Fast analysis of collective Thomson scattering spectra on Wendelstein 7-X
J. van den Berg, I. Abramovic, N. J. Lopes Cardozo, and D. Moseev

Citation: Review of Scientific Instruments 89, 083507 (2018); doi: 10.1063/1.5035416
View online: https://doi.org/10.1063/1.5035416
View Table of Contents: http://aip.scitation.org/toc/rsi/89/8
Published by the American Institute of Physics

Articles you may be interested in
Tomography diagnostic of plasma radiated power on the spherical tokamak Globus-M
Review of Scientific Instruments 89, 083509 (2018); 10.1063/1.5039801

Description of complex viewing geometries of fusion tomography diagnostics by ray-tracing
Review of Scientific Instruments 89, 083506 (2018); 10.1063/1.5031087

Development and characterization of a helicon plasma source
Review of Scientific Instruments 89, 083508 (2018); 10.1063/1.5030624

Modelling of nitrogen seeding experiments in the ASDEX Upgrade tokamak
Physics of Plasmas 25, 032506 (2018); 10.1063/1.5019913

Influence of neutral beam attenuation on beam emission spectroscopy and charge exchange recombination spectroscopy
Review of Scientific Instruments 89, 073503 (2018); 10.1063/1.5028205

Design, capabilities, and first results of the new laser blow-off system on Wendelstein 7-X
Review of Scientific Instruments 89, 073505 (2018); 10.1063/1.5037543
Fast analysis of collective Thomson scattering spectra on Wendelstein 7-X

J. van den Berg, I. Abramovic, N. J. Lopes Cardozo, and D. Moseev

Department of Applied Physics, Eindhoven University of Technology, De Zaale, 5612 AJ Eindhoven, The Netherlands
Max-Planck-Institut für Plasmaphysik, Greifswald 17491, Germany

(Received 15 April 2018; accepted 9 July 2018; published online 17 August 2018)

Two methods for fast analysis of Collective Thomson Scattering (CTS) spectra are presented: Function Parametrization (FP) and feedforward Artificial Neural Networks (ANNs). At this time, a CTS diagnostic is being commissioned at the Wendelstein 7-X (W7-X) stellarator, with ion temperature measurements in the plasma core as its primary goal. A mapping was made from a database of simulated CTS spectra to the corresponding ion and electron temperatures ($T_i$ and $T_e$). The mean absolute mapping errors are 4.2% and 9.9% relative to the corresponding $T_i$ for the ANN and FP, respectively, for spectra with Gaussian noise equivalent to 10% of the average of the spectral maxima in the database at 650 sampling points per GHz and within a limited parameter space. Although FP provides some insight into the information contents of the CTS spectra, ANNs provide a higher accuracy and noise robustness, are easier to implement, and are more adaptable to a larger parameter space. These properties make ANN mappings a promising all-round method for fast CTS data analysis. Addition of impurity concentrations to the current parameter space will enable fast bulk ion temperature measurements in the plasma core region of W7-X. https://doi.org/10.1063/1.5035416

I. INTRODUCTION

In the quest for energy generation by controlled nuclear fusion, experiments are conducted in devices that confine hot (hundred million C) plasma of hydrogen isotopes in a toroidal geometry by means of a strong magnetic field. The two leading magnetic confinement concepts are the tokamak and the stellarator. Both feature a helical magnetic field in a toroidal geometry, but where the tokamak is characterized by a strong electric current inside the plasma, the field in the stellarator is entirely generated by external coils. While this comes at the cost of a more complex design, the potential advantages are that the stellarator is intrinsically steady state and has more resilience against plasma instabilities. The most recent large stellarator experiment is the Wendelstein 7-X (W7-X) device in Greifswald (Germany), which saw its first successful campaign in 2016. It aims at demonstrating the reactor potential of the so-called optimized stellarator configuration. In order to gain information on the plasma properties in fusion reactors, a variety of diagnostic techniques are employed.

This paper focuses on Collective Thomson Scattering (CTS), a diagnostic that measures the scattered waves from a narrow, collimated beam of high power mm-waves that is launched into the plasma. The collected waves contain information on the ions in the plasma and can in principle be used to obtain ion-related plasma parameters, such as the bulk ion temperature, ion velocity distribution, and plasma composition.

To describe the scattering process and compute the scattered spectra, the forward code eCTS is available. For the interpretation of measured spectra, the eCTS code is integrated in the generic Bayesian data analysis framework Minerva, implemented at W7-X.

