

Fast Analysis of Collective Thomson Scattering Spectra on Wendelstein 7-X

J. van den Berg,^{1, a)} I. Abramovic,¹ N.J. Lopes Cardozo,¹ and D. Moseev²

¹⁾*Department of Applied Physics, Eindhoven University of Technology,
De Zaale, 5612 AJ Eindhoven, The Netherlands*

²⁾*Max-Planck-Institut für Plasmaphysik, Greifswald 17491,
Germany*

(Dated: 26 June 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 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 in the information contents of the CTS spectra, ANNs provide a higher accuracy and noise robustness, are easier to implement and 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.

^{a)}JvandenBerg@pm.me

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 2016¹. 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 available². For the interpretation of measured spectra, the eCTS code is integrated in the generic Bayesian data analysis framework Minerva, implemented at W7-X³. 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 analysing a data set of simulated spectra that covers the expected operational space of the experiment. We compare the so-called Function Parametrization (FP)^{4,5} which defines an inverse mapping based on functions that are used to interpolate between the points in the data base, and the more generic approach of Artificial Neural Networks (ANNs)⁶. 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 set up 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 generic and have been applied to other diagnostics^{5,7,8}. 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 sections II A and II B, then discuss the spectrum information content, dimension reduction and the example database in section II C. The principles of the FP and ANN techniques and their implementations are recalled in sections II D and II E, respectively. The mapping results are described in section III and the relative merits of both methods are discussed in section IV.

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 $(\lambda_D k^\delta)^{-1} > 1$ is satisfied⁹, where λ_D is the Debye length, $k^\delta = |\mathbf{k}^\delta|$ and $\mathbf{k}^\delta = \mathbf{k}^s - \mathbf{k}^i$ specifies the direction along which the fluctuations are resolved, with $\mathbf{k}^{i,s}$ being the wave vectors of incident 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 distributions¹⁵. 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.

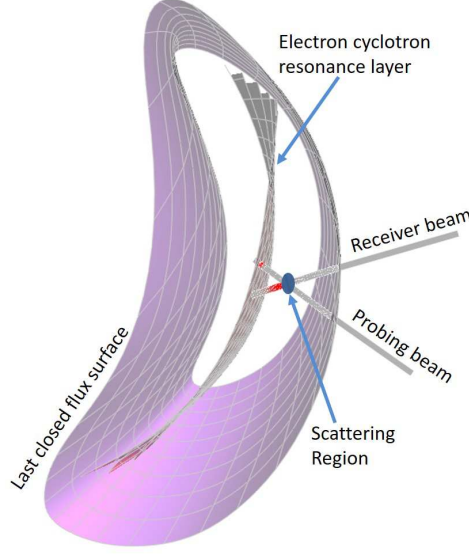


FIG. 1. Schematic of the CTS geometry on Wendelstein 7-X. The scattering angle $\theta = 130^\circ$, the observation angle $\phi = 110^\circ$. The second harmonic electron cyclotron resonance layer for 140 GHz is shown. The probing beam absorption region is in the resonance layer (depicted in red). The red region on the receiver beam path depicts a region of electron cyclotron emission at 140 GHz. The blue ellipsoid shows the location of the scattering volume, where the probing and viewing beams intersect.

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 cut-offs, resonances and refraction may occur. Locally, the CTS scattering geometry is determined by the following two angles: the observation angle $\phi = \angle(\mathbf{B}, \mathbf{k}^\delta)$ and the scattering angle $\theta = \angle(\mathbf{k}^s, \mathbf{k}^i)$. 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 figure 1.

Inference of the parameter values from the measured spectra is typically done using the Bayesian formalism in conjunction with a forward code of the scattering¹⁰. The developed forward model of CTS (eCTS¹⁶) calculates the spectral power density of the scattered radi-

ation, as given by¹⁷:

$$\frac{\partial P^s}{\partial \omega} = P_{in} O_b \frac{\omega^i \omega^s}{2\pi c^2} r_e^2 n_e G S(\mathbf{k}^\delta, \omega) \quad (1)$$

with P_{in} the probing beam power, O_b the overlap volume, $\omega^{i,s}$ the angular frequencies of incident and scattered radiation, respectively, $\omega = \omega^s - \omega^i$, r_e the classical electron radius, n_e the local electron density, G the so-called geometrical form factor and finally $S(\mathbf{k}^\delta, \omega)$ the spectral density function, which governs the shape of the CTS spectra, and is in turn influenced by a variety of plasma parameters¹⁰. Note that the spatial power density, n_e , G , \mathbf{k} and $S(\mathbf{k}^\delta, \omega)$ 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 of fluctuations in the electron density is taken into account². The spectral density function is given by¹⁸:

$$S(\mathbf{k}^\delta, \omega) = S_e(\mathbf{k}^\delta, \omega) + \sum_i S_i(\mathbf{k}^\delta, \omega), \quad (2)$$

