In-situ spectral calibration from plasma measurements of the W7-X Thomson scattering diagnostic

Plasma Physics and Controlled Fusion

OPEN ACCESS

RECEIVED
17 September 2025

REVISED
11 December 2025

ACCEPTED FOR PUBLICATION
7 January 2026

PUBLISHED
23 January 2026

PAPER

In-situ spectral calibration from plasma measurements of the W7-X Thomson scattering diagnostic

G Fuchert[1,*], S A Bozhenkov[1], K J Brunner[1], J Knauer[1], E Pasch[1], A Raak[1,2], J Wagner[1], M Hirsch[1], R C Wolf[1,2] and the W7-X team[3]

[1] Max-Planck-Institut für Plasmaphysik (IPP), 17491 Greifswald, Germany
[2] Technische Universität Berlin, 10623 Berlin, Germany
[3] See Grulke *et al* 2024 (https://doi.org/10.1088/1741-4326/ad2f4d) for the W7-X Team.
[*] Author to whom any correspondence should be addressed.

E-mail: golo.fuchert@ipp.mpg.de

Keywords: Thomson scattering, calibration, diagnostics, optimizer

Abstract

The quality of Thomson scattering (TS) profiles relies on high-quality calibrations since each spatial point is essentially an independent measurement. For the spectral calibration of TS diagnostics, several techniques exist. Typically, these include a known light source and involve changes to the optical setup to observe that light source. Two assumptions are frequently made: The spectral response of the diagnostic is stable in time and the required changes to the optical setup only lead to negligible errors in the calibration. The temporal stability is questionable for larger fusion experiments, where the harsh environment leads to coating and degradation of optical components. Both the stability and validity of the spectral calibration can be checked by in-situ calibrations. Here, the calibration is performed either during operation or in between experiments without any modifications to the diagnostic. So far, two promising candidates have been proposed: dual-wavelength TS and Rayleigh scattering on, for example, argon gas using a tunable optical parametric oscillator. In this contribution, we introduce a new method for an in-situ spectral calibration using experimental data from plasma measurements. This method was developed for polychromator-based TS diagnostics, but can probably also be adapted to spectrometer-based systems. A forward model is used to predict the signals measured by the diagnostic (assuming a certain spectral calibration) and these predictions are then compared with experimental data the diagnostic has acquired. Using an optimizer, the spectral calibration is varied until predicted and measured data agree. Rough knowledge of the polychromator design is sufficient to find an estimate for the calibration of the diagnostic. We demonstrate, using polychromators from Wendelstein 7-X as example, that even data measured in the past can be calibrated with this technique.

1. Introduction

Thomson scattering (TS) is frequently employed in plasma physics to measure the electron density and temperature. Essentially measuring the spectrum of the scattered light, TS diagnostics require some kind of spectrometer. Due to typically low signal-to-noise ratios, many TS diagnostics use so-called *polychromators* [1], which are integrating the spectrum over a number of spectral channels, each representing a specific wavelength range.

This integration is achieved by a set of band-pass filters, covering the wavelength-range of interest, installed in front of a sensitive detector (e.g. an avalanche photodiode, APD). The entire spectrum is then represented by a set of scalar quantities, comprising the integrated photon counts over the different spectral channels. Since the shape of the spectrum can be modeled, the set of integrals is sufficient to infer the actual spectrum and, hence, the electron temperature and density. A suitable model for a wide temperature range (including high electron temperatures for which relativistic effects have to be considered) is the analytical *Naito formula* [2].

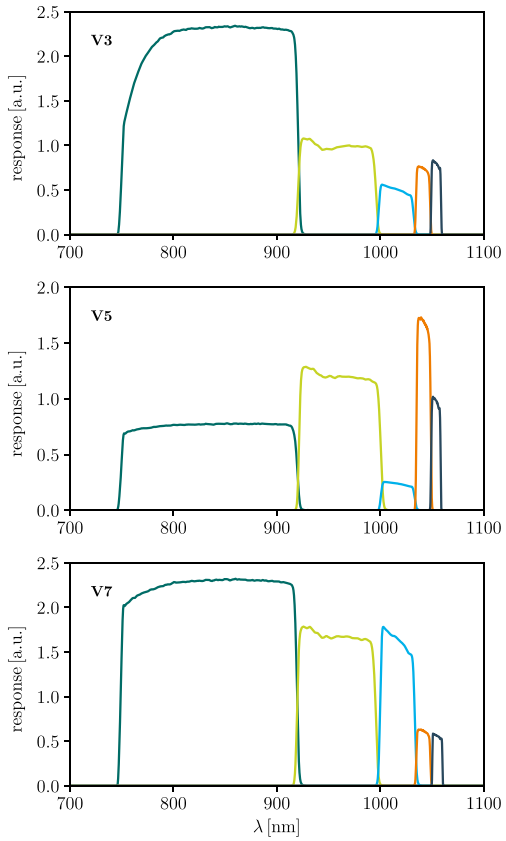

Figure 1. Three exemplary spectral calibrations for the W7-X scattering volumes 3, 5 and 7. While the general shape of the spectral channels is rather similar, the relative channel sensitivity can vary substantially, mainly due to different sensitivities of the APDs.

In order to connect the measured data with a modeled spectrum, the spectral response of the detector has to be known. This response, also referred to as the *spectral calibration*, is determined by dedicated calibration measurements. The spectral calibration is dominated by the combined transmission of all optical components (windows, observation lenses, optical fibers, the band-pass interference filters, etc) and the wavelength-dependent sensitivity of the detector.

An example from Wendelstein 7-X (W7-X) is shown in figure 1, displaying the spectral response for three different polychromators. Since each polychromator observes a small volume along the laser beam path (*scattering volume*), the calibrations are indicated by the numerical id of the observed volume.

The polychromators at W7-X feature five spectral channels (displayed by the different colors). Each channel consists of an APD and an interference band-pass filter, selecting the wavelength-range detected by that APD [3]. The bias voltage of the APDs is chosen such that the gain is maximized while avoiding spontaneous signal bursts close to the breakdown voltage. Having different gains for each APD explains the large variations in the relative channel height in figure 1.

For each laser pulse, each spectral channel integrates the observed TS spectrum over the respective wavelength-range. As already mentioned, knowing the spectral calibration, the electron temperature can be reconstructed from these integrals (the absolute scale is only important for the density and requires an additional absolute calibration). The same physical model can also be used in the inverse direction: for a given density, temperature and calibration, the five integrals can be calculated (and also here the density is only a scaling factor for the relative spectral shape determined by the temperature).

Example spectra and integrals are shown in figure 2. Here, the spectral density has been evaluated for the geometry of this particular scattering volume (V7) to obtain the number of photons per wavelength, density and laser energy. Spectra are shown for 0.5 keV and 2.0 keV, respectively. The integrated signal for the five spectral channels is shown in the legend for both temperatures. For this particular spectral calibration, these five numbers are a unique representation of the temperature. Naturally, in each practical setting, background signals, intrinsic shot noise and detector noise introduce an uncertainty in the integral values and, hence, the derived temperature value.

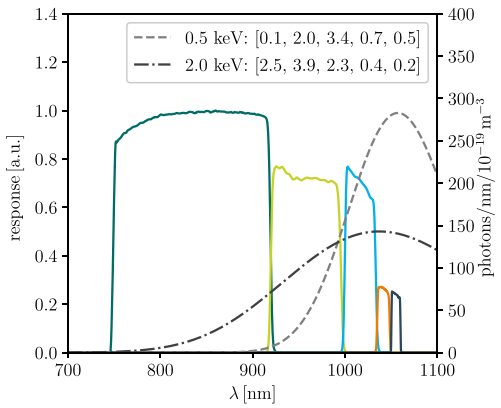

Figure 2. The TS spectra are integrated over the wavelength range of each of the five spectral channels. Here, the measured spectral calibration for V7 is shown together with two example spectra. The resulting integral values (in arbitrary units) are shown in the legend.

Conceptually, the measurements required to determine the spectral calibration are simple. A tunable light source of sufficient wavelength resolution scans the wavelength range covered by the band-pass filters and is observed by the optical and electrical systems comprising the diagnostic. In practice, however, there are several issues with this approach. Firstly, many techniques require changes to the diagnostic setup in order to observe the calibration light source. Ideally, these changes only have a negligible impact on the calibration. However, it is challenging to verify this assumption. Secondly, often these changes require human intervention or remote handling tools to prepare the diagnostic for the calibration. These interventions can be difficult to integrate into a tight experimental schedule especially if indispensable components fail and have to be replaced. Thirdly, with conventional methods, there is no straightforward way to determine temporal changes to the diagnostic other than frequent repetitions of the calibration.

Changes in the spectral calibration can occur over time by coating or degradation of optical components and, potentially, hardware issues with the detector (e.g. an unstable power supply). Naturally, hardware issues can be absent for a long time and then appear suddenly without necessarily being noticed immediately. In the worst case, this can lead to valuable experimental data being challenging or impossible to interpret.

Regardless, due to the lack of alternatives, measured calibrations often have to be treated as *ground truth* without further verification. The measurements themselves suffer from experimental errors (some of which could be systematic), possibly do not fully reflect the diagnostic state during experimental operation and are not able to track changes over time until a new calibration measurement is performed.

Calibrations that do not require changes to the diagnostic are commonly referred to as *in-situ* calibrations. An in-situ calibration under development employs Rayleigh scattering from a tunable optical parametric oscillator (OPO) [4, 5]. With the need to fill gas (e.g. argon) into the plasma vessel, frequent recalibrations are incompatible with tight experimental schedules and, hence, changes over time can only be tracked insufficiently. To solve this issue, calibration data has to be measured during experimental operation (or right before or after an experiment). This can either be a dedicated calibration signal, or in certain cases even the actual experimental data of the diagnostic. The latter is also known as a *self-calibration*. So far, two self-calibration techniques were discussed for TS diagnostics. By observing the same spectrum under two angles (*dual-angle TS* [6, 7]) or with two different laser wavelengths (*dual-wavelength TS* [8, 9]) the redundant information contained in the measurements of the same spectrum under different angles or with different wavelengths facilitates the determination of the temperature or relative spectral calibration.