However, finding the best fit using the eCTS code requires many runs, which takes up to several minutes of computation per spectrum. In this paper, we explore the possibility to develop a custom-made inverse mapping, by analyzing a data set of simulated spectra that covers the expected operational space of the experiment. We compare the so-called Function Parametrization (FP) which defines an inverse mapping based on functions that are used to interpolate between the points in the database and the more generic approach of Artificial Neural Networks (ANNs). Both techniques require the construction of a data set of spectra, a form of data reduction such as binning or Fourier analysis, and result in an inverse mapping that can be applied to the data in real time. Moreover, the fast inverse mapping can be used to do optimization studies, e.g., to tune the diagnostic setup for a particular parameter, or sensitivity analysis, e.g., to investigate the feasibility of the measurement of additional quantities, such as the radial electric field, for given experimental noise. These techniques are general and have been applied to other diagnostics. We applied them to the specific case of CTS at W7-X, for the given set of parameters. Once the methodology has been set up, creating a mapping for different parameters is a relatively simple action.

We first give a short description of the W7-X environment and the CTS diagnostic and the mapping parameter space in Secs. II A and II B and then discuss the spectrum information content, dimension reduction, and example database in Sec. II C. The principles of the FP and ANN techniques and their implementations are recalled in Secs. II D and II E, respectively. The mapping results are described in Sec. III, and the relative merits of both methods are discussed in Sec. IV.

0034-6748/2018/89(8)/083507/9/$30.00
II. METHODS

A. Collective Thomson scattering on W7-X

Collective Thomson scattering is the scattering of electromagnetic radiation off of fluctuations in the electron density, electric field, magnetic field, and current density. Scattering of the probing radiation is collective provided that the condition \( \left( \lambda_D k^6 \right)^{-1} > 1 \) is satisfied,\(^9\) where \( \lambda_D \) is the Debye length, \( k^6 = ik^6 \), and \( k^6 = k^6 - k^6 \) specifies the direction along which the fluctuations are resolved, with \( k^6 \) being the wave vectors of incident radiation and scattered radiation, respectively. The main contribution to the bulk-ion part of the scattering spectrum usually comes from the electron density fluctuations. It is a powerful mm-wave diagnostic that has successfully been applied on a number of fusion experiments.\(^10-14\) CTS spectra are sensitive to a broad range of parameters, including the electron density, ion species densities, magnetic field, and ion velocity distributions.\(^15\) With a broad range of sensitivities comes a broad range of potentially measurable parameters, including the ion temperature, isotope ratio, fast ion velocity distribution, plasma rotation, and MHD activity. A CTS system is being commissioned for the upcoming experimental campaign on the stellarator W7-X. The first CTS measurements on W7-X will be aimed at diagnosing the bulk ions. To this end, one of the 140 GHz electron cyclotron resonance heating gyrotrons, with the output power of approximately 500 kW, will be used as the source of the probing beam. High power of the probing beam is necessary due to the small scattering cross section of CTS. The local scattering process depends on geometric parameters including the direction and polarization of the incident beam, the local direction of the magnetic field, and local plasma properties such as density and temperature. In addition, the incoming and scattered waves propagate through a dispersive plasma in which cutoffs, resonances, and refraction may occur. Locally, the CTS scattering geometry is determined by the following two angles: the observation angle \( \phi = \angle \left( \mathbf{B}, k^6 \right) \) and the scattering angle \( \theta = \angle \left( k^6, k^6 \right) \). In this work, the geometry \( \phi = 110^\circ \) and \( \theta = 130^\circ \) in the bean-shaped plasma cross section was used, of which a schematic representation is shown in Fig. 1.

Inference of the parameter values from the measured spectra is typically performed using the Bayesian formalism in conjunction with a forward code of the scattering.\(^16\) The developed forward model of CTS (eCTS)\(^16\) calculates the spectral power density of the scattered radiation, as given by\(^17\)

\[
\frac{\partial P_s}{\partial \omega} = P_{in} O_b \frac{\omega^3}{2\pi c^2} r_e^2 n_e G S(k^6, \omega),
\]

with \( P_{in} \) as the probing beam power, \( O_b \) as the overlap volume, \( \omega^3/\omega \) as the angular frequencies of incident and scattered radiation, respectively, \( \omega = \omega - \omega' \), \( r_e \) as the classical electron radius, \( n_e \) as the local electron density, \( G \) as the so-called geometrical form factor, and finally \( S(k^6, \omega) \) as the spectral density function, which governs the shape of the CTS spectra, and is in turn influenced by a variety of plasma parameters.\(^10\) Note that the spatial power density, \( n_e, G, k, \) and \( S(k^6, \omega) \) are assumed to be constant over the overlap volume, simplifying the overlap integral of incident and observed radiation to a multiplication of scattered power and volume. In the eCTS code, the spectral density function is calculated in the electrostatic approximation, where only the scattering off fluctuations in the electron density is taken into account.\(^2\) The spectral density function is given by\(^18\)

\[
S(k^6, \omega) = S_e(k^6, \omega) + \sum_i S_i(k^6, \omega),
\]

with \( S_e \) as the electron scattering term and \( S_i \) as the scattering terms due to electrons screening ion species \( i \).