with S_e the electron scattering term and S_i 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} , ω^i , 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 ϕ , and the angle between the probing and scattered waves $\theta = \angle(\mathbf{k}^s, \mathbf{k}^i)$. Eight CTS spectra computed by the eCTS code, at different T_i and fixed scattering geometry, are shown in figure 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 set-up 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. Firstly, there is an uncertainty in scattering geometry. Although the placement of the diagnostic parts is fixed during measurements, the refraction of probing and scattered radiation in the plasma introduces variations in θ and ϕ , which are expected to be a few degrees. The magnetic field in W7-X is known as a function of scattering location, but due to uncertainty in this location, it might slightly vary: $B = 2.2 \pm 0.05 T$. During operational phase 1.2 (OP1.2), T_i is not expected to exceed $4 keV$ and $T_e \lesssim 8 keV$. Both temperatures were taken as free parameters in the range $0.5 \dots 10.0 keV$. In this work, Deuterium (ion mass $m_D = 2u$ and charge number $Z_D = 1$) was chosen as the main ion species. The impurity concentrations are not known. However, they affect the CTS spectrum in several ways. Firstly, by dilution: $n_e = \sum_i n_i Z_i$. This was simulated in a simplified way: by adding the main ion species density to the mapping parameter space. Since the electron density is considered a scaling parameter, the ratio n_D/n_e was used as input parameter, in the range $0.2 \dots 1$. In a very clean plasma, the impurity concentrations can be so small that n_i comes close to n_e . To simulate such a clean plasma, another two mappings were made with the density ratio in the range $0.9 \dots 1$.

Plasma impurities also affect the shape of the CTS spectrum. As stated in equation 2, each ion species has a contribution S_i to the total scattering. These terms do not only depend on general parameters (θ, ϕ, B, ω) and species-specific properties (m_i, Z_i, n_i, T_i), but also on the properties of other plasma species. A quantitative analysis of measured CTS spectra requires a quantitative inclusion of impurities in the data analysis model. The eCTS code used for this work did not include impurity concentrations, but its results do provide a methodological comparison of the two fast analysis methods.

The total mapping parameter space spans six dimensions, plus the sampling frequency ν_s . The expected parameter values are summarized in table I, together with the actual parameter ranges used for the mapping. These ranges were chosen larger than or equal to the expected values.

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

parameter	expected value	mapping parameter range
T_i	$< 4 \text{ keV}$	0.5 - 10 keV
T_e	$\lesssim 8 \text{ keV}$	0.5 - 10 keV
θ	$130 \pm 2^\circ$	$125 - 135^\circ$
ϕ	$110 \pm 2^\circ$	$105 - 115^\circ$
B	$2.2 \pm 0.05 \text{ T}$	2.1 - 2.3 T
n_i/n_e	0.2 - 1	0.2 - 1.0

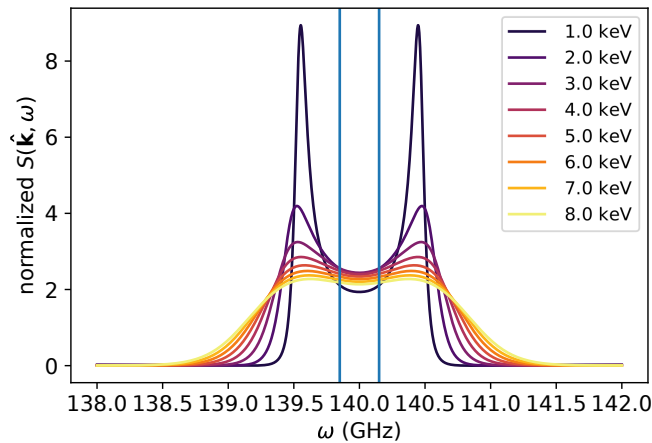


FIG. 2. Collective Thomson scattering spectra as calculated by the eCTS code², normalized to their mean value. The varied parameter is the ion temperature. The other input parameters were $T_e = 5 \text{ keV}$, $\theta = 130^\circ$, $\phi = 110^\circ$, $B = 2.2 \text{ T}$ and $n_i = n_e$.

C. Database and dimension reduction

The goal of this work is to find a fast mapping from the measured spectra \mathbf{q} to the input parameters \mathbf{p} of the eCTS model. The former is presented in figure 2 and the latter in table I. In this work, the density of sampling frequencies was taken to be $\rho_s = 650 \text{ GHz}^{-1}$ (corresponding to a spectral resolution of 1.538 MHz), in the range $138 < \nu_s < 142 \text{ GHz}$, which makes for an initial number of parameters $N_{q,0} = 2600$. In order to retain computational tractability of the database building, dimension reduction of \mathbf{q} is necessary. In this process,