Dual-angle and dual-wavelength TS are playing a crucial role for ITER-like experiments. They do require dedicated hardware modifications (additional optics for dual-angle and second wavelength laser for dual-wavelength TS), however, and come with their own uncertainties. Apart from noise, the accuracy of dual-angle TS depends on the realized geometry (the two scattering angles) [10], which may be subject to limitations from the overall machine design. Dual-wavelength TS requires a minimum temperature in order for both spectra to be observed by the detector and above that temperature the impact of noise decreases with increasing temperature (see e.g. [11]). Dual-wavelength TS is, hence, most suitable for the high-temperature regions of the plasma and it would be beneficial to have

an independent way to confirm the spectral calibration (either measured or derived from dual-angle or dual-wavelength TS). Such an additional method could also help to improve the accuracy of dual-angle and dual-wavelength TS by restricting the space of possible solutions (possibly even in a combined analysis).

Furthermore, with such a verification it becomes possible to validate or recalibrate the diagnostic retrospectively, even if substantial changes have been made to the diagnostic in the meantime (e.g. the replacement of a damaged diagnostic component or even the decommissioning of the experiment).

As we are going to show in this paper, finding an estimate of the spectral calibration from the actual experimental data alone is indeed possible with only rough knowledge of the wavelength range of the spectral channels and the general shape of the calibration. For this we will make use of a simple model for the polychromators. If measured calibrations are available, using them instead of the simple model is obviously preferable, since it improves the accuracy. The simple model is useful if no calibration measurements are possible. Furthermore, the comparison of results obtained with the simple model and actual calibrations (section 8) indicate that estimates for the electron temperature can still be found, even if the spectral dependence of the assumed calibration is not matching reality perfectly (which is also relevant for wavelength-dependent transmission changes due to degradation).

The observed wavelength ranges of the spectral channels are predominantly determined by the band-pass filters. As long as the filters are accessible, the filter curves can be measured precisely. However, due to the broad shape of the TS spectra (see e.g. figure 2), even rough knowledge of the filter curves is sufficient. In order to show that the procedure is applicable even after the decommissioning of an experimental device, in the following we are only relying on manufacturer documentation (filter curves and APD response).

For the simple polychromator model we assume that the filter curves are perfect rectangles with a transmission of 1 in the passband and 0 outside. Whether or not the actual peak transmission is indeed close to 1 is not important, as a lower value will be absorbed by the *gain factors* introduced in the following. All other optical components are assumed to have a perfect transmission of 1. The model is improved by considering the wavelength-dependence of the APD response. In this simple model, the relative sensitivity of the different spectral channels arises mainly from the electronics (in particular individually different APD sensitivities). Without attempting to fully understand these differences, they can be summarized in what we will be referring to as *gain factors*.

As mentioned above, the gain factors also absorb relative differences in the filter transmission, since these lead to a similar effect as differences in the detector sensitivity. This also applies to transmission changes of the optics due to coating or irradiation. If the resulting changes in transmission are relatively smooth (i.e. no drastic changes over the spectral width of the filters), such changes are directly absorbed by the gain factors and, hence, taken into account by the model. If the changes in transmission are more complex, the optimizer will find an effective correction, which is not perfect, but can still lead to a more accurate estimate of the temperature than the initially measured spectral calibration. Furthermore, if the spectral dependence of transmission changes are known (e.g. from theoretical models or by measuring the calibration before and after an experimental campaign), they can be included in the polychromator model with additional free parameters describing the gradual progression of these changes.

A graphical representation of the individual components and the resulting simple model polychromator is shown in figure 3. For all spectral channels ($i \in \{1, \ldots, 5\}$) a gain factor of $g_i = 1$ is assumed. Also, for channel 1, the dashed gray curve indicates the effect of a change in that gain factor. Assuming $g_1 = 0.5$, the relative channel sensitivity drops by one half. In practice, the gain factors can vary substantially between the different spectral channels. For a full description of a specific polychromator, in principle, five gain factors have to be known. However, since the spectral calibration is a relative calibration, the gain factor of one channel can arbitrarily be set to 1 and four free parameters are sufficient to describe the entire spectral calibration. The absolute scale has to be set with a separate absolute calibration.

Using the W7-X TS model [3], we can predict the signals (integrals) the simple model polychromator with $g_i = 1$ would see for different temperatures. The resulting curves are shown in figure 4 for all spectral channels. Additionally, for illustration purposes, the gain factor of the first filter is set to 0.5 (instead of 1). As expected, the signal of channel 1 drops exactly by a factor of 2 (i.e. the signal levels depend linearly on the gain factors). From this graph it is already clear that one single data sample (five integrals) would already fix all gain factors if there was no noise and if the corresponding temperature was known precisely. Consequently, using the simple polychromator model, a *cross calibration* with other diagnostics is trivial.

Furthermore, it is clear that a given data set of integrals can easily rule out several combinations of gain factors, even if the temperature is not known. For example, the data sample $[1.0, 0.6, 0.3, 0.4, 0.2]$

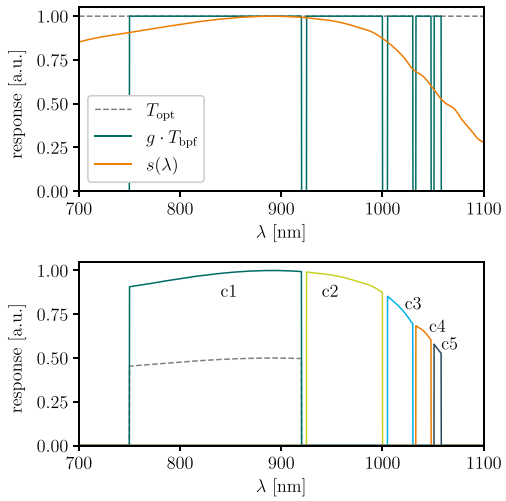

Figure 3. A simple model of a W7-X polychromator. Top: The transmission of the band-pass filters, T_{bpf} is assumed to be perfectly rectangular. The filter edges are the $T = 50\,\%$ points of the real filters (either measured or taken from manufacturing documentation). Note that the wavelength gaps between the filters have been artificially increased in this graph for illustrative purposes. The wavelength-dependent sensitivity of the APD is taken from its data sheet. For all spectral channels, a relative gain factor $g_i = 1$ is assumed. All other optical components are assumed to have a perfect transmission $T_{opt} = 1$. Bottom: The combination of the different components leads to the total spectral response of the polychromator with its five spectral channels (channel 1 to 5 from low to high wavelengths).

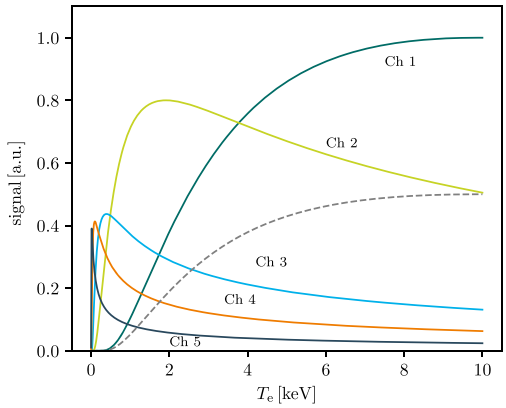

Figure 4. The integrated signals for the five different spectral channels of the simple model polychromator as a function of T_e (at constant density). The dashed gray line shows the signal of channel 1 for a gain factor $g_1 = 0.5$ instead of 1.

is not compatible with all gain factors $g_i = 1$: according to figure 4, there is no temperature for which channel 1 shows the maximum signal and channel 4 is larger than channel 3.

The observation that the signals for the different spectral channels are not independent from each other has led to the development of correction algorithms to deal with errors or changes in the spectral calibration. According to our knowledge, the earliest work on such algorithms has been done by Salzmann *et al.* In [12], an analysis is presented in which correction factors are introduced to the different channel sensitivities of the detector. These correction factors are derived from the experimental TS data in an iterative process. In [13] it is noted that this method tends to produce more accurate results for channels closer to the laser wavelength and less accurate results for channels further away from it. According to our analysis, it is likely that the reason for their observation is experimental noise.

As will be shown in detail in section 4, inevitable noise in the experimental data affects the resulting gain factors for the different spectral channels differently. This can qualitatively be understood from figure 4: At sufficiently high temperature (a few keV), the relative amplitudes of the low-temperature channels 3, 4 and 5 are rather insensitive to changes in the temperature (and could almost be obtained by eye from the experimental data), while channels 1 and 2 still show substantial variations.

Consequently, the accuracy, with which the gain factors are determined, is highest for the low temperature channels and lowest for the high temperature channels. This also coincides with how far a spectral channel is away from the laser wavelength (here 1064 nm). The further away a channel is from this wavelength, the lower the accuracy (since it measures the higher temperature parts of the spectrum). In our nomenclature, the accuracy should increase with the channel number (i.e. channel 1 should have the lowest and channel 5 the highest accuracy). As we are going to show in this paper, this effect can be mitigated by averaging the resulting gain factors for several independent data sets.

Furthermore, by comparing results for the simple polychromator model to the results obtained by correcting measured calibration curves, we will show that the exact details of the spectral dependence of the channel response do not play a large role. This is important, as impurity coating or irradiation damage of optical components can impact the spectral response of the optical system and introduce additional systematic errors. All that is needed to find such estimates for the spectral calibration is a sufficiently large data set of various different (unknown) temperatures. The data set can be assembled either from a single plasma experiment with a temperature scan (variation of plasma density or heating power) or from a number of different plasmas. In the latter case, the integrals of the various channels can be used to assemble the data set without the need for any information on the temperature.

One way to achieve a self-calibration from a data set of different (unknown) temperatures is to turn the calibration into an optimizer problem. In this paper, we demonstrate the self-calibration of our TS diagnostic with an optimizer-based solution using *differential evolution (DE)*.