The input parameters of eCTS are the incident radiation mode, scattered radiation mode, \( P_{in}, \omega, \omega', O_b, n_e \), ion species densities \( n_i \), electron temperature \( T_e \), ion temperature \( T_i \), strength of the magnetic field \( B \), observation angle \( \phi \), and the angle between the probing and scattered waves \( \theta = \angle \left( k^6, k^6 \right) \). Eight CTS spectra computed by the eCTS code, at different \( T_i \) and fixed scattering geometry, are shown in Fig. 2.

B. The mapping domain

The product of this work is a mapping function that, for a given CTS spectrum, provides the corresponding plasma parameters \( T_i \) and \( T_e \). This mapping is obtained by analysis of a database containing synthetic CTS spectra and their corresponding eCTS input parameters. The domain of the mapping is determined by the range of input parameters covered in the database. To minimize the number of dimensions in the mapping, it was decided to only include input parameters that influence the shape of the CTS spectrum.

The electron density, \( P_{in}, O_b \), and the attenuation of the collection system are assumed to uniformly influence the measured spectral power density. These scaling factors were
excluded from the mapping by normalizing the spectra to their mean value.

The domain of the other parameters was limited by their possible variation at a specific setup of the CTS diagnostic. That is, this mapping does not require input from external measurements and its results can be applied for CTS as a standalone diagnostic. First, there is an uncertainty in scattering geometry. Although the placement of the diagnostic parts is fixed, this mapping does not require input from external measurements and its results can be applied for CTS as a standalone diagnostic. First, there is an uncertainty in scattering geometry. Although the placement of the diagnostic parts is fixed, the refraction of probing and scattered light. These terms do not only depend on the properties of other plasma species. A quantitative analysis of measured CTS spectra requires a quantitative inclusion of impurities in the input parameter space. Two methods of dimension reduction are considered: binning and the Discrete Fourier Transform (DFT).

For binning, the spectrum is divided in bins, the output of which is given by the average spectral power density in each bin. The noise of the binned spectrum is given by \( \sigma_{\text{noise}} = \sigma_{\text{signal}} / \sqrt{BF} \), with \( BF \) as the number of sampling points in each bin.

Figure 3 shows the DFTs of several noiseless CTS spectra with varying bulk ion temperature. In this work, only the magnitude, not the phase, of the Fourier components was considered. The amplitude of the Fourier components strongly decreases with the index, and (for \( T_i > 1 \text{ keV} \)) only the first 10–20 components can carry information distinguishable from noise. In the Fourier space, the process of binning is very similar to removing the higher Fourier components. To show this, a CTS spectrum with \( \rho_s = 650 \text{ GHz}^{-1} \) was binned with several binning factors, followed by a DFT. The normalized amplitudes of these DFTs are shown in Fig. 4. First, the differences in the Fourier components above the noise level are negligible for \( \rho_s \leq 12.5 \text{ GHz}^{-1} \). Therefore, with the given parameters and noise level, binning down to 12.5 GHz\(^{-1}\) does not significantly reduce the amount of useful information in the spectrum. Second, it can be seen that the noise level is equal for all binning factors.

### Table I. Expected values of the relevant plasma and geometry parameters during measurement and the parameter ranges that were used in the mapping.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Expected value</th>
<th>Mapping parameter range</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_i ) (keV)</td>
<td>&lt;4</td>
<td>0.5–10</td>
</tr>
<tr>
<td>( T_e ) (keV)</td>
<td>( \leq 8 )</td>
<td>0.5–10</td>
</tr>
<tr>
<td>( \theta ) (deg)</td>
<td>130 ± 2</td>
<td>125–135</td>
</tr>
<tr>
<td>( \phi ) (deg)</td>
<td>110 ± 2</td>
<td>105–115</td>
</tr>
<tr>
<td>( B(T) )</td>
<td>2.2 ± 0.05</td>
<td>2.1–2.3</td>
</tr>
<tr>
<td>( n_i/n_e )</td>
<td>0.2–1</td>
<td>0.2–1.0</td>
</tr>
</tbody>
</table>

---

**FIG. 2.** Collective Thomson scattering spectra as calculated by the eCTS code, normalized to their mean value. The varied parameter is the ion temperature. The other input parameters were \( T_e = 5 \text{ keV}, \theta = 130^\circ, \phi = 110^\circ, B = 2.2 \text{ T}, \) and \( n_i = n_e \).
tra, however, scales linearly with $\rho$. Spectral densities covering the expected domain of parameter space (described in Table I).

The database was built by doing $N = 25\,000$ simulations with the eCTS model, yielding vectors $p$ and $q$ with $(\alpha = 1, \ldots, N)$. $p^\alpha$ were picked from a uniform random distribution covering the expected domain of parameter space (described in Table I).