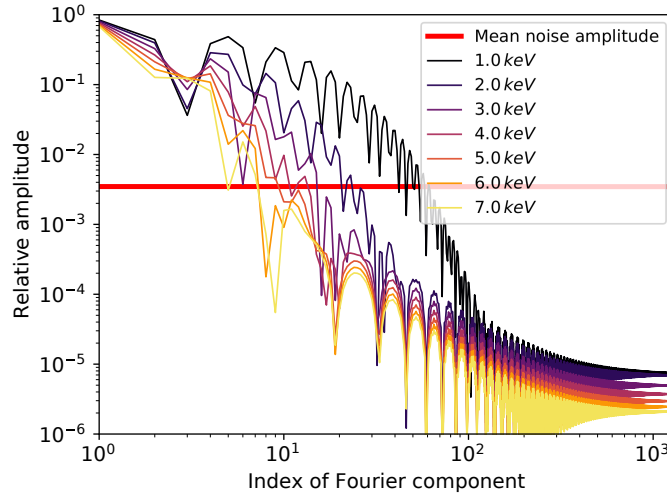


FIG. 3. Discrete Fourier transforms of noiseless spectra at $T_e = 5.0 \text{ keV}$, $\theta = 130^\circ$, $\phi = 110^\circ$, $B = 2.2T$, $n_i = n_e$ and varying T_i . Also shown is the average amplitude of the DFT of the noise $\sigma_s = 10\%$ at $\rho_s = 650 \text{ GHz}^{-1}$. Only components above this level are useful for the approximation.

the sensitivity of the result with regard to \mathbf{p} should be retained while removing the largest possible amount of unnecessary information.

Without noise, only five variables with finite variance in the parameter range are required to determine five independent parameters. However, the measured spectra contain noise which is here assumed to be Gaussian: $\sigma_s = 10\%$ at $\rho_s = 650 \text{ GHz}^{-1}$, defined in relation to the average maximum spectral power density 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 $\sigma_b = \sigma_s / \sqrt{BF}$, with BF 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 DFT's are

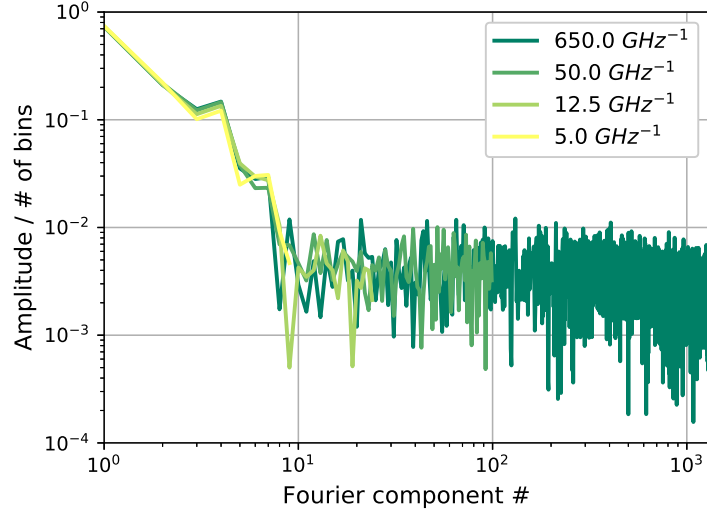


FIG. 4. Discrete Fourier transforms of a CTS spectrum at $T_i = T_e = 5 \text{ keV}$, 2600 bins ($\rho_s = 650 \text{ GHz}^{-1}$) and artificial noise $\sigma_s = 10\%$, that was binned to different spectral densities. Spectral densities $> 12.5 \text{ GHz}^{-1}$ can be used to reduce computation time during database building and application of the FP.

shown in figure 4. Firstly, 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. Secondly, it can be seen that the noise level is equal for all binning factors. The computation time required to produce the spectra, however, scales linearly with ρ_s . Thus, the lower value $\rho_s = 12.5 \text{ GHz}^{-1}$ was used for database building. The artificial noise was changed accordingly:

$$\sigma_s = \sqrt{\frac{\rho_s}{\rho_0}} \times \sigma_{s,650}, \quad (3)$$

where $\sigma_{s,650}$ is defined as the equivalent spectral noise at $\rho_s = 650 \text{ GHz}^{-1}$.

The database was built by doing $N = 25,000$ simulations with the eCTS model, yielding vectors \mathbf{p}^α and \mathbf{q}^α with $(\alpha = 1, \dots, N)$. \mathbf{p}^α 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 \mathbf{p}^α and the 50 sampling frequencies in the range $140 \pm 2 \text{ GHz}$. The eCTS code then produces the corresponding spectrum. As stated in section IIB, 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)⁷.