The paper is structured as follows. Since we need a reference to show the successful self-calibration of the diagnostic, section 2 briefly summarizes how the spectral calibration is measured at W7-X at present. Then, in section 3, we are discussing the setup of the optimizer used for the self-calibration and apply it to synthetic data in section 4. In section 5 the optimizer is used to estimate the calibration using experimental data. A reference experiment has been introduced at W7-X to obtain a sufficiently large data set of various temperatures from just one experiment. The resulting spectral calibrations and their accuracy will be discussed in detail. With the current state of the optimizer, a number of those reference experiments are needed to achieve a sufficiently accurate calibration. However, in section 6 we are showing that a validation of an existing calibration (measured or averaged over a number of self-calibrations) is possible for each individual data set. In a sense, given that other calibration methods exist, the ability to validate an existing calibration is even more important than the self-calibration itself. Furthermore, in most cases, a better guess for the calibration exists than just the simple polychromator model. In section 7 we are showing that this information can be integrated into the optimization by rescaling an existing spectral calibration. Next, in section 8 we are discussing the impact of different calibrations on the determined electron temperature. Finally, a summary will be given together with a conclusion in section 9.

2. Spectral calibration at W7-X

At present, the calibration method trusted the most at W7-X, is a standard method in TS where a white scattering disk is placed on the two observation windows (see [14] for a general overview of the diagnostic). Using a monochromator, monochromatic light can be produced from a white laser-like *supercontinuum source* (’*SuperK*’) over the relevant wavelength range (700 to 1065 nm). This light is then guided to the scattering disk with optical fibers and is finally detected by the observation optics of the TS diagnostic. The spectral calibrations used as reference calibrations throughout this paper (including the ones shown in figure 1) have been obtained with this method. The SuperK method allows for long measurements with high signal-to-noise ratio and, consequently, very low statistical error. The slight variations in the flat-top regions of the spectral channels in figure 1 indicate that the statistical error is typically only a few percent and can be neglected for practical purposes. However, this calibration method also has two major disadvantages: 1) Since the scattering disk has to be placed on top of the observation windows, these windows themselves are not included in the calibration. So far, separate measurements have not shown a strong wavelength-dependence of the window transmission, but it is a potential source of a systematic error. 2) As it is currently implemented, the placement of the scattering disks and the preparation of the optics requires access to the torus hall and several manual interventions. Hence, the calibrations have to be integrated into the time planning for W7-X and require several days to finish. While this is acceptable at present, shorter calibration times with only remote preparations would be preferable.

Developing fully remote techniques is, furthermore, a necessary preparation for even larger experiments like ITER.

As a consequence, additional calibration methods are being developed at W7-X (an overview is given in [15]). Most promising seems to be Rayleigh scattering from an optical parametric oscillator. Rayleigh scattering occurs at the wavelength of the exciting light. With, for example, argon gas in the plasma vessel, a tunable laser-like source like the OPO can be used as light source, matching exactly the optical properties of the TS setup. Initial tests show promising results [5], but also revealed experimental challenges that still need to be overcome. During these tests, however, it was also found that just the OPO stray-light, originating from windows and other optical components along the beam path, already provides a fairly good estimate of the spectral calibration [15]. While being less accurate, this method requires very little preparation (especially no gas in the plasma vessel) and could provide a very fast, fully-remote overview calibration.

So far, very little trust has been put in the spectral calibrations determined from this stray-light method. However, we will use it as a second reference in this paper, especially with regard to observations where the optimizer solution shows deviations from the reference calibration, but agrees with the stray-light method.

3. Optimizer setup

As discussed in the introduction, for a known set of temperatures, it would be trivial to determine the gain factors from experimental TS data, as long as the rough shapes of the transmission curves are known for the different spectral channels. This would essentially be a regression, matching the experimental data with the signal curves displayed in figure 4. This also means that it is equivalent to determine either the gain factors or the electron temperatures. We will show in the following that the gain factors and temperatures can be determined simultaneously by minimizing a suitable optimizer metric.

The basic idea is to vary the gain factors as free parameters. In each optimizer iteration, the current gain factors update the spectral calibration and using this calibration the temperatures, $\mathbf{T}^\star$, are calculated from the raw data, $\mathbf{D}$. If the current guess for the calibration does not resemble the real one, also $\mathbf{T}^\star$ will not match the actual temperatures, $\mathbf{T}$. More importantly, there is an inconsistency between $\mathbf{D}$ (the measured raw data), and $\mathbf{D}^\star$, the data that *would* be observed for $\mathbf{T}^\star$ with the assumed spectral calibration.

To quantify this mismatch, we use the TS model to predict $\mathbf{D}^\star$ from $\mathbf{T}^\star$, using the spectral calibration of each iteration. Once the optimizer finds the correct spectral calibration, $\mathbf{D}$ and $\mathbf{D}^\star$ as well as $\mathbf{T}$ and $\mathbf{T}^\star$ should be similar (though $\mathbf{T}$ is, of course, generally not known).

We construct the following optimizer metric (cost function), M, as a *relative squared error*:

$$M = \sum_{i=1}^{N} \sum_{j=1}^{5} \left(\left(D_{ij} - D_{ij}^\star \right) \Big/ \left(D_{ij} + D_{ij}^\star \right) \right)^2 . \tag{1}$$

Here, i indexes the N data samples and j the five spectral channels. The D_{ij} and $D_{ij}^\star$ then represent the experimental and predicted signal values, respectively, for spectral channel j of data sample i.

The optimizer varies four of the five gain factors. As discussed in the introduction, since the spectral calibration is a relative calibration, one gain factor is not a free parameter. The gain factor of the first channel, g_1, is set to 1. Furthermore, we have found that the optimizer converges much faster if we are not using the gain factors g_i themselves as free parameters, but their ratios (i.e. $r_i = g_i/g_{i-1}$). In order to constrain the search space, we assume that the relative sensitivity between the channels does not vary by more than a factor of 10 (which is already more than we see in practice or would accept for a useful polychromator). This is ensured by the constraint that $r_i \in [0.1, 10]$ and $g_i \in [0.1, 10]$ (in principle, the latter constraint is already sufficient, but the additional constraint on the r_i leads to fewer rejected iterations and, hence, a faster runtime).

Evaluations of M using randomly sampled values for r_i revealed that M has shallow gradients around its minimum, but large gradients farther away. This is an issue since the integrals representing the TS data are often noisy, making the search for the exact minimum of M unreliable. An appropriate optimizer choice in such cases is *DE* [16]. In DE, there is not just one candidate solution that is varied in each iteration of the optimizer. Instead, there is a *population* of candidate *individuals*. An initial population is formed by generating a number of initial guesses. Then, new candidate solutions are proposed in each iteration and compete with the existing individuals for survival. Following a set of rules, these new candidates are composed from the individuals of the existing population.

There are different rule sets for how exactly the new candidates are determined. But generally, for each individual $\mathbf{x}$ (all individuals are denoted as vector-like quantities, representing one set of the free

parameters to be optimized), three other individuals, $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$, are randomly chosen from the population and combined in the *mutation* step as

$$\mathbf{y} = \mathbf{a} + F(\mathbf{b} - \mathbf{c}) \,. \tag{2}$$

F is called the *differential weight* and $F \in [0, 2]$. In some rule sets, the so-called *base vector* $\mathbf{a}$ is not chosen randomly, but selected as the individual with the lowest value of M in the current population. Before selection, the new candidate is modified further in the *crossover* step: each element of $\mathbf{y}$ is either retained or replaced with the corresponding element of $\mathbf{x}$, based on a predefined *crossover probability*. Finally, $\mathbf{x}$ and $\mathbf{y}$ compete for survival. The new candidate $\mathbf{y}$ replaces $\mathbf{x}$ in the population if $M(\mathbf{y}) \leqslant M(\mathbf{x})$. After all new candidates have been compared to their respective individuals from the current population and have either replaced them or were discarded, the next iteration begins with the updated population.

DE has two important properties that help to overcome the noise issue. Firstly, its search strategy does not utilize gradients and secondly, since several candidate solutions compete with each other simultaneously, the structure of M is typically mapped out rather well around its minima.

As we are going to show in section 4, the exact minimum found by the optimizer is not necessarily a reliable representation of the spectral calibration. As has already been mentioned, M is very flat around the minimum and with noisy data it cannot be expected to find the perfect solution. In order to deal with this we have slightly modified the determination of the optimizer result. Without a break condition, DE will continue until all individuals settle in a minimum (not necessarily the same one). In this case, the population is not updated further or only by insignificant variations around the current solutions. Hence, a break condition is required to end the optimization. Standard break conditions mostly focus on changes in the population over time, measuring the distance between the population before and after the latest iteration. In our approach, we want to retain information about the behavior of M around the minimum. Hence, the break condition is defined based on the variability within the population. As soon as the maximum M of the current population is less than $10\,\%$ higher than the minimum M for the same population, the optimization is stopped.

We have found empirically that the measured reference calibration is approximated best if we determine the final r_i by averaging the minimum and maximum r_i for all individuals with M less than $2.5\,\%$ above the minimum of M for that final population. Note that over the course of the optimization, M changes over several orders of magnitude. Hence, the $2.5\,\%$ represent very small changes in M. Naturally, it would be preferable to modify the optimization metric such that no such averaging and choice of threshold are required. Finding such modifications is still work in progress.

For this work, we are employing the SciPy [17] implementation of DE. When setting up DE, a number of parameter choices have to be made that impact the speed and the robustness of the optimizer. The most important are the population size, the differential weight F and the crossover probability (see above for an explanation of these parameters). Furthermore, the exact rule set for determining the new candidates has to be defined (often referred to as *strategy*).

For the population size, there is a rule of thumb of ten times the number of free parameters (i.e. in our case $10 \cdot 4 = 40$). However, since we are using the final population as a map of M around the minimum, we use a population size of 120 to increase the accuracy of the final averaging of the free parameters. For the other parameters we use $F = 1.5$ and a crossover probability of 0.8 (i.e. each parameter of the new candidate has an $80\,\%$ chance of being accepted by the mutated vector $\mathbf{y}$).

These parameters had to be tuned initially to find the right balance between exploration of the parameter space and convergence towards promising solutions. This is mostly a process of trial-and-error and is one of the weaknesses of DE. The exact parameter choice usually affects primarily the speed with which the optimizer converges and not the final solution. For cost functions with local minima, premature convergence can prevent DE from finding the global minimum. Using synthetic TS data to evaluate equation (1), however, we have not observed any local minima and saw no significant impact of the parameter choice on the resulting solution. Details and best practices about how to setup a DE optimizer can be found in [18].