The input to each simulation consists of the input parameters $p^\alpha$ and the 50 sampling frequencies in the range of $140 \pm 2$ GHz. The eCTS code then produces the corresponding spectrum. According to Sec. II B, the scaling factor was removed by normalizing each spectrum to its mean value. A final step in the simulation was to apply a notch filter. The notch width is assumed to be 300 MHz, centered at the probing frequency 140 GHz. In principle, the notch can be applied by cutting out the corresponding part of the spectrum. But since this would introduce artefacts in the DFT of the spectra, the notch was applied by replacing the values within the notch by a linear interpolation between the spectral values bordering the notch.

Next to the database built for learning, several test sets were built. These had equal parameter domains, except for the temperatures. Considering that the W7-X plasma will be heated through the electron cyclotron resonance, the temperature range was limited to $\frac{1}{10} < T_i/T_e < 3$.

1. Principal component analysis

Depending on the mapping method, 50 variables might still be too much for a stable mapping. On the other hand, the binned or Fourier-transformed spectrum might still contain information that is unnecessary for an adequate mapping. The number of dimensions can be further reduced through a Principle Component Analysis (PCA).

In a PCA, a set of observations of a number of possibly correlated variables is transformed to a different set of values of uncorrelated variables, called the principle components. In this case, the set of observations are the database spectra $q^\alpha$. The principle components can be sorted with respect to their variance in the database and thus with respect to their relative importance in the database. Because the principle components are uncorrelated, one can use PCA to reduce the number of variables in the data without discarding the most significant patterns in the database.

The process of a PCA on a database $q^\alpha$ is as follows: first, the database is normalized: for each element $j$ in $q^\alpha$, $q'^j_j = (q^\alpha_j - \langle q^\alpha_j \rangle)/\sigma_j$, with $\sigma_j$ as the standard deviation of variable $j$ and $\langle q^\alpha_j \rangle$ as the mean value of $q^\alpha_j$ in the database. Then, the covariance matrix $D^q$ of the normalized variables is calculated, the elements of which are given by

$$D^q_{ij} = \frac{1}{N-1} \sum_\alpha q'^j_\alpha q'^i_\alpha.$$  

(4)

The matrix $D^q$ contains the cross correlations between the elements of $\tilde{q}$. The eigenvalues $\lambda_j$ of $D^q$ are a measure of the regressive significance of the corresponding eigenvectors $D^q e_j = \lambda_j e_j$. The principal components of each spectrum are then given by $\tilde{q}^\alpha = e_j \cdot \tilde{q}^\alpha$. If $\lambda_j > (3\sigma_j)^2$, with $\sigma_j$ as the measurement noise, $\tilde{q}_j$ is considered a significant component.

The eigenvalues and corresponding eigenvectors are sorted such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, with $N_\alpha$ as the number of significant eigenvectors. One can now make a mapping from the significant components in $\tilde{q}$ to $p$. The application of the PCA in the function parametrization is described in Subsection II D.

D. Function parametrization

After the dimension reduction is complete, a mapping from the processed variables $q_1$ to the eCTS input parameters $p$ can be made. The first mapping method considered in this work is function parametrization (FP). A FP, a regression is performed on a database of examples, in which the error of a certain mapping function $p = M(q_1)$ is minimized.

For the regression in FP, one has to select a function $f$ that can adequately describe the trends in the model.

\begin{align*}
\end{align*}
eCTS model is non-linear, so the regressor must be too if the whole parameter range is to be covered. Multivariate Adaptive Regression Splines (MARS)\textsuperscript{19} is a regression algorithm that uses non-linear hinge functions as basis for a multivariate model with a multi-dimensional output. The regression consists of a forward pass where terms are added to the model and a backward pass where the least useful terms are removed from the model one by one (pruning). Because it minimizes the complexity of the model, the backward pass tends to prevent overfitting. However, if the model obtained in the forward pass cannot fully capture the patterns in the database, pruning is unlikely to yield much results.

The function parametrization in this work was obtained by performing a MARS regression on the principle components q of the discrete Fourier transform of the spectra in the database. Only the first 10 principle components were used because it was found that the computation time and systematic errors of the resulting mapping increase drastically for N\textsubscript{q} \geq 10. Moreover, the maximum number of terms in the model was limited to 80, each with a maximum degree of three (corresponding to the product of three hinge functions). Gaussian noise was added to the database spectra in order to stabilize the MARS regression. The amplitude of this noise was equal to 0.1\% of the mean of the maximum values of each spectrum in the database.

All computation for this research was performed on an Intel i5-6300U dual-core processor along with 20 GB of RAM.

The obtained FP was applied to test sets with different amounts of noise and ρ\textsubscript{s} = 12.5 GHz\textsuperscript{-1}. The results for σ\textsubscript{s,650} = 10\% are presented in Sec. III A. The results for tests at different noise levels (σ\textsubscript{s,650} = 1\%, 5\%, 10\%, 30\%, and 100\% at ρ\textsubscript{s} = 12.5 GHz\textsuperscript{-1}) are presented in Sec. IV.