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 \mathbf{q}^α . 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 \mathbf{q}^α is as follows: first, the database is normalized: for each element j in \mathbf{q}^α , $\bar{q}_j^\alpha = (q_j^\alpha - \langle q \rangle_j)/\sigma_j$, with σ_j the standard deviation of variable j and $\langle q \rangle_j$ the mean value of q_j in the database. Then, the covariance matrix $\mathbf{D}^{\bar{\mathbf{q}}}$ of the normalized variables is calculated, the elements of which are given by:

$$D_{ij}^{\bar{\mathbf{q}}} = \frac{1}{N-1} \sum_{\alpha} \bar{q}_i^\alpha \bar{q}_j^\alpha. \quad (4)$$

The matrix $\mathbf{D}^{\bar{\mathbf{q}}}$ contains the cross-correlations between the elements of $\bar{\mathbf{q}}$. The eigenvalues λ_j of $\mathbf{D}^{\bar{\mathbf{q}}}$ are a measure of the regressive significance of the corresponding eigenvectors $\mathbf{D}^{\bar{\mathbf{q}}}\mathbf{e}_j = \lambda_j\mathbf{e}_j$. The principal components of each spectrum are then given by $\tilde{q}_j^\alpha = \mathbf{e}_j \cdot \bar{\mathbf{q}}^\alpha$. If $\lambda_j > (3\sigma_s)^2$, with σ_s the measurement noise, \tilde{q}_j is considered a significant component. The eigenvalues and corresponding eigenvectors are sorted such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{N_s}$ with N_s the number of significant eigenvectors. One can now make a mapping from the significant components in $\tilde{\mathbf{q}}$ to \mathbf{p} . The application of the PCA in the function parametrization is described in the next subsection.

D. Function parametrization

After the dimension reduction is complete, a mapping from the processed variables \mathbf{q}_1 to the eCTS input parameters \mathbf{p} can be made. The first mapping method considered in this work is function parametrization (FP)^{4,7}. In a FP, a regression is performed on a database of examples, in which the error of a certain mapping function $\mathbf{p} = M(\mathbf{q}_1)$ is minimized.

For the regression in FP, one has to select a function f that can adequately describe the trends in the model. The eCTS model is non-linear, so the regressor must be too if the whole parameter range is to be covered. Multivariate Adaptive Regression Splines (MARS¹⁹) 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 $\tilde{\mathbf{q}}$ 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 $N_{\mathbf{q}_1} \gtrsim 10$. 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 CPU along with 20 GB of RAM.

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

1. Error analysis

The simplest error analysis is the comparison of the approximation \mathbf{p}_{inv} with the eCTS input parameters \mathbf{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:

$$\epsilon_i \equiv \frac{1}{N_{\text{test}}} \sum_{\beta} \frac{|p_{i,\text{inv}}^{\beta} - p_i^{\beta}|}{p_i^{\beta}}, \quad (5)$$

with β the test spectrum index and N_{test} the number of spectra in the test set.

During operation, \mathbf{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 \mathbf{p}_{out} , with the observed spectrum. This was implemented as follows:

$$\chi^2 \equiv \sum_{n=1}^N \left(\frac{q_n - S(\omega_n, \mathbf{p}_{\text{out}})}{\sigma_s} \right)^2, \quad (6)$$

where q_n are the normalized spectral power densities measured at frequencies ω_n . S is the spectral power density as calculated by the forward model, including artificially added noise with magnitude equal to the measurement noise σ_s . The vector \mathbf{p}_{out} contains the measured parameters. Reduced chi-squared is then $\chi_{\text{red}}^2 = \chi^2/K$, with K the number of degrees of freedom. The value of χ_{red}^2 can be interpreted in the following way: The closer χ_{red}^2 is to

1, the better the fit. Values of $\chi_{red}^2 > 1$ represent an underfit whereas $\chi_{red}^2 < 1$ represents overfitting. However, it is not possible to use χ_{red}^2 as a quantitative measure of the goodness of fit²⁰. χ_{red}^2 has an intrinsic probability distribution function, which can only be known for the true model having the true parameter values.

χ_{red}^2 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_{red}^2 of the most accurate measurements should follow the χ_{red}^2 -distribution and thus have a mean value of 1. Moreover, a correlation between ϵ and χ_{red}^2 might provide a rough error estimation for measurements. The results of this analysis are presented in section III A 1.

E. Artificial neural networks

An artificial neural network (ANN) is a computing system that consists of a network of artificial neurons: mathematical function that mimic the behaviour 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 Rosenblatt⁶, 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:

$$o_i^k = \phi(net_i^k) = \phi\left(\sum_{l=1}^{N_i} o_{i-1}^l w_{i-1}^{lk} + b_i^k\right), \quad (7)$$

where i and k denote the layer and neuron indices, respectively. net_i^k is the net input of neuron k in layer i , formed by the sum of all N_i neuron outputs of the previous layer, multiplied by the corresponding weights w_{i-1}^{lk} , plus b_i^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 $(-\infty, \infty)$. Many activation functions can be used, but for this work, the

standard logistic function and its convenient derivative were used:

$$\begin{aligned}\phi(x) &= \frac{e^x}{1 + e^x} \\ \frac{d\phi}{dx} &= \phi(x)(1 - \phi(x))\end{aligned}\tag{8}$$

1. *Training procedure*

An MLP can be trained by repeatedly offering (possibly pre-processed) simulated spectra \mathbf{q}_1^α to the first neuron layer \mathbf{o}_{in} 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, $\delta_i^k = \frac{\partial E}{\partial w_i^{kl}}$. δ_i^k can be calculated in an algorithm called backpropagation. Following the gradient descent optimization algorithm, the weight adjustments are then given by $\Delta w_i^{kl} = \eta \delta_i^k o_i^k$, with η 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 α^{21} :

$$\Delta w_i^{kl}(t+1) = (1 - \alpha)\eta \delta_i^k o_i^k + \alpha \Delta w_i^{kl}(t).\tag{9}$$

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, similarly 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:

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

with N_{test} the number of examples in the test set, i the index of \mathbf{p} , and the subscripts denoting the true, predicted and mean parameter values, respectively. 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 testset does not, the MLP might be overfitted to the dataset, and not generalisable to new data.