Furthermore, in order to speed up the convergence, we are using the *DE/best/1/bin* strategy as our rule set to compose the new candidate individuals in DE (see [16] for the naming convention and introduction of different strategies). This rule set essentially follows the basic rules as outlined above and uses the currently leading individual (lowest value of M) as the base vector $\mathbf{a}$ in equation (2). This strategy has the advantage of quickly steering towards a solution, but at the risk of getting stuck in a local minimum for multi-modal optimization metrics. This does not seem to be an issue in our case, as we have verified by comparing the final solution with those obtained from strategies with randomly chosen base vectors (such as *DE/rand/1/bin*). Such strategies are more robust but at the cost of slower convergence.

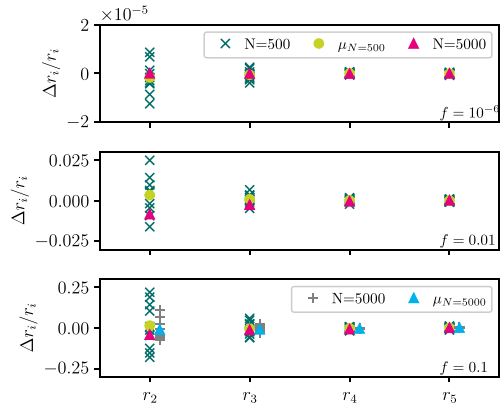

Figure 5. For each noise level ($f = 10^{-6}$, 0.01 and 0.1, from top to bottom), the deviation of the optimizer solution to the model input ($\Delta r_i/r_i$) is shown for the different r_i for all analyzed data sets. For $N = 500$, ten data sets have been analyzed. The individual results are shown together with the average over all ten runs. For the cases with low or no noise, only one data set has been analyzed for $N = 5000$, since this represents the same number of total data samples. For the high-noise case, an additional analysis with ten $N = 5000$ data sets has been included to illustrate the impact of the sample size.

4. Application to synthetic data

By creating synthetic data from a known calibration, it can be shown that the described algorithm retrieves that input calibration. For this, we use the simple polychromator model introduced above. For each data set, a temperature range is defined by sampling a lower and upper temperature boundary (T_{min} and T_{max}). To loosely resemble W7-X, T_{min} is drawn from a uniform distribution between 0.5 and 1.0 keV and T_{max} from a uniform distribution between 2.0 and 5.0 keV. The N temperature values for each data set are drawn from a uniform distribution between T_{min} and T_{max} for that respective data set.

Using the W7-X TS model together with the simple polychromator model, the synthetic TS data for all of the sampled temperature values can be simulated (using a scattering angle of 118.5° as an example). Finally, we add noise to each individual data sample. This synthetic error is sampled from a Gaussian distribution with $\sigma = fI$. Here, f is a constant factor and I is the respective signal value in a spectral channel. In the case of W7-X, it has been observed that $\sigma = fI$ with $f = 0.1$ is a good estimate for the statistical error over a broad signal range. This can be studied, for example, during the Raman calibration.

In order to investigate the impact of noise, we investigate three cases. No noise ($f = 10^{-6}$, $f = 0$ is not possible for numerical reasons, as it would lead to a division by zero in the error analysis of the TS evaluation), *low noise* ($f = 0.01$) and *high noise* ($f = 0.1$).

In summary, for each of the temperature samples T_{i}, the ideal TS data for each spectral channel is simulated, $[I_1, I_2, I_3, I_4, I_5]$, and then from this the synthetic measurement is created as $[I_1 + G(fI_1), I_2 + G(fI_2), I_3 + G(fI_3), I_4 + G(fI_4), I_5 + G(fI_5)]$. Here, $G(\sigma)$ represents one random sample from a Gaussian distribution with standard deviation σ, $G(\sigma) \sim \mathcal{N}(0, \sigma^2)$.

As will be discussed in section 5, a typical experiment at W7-X yields several hundred to a few thousand TS profiles. These can either be analyzed in one combined data set or individually for each experiment (or possibly also in even smaller batches). In order to investigate whether one large data set or several smaller ones result in a more accurate estimate of the calibration, for each noise level, ten data sets are created with 500 samples and additionally one data set with 5000 samples. This way, both test sets have the same number of total samples and need roughly the same computational time to process. The gain ratios are determined for each of the data sets using the optimization algorithm. In this analysis, the correct gain ratios are known since they were used to calculate the synthetic data and the success of the optimizer can be assessed by the difference of the model input, r_{in}, and the optimizer output r_{out}, normalized to the model input (i.e. for gain ratio i we obtain $\Delta r_i/r_i = (r_{\mathrm{in}} - r_{\mathrm{out}})/r_{\mathrm{in}}$). A result of $\Delta r_i/r_i = 0$ describes a perfect reconstruction of r_i by the optimizer. Figure 5 shows the resulting $\Delta r_i/r_i$ for all of the data sets created as described above and the corresponding numerical values are stated in table 1.

As expected (see the discussion of figure 4), the largest uncertainty is observed in r_2 for all cases, followed by r_3. Furthermore, the uncertainty in all gain ratios increases with the level of noise. In the case

Table 1. $\Delta r_i/r_i$ and $\sigma_{\bar{r}_i}/\bar{r}_i$ for the different data sets in figure 5.

Data set	$\Delta r_2/r_2$	$\Delta r_3/r_3$	$\Delta r_4/r_4$	$\Delta r_5/r_5$
$N = 5000, f = 0$	5×10^{-8}	-2×10^{-8}	-6×10^{-8}	4×10^{-8}
$10 \times 500, f = 0$	-2×10^{-6}	-6×10^{-7}	-1×10^{-9}	-1×10^{-7}
$\sigma_{\bar{r}_i/\bar{r}_i, 10\times500, f=0}$	2×10^{-6}	6×10^{-7}	1×10^{-7}	1×10^{-7}
$N = 5000, f = 0.01$	-0.008	-0.003	4×10^{-4}	2×10^{-4}
$10 \times 500, f = 0.01$	0.003	0.001	2×10^{-4}	3×10^{-6}
$\sigma_{\bar{r}_i/\bar{r}_i, 10\times500, f=0.01}$	0.004	0.001	3×10^{-4}	2×10^{-4}
$N = 5000, f = 0.1$	-0.042	-0.012	-0.008	0.001
$10 \times 500, f = 0.1$	0.014	-0.006	-0.003	3×10^{-5}
$\sigma_{\bar{r}_i/\bar{r}_i, 10\times500, f=0.1}$	0.043	0.013	0.003	0.002
$10 \times 5000, f = 0.1$	-0.010	-0.009	-0.004	0.001
$\sigma_{\bar{r}_i/\bar{r}_i, 10\times5000, f=0.1}$	0.009	0.003	4×10^{-4}	0.001

of no noise, the remaining uncertainty is well below the percentage range and would be of no practical importance. In the low-noise case, the uncertainty in r_2 is in the low percentage range for the individual cases with 500 samples. For the high-noise case (representing what is observed at W7-X) the uncertainty in r_2 is substantially higher (few tens of percent). However, the resulting gain ratios seem to scatter around the correct value used to create the synthetic data ($\Delta r_i/r_i = 0$). Their mean value should, hence, represent a better estimate for the actual calibration as the individual solutions and their spread also allows for an initial error estimate. The noise-induced uncertainty in the gain ratios (averaged for several independent data sets) can be estimated by the standard deviation of the mean of the different r_i (i.e. $\sigma_{\bar{r}_i} = \sigma(r_i)/\sqrt{n}$, where n is the number of analyzed data sets). An additional observation is that at least for the analyzed data set, as soon as noise is present, the average over the results from all ten sets of 500 samples is closer to the actual solution than the optimizer result for one data set of 5000 samples. Since noise will always be present, it seems favorable to divide an existing data set into subsets and analyze each of them individually (reducing the impact of the noise and allowing for an error estimation).

To study the impact of the sample size, we have analyzed an additional data set for the most relevant high-noise case with ten times 5000 samples (also shown in figure 5 and table 1). The most obvious difference to the 500 sample case is the reduced $\sigma_{\bar{r}_i}/\bar{r}_i$ (lower scatter). The averaged solution is also slightly closer to the model input in r_2, but not the other r_i. Given that the total number of samples and computational effort was ten times higher than with 500 samples per data set, also this result indicates that it is more efficient to have a larger number of data sets with a smaller sample size.

Further studies are required to find the ideal sample size and number of data sets providing the optimum between highest accuracy in the final result and lowest number of data samples and computational effort. Also, using $\sigma_{\bar{r}_i}$ as error estimate for r_i assumes that the individual solutions are normally distributed around the actual solution. This needs to be confirmed in future studies with a large number of analyzed data sets or the error estimate has to be revised.

5. Application to W7-X data

The easiest way to generate a sufficient data set for the self-calibration is to run a density ramp (i.e. a linear increase of the plasma density over a wide range within just a few seconds). The TS diagnostic operates with three lasers at 30 Hz repetition rate each. Hence, over one second of plasma operation, 90 full profiles are recorded. Following the analysis of synthetic data in section 4, we aim to collect around 500–1000 samples per experiment and analyze each density ramp individually. As the predicted signals shown in figure 4 illustrate, a selection of different temperatures is possible just from the raw data without any knowledge on the actual temperature (which, obviously, is a prerequisite for a self-calibration). Collecting data from different experiments would be possible as well, but for W7-X a TS reference experiment has been established, where the line-integrated density is ramped by gas-puffing from around 4 to $9 \cdot 10^{19}\,\mathrm{m}^{-2}$ over a time span of around 8 s (at a constant ECRH heating power of 2 MW). Using feedback control for the gas-puff based on the line-averaged density (measured by a dispersion interferometer [19]), such a scenario can be run in W7-X very reproducibly.