1. Error analysis

The simplest error analysis is the comparison of the approximation p\textsubscript{inv} with the eCTS input parameters p. This can be done for each individual example to get a grasp of the estimation error in different parts of the parameter subspace. An error indication can also be obtained for the FP as a whole, defined as follows:

\[ ε_i ≡ \frac{1}{N\text{test}} \sum_{β} \left| \frac{p_{β,\text{inv}}}{p_β} \right|, \]

with β as the test spectrum index and N\text{test} as the number of spectra in the test set.

During operation, p will not be available. Reduced chi-squared analysis is proposed as a self-consistent method to provide an indication of the goodness of fit of a spectrum corresponding to the measured p\textsubscript{out}, with the observed spectrum. This was implemented as follows:

\[ χ^2 ≡ \sum_{n=1}^{N} \left( \frac{q_n - S(ω_n, \text{Out})}{σ_s} \right)^2, \]

where q\textsubscript{n} are the normalized spectral power densities measured at frequencies ω\textsubscript{n}. S is the spectral power density as calculated by the forward model, including artificially added noise with a magnitude equal to the measurement noise σ\textsubscript{s}. The vector p\textsubscript{out} contains the measured parameters. Reduced chi-squared is then \( χ^2_{\text{red}} = χ^2/K \), with K as the number of degrees of freedom. The value of \( χ^2_{\text{red}} \) can be interpreted in the following way: The closer the \( χ^2_{\text{red}} \) is to 1, the better the fit is. Values of \( χ^2_{\text{red}} > 1 \) represent an underfit, whereas \( χ^2_{\text{red}} < 1 \) represents overfitting. However, it is not possible to use \( χ^2_{\text{red}} \) as a quantitative measure of the goodness of fit.\textsuperscript{20} \( χ^2_{\text{red}} \) has an intrinsic probability distribution function, which can only be known for the true model having the true parameter values.

\( χ^2_{\text{red}} \) can still provide an indication of self-consistency: under the assumption that the approximation (be it FP or a neural network) follows the true model, \( χ^2_{\text{red}} \) of the most accurate measurements should follow the \( χ^2_{\text{red}} \) distribution and thus have a mean value of 1. Moreover, a correlation between ε and \( χ^2_{\text{red}} \) might provide a rough error estimation for measurements. The results of this analysis are presented in Sec. III A 1.

E. Artificial neural networks

An artificial neural network (ANN) is a computing system that consists of a network of artificial neurons: a mathematical function that mimics the behavior of biological neurons. Neurons mutually transmit signals which depend on the signals they receive from other neurons. Although ANNs are much simpler than biological neural networks, they share the ability to learn new patterns through the consideration of examples.

In this research, a specific class of feedforward neural network was used: the multi-layer perceptron (MLP). The perceptron is a class of ANNs that was invented by Rosenblatt,\textsuperscript{8} initially as a hardware single binary classifier but later generalised to a computation system consisting of neurons arranged in layers, which can be activated by inputs from all neurons in the preceding layer and transmit their output o to the succeeding layer. This is summarised in the following recurrent formula:

\[ o_k^l = φ(\text{net}_k^l) = φ\left( \sum_{i=1}^{N_l} o_{i-1}^l w_{i-1,k}^l + b_k^l \right). \]

where i and k denote the layer and neuron indices, respectively, net\textsubscript{k} is the net input of neuron k in layer i, formed by the sum of all N\textsubscript{i} neuron outputs of the previous layer, multiplied by the corresponding weights w\textsubscript{i-1,k} and plus b\textsubscript{k}, the bias of neuron k. Finally, φ describes the “activation” of a neuron based on the information it receives. This so-called activation function should be a strictly increasing and non-linear function between 0 and 1 in the domain (−∞, ∞). Many activation functions can be used, but for this work, the standard logistic function and its convenient derivative were used,

\[ φ(x) = \frac{e^x}{1 + e^x}, \]

\[ \frac{dφ}{dx} = φ(x)(1 − φ(x)). \]

1. Training procedure

An MLP can be trained by repeatedly offering (possibly pre-processed) simulated spectra q\textsubscript{i} to the first neuron layer
and adjusting the weights and biases of the network for minimization of the loss function, \( E = \frac{1}{2} | \mathbf{o}_{\text{out}} - \mathbf{p} |^2. \) The influence of each weight on the loss function \( \delta^k = \frac{\partial E}{\partial w^k}. \) \( \delta^k \) can be calculated in an algorithm called backpropagation. Following the gradient descent optimization algorithm, the weight adjustments are then given by \( \Delta w^k(t + 1) = (1 - \alpha) \eta \delta^k \sigma^k + \alpha \Delta w^k(t), \) with \( \eta \) as the learning rate, which can be adapted during training. Also, to avoid ending up in a local minimum of \( E, \) momentum can be added through the inertia term \( \alpha^2 \) \[ \Delta w^k(t + 1) = (1 - \alpha) \eta \delta^k \sigma^k + \alpha \Delta w^k(t). \] (9)

Backpropagation is normally performed on a batch of examples concurrently. The larger the batch, the more general the trends that can be captured. A database can be chopped up in a number of batches, which can be offered one by one to form one epoch of training. In the next epoch, the database can be chopped up again (using a possibly different order of examples) for the next epoch. Depending on the application, the training is continued for one, several, or many epochs, or until a certain desired error margin is reached.