In this work, the MLP regressor class from the SciKit-Learn package²² 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 $[N_{q,1}, N_1, N_2, N_3, N_p] = [50, 150, 100, 50, 6]$. 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 $\alpha = 5 \cdot 10^{-5}$ and training was done with batches of size 250.

2. *Error propagation through sensitivity analysis*

As described in section IID 1, observed measurement errors and χ_{red}^2 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 MLP : $\mathbf{p}_{inv} = MLP(\mathbf{q}_1)$ with \mathbf{p}_{inv} the retrieved system parameters and \mathbf{q}_1 the preprocessed spectra, or $\mathbf{o}_{out} = MLP(\mathbf{o}_{in})$ with in and out depicting the first and last layer of the MLP. Analogous to the partial derivative $\frac{\partial E}{\partial w_{kl}^{ij}}$ in backpropagation, one can calculate $\left. \frac{\partial o_{out}^i}{\partial o_{in}^j} \right|_{\mathbf{q}}$ for all neuron combinations i and j and a given \mathbf{q}_1 ²³. 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 parameters²⁴. 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}^p . The elements of this matrix are

given by:

$$D_{ij}^p = \sum_{k=1}^{N_q} \sum_{l=1}^{N_q} J_{ik} D_{kl}^q J_{jl}, \quad (11)$$

with D^q 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 σ_s^2 . 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}^q 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 $\sigma_{s,650} = 11 \%$, \mathbf{D}^p was evaluated on a grid of points in parameter space, at fixed geometry $\theta = 130^\circ$, $\phi = 110^\circ$, $B = 2.2 T$, $n_i = n_e$ and temperatures in the range $0.5 \leq T_i, T_e \leq 10.0 \text{ keV}$.

III. RESULTS

A. Function parametrization results

The computation time required to make the FP mapping was 71 minutes. The results of the function parametrization for a test set with $\rho_s = 12.5 \text{ GHz}^{-1}$ and $\sigma_{s,650} = 10 \%$ are shown in figure 5. In the MARS regression process, the maximum number of 80 terms was reached: three terms of the first degree, four terms of second degree and the rest of third degree. During pruning, none of these terms was removed. This indicates an underfit of the training data. The results for varying noise levels are shown in figures 9 and 10 for n_D/n_e in the ranges $0.2 - 1$ and $0.9 - 1$, respectively.

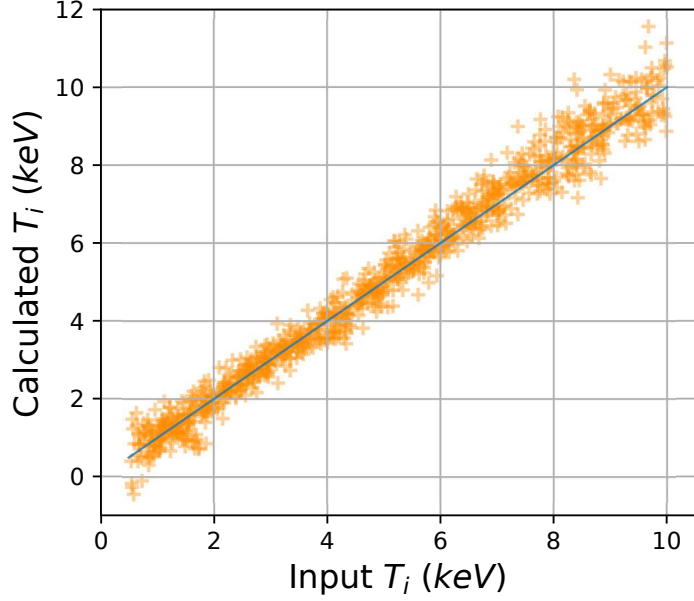


FIG. 5. Input (real) and output (approximated) ion temperature of the application of the FP on a test set with 1000 samples, 1.4 % noise and $\rho_s = 12.5 \text{ GHz}^{-1}$ ($\sigma_{s,650} = 10\%$). The average absolute deviation for T_i is 9.9 % of the input temperatures.

1. *Chi-squared analysis results*

The χ_{red}^2 values of the FP applied to a test set with $N_{test} = 1000$ for different noise levels are shown in figure 6. With a high noise level, almost no correlation between χ_{red}^2 and ϵ can be recognized and the median χ_{red}^2 -value approaches 1.

