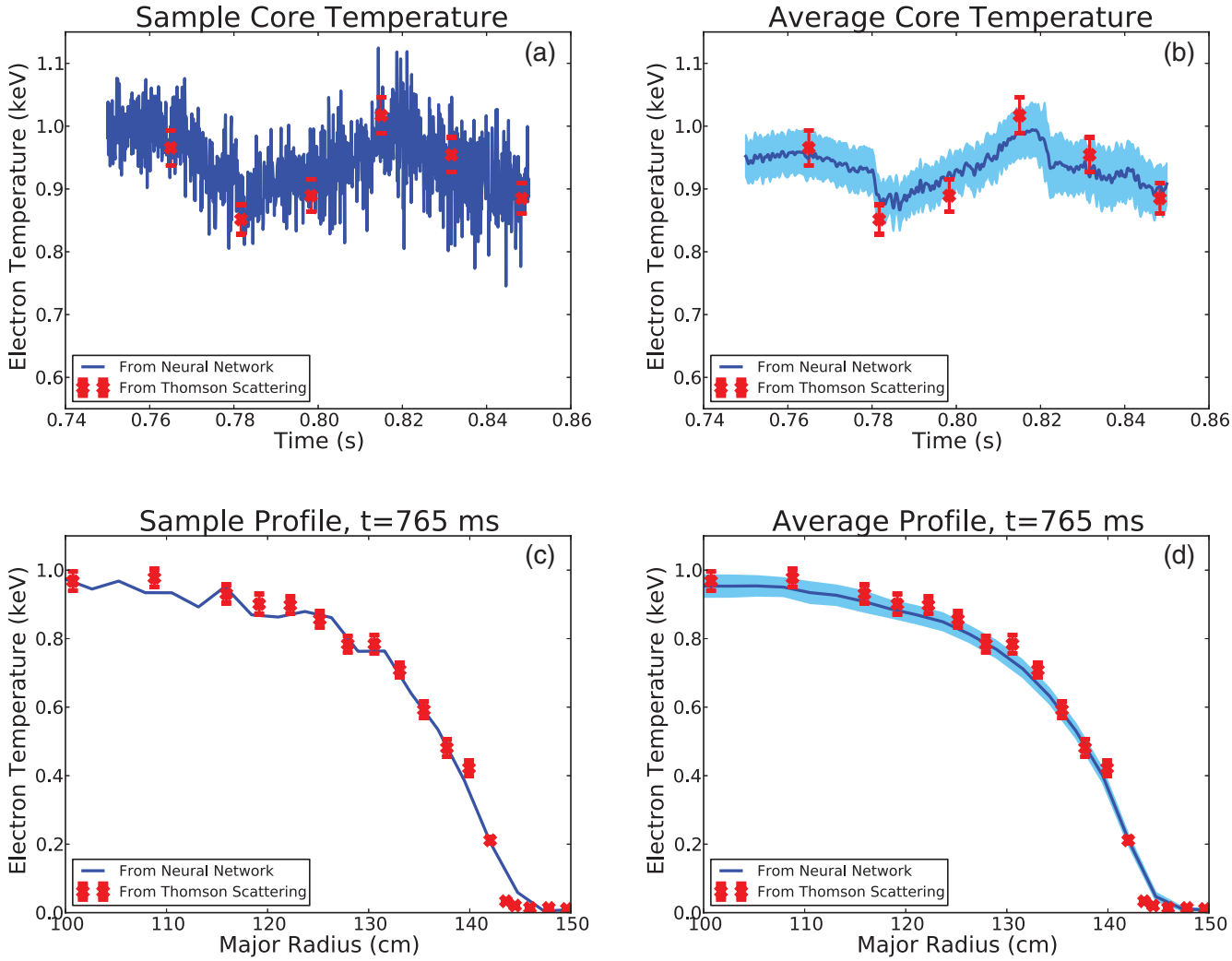

Figure 7. Results from figure 6 are compared to Thomson measurements. (*a*) Time evolution of T_e at the core. (*b*) Average T_e at the core, resulting from retraining the network 100 times, smoothed to 1 kHz. (*c*) Radial T_e profile at 765 ms. (*d*) Average radial profile from retraining the network 100 times. The shaded regions indicate the standard deviation.

than the frequency of the Thomson scattering measurements. This is highlighted in figure 7(*b*), an averaged, 1 kHz smoothed version of the raw 10 kHz time trace of core electron temperature obtained from the network and plotted in figure 7(*a*). It can be seen that a large drop in temperature, occurring between Thomson scattering measurements at 765 and 782 ms, is due to a 1 ms temperature crash at 780 ms. Figure 7(*c*) shows a close fit between a temperature profile from the neural network and from measurement, with a typical RMS error <5%. The network was retrained 100 times, each time with new, randomly generated initial weights, and the averages and standard deviations of the outputs of these 100 networks are shown in figures 7(*b*) and (*d*). The variation in the output of the networks is of the same order as the experimental error in the Thomson scattering measurements.