Each run of this reference experiment yields around 700 samples (90 Hz repetition rate for 8 s during the density ramp). Figure 6 shows time traces of the ECRH power, the line-integrated density and TS temperatures for three different spatial locations (scattering volumes). The temperatures are evaluated

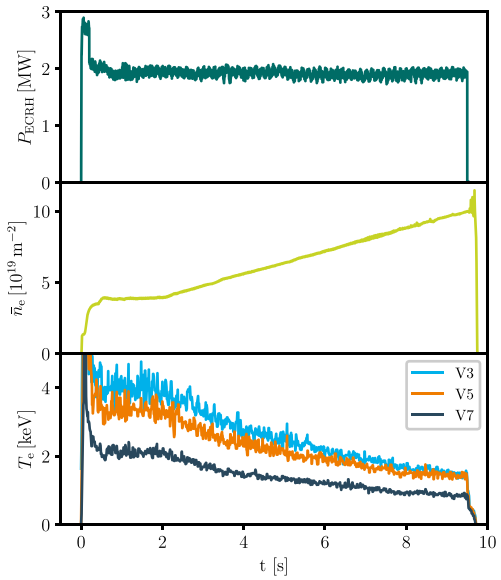

Figure 6. To ensure reproducible plasma conditions, a reference experiment has been established for the TS self-calibration at W7-X. The top graph shows the ECRH power, the middle graph the line-integrated density and the bottom graph shows electron temperatures measured by TS and evaluated with the *SuperK* reference calibration (see section 2) for three exemplary radial locations from core to half-radius (scattering volumes V3, V5 and V7).

with the measured reference calibration for the example scattering volumes 3, 5 and 7 shown previously. These volumes represent roughly the plasma center (highest temperature), one quarter of the minor radius (intermediate temperature in the graph) and mid-radius (lowest temperature in the graph). The temperature range covered is roughly between 1.5 and 4 keV for volume 3, 1.5 and 3.5 keV for volume 5 and 0.8 and 2.2 keV for volume 7. These temperatures are not known to the optimizer and are only displayed to show the typical temperature range covered.

In the experimental campaign of 2025, OP2.3, the reference experiment was repeated six times. The optimizer calibration has been attempted for all six experiments individually. The resulting gain factor ratios are shown in figure 7 for the three example volumes. In order to make differences in precision easily visible between the different ratios, each resulting ratio r_i is normalized by the average of r_i over all six experiments. Furthermore, for each resulting calibration, the temperatures are evaluated and the average temperatures between $t = 2$ and 9 s are shown in the graph to characterize differences between the individual results. In order to be robust against outliers and slight differences between the different experiments, the median has been chosen over the mean.

It is visible that there is a substantial scatter in the resulting calibration parameters (represented by the $r_i/\langle r_i \rangle$) and the median temperatures $\bar{T}_e$. The variation seen in $\bar{T}_e$ is a direct consequence of the scatter in r_2. As the analysis of the synthetic data shows, this is caused by the noise of the experimental data (note how similar the scatter in the different gain ratios looks to figure 5). Following the discussion in section 4, we average the resulting gain ratios for the different data sets to find the final estimate for the spectral calibration and, in turn, the correct electron temperature.

As in the analysis of the synthetic data, the highest uncertainty is present in r_2 (i.e. the two gain factors g_1 and g_2). The ratio r_3 is already determined more precisely and the scatter in r_4 and r_5 is small (except for r_5 in V3, which will be discussed further in section 6). The resulting $\sigma_{\bar{r}_i}/\bar{r}_i$ are shown in table 2.

In this procedure, averaging has taken place twice. Firstly over the final population for each experiment and secondly over the individual optimizer runs. We are referring to this procedure as *two-level averaging*. The final calibrations after the two-level averaging are shown in figure 8 and compared to the measured reference calibrations (see section 2). The error band in the graph is constructed as follows: assuming that $g_1 = 1$, each other g_i can be constructed as $g_i = g_{i-1}/r_i$. Two calibrations can be constructed from $r_i + \sigma_{\bar{r},i}$ and $r_i - \sigma_{\bar{r},i}$. This way, however, different calibrations would all show the same sensitivity for channel 1 ($g_1 = 1$), and all other channels would be scaled up or down depending on the actual sensitivity of channel 1. In practice, the absolute calibration fixes the scale of the spectral calibration. For the temperature, only the relative differences between the channels are relevant and they remain

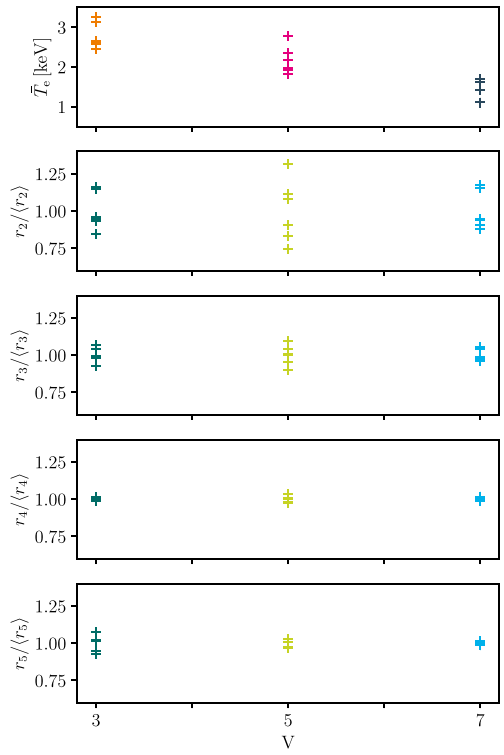

Figure 7. Using the simple polychromator model, the self-calibration has been attempted for six identical reference experiments. The uncertainty between the individual experiments is characterized by the resulting median temperature, $\bar{T}_e$ and the variations in the individual gain factor ratios (normalized to the average gain factors for all six experiments).

Table 2. $\sigma_{\bar{r}_i}/\bar{r}_i$ for the three example volumes analyzed with the simple polychromator model.

Volume	$\sigma_{\bar{r}_2}/\bar{r}_2$	$\sigma_{\bar{r}_3}/\bar{r}_3$	$\sigma_{\bar{r}_4}/\bar{r}_4$	$\sigma_{\bar{r}_5}/\bar{r}_5$
V3	0.048	0.018	0.005	0.020
V5	0.079	0.025	0.008	0.009
V7	0.048	0.014	0.004	0.003

unchanged by any renormalization. Hence, purely for visualization, with no further effects on the analysis, we normalize each channel by the average of g_3 and g_4 ($g_i^* = g_i/(g_3 + g_4)$). This makes it possible to visually asses the variation in all channels.

The following observations can be made from figure 8:

1. Overall, the reference calibrations are well reproduced by the optimizer using the simple polychromator model. The agreement is almost perfect for V5, V3 shows minor differences and for V7 these are more pronounced.
2. For V3, one optical element seems to have a higher wavelength cut-off in the range between 700 and 800 nm. Since this behavior is not included in the simple model, it cannot be reproduced. However, the resulting optimizer solution seems to lower the overall sensitivity of the first channel to keep the integral over the different TS spectra in-line with the expectations.
3. Also in V3, g_5 seems to be around 10 % higher for the optimizer calibration compared to the reference calibration. This seems odd, since it is expected that g_3, g_4 and g_5 show much less scatter than g_1 and g_2. This could indicate that the observed feature is true, which will be discussed further in the following.
4. For V7, the g_3, g_4 and g_5 are well reproduced, while g_1 and g_2 show noticeable differences. Initially, we have attributed these deviations to the higher uncertainty of the optimizer solution in these two parameters. Further analysis revealed, however, that the other experimental calibration methods developed for W7-X agree with the optimizer solution rather than the reference calibration. This will also be discussed further below.

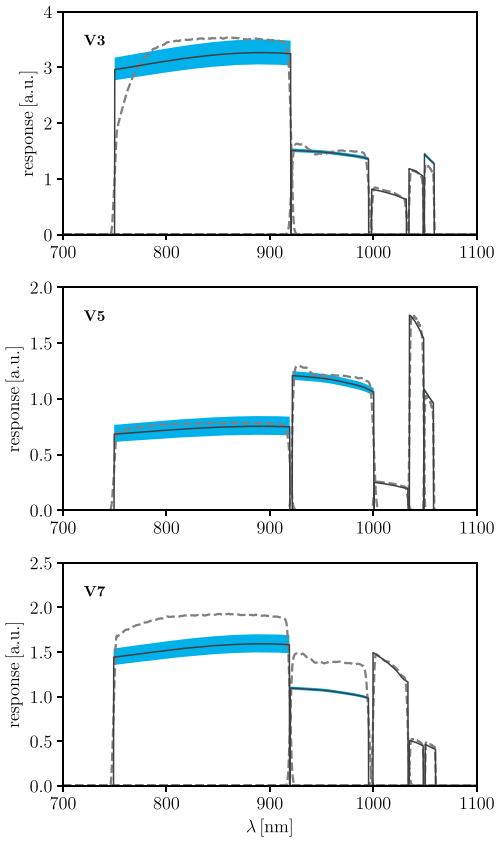

Figure 8. Comparison of the measured reference calibration (gray dashed line) with the optimizer solution (gray line with blue error band) for three different example volumes (V3, 5 and 7). Counting from left to right, channels 1 and 2 are less well reproduced than channels 3 to 5. The error band represents the standard deviation of the mean of the different r_i.

Overall, the self-calibration obtained from the optimizer results shows a good agreement with the measured reference calibration. As has just been discussed, noticeable differences are observed in channel 5 for V3 and channels 1 and 2 for V7. Given the scatter observed for the optimizer solutions between the different reference experiments, it seems tempting to attribute these differences purely to uncertainties in the optimizer solution. In order to get a better understanding for these deviations, both the self-calibration and the reference calibration are also compared to the *stray-light calibration* introduced briefly in section 2 and discussed in more detail in [15].