At lower noise levels, there is a certain correlation, and probabilities of certain ϵ for given χ_{red}^2 can be provided. For example, in the test set of the $\sigma_{s,650} = 10\%$ in figure 6: $P(\epsilon > 0.1 | \chi_{red}^2 < 10) = 0.21$. In principle, a probability distribution $P(\epsilon | \chi_{red}^2)$ can be obtained in this way. However, since the theoretical uncertainty for a given χ_{red}^2 can only be known for a perfect model and ϵ equal to zero²⁰, this method cannot be used to provide sound quantitative error estimations.

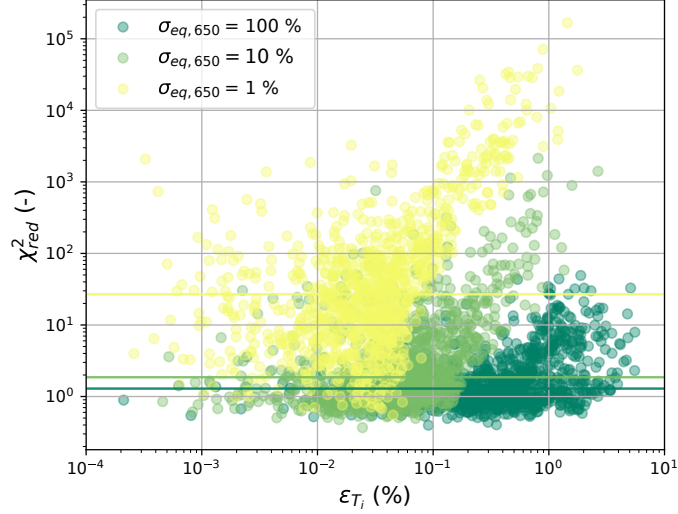


FIG. 6. Scatter plots of χ^2_{red} versus the estimation error ϵ_{T_i} for a set of $N_{test} = 1000$ tests on the FP mapping, at different noise levels. The horizontal lines show the median values of χ^2_{red} : 26.8, 1.9 and 1.2 for equivalent noise levels of 1 %, 10 % and 100 %, respectively. For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

B. Artificial neural networks

The training progress for a dataset with $\sigma_{s,650} = 10\%$ is shown in figure 7. The corresponding estimation error $\epsilon_{T_i} = 4.2\%$. The results for varying noise levels are shown in figures 9 and 10 for n_D/n_e in the ranges $0.2 - 1$ and $0.9 - 1$, respectively.

1. Sensitivity analysis

The obtained uncertainties were compared with the absolute errors of a small test set, the results of which are shown in figure 8. It can be seen in figure 8 that for both temperatures, the majority of tests have $\epsilon_i < \sigma_i(T_i, T_e)$. Also, the average standard deviation of the sensitivity grid corresponds well with the average estimation mapping error of the ANN: $\epsilon_{T_i} = 4.2\%$. 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.

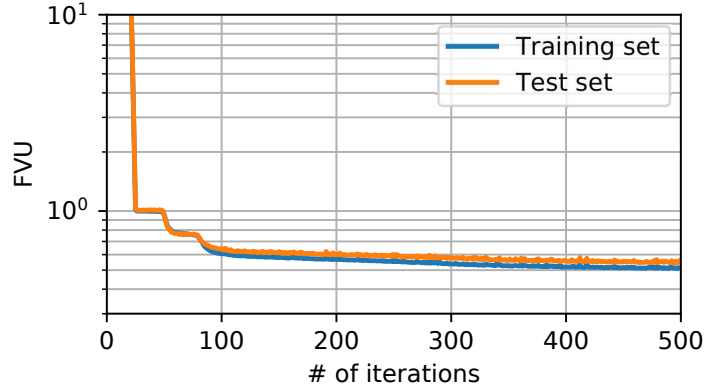


FIG. 7. Neural network trianing progress for the assumed parameter space described in table I. Each training iteration was performed with a batch of 250 spectra. With a database of 25,000 spectra, an epoch is completed after 100 iterations. FVU is the Fraction of Variance Unexplained, defined in equation 10.

IV. DISCUSSION

In figure 9, it can be seen that both FP and ANNs can provide an adequate mapping of the ion temperature, with mean errors $\epsilon_{T_i} = 9.9\%$ 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 $\sigma_{s,650} \gtrsim 30\%$, whereas $\epsilon_{T_i} = 14.3\%$ for an ANN trained at $\sigma_{s,650} \gtrsim 100\%$. 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_D/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_D/n_e 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 figure 10.