6. Conclusions

In summary, simple neural networks have proven capable of producing electron temperature profiles from multi-energy SXR measurements, when trained with a database of T_e profiles from Thomson scattering measurements. Reasonable results can be generated with networks trained with about one run day worth of data. The accuracy of the network increases with a decrease in the number of nodes in the hidden layer, at the cost of radial detail. Using synthetic x-ray data, it was found that at least two ME-SXR arrays were needed to distinguish changes in temperature. While a network trained with ME-SXR data alone was adequate in most cases, the addition of spectroscopic data improved performance by as much as 50%, particularly when there were large changes in impurity concentrations. Additional data, such as line-integrated electron density from a single-chord FIReTIP diagnostic, further improved performance. An example has been presented in which a network was applied to real ME-SXR data, and time-resolved T_e profiles were obtained that compared favorably to Thomson scattering measurements, matching within 5%. In NSTX-U, perturbative electron heat transport measurements will be performed by measuring the propagation of cold pulses with a new, in-vessel ME-SXR diagnostic, and tracking the changing T_e profile with the use of

neural networks. These networks may also be incorporated in the NSTX-U control system for real-time feedback, following a study of the applicability of these networks to data from separate experimental campaigns.

Acknowledgment

This work was supported by the United States DoE contract number DE-FG02-09ER55012.

References

[1] Ono M *et al* and the NSTX Team 2000 *Nucl. Fusion* **40** 557
[2] LeBlanc B P, Bell R E, Johnson D W, Hoffman D E, Long D C and Palladino R W 2003 *Rev. Sci. Instrum.* **74** 1659
[3] Delgado-Aparicio L F, Stutman D, Tritz K, Finkenthal M, Bell R, Hosea J, Kaita R, LeBlanc B, Roquemore L and Wilson J R 2007 *J. Appl. Phys.* **102** 073304
[4] Delgado-Aparicio L F *et al* 2007 *Plasma Phys. Control. Fusion* **49** 1245
[5] Kiraly J *et al* 1987 *Nucl. Fusion* **27** 397
[6] Murari A, Franz P, Zabeo L, Bartiromo R, Carraro L, Gadani G, Marrelli L, Martin P, Pasqualotto R and Valisa M 1999 *Rev. Sci. Instrum.* **70** 581
[7] Franz P *et al* 2006 *Rev. Sci. Instrum.* **77** 10F318
[8] Baião D, Medina F, Ochando M, McCarthy K, Tabarés F, Pastor I and Varandas C 2012 *Rev. Sci. Instrum.* **83** 053501
[9] Delgado-Aparicio L F, Stutman D, Tritz K, Finkenthal M, Kaita R, Roquemore L, Johnson D and Majeski R 2004 *Rev. Sci. Instrum.* **75** 4020
[10] Tritz K, Clayton D J, Stutman D and Finkenthal M 2012 *Rev. Sci. Instrum.* **83** 10E109
[11] Menard J E *et al* and the NSTX Team 2012 *Nucl. Fusion* **52** 083015
[12] Lister J B and Schnurrenberger H 1991 *Nucl. Fusion* **31** 1291
[13] Hernandez J V, Vannucci A, Tajima T, Lin Z, Horton W and McCool S C 1996 *Nucl. Fusion* **36** 1009
[14] Cannas B, Fanni A, Pautasso G, Sias G and Sonato P 2010 *Nucl. Fusion* **50** 075004
[15] Demeter G 1997 *Rev. Sci. Instrum.* **68** 1438
[16] Morgan W L, Larsen J T and Goldstein W H 1994 *J. Quant. Spectrosc. Radiat. Transfer* **51** 247
[17] Bishop C M and Roach C M 1992 *Rev. Sci. Instrum.* **63** 4450
[18] Baker D R, Groebner R J and Burrell K H 1994 *Plasma Phys. Control. Fusion* **36** 109
[19] Svensson J, von Hellermann M and König R W T 1999 *Plasma Phys. Control. Fusion* **41** 315
[20] Svensson J, Mohanti R, Lawson K, von Hellermann M and Meigs A 2001 *Plasma Phys. Control. Fusion* **43** 405
[21] Allen L and Bishop C M 1992 *Plasma Phys. Control. Fusion* **34** 1291
[22] Barana O, Murari A, Franz P, Ingesson L C and Manduchi G 2002 *Rev. Sci. Instrum.* **73** 2038
[23] Svensson J, von Hellermann M and König R W T 2001 *Plasma Phys. Control. Fusion* **43** 389
[24] Bishop C M 1994 *Rev. Sci. Instrum.* **65** 1803
[25] Schaul T, Bayer J, Wierstra D, Sun Y, Felder M, Sehnke F, Rückstieß T and Schmidhuber J 2010 *J. Mach. Learn. Res.* **11** 743
[26] Igel C and Hüsken M 2003 *Neurocomputing* **50** 105
[27] Bell R E and Feder R 2010 *Rev. Sci. Instrum.* **81** 10D724
[28] Lepson J K, Beiersdorfer P, Clementson J, Gu M F, Bitter M, Roquemore L, Kaita R, Cox P G and Safronova A S 2010 *J. Phys. B: At. Mol. Opt. Phys.* **43** 144018
[29] Kumar D, Finkenthal M, Stutman D, Bell R E, Clayton D J, Diallo A, LeBlanc B P, Podesta M and Tritz K 2012 *Plasma Phys. Control. Fusion* **54** 065010
[30] Dere K P, Landi E, Mason H E, Monsignori Fossi B C and Young P R 1997 *Astron. Astrophys. Suppl. Ser.* **125** 149
[31] Landi E, Del Zanna G, Young P R, Dere K P and Mason H E 2012 *Astrophys. J.* **744** 99
[32] Kumar D, Stutman D, Tritz K, Finkenthal M, Tarrio C and Grantham S 2010 *Rev. Sci. Instrum.* **81** 10E507
[33] Park H K, Domier C W, Geck W R and Luhmann N C Jr 1999 *Rev. Sci. Instrum.* **70** 710