As can be seen in figure 9, both in V3 and V7, the stray-light calibration agrees better with the optimizer solution than the reference calibration: For V3, the simple polychromator model is obviously unable to recover the reduced sensitivity in channel 1 below 800 nm. But together with the stray-light calibration it indicates that the overall sensitivity of channel 1 could be overestimated by the reference calibration. Furthermore, both the optimizer solution and the stray-light calibration show an increased sensitivity in channel 5 of the same volume. For V7, the optimizer solution and the stray-light calibration agree on the reduced sensitivity of channel 1 and 2. As has been discussed in section 2, the accuracy of the stray-light calibration is considered to be low, since assumptions on the stray-light spectrum have to be made. So far, it is not possible to measure the stray-light spectrum entering the optical fibers of the TS diagnostic. It is, hence, assumed that the stray-light source has no spectral dependence other than energy spectrum of the OPO (see [15] for details). Consequently, differences between the stray-light calibration and the reference calibration can partially be explained by the stray-light source having a more complex spectral dependence. An incorrect assumption of the stray-light spectrum could explain differences between the reference and the stray-light calibration like the one seen in the transmission band of channel 1 in V7 (between 750 and 920 nm). Features like the reduced sensitivity of channel 2 in the same volume (which are more relevant for the temperature and the discussion in this paper), however, cannot be explained by an incorrect stray-light spectrum. It is very unlikely that the stray-light spectrum is reduced drastically exactly in the wavelength range of channel 2 (between 920 and 1000 nm) and only for V7, but not the others.

But if these differences are real, what could cause such disagreements? Obviously, all calibrations have been measured or derived with very different methods and from completely independent data sets. There can be optical effects that change the effective sensitivity of individual channels. One possible example for such an effect is an internal misalignment of the polychromators. For a perfectly aligned polychromator, the entire collection cone (defined by the numerical aperture) is focused onto the APDs and the size (angular extent) of the light source should not affect the spectral calibration. If some polychromators were not aligned perfectly, however, the calibration could depend on the size of the observed light source. Note, however, that this is just one example to illustrate why the spectral calibration could depend on the spot size. So far, the reason for the observed deviations is not known and under investigation.

Another important difference that could explain the observations: The different calibrations represent the diagnostic at different times. While the reference and the stray-light calibrations were measured before the experimental campaign (and the reference calibration being the oldest), the reference experiments for the self-calibrations were all executed during the campaign. Hence, it could also be the case that something has changed the state of the detector between the reference calibration and the stray-light calibration, but then remained stable during the campaign. Possible explanations are under investigation, but since APDs are sensitive to changes in the ambient temperature and the applied bias-voltage, the power supply and temperature control of the affected polychromators are obvious candidates to investigate. In particular, issues with the APD stability are likely to explain the observed scatter in r_5 for V3. While seemingly the optimizer solution and the stray-light calibration agree well, the scatter in r_5 cannot be explained by detector misalignment. Our current interpretation is that the gain factor g_5 of this particular polychromator seems to be drifting. We come to this conclusion since r_5 shows a scatter that is high both in comparison to the other volumes (V5 and V7) and also r_4 for the same volume (V3).

This leads us to an important issue, however. One of the main motivations of developing the optimizer-based self-calibration was to verify the stability of the calibration over time. And even though we have seen that a reliable calibration can be found by two-level averaging, the uncertainty between individual reference experiments seems to make the possibility of a validation of the time-stability of g_1 and g_2 questionable. As will be discussed in the following, however, the time-dependence can be assessed even if the actual calibration remains uncertain.

6. Validation of an existing calibration

As shown in the previous section, an optimizer-based self-calibration seems to be possible, which was one of the goals of this project. However, due to the inevitable noise on the raw data and low-sensitivity of the metric around the minimum, the gain ratio r_2 (and to a smaller degree r_3) showed a large scatter between individual experiments and two-level averaging had to be employed to achieve a satisfactory accuracy. In contrast, due to their low temperature-sensitivity above 1 keV, the temporal stability of r_4 and r_5 can be assessed directly.

One might argue, however, that the validation of the stability of the calibration is even more important than the calibration itself: There are proven methods to measure the spectral calibration directly. With the optimizer-based self-calibration, systematic errors and time-dependencies can be identified (or excluded) in existing calibrations exactly at the time of plasma operation.

As we show in the following, a validation is possible for individual experiments if there is a reproducible way to obtain the data set for the self-calibration. This is achieved either by a reproducible reference experiment or by the input of other diagnostics. These can either be local measurements for a cross-calibration, or global diagnostics (heating power, line-averaged densities, etc) in order to assemble a data set of comparable plasma states.

Here, we focus on the use of a reference experiment in order to show that a validation is possible using the TS diagnostic alone. In practice, it is always preferable to include further diagnostics to ensure reproducible conditions.

As a proof-of-principle, we assume that the median temperature $\bar{T}_e$ for the different reference experiments is identical. Since the uncertainty in r_2 and T_e are strongly coupled, keeping $\bar{T}_e$ constant will also fix r_2, even if the resulting value may not be the correct one. This can, for example, be achieved by introducing a penalty term in the optimizer metric.

First, an already existing calibration (either measured or from the optimizer) is used to evaluate the temperatures of one of the reference experiments and the median temperature $\bar{T}_e$ is recorded. Second, the optimizer is used for each individual data set. Whenever the metric is evaluated, the currently analyzed calibration candidate is used to calculate the temperatures and determine $\bar{T}_e$. This current estimate for $\bar{T}_e$ is compared to the recorded reference value. If it agrees within $\pm 5\,\%$, the resulting value of M is

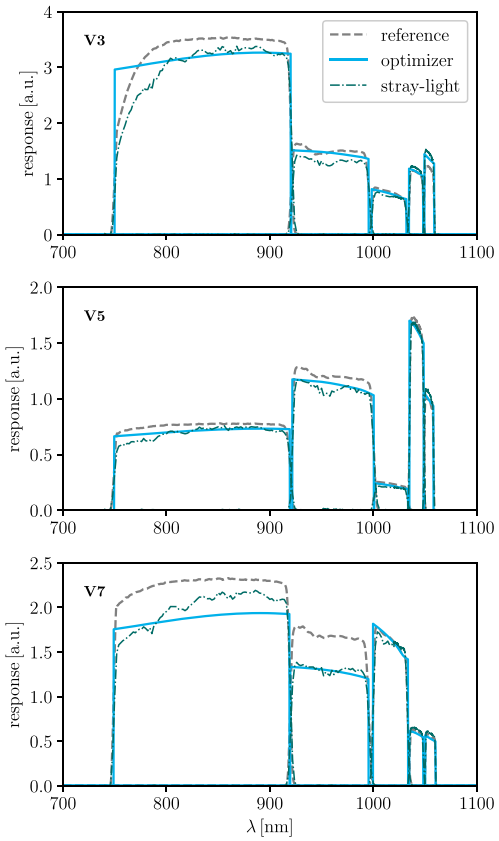

Figure 9. Comparison of the measured reference calibration (dashed gray line) with the optimizer solution (blue line) and the stray-light calibration (dash-dotted dark green line) for three different example volumes (V3, 5 and 7).

returned without modification. If the disagreement is larger, a severe penalty value is multiplied to M. In this example, we use a penalty value of 10, but since the metric is very flat around its minimum, the exact choice of this penalty value does not matter much (though a tapered penalty that increases with the disagreement in $\bar{T}_e$ can help the optimizer to converge faster).

In the following it will be demonstrated that such a penalty term indeed reduces the scatter in r_2 and r_3 observed earlier. Furthermore, the ability of this method to detect changes in r_2 can be shown directly. As discussed in the introduction of the simple polychromator model, the detected signals are directly proportional to the gain factors. Hence, a change in r_2 (or any other ratio) can be simulated by modifying the raw-data accordingly. A reduction of 10 % in g_1 is equivalent to a reduction of all signals from that channel by 10 %. This can easily be simulated by manipulating one of the data sets (multiplication of all signals from channel 1 with 0.9). Everything else being equal, the optimizer should find an increase in $r_2 = g_2/g_1$ of roughly 11 % ($1/0.9 \approx 1.11$).

The analysis of all reference experiments has been repeated with the temperature-penalty term and for one of the data sets, a reduction in g_1 by 10 % was simulated as discussed above. Figure 10 shows the resulting gain ratios. The optimizer solutions for all unmodified data sets are shown with '+' markers, while the solution for the modified data set is shown as purple 'x'. As expected, the scatter in r_2 and r_3 is drastically reduced and the modified data set was clearly identified with a change in r_2 of the right order of magnitude. A small remaining scatter in the different parameters is seen and is likely due to small changes in the actual median temperature between the different reference experiments. This observation indicates that, whenever possible, further diagnostics should be added to ensure a reproducible data set and, in turn, increase the achieved precision.

Furthermore, it is also clearly visible that the scatter in r_5 is still present in V3, reinforcing our concern that the gain of this particular channel seems to be drifting.

The fact that both the simulated change in g_1 (all volumes) and the suspected variability in g_5 (only V3) have been detected, demonstrates that the stability of the calibration can indeed be investigated, even with the uncertainties observed for individual experiments. With this, the two main goals of the optimizer calibration are achieved. With only minimal knowledge of the polychromators, even a simple

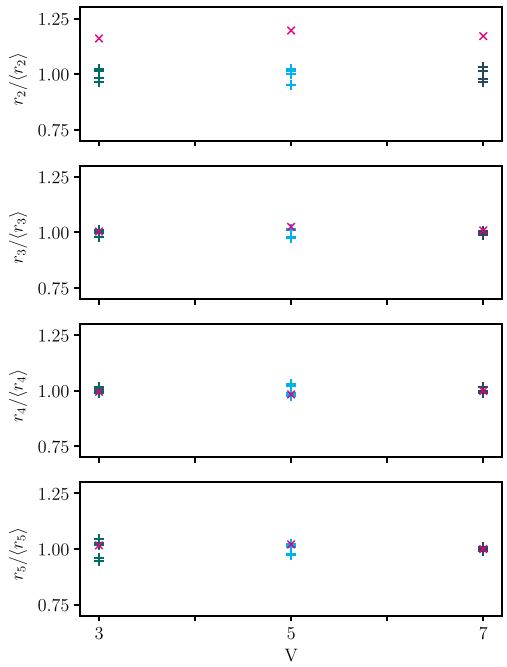

Figure 10. Normalized gain ratios r_i for the different reference experiments, using a penalty term enforcing $\bar{T}_e$ to change less than $\pm 5\,\%$ between the individual data sets. The results from the individual reference experiments are indicated by the '+' markers and show a low scatter (good temporal stability) for most of the calibration parameters. The purple 'x' represents the solution for a data set that has been manipulated intentionally to simulate a lower g_1 (higher $r_2 = g_2/g_1$).

model is sufficient to find a rough estimate of the of the spectral calibration and the resulting temperature values. Furthermore, as long as a reproducible data set can be constructed, the temporal stability of the diagnostic can be assessed.