In figure 9, the following can be also recognized: for the FP, the error increases linearly with noise, whereas the MLP mapping error levels off just above 30% . This highlights a

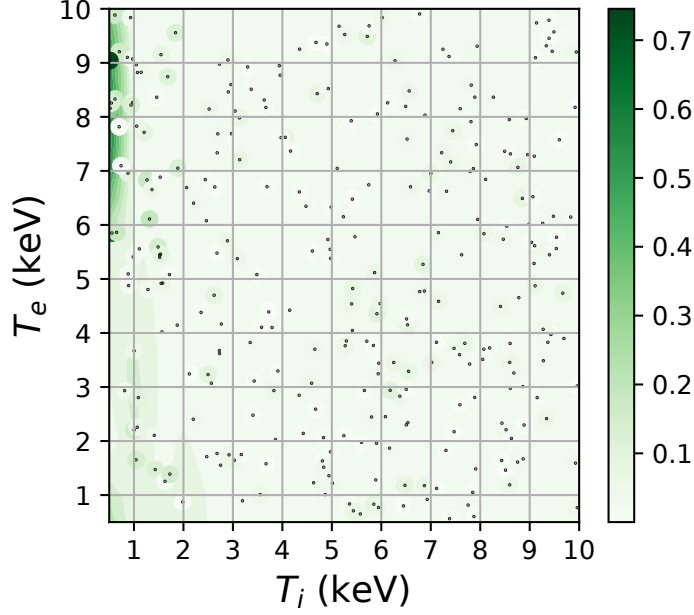


FIG. 8. Contour plot of the normalized standard deviation σ_{T_i}/T_i , as obtained from the covariance matrix D^p . This matrix was evaluated on an equidistant grid with T_i and $T_e = [0.5, 1.0, \dots, 10]$, at fixed geometry $\theta = 130^\circ$ and $\phi = 110^\circ$, $B = 2.2T$, $n_i = n_e$, using an MLP trained with $\rho_s = 12.5 \text{ GHz}^{-1}$, $\sigma_{s,650} = 10\%$ and $2 < n_D/n_e < 1$. The mean standard deviation over the grid is $\sigma_{T_i}/T_i = 4.3\%$. On top of the contour plot is a scatter plot of $N_{test} = 500$ actual ion and electron temperature estimations where $T_{i,e}$ are in the range $0.5 \dots 10 \text{ keV}$. These are depicted by black dots surrounded by a circle, the colour of which represents the normalized deviation on the same colour scale.

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 map 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 only provide output 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 NN 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

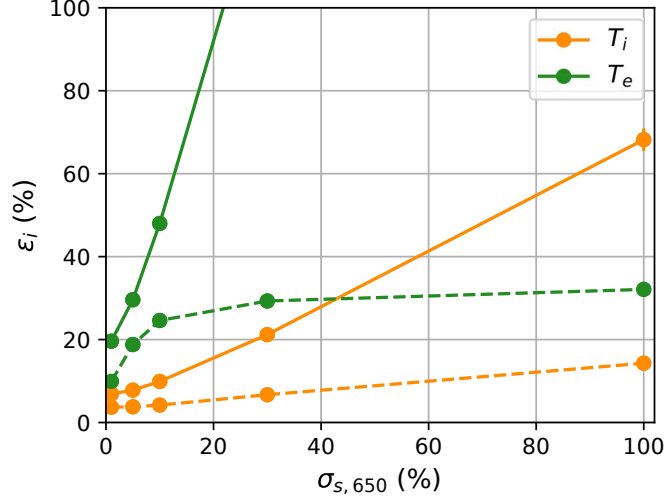


FIG. 9. Mean absolute deviations of T_i and T_e as calculated with function parametrization (full lines) and an artificial neural network (dashed lines) at $\rho_s = 12.5 \text{ GHz}^{-1}$, for different equivalent noise levels and $0.2 < n_D/n_e < 1$.

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 input and output, thus providing insight in 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 coarse 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.

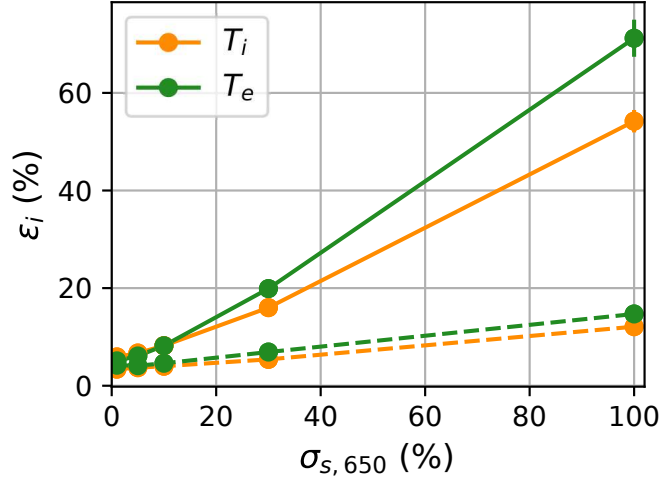


FIG. 10. Mean absolute deviations of T_i and T_e as calculated with function parametrization (full lines) and an artificial neural network (dashed lines) at $\rho_s = 12.5 \text{ GHz}^{-1}$, for different equivalent noise levels and $0.9 < n_D/n_e < 1$.