During training, the backpropagation is performed on examples from the database. However, similar to FP, it is better to test the resulting MLP on new test data. The fraction of variance unexplained (FVU) can provide a measure of how well \( \mathbf{p} \) is replicated by the model and is defined as the ratio of the variance in the error over the total variance,

\[
FVU_i = \frac{\sum_{k=1}^{N_{\text{test}}} (p_{i,\text{pred.}} - p_{i,\text{true}})^2}{\sum_{k=1}^{N_{\text{test}}} (p_{i,\text{true}} - p_{i,\text{mean}})^2},
\] (10)

with \( N_{\text{test}} \) denoting the number of examples in the test set, \( i \) denoting the index of \( \mathbf{p} \), and the subscripts denoting the true, predicted, and mean parameter values. The FVU can be calculated after each batch to quantify the training progress. This was done during training for both the training database and the test set. If the FVU of the database keeps decreasing but the FVU of the test set does not, the MLP might be overfitted to the data set and not generalisable to new data.

In this work, the MLP regressor class from the SciKit-Learn package \(^{22}\) was used to make the approximation. Three hidden layers were used, containing 150, 100, and 50 neurons, respectively. The MLP was trained with the same databases as used for the FP. The DFT and PCA were omitted, \( \mathbf{q}_1 = \mathbf{q}. \) The total layout of the MLP is then given by \( [N_{\mathbf{q}_1}, N_1, N_2, N_3, N_p] = [50, 150, 100, 50, 6]. \) Instead of training once on low-noise spectra and applying the resulting model on noisy spectra, the MLP was trained for each desired noise level separately. The MLP regressor was set to use adaptive training and an inertia term \( \alpha = 5 \times 10^{-3} \), and training was done with batches of size 250.

### 2. Error propagation through sensitivity analysis

As described in Sec. II D 1, observed measurement errors and \( \chi^2 \text{red} \) cannot provide a quantitative uncertainty of the mapping output during new measurements. A tool that can provide such quantifications is the sensitivity analysis.

The trained network can be represented as a function \( \mathbf{MLP}, \mathbf{p}_{\text{inv}} = \mathbf{MLP}(\mathbf{q}_1), \) with \( \mathbf{p}_{\text{inv}} \) denoting the retrieved system parameters and \( \mathbf{q}_1 \) denoting the preprocessed spectra, or \( \mathbf{o}_{\text{out}} = \mathbf{MLP}(\mathbf{q}_1), \) with in and out depicting the first and last layers of the MLP. Analogous to the partial derivative \( \frac{\partial E}{\partial w^k} \) in backpropagation, one can calculate \( \frac{\partial \mathbf{o}_{\text{out}}}{\partial \mathbf{p}_{\text{inv}}} \) for all neuron combinations \( i \) and \( j \) and a given \( \mathbf{q}_1. \) Together, the derivatives form the Jacobian matrix \( \mathbf{J}. \)

One could use these sensitivities as an indication of the relative importance of different spectral regions in the estimation of each \( p_i. \) However, it has been demonstrated that such indications are unreliable for dependent input parameters. \(^{24}\) Namely, there are many possible networks that describe the same patterns in the data while having different distributions of the sensitivity.

The sensitivities can also be used for error propagation. In general, an error indication can be obtained from the covariance matrix of the output, \( \mathbf{D}^o. \) The elements of this matrix are given by

\[
\mathbf{D}^o_{ij} = \sum_{k=1}^{N_q} \sum_{l=1}^{N_p} J_{ik} J_{jl} \mathbf{D}^q_{kl},
\] (11)

with \( \mathbf{D}^q \) denoting the covariance matrix of the input. At a fixed point \( \mathbf{p} \) in parameter space, the uncertainties in \( \mathbf{q} \) can be assumed to be uncorrelated and equal to the spectral noise. In that case, \( \mathbf{D}^q \) is a diagonal matrix with constant value \( \sigma^2_q. \) The synthetic data indeed originate in a single point \( \mathbf{p}, \) but due to the finite scattering volume and signal integration time, the collected radiation is given by the integral of scattered power over a small subspace in \( p\)-space and time. Although incorporation of the finite scattering volume in the mapping would be overly complicated, its integration in the uncertainty estimation is possible. For each point at which \( \mathbf{D}^p \) is evaluated, one could evaluate \( \mathbf{D}^o \) over a set of spectra which was picked from the expected \( p\)-subspace.