7. Rescaling measured calibrations

So far, we have assumed that only rudimentary knowledge on the spectral calibration is available. A simple polychromator model is sufficient to approximate an unknown calibration and, hence, determine the electron temperature from experimental TS data alone. In most fusion experiments, however, calibration measurements are possible (at least infrequently) and, therefore, the relative spectral calibration is well known.

A self-calibration based on plasma measurements enables the verification of the spectral calibration, ensuring high-quality profile data. If the diagnostic has undergone changes, the self-calibration not only identifies invalid spectral calibrations but also enables their correction through rescaling. This preserves the usability of affected raw data, saving valuable experimental time by reducing the need to repeat experiments with insufficient profile quality. Such an approach resembles correction algorithms like the one discussed in [12, 13].

For this rescaling, the measured spectral calibration is taken as input. Each spectral channel is normalized to its peak response and this *normalized spectral calibration* replaces the simple polychromator model in the optimization. The resulting median temperatures and gain factor ratios (normalized to their means) are shown in figure 11.

Essentially, the same observations can be made as with the simple model. For the individual runs, there is a relatively high uncertainty in $\bar{T}_e$ and r_2, less uncertainty in r_3 and almost no uncertainty in r_4 and r_5. Also the scatter in r_5 for V3 is retrieved, giving further evidence that this could be caused by an unstable gain for that particular APD.

Following the same procedure as before, the resulting calibration parameters are averaged to account for the low precision found in the analysis of the different reference experiments. This averaged solution is compared to the measured reference calibration in figure 12 and the numerical values of the resulting $\sigma_{\bar{r}_i}$ are listed in table 3.

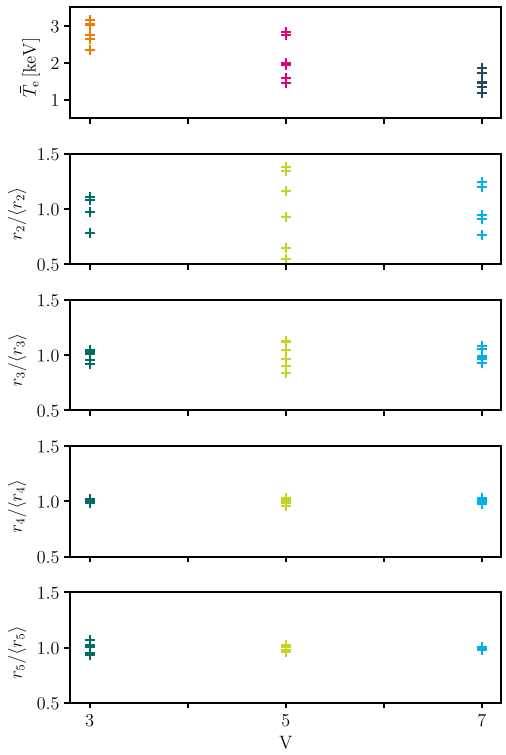

Figure 11. Same representation as in figure 7, but with a rescaling of the measured reference calibration as a more realistic model compared to the simple polychromator model used before.

Interestingly, the somewhat lower sensitivity of channel 1 in V3, found earlier by the simple model and also the stray-light calibration method (see figure 9), seems to be confirmed by the rescaling of the reference calibration using the optimizer. Even though the difference is small, a thorough analysis of the stability of certain APDs should also be extended to this channel.

Overall, the rescaling of the reference calibrations leads to a similar calibration as the simple polychromator model. The rescaling is certainly more accurate in cases where the simple model is not sufficient to describe the observed calibration (e.g. channel 1 in V3). Both methods show slight variations from the measured reference calibrations. These deviations between the different methods are primarily caused by 1.) changes in the actual diagnostic state (part of the motivation for the self-calibration), 2.) the imprecision of the temperature reconstruction of the optimizer (in particular affecting channel 1 and 2), and 3.) differences between the shape of the actual spectral calibration and the assumed shape of the simple polychromator model. In the next section, the impact of these deviations on the derived electron temperatures is studied.

8. Impact on the temperature

Concerning the accuracy of the electron temperature determined by TS, two competing effects have been found so far: While the actual spectral calibration can be incorrect due to systematic errors or changes over time, the self-calibration can be affected by inaccuracies in the optimizer solution in g_2. Furthermore, it is *a priori* unclear how much uncertainty is caused by using the simple polychromator model, since the filter edges and possible deviations from the expected wavelength dependence (e.g. channel 1 in V3) are not taken into account. The exact impact of these effects depends on the temperature, which channels are affected by possible sensitivity changes and the magnitude of those changes. Hence, no general statement can be made about their impact on the temperature.

Nevertheless, we can use the examples analyzed in this paper to get an idea how large the impact of such deviations could be. For that, we reanalyzed the temperatures of the reference experiment shown in figure 6 with the optimizer calibrations found using the simple polychromator model (figure 8) and by rescaling the reference calibration (figure 12). On the x-axis of figure 13, the electron temperature analyzed with the measured reference calibration (*SuperK* method) is shown as reference temperature T_e. The y-axis shows the ratio of the temperatures evaluated using the optimizer solutions, $T_{e,opt}$

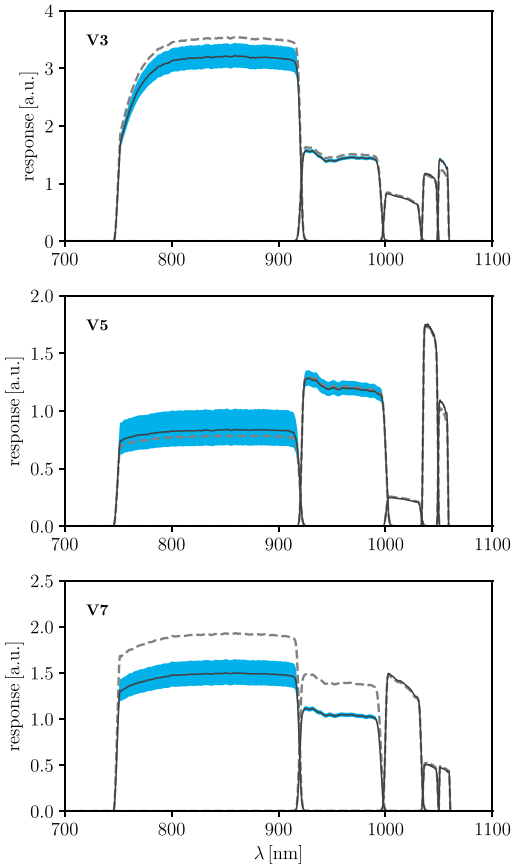

Figure 12. Comparison of the measured reference calibration with the averaged optimizer solution, rescaling the individual channel sensitivities of the same reference calibration (same representation as in figure 8).

Table 3. $\sigma_{\bar{r}_i}/\bar{r}_i$ for the rescaling of the measured calibration for the three example volumes.

Volume	$\sigma_{\bar{r}_2}/\bar{r}_2$	$\sigma_{\bar{r}_3}/\bar{r}_3$	$\sigma_{\bar{r}_4}/\bar{r}_4$	$\sigma_{\bar{r}_5}/\bar{r}_5$
V3	0.045	0.018	0.005	0.019
V5	0.132	0.044	0.009	0.010
V7	0.069	0.021	0.007	0.004

(simple model or rescaled calibration) and the reference temperature. The error bars show the effect of the uncertainty in the r_i on the temperatures obtained with the optimizer (by creating alternative calibrations as $r_i^* = r_i \pm \sigma_{\bar{r}_i}$).

The most obvious observation from figure 13 is that the relative error between the reference temperature and the optimizer solution increases with temperature. This can easily be understood: the higher uncertainty in the gain factors for the first two spectral channels mainly affects higher temperatures, while lower temperatures feature a more narrow TS spectrum and are mostly determined by the remaining spectral channels 3, 4 and 5. Furthermore, apart from a mismatch in the absolute value of the temperature, the traces for the simple polychromator model and the rescaled measured calibrations look remarkably similar (note, for example, the observed behavior below 1 keV). It seems that the remaining uncertainty in g_1 and g_2 (changing the absolute value especially at higher temperatures) plays a larger role than the difference between the two models. This can be understood by the broadness of the TS spectra (see the examples shown in figure 2). If the transmission of the band-pass filters has reasonably steep edges, the simple model gives a fairly accurate estimate of the signal observed in the different spectral channels. V3 also shows that the simple model is able to give a reasonable estimate for the temperature, despite not resembling the reduced sensitivity below 800 nm, which is relevant in situations were coating or irradiation changes the spectral calibration in a noticeable way (i.e. with complex features narrower than the typical filter width).

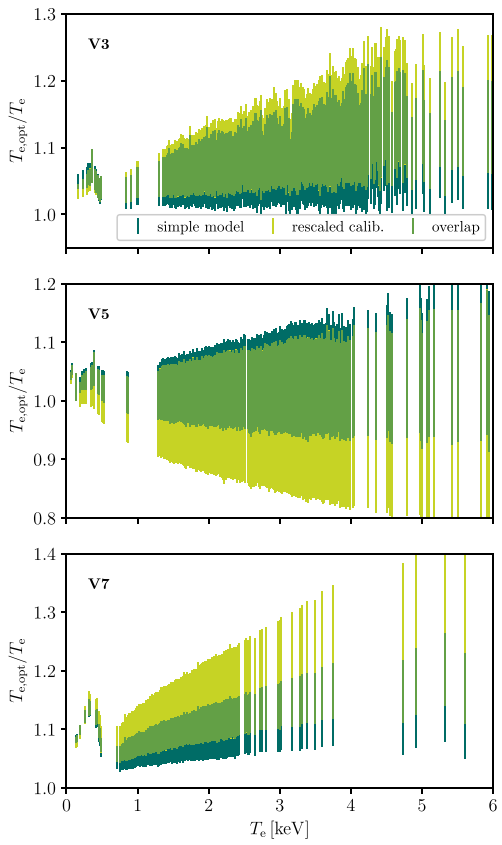

Figure 13. Comparison of the electron temperature evaluated with the measured reference calibration (*SuperK* method), T_e, with the ones evaluated with the optimizer solutions, $T_{e,opt}$ for either the simple polychromator model (dark green) or the rescaled calibration (light green). Depicted are the relative deviations $T_{e,opt}/T_e$ as a function of the reference temperature for the three example volumes. The error bars represent the calibrations constructed from $r_i^* = r_i \pm \sigma_{r_i}$ (an overlap in the error bars is indicated by mid-green color).