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 measured spectra in W7-X. As stated in section 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 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 $\rho_s = 12.5 \text{ GHz}^{-1}$ 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 by 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 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

I would like to thank the W7-X team for their efforts on 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 20142018 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the

REFERENCES

- ¹R. C. Wolf, C. D. Beidler, A. Dinklage, P. Helander, H. P. Laqua, F. Schauer, T. S. Pedersen, and F. Warmer, *IEEE Transactions on Plasma Science* **44**, 1466 (2016).
- ²I. Abramovic and Team, “Forward modelling of collective thomson scattering for wendelstein 7-x plasmas: Electrostatic approximation,” (2017), submitted to *Computer Physics Communications*, 2017.
- ³J. Svensson and A. Werner, in *2007 IEEE International Symposium on Intelligent Signal Processing* (2007) pp. 1–6.
- ⁴B. Braams, W. Jilge, and K. Lackner, *Nuclear Fusion* **26**, 699 (1986).
- ⁵B. van Milligen, H. Soltwisch, and N. J. Lopes Cardozo, *Nuclear Fusion* **31**, 309 (1991).
- ⁶F. Rosenblatt, *Principles of neurodynamics: perceptrons and the theory of brain mechanisms* (Spartan Books, 1962).
- ⁷B. P. van Milligen and N. J. Lopes Cardozo, *Computer Physics Communications* **66**, 243 (1991).
- ⁸A. Sengupta, P. McCarthy, J. Geiger, and A. Werner, *Nuclear Fusion* **44**, 1176 (2004).
- ⁹E. E. Salpeter, *Phys. Rev.* **120**, 1528 (1960).
- ¹⁰M. Salewski, F. Meo, M. Stejner, O. Asunta, H. Bindslev, V. Furtula, S. Korsholm, T. Kurki-Suonio, F. Leipold, F. Leuterer, P. Michelsen, D. Moseev, S. Nielsen, J. Stober, G. Tardini, D. Wagner, and P. Woskov, *Nuclear Fusion* **50**, 035012 (2010).
- ¹¹M. Nishiura, S. Kubo, K. Tanaka, R. Seki, S. Ogasawara, T. Shimoizuma, K. Okada, S. Kobayashi, T. Mutoh, K. Kawahata, T. Watari, L. E. Group, T. Saito, Y. Tatematsu, S. Korsholm, and M. Salewski, *Nuclear Fusion* **54**, 023006 (2014).
- ¹²M. Stejner, S. K. Nielsen, S. B. Korsholm, M. Salewski, H. Bindslev, V. Furtula, F. Leipold, F. Meo, P. K. Michelsen, D. Moseev, A. Bürger, M. Kantor, and M. De Baar, *Review of Scientific Instruments* **81** (2010), 10.1063/1.3475540.
- ¹³D. H. Froula, J. S. Ross, L. Divol, N. Meezan, A. J. Mackinnon, R. Wallace, S. H. Glenzer, D. H. Froula, J. S. Ross, L. Divol, N. Meezan, and A. J. Mackinnon, *Physics of Plasmas* **13** (2014), 10.1063/1.2203232.

- ¹⁴H. Bindslev, S. K. Nielsen, L. Porte, J. A. Hoekzema, S. B. Korsholm, F. Meo, P. K. Michelsen, S. Michelsen, J. W. Oosterbeek, E. L. Tsakadze, E. Westerhof, P. Woskov, and the TEXTOR Team, *Plasma Physics and Controlled Fusion* **49**, B551 (2007).
- ¹⁵S. K. Nielsen, P. K. Michelsen, S. K. Hansen, S. B. Korsholm, F. Leipold, J. Rasmussen, M. Salewski, M. Schubert, M. Stejner, J. Stober, and D. Wagner, *Physica Scripta* **92**, 024001 (2017).
- ¹⁶I. Abramovic, A. Pavone, J. Svensson, D. Moseev, M. Salewski, H. Laqua, N. L. Cardozo, and R. Wolf, *Journal of Instrumentation* **12**, C08015 (2017).
- ¹⁷H. Bindslev, *On the Theory of Thomson Scattering and Reflectometry in a Relativistic Magnetized Plasma On the Theory of Thomson Scattering and Reflectometry in a Relativistic Magnetized Plasma*, Ph.D. thesis, University of Oxford (1992).
- ¹⁸T. Hughes and S. Smith, *Nuclear Fusion* **28**, 1451 (1988).
- ¹⁹J. H. Friedman, *The Annals of Statistics* **19**, 1 (1991).
- ²⁰R. Andrae, T. Schulze-Hartung, and P. Melchior, *ArXiv e-prints* (2010), arXiv:1012.3754 [astro-ph.IM].
- ²¹Y. Li, Y. Fu, H. Li, and S. W. Zhang, in *2009 International Conference on Computational Intelligence and Natural Computing*, Vol. 1 (2009) pp. 73–76.
- ²²F. Pedregosa, R. Weiss, and M. Brucher, *Journal of Machine Learning Research* **12**, 2825 (2011).
- ²³M. Gevrey, I. Dimopoulos, and S. Lek, *Ecological Modelling* **160**, 249 (2003).
- ²⁴M. A. Mazurowski and P. M. Szecówka, in *2006 IEEE International Conference on Computational Cybernetics, ICC3* (2006).