Using the network trained with spectra at \( \rho_s = 12.5 \text{ GHz}^{-1} \) and \( \sigma_s = 11 \%, \) \( \mathbf{D}^p \) was evaluated on a grid of points in parameter space, at fixed geometry \( \theta = 130^\circ, \phi = 110^\circ, B = 2.2 \text{ T}, \) and \( n_s = n_p \) in temperatures of the range \( 0.5 \leq T_s, T_e \leq 10.0 \text{ keV}. \)

### III. RESULTS

#### A. Function parametrization results

The computation time required to make the FP mapping was 71 min. The results of the function parametrization for a test set with \( \rho_s = 12.5 \text{ GHz}^{-1} \) and \( \sigma_s = 10 \% \) are shown in Fig. 5. In the MARS regression process, the maximum number of 80 terms was reached: three terms of the first degree, four terms of the second degree, and the rest of the third degree. During pruning, none of these terms were removed. This indicates an underfit of the training data. The results for varying noise levels are shown in Figs. 9 and 10 for \( n_{D}/n_{E} \) in the ranges of 0.2-1 and 0.9-1, respectively.

#### 1. Chi-squared analysis results

The \( \chi^2\text{red} \) values of the FP applied to a test set with \( N_{\text{test}} = 1000 \) for different noise levels are shown in Fig. 6. With a
FIG. 5. Input (real) and output (approximated) ion temperatures of the application of the FP on a test set with 1000 samples, 1.4% noise, and $\rho_s = 12.5 \text{ GHz}^{-1}$ ($\sigma_{s,650} = 10\%$). The average absolute deviation for $T_i$ is 9.9% of the input temperatures.

high noise level, almost no correlation between $\chi^2_{\text{red}}$ and $\epsilon$ can be recognized and the median $\chi^2_{\text{red}}$-value approaches 1.

At lower noise levels, there is a certain correlation, and probabilities of certain $\epsilon$ for given $\chi^2_{\text{red}}$ can be provided. For example, in the test set of $\sigma_{s,650} = 10\%$ in Fig. 6, $P(\epsilon > 0.1 | \chi^2_{\text{red}} < 10) = 0.21$. In principle, a probability distribution $P(\epsilon | \chi^2_{\text{red}})$ can be obtained in this way. However, since the theoretical uncertainty for a given $\chi^2_{\text{red}}$ can only be known for a perfect model and $\epsilon$ is equal to zero, this method cannot be used to provide sound quantitative error estimations.

**B. Artificial neural networks**

The training progress for a data set with $\sigma_{s,650} = 10\%$ is shown in Fig. 7. The corresponding estimation error $\epsilon_T = 4.2\%$. The results for varying noise levels are shown in Figs. 9 and 10 for $n_D/n_e$ in the ranges of 0.2-1 and 0.9-1, respectively.

![Image](image-url)

FIG. 6. Scatter plots of $\chi^2_{\text{red}}$ versus the estimation error $\epsilon_T$ for a set of $N_{\text{test}} = 1000$ tests on the FP mapping, at different noise levels. The horizontal lines show the median values of $\chi^2_{\text{red}}$: 26.8, 1.9, and 1.2 for equivalent noise levels of 1%, 10%, and 100%, respectively. For measurements with a low error in their category, the values of $\chi^2_{\text{red}}$ are distributed around a value slightly larger than one, indicating a modest underfit of mapping to data.

**1. Sensitivity analysis**

The obtained uncertainties were compared with the absolute errors of a small test set, the results of which are shown in Fig. 8. It can be seen in Fig. 8 that for both temperatures, the majority of tests have $\epsilon_T < \sigma_T(T_i, T_e)$. Also, the average standard deviation of the sensitivity grid corresponds well with the average estimation mapping error of the ANN, $\epsilon_T = 4.2\%$. Thus, the sensitivity is a reasonable measure of the mapping accuracy of a specific ANN and can be used to obtain uncertainties of the mapping results for new measurements within the parameter domain covered during training.

**IV. DISCUSSION**

In Fig. 9, it can be seen that both FP and ANNs can provide an adequate mapping of the ion temperature, with mean errors $\epsilon_T = 9.9\%$ and 4.2%, respectively, at the expected noise level $\sigma_{s,650} = 10\%$. However, the ANN is much more robust to noise:
if the expected estimation error is to be kept below 20%, FP will not be useful for \( \sigma_{T_{\rho}} \geq 30 \% \), whereas \( \epsilon_{T_{\rho}} = 14.3 \% \) for an ANN trained at \( \sigma_{T_{\rho}} \geq 100 \% \). Because the CTS probing beam is also used for heating and there are several other heating gyrotrons operating in the machine at the probing frequency, there will be large amounts of stray radiation in the frequency range of the CTS spectra. Therefore, the mapping method with maximal noise robustness is preferable.