Concerning a quantitative comparison, it can be seen that the two optimizer models typically agree within 10 % (probably indicative of the current accuracy of the optimizer method itself rather than differences due to the choice of the calibration model), while both optimizer solutions also seem to agree with the reference temperature within ±10 %. A higher disagreement is observed in V3 for temperatures above 2 keV and in V7. These are, however, also the polychromators for which at least some (if not all) of the observed disagreements could be caused by an inaccurate reference calibration. The analysis for V5 indicates that the averaged optimizer solution (see figures 8 and 12) can reproduce the temperature of the measured calibration within ±10 %. This could probably be improved further by including more independent data sets to the analysis (reducing the error in the gain factors). In other words, the temperatures determined from the optimizer solution for V3 and V7 are probably closer to the actual electron temperature, indicating the importance of self-calibration methods to identify and correct detector misalignment or time-dependent effects in the calibration.

It is also clear, however, that a further increase in the accuracy and precision of the optimizer solution would be beneficial. This may involve a modification of the cost function M, but can also partially be achieved simply by analyzing an even larger set of experiments.

9. Summary and conclusion

Calibrating diagnostics for fusion experiments is not a trivial task. The need for remote-handling and tight experimental schedules conflict with having to validate and recalibrate to account for changes arising from the harsh conditions of such experiments. In-situ self-calibrations can augment current calibration techniques by improving their accuracy while simultaneously prolonging the time between necessary recalibrations. In the case of the TS diagnostic of W7-X, such a self-calibration was possible by combining a model of the diagnostic with an optimizer. Given sufficiently large data sets from actual plasma measurement, the optimizer is able to modify the free parameters of a spectral calibration model

until the experimentally measured data is reconciled with expectations for the modeled calibration. This is useful to identify issues with the diagnostic or systematic differences in the response of the diagnostic during plasma operation and calibration measurements. It also makes it possible to reassess the spectral calibration retrospectively when calibration measurements are no longer possible. Possible applications are, for example, the correction of hardware failures or the validation of historic data (measured before substantial changes to the diagnostic or decommissioning of the experiment). In the future, a similar optimizer approach can also be combined with advanced TS techniques like dual-wavelength TS to improve their accuracy.

The biggest challenge in this procedure is that the cost function (or metric) the optimizer is trying to minimize is very flat around the minimum. This has a number of consequences: Firstly, gradient-based optimization algorithms may fail. Hence, we use DE to find the most likely solution. Secondly, due to the inevitable noise on the raw data, a substantial uncertainty remains in the electron temperature and the calibration parameters that are sensitive to it. For one individual data set, this uncertainty was found to be too large for practical purposes. The analysis of synthetic data revealed, however, that individual solutions seem to scatter around the correct calibration parameters, such that the average over different data sets yields a good approximation of the actual spectral calibration. Thirdly, for the same reason it is difficult to validate the temporal stability of the most temperature-sensitive calibration parameters. This can be solved by compiling data sets in a reproducible way, such that all data sets feature the same average temperature.

Using the calibration averaged over several data sets, this average temperature can be calculated. With a penalty term in the optimizer metric, solutions resulting in a different average temperature can be discarded. This way, the optimizer should find the same solution for all data sets. This means that the temporal stability of the calibration can be assessed independently from how accurate the optimizer was able to estimate the actual spectral calibration for each individual data set.

This ability to validate the stability of the calibration is even more important than finding the calibration itself. For example, an experimental campaign could be started with a measured calibration (using traditional calibration methods) to determine the temperatures of a suitable reference experiment as early as possible. This reference experiment could be repeated throughout the campaign and the optimizer could be used to show that the calibration is still valid or to rescale it if necessary. A different approach could be to use input from other diagnostics to compile reproducible data sets without the need for a dedicated reference experiment.

In this work, as a proof of concept, we followed the approach of using a reference experiment together with a temperature penalty term to investigate the stability of the spectral calibration of the TS diagnostic at Wendelstein 7-X. Overall, the measured calibrations are well reproduced by the optimizer self-calibration. However, individual channels showed clear differences in sensitivity. Some of these differences seemed to be stable, as if the diagnostic underwent some kind of state change, while one other channel seemed to show a drifting behavior. Understanding these observations is work in progress, but stability of the bias voltage, temperature control of the APDs, and internal detector misalignment are obvious starting points for an investigation. These findings underline why optimizer-based self-calibrations can be a useful tool.

The current implementation of the optimizer-based calibration goes beyond a proof-of-principle. As has been demonstrated in this paper, the current approach is already useful to monitor the stability of the TS diagnostic at W7-X. However, further developments are needed to improve the accuracy and precision of the self-calibration. Another future improvement can be the inclusion of expected degradations of the diagnostic (e.g. neutron damage or surface coating). As seen in our analysis for volume 3, small deviations in the wavelength dependence between the modeled and the actual calibration do not play a large role (due to the broad nature of the TS spectra). Nevertheless, adding models or the measured impact of possible degradations to the model (with additional free parameters in the optimization) could improve the accuracy of the resulting temperatures further, while also extending the time between necessary recalibrations.

In summary, while further improvements are desirable, an in-situ self-calibration from plasma measurements could be demonstrated. If data sets can be assembled reproducibly (e.g. with a reference experiment), even relatively small sample sizes are sufficient to validate the temporal stability of the diagnostic. For larger fusion experiments with limited access, such a self-calibration is a powerful tool to improve the data quality and reduce the need for frequent recalibrations. As long as the diagnostic output can be modeled accurately and the space of possible measurements is restricted enough to fix the free parameters of the diagnostic model, a similar approach should be feasible for other diagnostics as well.

Data availability statement

The data cannot be made publicly available upon publication because access is regulated by the W7-X data policy. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Program (Grant Agreement No. 101052200-EUROfusion). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Author contributions

G Fuchert 0000-0002-6640-2139
Conceptualization (lead), Data curation (equal), Investigation (lead), Software (equal), Writing – original draft (lead)

S A Bozhenkov 0000-0003-4289-3532
Software (equal)

K J Brunner 0000-0002-0974-0457
Conceptualization (supporting), Data curation (equal), Software (supporting)

J Knauer 0000-0001-7359-6472
Project administration (equal), Resources (equal)

E Pasch
Project administration (supporting), Resources (supporting)

A Raak 0009-0009-8561-096X
Conceptualization (supporting), Methodology (supporting)

J Wagner 0009-0001-1302-1498
Data curation (supporting), Resources (supporting)

M Hirsch 0000-0002-7120-6087
Project administration (supporting), Supervision (supporting)

R C Wolf 0000-0002-2606-5289
Project administration (lead), Supervision (lead)

References

[1] Röhr H, Steuer K-H, Schramm G, Hirsch K and Salzmann H 1982 *Nucl. Fusion* **22** 1099
[2] Naito O, Yoshida H and Matoba T 1993 *Phys. Fluids B* **5** 4256–8
[3] Bozhenkov S A *et al* 2017 *J. Instrum.* **12** 10004
[4] Bozhenkov S A, Heym S J, Beurskens M N A, Fuchert G, Pasch E, Scott E R and Wolf R C 2019 *Rev. Sci. Instrum.* **90** 033505
[5] Scott E R, Beurskens M N A, Bozhenkov S A, Fuchert G, Hirsch M, Nelde P, Pasch E and Wolf R C 2019 *J. Instrum.* **14** C10033
[6] Yatsuka E *et al* 2010 *J. Plasma Fusion Res. ser.* **9** 12 (available at: www.jspf.or.jp/JPFRS/PDF/Vol9/jpfrs2010_09-012.pdf)
[7] Tojo H, HATAE T and ITAMI K 2011 *Plasma Fusion Res.* **6** 1302018
[8] Smith O R P, Gowers C, Nielsen P and Salzmann H 1997 *Rev. Sci. Instrum.* **68** 725–7
[9] McCormack O, Giudicotti L, Fassina A and Pasqualotto R 2017 *Plasma Phys. Control. Fusion* **59** 055021
[10] Tojo H, Hatae T and Itami K 2012 *J. Instrum.* **7** C05004
[11] Giudicotti L and Pasqualotto R 2014 *Nucl. Fusion* **54** 043005
[12] Fajemirokun H, Gowers C, Nielsen P, Salzmann H and Hirsch K 1990 *Rev. Sci. Instrum.* **61** 2849–51
[13] International Atomic Energy Agency, Vienna (Austria) 1991 LIDAR Thomson scattering *IAEA Technical Committee Meeting on LIDAR Thomson Scattering* (Abingdon, United Kingdom)
[14] Pasch E, Beurskens M N A, Bozhenkov S A, Fuchert G, Knauer J and Wolf R C 2016 *Rev. Sci. Instrum.* **87** E729
[15] Fuchert G *et al* 2024 *Rev. Sci. Instrum.* **95** 083533
[16] Storn R and Price K 1997 *J. Global Optim.* **11** 341–59
[17] Virtanen P *et al* 2020 *Nat. Methods* **17** 261–72
[18] Price K, Storn R M and Lampinen J A 2005 *Differential Evolution: A Practical Approach to Global Optimization* (Springer)
[19] Brunner K J, Akiyama T, Hirsch M, Knauer J, Kornejew P, Kursinski B, Laqua H, Meineke J, Trimiño Mora H and Wolf R C 2018 *J. Instrum.* **13** 09002