The electron temperature mapping accuracy was much smaller for both methods. The reason for this is that \( n_{\rho}/n_{e} \) and \( T_{e} \) appear very similarly in the CTS equations, and thus their influence on the CTS spectra is comparable. It should be noted that with a small range of \( n_{\rho}/n_{e} \), the in the mapping parameter space, \( T_{e} \) can be estimated just as well as \( T_{i} \) by both the FP and ANN mappings, as shown in Fig. 10.

In Fig. 9, the following can also be recognized: for the FP, the error increases linearly with noise, whereas the MLP mapping error levels off just above 30%. This highlights a fundamental difference between the FP and the MLP. The MARS model does not have parameter space boundaries. It was made to fit best to the examples in the training database but can be extrapolated for anomalous input spectra and thus mapped outside of the training parameter space. The MLP, on the other hand, was trained to provide a mapping within the training database only and will provide output only in that range. This is in the nature of the activation function, which is limited between zero and one for all input values.

The chi-squared analysis can be applied to both FP and ANN to provide an indication of the goodness of fit of real measurements. A quantitative uncertainty for new measurements can be obtained from a sensitivity analysis. For the MLP, this uncertainty was in good agreement with actual results from predictions on synthetic data. In principle, FP also allows sensitivity analysis. But, because the phase information of the DFT is not used in the FP, the derivatives and thus the sensitivity of the FP are more difficult to interpret.

An advantage of FP over ANNs is that the terms in the MARS functions form direct links between the input and the output, thus providing insight into the mapping. The maximum number of inputs to MARS poses the need for a stricter dimension reduction compared to the ANN. This can be achieved by two extra steps: DFT and PCA. In principle, these steps can provide insight into the system complexity and the amount of information that can be distinguished from noise. In the DFT, half of the information is removed: the half concerning the location of spectral changes (corresponding to phase in a DFT of a time-dependent signal). In future expansions of the mapping parameter space, this information might be necessary, further complicating the FP.

The current ANN implementation considers the (binned) spectra as a whole. This method seems more coarser than the dimension reduction in the FP. However, one can also train a new MLP on the transformed variables that were input to MARS or any other set of variables obtained from the spectra. For example, an MLP can be trained with variable spectral density to emphasize certain spectral regions while reducing the required complexity of the neural network.

In fact, except for the increased computation time required for training, dimension reduction is no principal requirement for the MLP. It has the ability to learn the most relevant patterns by itself. This also makes ANNs more flexible to changes in the parameter space, i.e., a larger range of scattering geometries or additional dimensions, like the plasma drift or properties of the (fast) ion velocity distribution.

Another important set of additional dimensions are the plasma impurity concentrations, which might have significant influence on the shape of the measured spectra in W7-X. As stated in Sec. II B, the parameter space used in this work does not account for impurities. However, once the addition of impurities is implemented in the eCTS code, a new mapping can be made with inclusion of the impurity concentrations.

More in general, the following should be noted: an ANN can only describe the (physics) patterns that were used for training. There are two ramifications of this. The first is that the physics model used to produce the synthetic database is wrong. The eCTS model provides an electrostatic estimation of CTS, which might not cover all processes under observation during measurements. The second aspect is the difference in parameter space between synthetic data and measurements. This includes the omission of impurity concentrations but also other differences that might or might not be foreseen. For example, due to the finite size of the scattering volume, the
measured spectra are formed by the addition of different scattering processes and thus span a finite subspace in parameter space, instead of a single point. Also, the synthetic background radiation is uniform white noise, whereas the real background originates in stray radiation from the gyrotron (modulation), electron cyclotron emission from the plasma in the line of sight, and any other radiation present in the concerning spectral region.

V. CONCLUSION

Both FP and ANNs can provide a mapping between synthetic CTS spectra and the bulk ion temperature. Without any additional input, the FP and ANN can estimate $T_i$ with mean absolute errors $\epsilon_{T_i} = 9.9\%$ and $4.3\%$ for spectra with $\rho_s = 12.5$ GHz$^{-1}$ and $\sigma_{s,650} = 10\%$. In a more limited parameter space where $n_D/n_e$ is almost fixed, $T_e$ can also be identified, $\epsilon_{T_e} = 8.4\%$ and $4.6\%$ for the FP and ANN, respectively.

The current mappings can be made into a ready-to-use data analysis method including relevant impurity concentration ranges in the eCTS code and mapping parameter space. Other significant differences between synthetic and measured spectra might appear as well, for example, due to background radiation in the plasma. These differences might harm the applicability of both the FP and the ANN. But, ANNs have several significant advantages: a much larger noise robustness, ease of implementation, and higher adaptability with respect to the parameter space. Thus, artificial neural networks are the method of choice to enable fast CTS ion temperature measurements and possibly other plasma properties in the future.

ACKNOWLEDGMENTS

The authors would like to thank the W7-X team for their efforts on the development of the CTS diagnostic and their help and cooperation in the preparation of this article. This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the EURATOM research and training programme 2014–2018 under Grant Agreement No. 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.